



VNIVERSIDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

Departamento de Biblioteconomía y Documentación

INDEXAÇÃO E CONTROLO DA TERMINOLOGIA EM BIBLIOTECAS
DO ENSINO SUPERIOR POLITÉCNICO EM PORTUGAL:
O SISTEMA NO INSTITUTO POLITÉCNICO DE PORTALEGRE

TESE DE DOUTORAMENTO

M^a Catarina Barradas Martins

Tese dirigida pela Professora Doutora Carmen Caro-Castro

Salamanca, 2013

MARTINS, M^a Catarina Barradas

Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal: o sistema do Instituto Politécnico de Portalegre. Tese de Doutoramento /Maria Catarina Barradas Martins; Tese dirigida pela Prof^a: Dr^a Carmen Caro-Castro. Salamanca: 2013.

391 p. ; il.

1. Caro-Castro, Carmen, dir. II. Título. 1. Controlo da terminologia. 2. Bibliotecas politécnicas. 3. Sistemas de indexação.

025.4(076)

027.7(469)



Departamento de Biblioteconomía y Documentación

INDEXAÇÃO E CONTROLO DA TERMINOLOGIA EM BIBLIOTECAS
DO ENSINO SUPERIOR POLITÉCNICO EM PORTUGAL:
O SISTEMA NO INSTITUTO POLITÉCNICO DE PORTALEGRE

M^a Catarina Barradas Martins

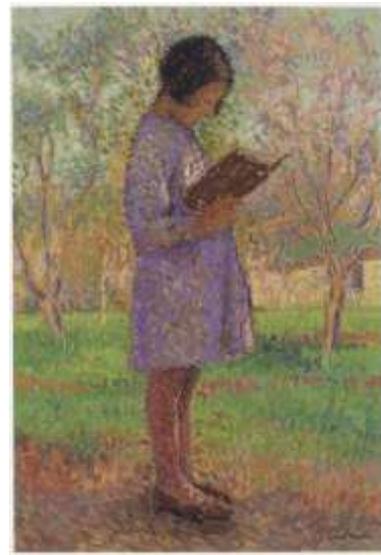
Tese de Doutoramento sob orientação da
Professora Doutora Carmen Caro-Castro
apresentada à Universidade de Salamanca.

Diretora: **Professora Doutora Carmen Caro-Castro**

Salamanca, 2013

A maior aventura de um ser humano é viajar, e a maior viagem que alguém pode empreender é para dentro de si mesmo. E o modo mais emocionante de realizá-la é ler um livro, pois um livro revela que a vida é o maior de todos os livros...

Augusto Cury



Henri Martin “Young girl reading”

Na leitura, a amizade é subitamente reconduzida à sua pureza primordial.

Com os livros, nada de amabilidades.

Se passamos a noite com esses amigos, é porque assim verdadeiramente nos apetece.

A eles, pelo menos, só a contragosto é que muitas vezes os deixamos [...]

A atmosfera desta pura amizade é o silêncio, mais puro do que a palavra.

Marcel Proust

Dedico este trabalho ao João, companheiro de todas as horas,
que desde o início me motivou com a sua paciência e espírito positivo,
a levar até ao fim este desafio.

Aos meus filhos, Manuel e Henrique, dedico o intenso trabalho destes anos,
como testemunho de que é possível concretizar sonhos,
com esforço e perseverança.

Agradecimentos

Agradeço especialmente à minha orientadora Prof.^a Carmen Caro-Castro, pela sua total disponibilidade, firmeza e dedicação na direção deste trabalho. A qualidade das suas sugestões e a exigência das correções, enriqueceram e motivaram a investigação realizada.

Na componente estatística devo agradecer às colegas Adelaide, Cristina e Paula, e ao amigo João Luís pelas suas ajudas.

Aos órgãos diretivos das Escolas Superiores do Instituto Politécnico de Portalegre, pela autorização concedida para pesquisar e divulgar dados bibliográficos dos catálogos das Bibliotecas.

Por fim, e de forma muito especial, aos meus pais, e aos amigos que compreenderam o meu afastamento e não deixaram de me ‘procurar’.

Resumo

O principal objetivo deste trabalho foi evidenciar a dificuldade no desenvolvimento do processo de indexação em bases bibliográficas. Os instrumentos normativos e de apoio a essa tarefa são dispersos e difíceis de utilizar para grande parte dos profissionais. Os termos de indexação são recolhidos de várias fontes, potenciando a sua inconsistência se não for feito o controlo de autoridades, pondo em causa a recuperação de informação eficaz por parte dos utilizadores dos catálogos.

A investigação divide-se em duas partes distintas: na primeira é identificado o uso de instrumentos normativos e de apoio à indexação nas bibliotecas politécnicas em Portugal. Na segunda parte são observadas as características dos índices de assunto num grupo de bibliotecas (Instituto Politécnico de Portalegre). A metodologia desenvolveu-se de acordo com os dois níveis de observação, no primeiro através de um inquérito às instituições, e no segundo através da análise de registos bibliográficos, em que foram definidos três eixos centrais: formato UNIMARC, sintaxe dos cabeçalhos e terminologia.

Através dos resultados obtidos, conclui-se que na maioria dos casos não são usados instrumentos normativos, e o sistema de indexação português SIPORbase é muito pouco utilizado. A recolha de termos é essencialmente de tesouros, e não é feito o controlo de autoridades, quer a nível da terminologia, quer da sintaxe dos cabeçalhos. A falta de consistência observada nos índices de assunto demonstrou que é necessário uma política de indexação, e com esse fim foi apresentado um modelo que se sugere utilizar.

Subject indexing and terminology control in the Portuguese polytechnic libraries: the indexing system in Instituto Politécnico de Portalegre.

Abstract

The main objective of this work was to evidence the existing difficulties in subject indexing in bibliographic databases. The normative instruments and tools that support the process are somewhat diffuse and hard to use for many professionals in the area. The indexing terms collected from different sources, may lead to the inconsistency of subject headings, unless the control of authorities is done. And inconsistent subject indexes compromise an accurate information retrieval in library catalogs.

This research is divided in two distinct parts: the first one was to identify the use of indexing instruments and tools in the Portuguese polytechnic libraries, and the second one was to analyze the catalogs' subject indexes in a group of libraries (Instituto Politécnico de Portalegre). The methodology was developed according to the defined levels of observation, namely through an inquiry to the institutions, and through the analysis of index features in the bibliographic records, where three central points were defined: the UNIMARC format, the subject headings syntax and the terminology.

The obtained results led us to the conclusion that the situation of indexing without normative support is dominant, and the use of the Portuguese indexing system SIPORbase is quite slight. The indexing terms are essentially collected from thesauri, and are not controlled, the same happening to the subject headings in their syntax.

The lack of consistency in the bibliographic catalogs revealed that an indexing policy is necessary, and with the aim of establishing a guiding one, a simplified indexing model is proposed.

Sumário

Índice de figuras.....	15
Índice de tabelas.....	16
Lista de Siglas e Acrónimos	17
1. Indexação e acesso por assuntos nos catálogos bibliográficos.....	33
1.1 O processo de indexação: definição e finalidade.....	36
1.1.1 Fases de desenvolvimento do processo.....	37
1.2 Tipologia e características das linguagens de indexação.....	45
1.2.1 Semântica e sintaxe dos vocabulários controlados	54
1.2.1.1 Sintaxe dos cabeçalhos de assunto	59
1.2.2 Sistemas de indexação de referência.....	62
1.2.2.1 O Sistema de Indexação em Português - SIPORbase.....	66
1.2.3 Evolução normativa dos vocabulários controlados.....	71
1.2.4 Alternativas e tendências de indexação.....	81
1.3 Acesso por assuntos nos catálogos bibliográficos	101
1.3.1 O catálogo bibliográfico e de autoridades.....	106
1.3.2 Controlo de autoridades	115
1.3.2.1 Referências de controlo de autoridades	122
1.3.3 Requisitos funcionais dos registos bibliográficos e de autoridade.....	130
1.3.4 Os novos princípios de catalogação e o código RDA	141
1.4 Avaliação da qualidade da indexação	145
1.4.1. Antecedentes de avaliação de Sistemas de Recuperação de Informação.....	146
1.4.2. Metodologias de avaliação	151
1.4.3. Medidas de avaliação	155
2. Contexto das bibliotecas do ensino superior politécnico	163

2.1 A função da biblioteca no ensino superior.....	165
2.2 Bibliotecas do ensino universitário e do ensino politécnico.....	173
2.3 O caso das Bibliotecas do Instituto Politécnico de Portalegre.....	176
2.3.1. Fontes de terminologia preferenciais	179
2.3.2 O sistema bibliográfico e de autoridades	182
2.3.3 O projeto do catálogo coletivo: necessidade de uniformização	189
3. Objetivos e Metodologia	199
3.1 Objetivos	202
3.2 Metodologia	204
3.2.1 Nível 1: Instrumentos de apoio à indexação nas Bibliotecas do Ensino Superior Politécnico	205
3.2.1.1 Recolha de dados	205
3.2.1.2 Seleção da amostra	206
3.2.1.3 Variáveis.....	207
3.2.1.4 Análise	209
3.2.2 Nível 2: Características da indexação nas Bibliotecas do Instituto Politécnico de Portalegre.....	210
3.2.2.1 Recolha e depuração de dados.....	210
3.2.2.2 Seleção da amostra	214
3.2.2.3 Variáveis.....	218
3.2.2.4 Análise	233
4. Resultados.....	237
4.1. Nível 1 - Instrumentos de apoio à indexação nas BESP.....	239
4.2. Nível 2 - Características da indexação nas BIPP	244
5. Discussão.....	265
5.1 Instrumentos normativos.....	267

5.2 Ferramentas de apoio: diversidade e uniformidade	270
6. Conclusões.....	275
6.1 Conclusões gerais	277
6.2 Conclusões específicas (BESP)	280
6.3 Conclusões específicas (BIPP)	281
6.3.1 Conclusões de utilização do SIPORbase.....	287
6.4 Sugestões para investigações futuras	288
7. Proposta de “Metodologia de Indexação”	291
7.1 Modelo de indexação	294
8. Bibliografia.....	309
9. Anexos.....	333
Anexo I. Tabelas abreviadas SIPORbase.....	335
Anexo II. Rede do Ensino Superior Politécnico em Portugal.....	353
Anexo III. Apresentação do estudo e pedido de resposta a inquérito.....	355
Anexo IV. Questionário <i>online</i>	357
Anexo V. Terminologia dos cabeçalhos de Nomes Comuns.....	361
Anexo VI. Terminologia dos cabeçalhos de Nomes Próprios	387

Índice de figuras

Fig. 1.1: Visão tradicional dos diferentes tipos de linguagens de indexação	46
Fig. 1.2: Exemplo de referências cruzadas Ver e Ver também num índice de assuntos.....	58
Fig. 1.3: Cabeçalhos simples e complexos, com subdivisões (lista LCSH).....	61
Fig. 1.4: Exemplo de linguagem natural na pesquisa de informação num resumo	82
Fig. 1.5: Exemplo de linguagem controlada no campo de assunto	84
Fig. 1.6: Resultado da indexação de um documento no HIVE.....	96
Fig. 1.7: “Linking Open Data cloud diagram”, (Richard Cyganiak e Anja Jentzsch)....	98
Fig. 1.8: “Important elements of SKOS” (Olensky 2010).....	99
Fig. 1.9: Registo de autoridade de nome de autor (visualização UNIMARC).....	112
Fig. 1.10: Registo de autoridade de nome de autor (visualização ISBD).....	113
Fig. 1.11: Nome de autor como assunto num registo bibliográfico	113
Fig. 1.12: Pesquisa por assunto com remissivas ‘Ver’ e ‘Ver Também’	114
Fig. 1.13: FRBR – Modelo Entidade-Relacionamento Grupo 1 e 2	134
Fig. 1.14: Relação FRSAD com FRBR (com adição da família de entidades FRAD)	137
Fig. 2.1: Interface de pesquisa do catálogo coletivo virtual	182
Fig. 2.2: Ficheiro de autoridades, com visualização de registos selecionados.....	184
Fig. 2.3: Registo bibliográfico sem e com controlo de autoridades	185
Fig. 2.4: Ligação do registo bibliográfico ao registo de autoridade	185
Fig. 2.5: Exemplo de pesquisa nos índices.....	194
Fig. 3.1: Registo bibliográfico em visualização UNIMARC	216
Fig. 4.1: Utilização de instrumentos normativos e de Tesouros ou léxicos.....	243
Fig. 4.2: Tipo de cabeçalhos, subdivididos e não subdivididos	245
Fig. 4.3: Termos 'não conforme' os índices da BNP.....	253
Fig. 4.4: Termos simples e termos compostos nos cabeçalhos 'não subdivididos'.....	255
Fig. 4.5: Não conformidade com o princípio do 'cabeçalho específico'.....	260
Fig. 4.6: Termos dos índices da BNP.....	262
Fig. 4.7: Termos dos tesouros /lista de descritores.....	263

Índice de tabelas

Tabela 1.1: Exemplos da reciprocidade das relações semânticas.....	57
Tabela 1.2: Uso da linguagem natural na indexação	88
Tabela 1.3: Correspondência de campos UNIMARC (Autoridades e Bibliográfico)...	111
Tabela 1.4: Definição das tarefas dos usuários nos três modelos.....	139
Tabela 1.5: Medidas de avaliação da indexação.....	158
Tabela 2.1: Termos existentes no ficheiro de autoridades de assuntos	187
Tabela 3.1: Grupo de variáveis do nível 1	209
Tabela 3.2: Bibliotecas e respetivas áreas temáticas	210
Tabela 3.3: Campos pesquisados nas bases para verificação de conteúdo.....	211
Tabela 3.4: Número de registos sem assunto e com assunto.....	212
Tabela 3.5: Registos com assunto (campo 606 e outros campos do bloco 6xx).	213
Tabela 3.6: Peso das bases na amostra de cabeçalhos	217
Tabela 3.7: Exemplos de cobertura das fontes de terminologia	230
Tabela 3.8: Grupo de variáveis do nível 2	232
Tabela 4.1: Uso de instrumentos normativos na prática da indexação.....	240
Tabela 4.2: Designação do instrumento normativo utilizado	240
Tabela 4.3: Uso de léxico ou tesouros no auxílio à indexação	241
Tabela 4.4: Designação de léxicos ou tesouros usados	242
Tabela 4.5: Utilização de ficheiro de autoridades de assunto.....	242
Tabela 4.6: Resultados nas variáveis UNIMARC (Nomes Comuns).....	247
Tabela 4.7: Resultados nas categorias da variável UNIMARC (Nomes Próprios).....	248
Tabela 4.8: Distribuição dos 'cabeçalhos subdivididos'.....	248
Tabela 4.9: Erros de sintaxe nos cabeçalhos de Nomes Comuns.....	249
Tabela 4.10: Terminologia 'não conforme' os índices BNP (cab. subdivididos)	252
Tabela 4.11: Terminologia 'não conforme' os índices BNP.....	252
Tabela 4.12: Distribuição de erros nas variáveis dos Nomes Próprios	257
Tabela 4.13: Soma de resultados por variável nos cabeçalhos de NC e NP.....	258
Tabela 4.14: Variáveis 'Cabeçalho específico' e 'Fontes de Terminologia'	264
Tabela 6.1 Ligação dos níveis de investigação:os índícios e as manifestações	279

Lista de Siglas e Acrónimos

APDIS – Associação Portuguesa de Documentação e Informação de Saúde
AACR2 – Anglo-American Cataloguing Rules, 2nd edition
BN – Biblioteca Nacional
B-ON – Biblioteca do Conhecimento Online
CLIP – Compatibilização de Linguagens de Indexação em Português
DECS – Descritores em Ciências da Saúde
ERIC – Education Resources Information Center
ESAE – Escola Superior Agrária de Elvas
ESE – Escola Superior de Educação
ESS – Escola Superior de Saúde
ESTG – Escola Superior de Tecnologia e Gestão
EUDISED – European Documentation and Information System for Education
FAO – Food and Agricultural Organization of the United Nations
FCCN – Fundação para a Computação Científica Nacional
FRAD – Functional Requirements for Authority Data
FRBR – Functional Requirements for Bibliographic Records
FRSAD – Functional Requirements for Subject Authority Data
GIPE – Grupo de Indexação para a Educação
GT BES – Grupo de Trabalho das Bibliotecas do Ensino Superior
IPP – Instituto Politécnico de Portalegre
IPQ – Instituto Português da Qualidade
ISBD – International Standard Bibliographic Description
LINCE – Linguagem de Indexação para as Ciências da Educação
MCES – Ministério da Ciência e do Ensino Superior
MCTES – Ministério da Ciência Tecnologia e do Ensino Superior
MESH – Medical Subject Headings
MFN – Master File Number
OPAC – Online Public Access Catalog
PORBASE – Base Nacional de Dados Bibliográficos

RDA – Resource Description and Access

RIDEPES – Rede de Informação e Documentação Educativa do Ensino Superior

SIPORBASE – Sistema de Indexação em Português

SKOS – Simple Knowledge Organization Systems

UMIC – Unidade de Missão Inovação e Conhecimento

ULMS – Unified Medical Language System

UNESCO – United Nations Educational Scientific and Cultural Organization

UNIMARC – Universal Machine Readable Cataloging

Introdução

É comum a imagem que temos das bibliotecas académicas tradicionais em que, apesar de se fazerem todas as adaptações tecnológicas para acompanhar as necessidades e exigências dos públicos, se continua a trabalhar com base nos mesmos preceitos e técnicas documentais que têm marcado as últimas décadas. Os processos de catalogação e de indexação de documentos exigem um conhecimento consistente para que seja aplicada uma prática verdadeiramente correta. A conjugação destas práticas com os atuais ambientes *www* na área da pesquisa e recuperação da informação, origina novas visões e desafios, em geral, nos novos modos de acesso à informação, e em particular ao nível da catalogação.

Também é comum perceber as seguintes situações dentro destas bibliotecas:

- os utilizadores mais jovens procuram as bibliotecas para aceder a informação essencialmente eletrónica, e começam a desprestigiar a informação em suporte de papel;
- os responsáveis institucionais sentem a necessidade premente de adaptarem as bibliotecas aos tempos atuais e atuam, cada vez mais, no sentido de ir ao encontro das necessidades dos utilizadores, sob ‘ameaça’ de as verem transformadas em ciberespaços, onde os catálogos não são consultados e os livros nas estantes se ‘enchem de pó’.

As novas gerações de utilizadores situam-se de facto em ciberespaços, e as bibliotecas tradicionais terão que fazer um esforço continuado de adaptação a esta nova realidade. As que encontram maiores dificuldades de se modernizarem, quer por questões económicas quer por outras questões conjunturais, deverão continuar a prestar serviços de qualidade aos seus utilizadores.



Ilustração 1: Biblioteca infinita de Job Koelewijn

Os sistemas de informação mais avançados privilegiam o acesso rápido e efetivo, e evoluem tecnologicamente a uma velocidade alucinante, impossível de acompanhar por parte de muitas bibliotecas tradicionais, que contêm milhares de livros em papel. É impossível tentar abrandar os comportamentos decorrentes da atual dinâmica da Sociedade da Informação, e competir com padrões de acessibilidade e usabilidade de *blogs, feeds, folksonomias, etc.*

Durante quanto tempo irão ser procurados os livros que enchem estas bibliotecas? Há discussões alargadas a este respeito, sem previsão de conclusões acertadas. Se nos interrogarmos quanto à possibilidade de potenciar o seu uso, aí sabemos que sim! Ainda que, à luz dos mais avançados centros de informação, se possam situar numa esfera de ‘bibliotecas retro’, o seu conteúdo continua a ser de grande valor para os amantes dos livros em papel.

As novas tecnologias têm sido geradoras de atividades e novos métodos na docência, no sentido de facilitar o processo de aprendizagem. O Espaço Europeu de Educação Superior (EEES), criado pela Declaração de Bolonha (1999) é um espaço comum centrado na aprendizagem, onde são favorecidas atitudes como a capacidade crítica, a atividade de pesquisa, o conhecimento e utilização de fontes originais, ou a capacidade de aprendizagem ao longo da vida. As instituições de ensino superior fazem frente ao repto de Bolonha através da criação de estruturas com a máxima disponibilidade e qualidade dos serviços, e os novos métodos de ensino e aprendizagem repercutem-se diretamente na biblioteca: maior uso dos recursos e serviços já disponíveis, exigência de novos serviços, sendo necessário reformular o modelo de organização e funcionamento da biblioteca universitária (Celestino Angulo 2007).

É neste contexto que surge o conceito dos CRAI (Centro de Recursos de Aprendizagem e Investigação), designação derivada da anglo-saxónica “*Learning Resources Centre*”, o novo modelo em que as bibliotecas se devem transformar, para apoiar o processo de educação superior. As novas formas de estudo, de leitura e de aprendizagem resultam na criação de novos modelos de bibliotecas universitárias. A

necessidade de adaptar as bibliotecas aos novos contextos e exigências de informação levou a que o modelo tradicional de biblioteca seja substituído pelo modelo CRAI. Este novo conceito de biblioteca implica uma adaptação à sociedade do conhecimento, com recursos para a aprendizagem e a investigação (Taladriz 2007; Delandés Palomares 2012).

Os novos modelos implementados em países avançados na inovação educativa como Estados Unidos, Reino Unido e Holanda, baseiam-se sobretudo em duas grandes ações: transformar a biblioteca presencial num centro com horários alargados, com recursos disponíveis para todo o tipo de aprendizagem, e numa forte componente tecnológica em equipamentos e serviços, com uma biblioteca digital que reúna os diversos sistemas de informação (Martínez 2004). O que está a acontecer não é só uma mudança tecnológica, mas uma total “rearranjo” de como abordamos a informação, ou seja, anteriormente colocava-se a questão: “a minha biblioteca tem este livro?” A questão que se coloca agora é: “este livre está disponível em algum sítio?” (Coyle 2007).

As propostas sobre os CRAI contemplam que desde um só ponto se possa aceder aos recursos do catálogo, a repositórios, revistas e livros eletrónicos subscritos, recursos *web*, etc. É neste contexto do acesso a uma variedade de recursos que é muito importante o controlo dos pontos de acesso à informação, para que os sistemas de pesquisa possam ir de encontro às expectativas dos usuários.

É sabido por quem se dedica a estas temáticas que uma pesquisa no Google¹ não serve os mesmos propósitos que uma pesquisa num catálogo de assuntos, que o utilizador que pretende fazer uma pesquisa exaustiva sobre um assunto utiliza a informação estruturada para o fazer, e neste caso coloca-se atualmente outra questão: também a informação estruturada está já disponível *online*, e de forma esmagadora; os sistemas mais avançados dispõem de estruturas informáticas em permanente atualização, presentes em muitas instituições com capacidade humana e tecnológica para as manter e para acompanhar os desafios de ponta.

A atual aplicação das novas tecnologias de representação e recuperação da informação, dá uma nova postura ao bibliotecário, no sentido de se comprometer com a construção do catálogo *online*, ferramenta crucial numa biblioteca, e essa responsabilidade é assumida perante a sua comunidade local e remota, cada vez mais exigente com a qualidade da informação (Boccatto 2009). A biblioteca académica é um ambiente conhecido por muitos leitores e investigadores, que recorrem diariamente aos catálogos em busca de referências que os orientem para a informação desejada. Os catálogos *online* atualmente disponíveis nas bibliotecas, são sistemas de recuperação de informação inseridos num contexto de áreas científicas especializadas que necessitam de constante atualização e manutenção para que sejam compatíveis com as necessidades da comunidade que servem (Boccatto, Fujita, Leiva 2011).

Há mais de uma década que se evidenciou a ideia de que é importante a mudança de mentalidade dos bibliotecários, no sentido de dar a devida importância ao catálogo, para

¹ Ver a perspetiva de Hjørland (2012) “*Is classification necessary after Google?*”, onde o autor considera que “à luz do novo entendimento teórico de classificação e de KO, os critérios de classificação são idênticos aos critérios de fornecer informação relevante”.

que passe a ser a ferramenta mais visível e mais eficaz da biblioteca, uma ferramenta de pesquisa e recuperação de informação que satisfaça as necessidades dos usuários, “se somos conscientes da importância do catálogo estamos na direção certa” (Villén-Rueda 2006). Segundo este autor, os sistemas de indexação e o controlo de autoridades de assuntos, continuam situações pendentes na maioria das instituições, ou talvez nunca tenham sido uma prioridade para a comunidade bibliotecária. A indexação e a recuperação por assuntos têm uma correspondência direta e as investigações mostraram que inadequadas políticas de indexação se repercutem na forma como os usuários utilizam o catálogo.

As bibliotecas de ensino superior constituem um apoio muito importante às atividades académicas e de investigação, sendo em muitos casos centros dinâmicos de pesquisa de informação. Os fundos documentais disponíveis nestas bibliotecas podem ter melhor divulgação pela associação de assuntos nos catálogos, através do enriquecimento dos registos bibliográficos, que correspondam à expectativa dos utilizadores em resultados de pesquisa. É imprescindível a adaptação das bibliotecas às atuais necessidades dos utilizadores, e o catálogo bibliográfico é um instrumento de pesquisa que carece de grande intervenção, para que seja realmente aproveitado o seu potencial, podendo ser uma ferramenta de grande utilidade.

É evidência comum que a comunidade académica reivindica o alargamento de horários de abertura das bibliotecas, e procura essencialmente espaços de estudo com acesso à internet e às ferramentas de pesquisa rápida aí disponíveis. É visível a concentração de interesses no acesso à informação eletrónica, e sabemos que no

contexto dos avanços tecnológicos a que se vem assistindo, os catálogos bibliográficos (OPACs) podem ser instrumentos de pesquisa valiosos. O facto de terem já potencialidades para integrar os documentos em texto integral associados às referências bibliográficas, reforça a sua importância.

A mais-valia destes catálogos é a informação estruturada aí representada, e é esta perspectiva que leva os usuários a reconhecer as suas vantagens em relação a outros instrumentos de pesquisa, em que a informação é recuperada de forma não controlada (Campos 2003).

Estreitamente ligada à recuperação da informação nos catálogos bibliográficos está a prática da indexação de documentos, e é evidente que da qualidade dessa tarefa resultam situações diferenciadas. Na área da indexação há várias abordagens possíveis, observando-se neste trabalho a perspectiva do uso da linguagem alfabética e da necessidade do controlo dos índices de assunto para acesso temático nos catálogos.

Estas questões levam-nos ao problema concreto da nossa investigação, que diz respeito à prática da indexação de documentos. O desenvolvimento de estudos e projetos nesta área manifesta-se a um ritmo de difícil acompanhamento, devido à sua complexidade e alguma dispersão; este facto aliado à escassez de recursos humanos especializados nas bibliotecas são os motivos que mais podem contribuir para a dificuldade no desenvolvimento do processo de indexação.

Os profissionais com responsabilidade nessa tarefa têm conhecimento de ‘uma e outra’ norma, de ‘um e outro’ sistema de indexação, e por vezes dão conta do avanço dos estudos na sua especificidade, sem que seja possível acompanhá-los. Quando se toma conhecimento de novos avanços nessa matéria, depara-se com uma quantidade de informação esmagadora, que logo desencoraja à leitura aprofundada. Nalguns casos, a estrutura dos documentos remete de uns pontos para outros, até se encontrarem respostas concretas às questões colocadas. Entretanto, como a dinâmica das tarefas diárias numa biblioteca é imparável, a execução desse trabalho de forma controlada é adiado, e os utilizadores, se por um lado, não têm a perceção das dificuldades de uma indexação de qualidade, por outro, exigem a disponibilização rápida da informação e resultados satisfatórios nas pesquisas.

Ao percebermos que o catálogo da biblioteca se está a tornar numa ferramenta de pesquisa que não dá resposta às questões solicitadas, e que não cumpre o seu papel de recuperar a informação lá registada, é altura de intervir. Então, os profissionais pretendem a todo o custo chegar à qualidade, rever as metodologias, controlar o processo de indexação, ou atualizar as ferramentas de trabalho. Isso exige-lhes tempo disponível, recolhimento, concentração, determinação e muito trabalho, até se poderem ver alguns resultados práticos. Trata-se de uma tarefa exigente e especializada, para a qual as instituições não estão sensibilizadas de forma a qualificar o sector das bibliotecas com os recursos humanos necessários. A proporção inadequada entre as exigências inerentes à prática da indexação, com a dispersão e complexidade de instrumentos normativos, aliados à falta de recursos especializados, foram fatores que

levaram à necessidade de intervenção nesta matéria em quatro bibliotecas do Instituto Politécnico de Portalegre.

No estudo de caso destas bibliotecas dispomos das seguintes evidências: é feita uma indexação *ad hoc*, pois cada uma evoluiu de forma isolada; não há um conhecimento aprofundado do método de análise de documentos para identificação dos conceitos e representação em termos de uma linguagem documental; as regras de controlo da terminologia são difíceis de aplicar; as fontes de terminologia são dispersas e contêm abordagens semânticas diversas. Nas quatro bibliotecas que constituem o estudo de caso e que pretendem disponibilizar os seus catálogos bibliográficos para um projeto de catálogo coletivo, deparamo-nos com as questões da falta de controlo de terminologia de assuntos.

A proposta de uma nova metodologia de trabalho deve conter recomendações que permitam orientar o processo de indexação, desde a fase de análise de documentos à construção do léxico de assuntos (com controlo de autoridades); a proposta irá servir como documento de apoio, sintetizando, de forma abreviada e sequencialmente lógica, os passos do processo: observação da Norma NP 3715, fontes temáticas recomendadas para recolha de termos, abordagem ao controlo de terminologia, e controlo de autoridades no UNIMARC.

Pensamos que, por questões conjunturais, estas dificuldades são comuns às bibliotecas do ensino politécnico, o que nos motivou a realizar um estudo exploratório para identificar a existência desse tipo de problemas nessas bibliotecas portuguesas. A

confirmação da existência dessas dificuldades no sistema bibliotecário politécnico, poderá contribuir para que haja um défice na prática de indexação controlada, relativamente ao avanço dos estudos nesta área, essencialmente por falta das condições ideais, quer ao nível de recursos especializados, quer ao nível da aplicabilidade dos instrumentos disponíveis.

São objetivos gerais da nossa investigação:

- a) Identificar dificuldades na utilização de ferramentas de apoio à indexação nas bibliotecas do ensino politécnico em Portugal;
- b) Ao nível de um grupo de bibliotecas, demonstrar a dificuldade da prática de uma indexação consistente, através do estudo dos seus índices de assunto, e contribuir para alterar a situação de ‘indexação *ad hoc*’, com a proposta de uma metodologia que permita orientar o controlo desse processo.

Ao nível teórico, o objetivo geral é clarificar as principais questões diretamente ligadas ao processo de indexação. Neste sentido será feita uma abordagem à teoria da indexação e da recuperação de informação, assim como à problemática dos catálogos bibliográficos *online*, e do seu contexto de utilização em bibliotecas do ensino superior.

As motivações para levar a cabo este trabalho de investigação são de carácter científico e pessoal. O estudo desta temática não está desenvolvido em Portugal, pelo que poderá ser dado um contributo nesta área, e eventualmente abrir caminhos para

outras bibliotecas realizaram novas investigações nesta matéria. A nível pessoal, o facto de ter desenvolvido atividades profissionais e docentes nesta área, desde há duas décadas no Instituto Politécnico de Portalegre, é uma forte motivação para dar resposta a um problema que se faz sentir nas suas bibliotecas, relativamente ao qual não existe um estudo teórico aplicado, e que poderá ser de grande utilidade no domínio da pesquisa e recuperação de informação.

■ Organização do trabalho

As fontes de informação selecionadas para pesquisa bibliográfica foram, para exploração do campo de estudo, as bases dados do recurso B-on² (Biblioteca do Conhecimento Online), como *Web of Knowledge*, *Elsevier*, *Springer*, a base LISA (*Library and Information Science Abstracts*), etc. Para recolha de informação específica sobre estudos efetuados em Portugal, foram feitas pesquisas nos sistemas *online* da Biblioteca Nacional de Portugal, concretamente no catálogo da BNP, no Catálogo Colectivo das Bibliotecas Portuguesas - PORBASE e na Bibliografia Nacional Portuguesa³ em linha. Deste levantamento bibliográfico foi recuperada pouca informação comparativamente à situação internacional.

² <http://www.b-on.pt/>

³ Neste recurso encontra-se a informação relativa às monografias e publicações periódicas editadas em Portugal, assim como teses e outros trabalhos de estabelecimentos de ensino superior portugueses sujeitos ao Depósito Legal desde 1986. Esta bibliografia está disponível em linha a partir do ano de 2003, e não foi feita pesquisa retrospectiva.

A contextualização teórica da nossa investigação é apresentada numa sequência lógica, relacionada com os pontos do processo de indexação. O capítulo 1, dedicado à indexação e acesso por assuntos nos catálogos bibliográficos, divide-se em quatro subcapítulos: o 1.1 refere-se ao processo de indexação e sua finalidade, e ao conhecimento das respetivas fases de prática normalizada; no 1.2 é abordada a tipologia e características das linguagens de indexação. É dada especial atenção à utilização da linguagem alfabética e aos vocabulários controlados, no que diz respeito à terminologia e relações semânticas, e à evolução normativa. No final deste capítulo são observadas alternativas de indexação e tendências, no que diz respeito à utilização da linguagem natural e à contribuição da indexação automática. Embora estas questões em particular não sejam aprofundadas pois estão fora do cerne da nossa investigação, pensamos que é essencial para os profissionais desta área conhecerem as problemáticas atuais e tendências futuras, que podem contribuir para a tomada de decisões conscientes nas suas instituições.

O subcapítulo 1.3 é dedicado ao acesso por assuntos nos catálogos bibliográficos, em que são observadas as características dos catálogos bibliográficos e de autoridades. É dada especial atenção ao controlo de autoridades, assim como aos novos desenvolvimentos no universo de descrição bibliográfica, como os modelos conceptuais, os novos princípios de catalogação e o código RDA. O capítulo 1 é encerrado com a reflexão sobre a avaliação da qualidade na indexação, no sentido em que é através da avaliação que se deve observar a efetiva aplicação de pressupostos anteriormente apresentados.

No capítulo 2 contextualiza-se a investigação à luz da missão e função das bibliotecas de ensino superior, referindo o contexto universitário e politécnico português. O subcapítulo 2.3 dedica-se à situação das bibliotecas que constituem o estudo de caso, apresenta-se o projeto do catálogo coletivo, as dificuldades no controlo de terminologia e as possibilidades de intervenção nas bases bibliográficas onde está em causa a qualidade dos índices de assunto.

Os capítulos 3, 4, 5 e 6 constituem a parte prática da investigação, onde são apresentados, respetivamente, os objetivos e metodologia, os resultados, a discussão, e as conclusões. A parte prática é apresentada de acordo com dois níveis de observação, um que diz respeito ao estudo exploratório efetuado nas bibliotecas do ensino superior politécnico, e outro que diz respeito às características da indexação nas bases bibliográficas das bibliotecas do Instituto Politécnico de Portalegre. Na sequência da observação do uso de instrumentos nas BESP, interessou aprofundar a análise da prática de indexação e observar as características dos índices do grupo de bibliotecas. Analisar a terminologia atribuída nos campos de assunto, permite a identificação de erros que podem por em causa a consistência da indexação ao nível do projeto do catálogo coletivo.

No capítulo 7 é proposta uma metodologia de apoio à indexação para as bibliotecas que fizeram parte do estudo de caso. O modelo que se propõe define uma política de indexação e apresenta um fluxograma com as atividades a desenvolver nas fases do processo. Este modelo é apoiado por tabelas abreviadas do sistema SIPORbase, com possibilidade de consulta interativa, apresentadas em anexo.

1. Indexação e acesso por assuntos nos catálogos bibliográficos

São vários os aspetos que estão ligados à recuperação da informação nos catálogos. Uma das formas de otimizar a precisão dos resultados de pesquisa é recorrer ao uso de linguagens controladas, e a indexação desempenha um papel fundamental para conseguir tanto a pertinência como a exaustividade desejadas em toda a pesquisa documental (Moscoso e Extremeño 1999). O acesso por assuntos nos catálogos bibliográficos está estreitamente relacionado ao processo de indexação de documentos. Na abordagem da prática profissional é fundamental conhecer as características desse processo tanto nas bases essenciais da perspectiva tradicional, quanto na visão das tendências atuais e futuras.

Neste capítulo são apresentadas as questões que sustentam essa abordagem, com ênfase especial na utilização da linguagem controlada, na sua normalização e controlo de autoridades nos catálogos. Numa sequência que nos parece lógica, são apresentadas no final possíveis linhas de avaliação, que permitem avaliar a qualidade da indexação e detetar eventuais falhas no desenvolvimento do processo.

1.1 O processo de indexação: definição e finalidade

A indexação de documentos consiste na operação destinada a representar, pelos elementos de uma linguagem documental ou natural, os resultados da análise do conteúdo de um documento, com o fim de facilitar a pesquisa e a recuperação da informação (NP 4285-3)¹. O objetivo deste processo é dar uma indicação concisa do conteúdo dos documentos, e permitir aos usuários recuperarem informação que corresponda às suas necessidades.

Na indexação efetuada por profissionais, essa operação implica a identificação do conteúdo dos documentos através de uma análise cuidada, e a seleção de termos apropriados para representar esse conteúdo, de forma a ser recuperado eficazmente numa pesquisa. Nesta perspectiva, o apoio normativo é fundamental, para que o processo seja desenvolvido de forma orientada e coerente.

Num sistema de recuperação de informação, as dificuldades envolvidas no processo de indexação são frequentes e afetam muitos profissionais e usuários. As dificuldades surgem para os indexadores no *input* de informação no sistema, quanto à compreensão dos assuntos a tratar, e refletem-se nos usuários no *output*, ao expressarem as suas necessidades de pesquisa. Para os usuários que fazem pesquisas, o sistema de informação deveria reduzir a probabilidade de erro tanto quanto possível, e o trabalho

¹ NP 4285-3:2000 – Informação e documentação. Vocabulário. Parte 3: Aquisição, identificação e análise de documentos e dados.

dos indexadores é um fator que pode atuar positivamente nesse sentido, quanto mais correto for esse trabalho, maior será a eficácia na recuperação de informação (Foskett 1997; Boccato e Fujita 2010).

O processo de indexação obedece a uma metodologia de desenvolvimento em três fases essenciais, das quais observamos alguns pontos que ajudam a refletir sobre a importância de cada uma delas.

1.1.1 Fases de desenvolvimento do processo

As três fases que devem ser consideradas no desenvolvimento do processo de indexação são, de acordo com a norma² NP 3715:

A análise do documento e definição do seu conteúdo;

A identificação e seleção dos conceitos que representam esse conteúdo;

A representação dos conceitos por termos de indexação.

Cada uma das fases é descrita tecnicamente na norma, no entanto, consideramos importante fazer algumas reflexões de carácter mais teórico.

² Norma Portuguesa NP 3715-1989, equivalente à ISO 5963-1985 (*Documentation – Methods for examining documents, determining their subjects, and selecting indexing terms*).

Para que as necessidades de informação dos usuários correspondam aos conteúdos dos documentos, é necessário que tenha sido feita uma análise e identificação de conceitos adequadas. A representação desses conceitos através de uma linguagem documental é outro ponto que deve ser assegurado na indexação. Uma indexação correta, logo a partir da fase da análise, conjugada com as possibilidades de utilização de estruturas semânticas existentes nos vocabulários controlados, são a garantia de maior consistência na pesquisa e melhores resultados na recuperação da informação (Bertrand, Cellier, Giroux 1996; Hjørland 1997).

Observando a fase da análise de uma forma objetiva, podemos dizer que se trata de uma análise de conteúdo³, a qual constitui um processo intelectual de analisar assuntos de um documento para subsequente expressão em dados que o representam. Neste contexto (Hjørland 1997) refere que a análise é levada a cabo dentro de um determinado contexto pragmático, quer seja profissional, cultural, histórico, etc., e que não há categorias permanentes ou rígidas de conhecimento, sendo o assunto de um documento o seu potencial informativo, “que implica uma interpretação do potencial do documento em relação aos interesses do conhecimento de um dado sistema de informação”.

A análise de assunto pode servir fins científicos ou académicos, profissionais, educacionais, ou objetivos comuns do dia-a-dia. O autor dá um exemplo simples desta conceção, que “um agricultor e um zoologista não classificam os animais da mesma maneira”. A análise pode ser mais específica e orientada para objetivos ou pode ser mais geral ou abstrata, como no caso de um serviço de informação geral, que tende a

³ Análise de conteúdo: “operação destinada a extrair, sob forma concisa e precisa, dados que caracterizem o conteúdo de um documento” (NP 4285-3).

usar esse tipo de descrição. Embora uma análise mais específica seja mais amigável, de relevância perceptível, uma classificação mais geral e básica será relevante para uma comunidade mais alargada, e conseqüentemente mais estável. Do resultado de uma análise podem-se selecionar termos de uso popular ou termos científicos, dependendo do tipo de usuário a quem se destinem (Montalvo 2011). Tendo em conta o potencial usuário devem ser vistas estas possibilidades, de generalizar ou especificar conceitos de forma equilibrada.

Na fase de identificação de conceitos, de acordo com Hjørland (1997), um documento não tem só um assunto verdadeiramente válido, tem várias “potencialidades epistemológicas” a que são dadas prioridades baseadas em pontos de vista disciplinares. Na sua “concepção pragmática de assunto” a representação de assunto tem esse objetivo prático: um usuário tem um problema para resolver, e uma necessidade particular de informação; essa informação é pesquisada em bibliotecas, nas bases de dados em que os documentos estão indexados por assunto. É nesse sentido que a indexação deve antecipar as necessidades do usuário para o qual a informação por assuntos tem uma função prática. Trata-se de uma relação instrumental entre um documento e uma necessidade do usuário. Nas palavras do autor:

“What should be chosen as the subject of the book? The answer is that different epistemic communities have different needs: an information service should analyze the subject according to the potential questions of its primary user group.”

Ligada a esta concepção é feita a distinção entre indexação orientada para o conteúdo, que é a descrição de assuntos concebida puramente como uma função dos atributos do documento, e a indexação orientada para o usuário ou necessidade, que é uma descrição

de assunto visto como a relação entre as propriedades do documento e uma antecipada necessidade do usuário (Soergel 1985). É nesta perspectiva que um documento poderá ter interesse para vários perfis de usuário, e é possível fazer essa ligação através da atribuição de termos adequados.

Quanto aos conteúdos, o indexador deve identificar os conceitos que os representam, não tendo de considerar necessariamente todos os conceitos identificados, mas escolher os que considere mais adequados ao fim para o qual vão ser utilizados, como por exemplo, a indexação de livros ou de artigos de periódicos. Neste contexto a exaustividade e a especificidade são as características da indexação mais suscetíveis de serem modificadas. A exaustividade é a característica ligada ao número de conceitos que foram considerados na indexação do documento, e a especificidade é a característica associada à exatidão com que um determinado conceito é representado por um termo de indexação. Estas características presentes no processo de indexação são indicadas por Foskett (1996) como os dois fatores que mais afetam a performance na recuperação de informação.

Um alto nível de exaustividade e especificidade, no conjunto constituem a profundidade⁴ da indexação. Quando um determinado conceito se encontra representado por um termo mais geral, há perda de especificidade, o que afeta a definição desse conceito. Nalguns casos pode ser mais adequado adotar um termo de indexação mais geral, como por exemplo no caso em que se considere que a utilização de um termo específico possa ter um efeito negativo no funcionamento do sistema. No entanto, em

⁴ Profundidade de indexação: grau de pormenor no qual um assunto pode ser representado num processo de indexação (NP 4285-3).

termos gerais, a falta de especificidade que pode decorrer também por erros de análise, dá origem a perda de informação pertinente (Wellisch 1994; Gavilán 2009).

É recomendável que haja um equilíbrio entre a exaustividade e a especificidade no processo de indexação, principalmente para o tratamento temático de livros (Boccatto e Fujita 2010). As autoras dão como exemplo:

“... um livro que trata especificamente de *Amálgama, Resina composta e Porcelana dentária*, e o seu tratamento temático reconheça o uso da exaustividade, ele vai contemplar também o termo genérico *Materiais dentários*. Nesse caso, se a necessidade de informação do usuário for referente ao assunto *Materiais dentários* de uma maneira geral, o documento será recuperado. Se ele precisar especificamente de um dos três materiais dentários, o documento também será recuperado.”

De um ponto de vista funcional, a necessidade da utilização de termos genéricos e específicos evitar-se-ia se o sistema de pesquisa possibilitasse de alguma maneira o uso das relações semânticas que há nos vocabulários controlados, ou no catálogo de autoridades, e desse ao usuário a possibilidade de selecionar termos ou de navegar pela estrutura hierárquica (Caro-Castro 2005), questões que abordamos mais adiante.

A última fase do processo de indexação consiste na representação dos conceitos através de uma linguagem documental, que deverá permitir estabelecer os pontos de acesso na pesquisa. Nesta fase trata-se de converter os conceitos identificados em termos de indexação, e a opção na utilização de uma linguagem documental terá implicações na forma como são expressos os conceitos, e como serão recuperados ao nível da pesquisa (Ribeiro 1996). Também na fase de representação de assuntos, se coloca a questão dos contextos envolventes, que podem ser de carácter cultural,

acadêmico, científico, histórico, etc., e é no enquadramento dessas perspectivas que o processo deve ser desenvolvido. Nesta última fase, deve igualmente ser identificado o potencial informativo dos assuntos em relação aos interesses de conhecimento do sistema de informação em que se aplica.

Em todo o processo de indexação, um dos fatores que mais afeta o seu desenvolvimento, mesmo partindo de uma prática normalizada, é a subjetividade por parte dos indexadores. A subjetividade é uma característica inerente à indexação e manifesta-se na sua consistência, que significa o grau de similitude entre os termos (cabeçalhos de assunto ou descritores) atribuídos ao mesmo documento por indexadores diferentes, ou pelo mesmo indexador em momentos diferentes. A consistência representa a coerência no tratamento de diferentes documentos em relação a uma mesma matéria⁵, e o que mais pode influenciar o grau de consistência são características como a formação dos indexadores, a sua experiência e outras características pessoais, a utilização de ferramentas de indexação e de linguagens documentais. As características dos indexadores estão ligadas à capacidade de abstração conceptual, aos graus de conhecimento sobre os assuntos a tratar e atualização das matérias, assim como a visões culturais e outros condicionamentos pessoais. A sua formação, experiência, e domínio das ferramentas são fatores que logo na fase da análise documental podem exercer influência na consistência da tarefa (Soler Monreal 2009; Méndez Rodríguez e Moreira González 1999; Rodríguez Yunta 1998).

⁵ Neste trabalho a expressão 'matéria' e 'assunto' pode ser usada com o mesmo significado.

A subjetividade faz parte do processo de indexação e não é possível eliminá-la. Contudo a observação de diretrizes que ajudem a estabelecer procedimentos e orientem numa abordagem mais objetiva é fundamental para um processo eficaz, e é também a qualidade da linguagem utilizada na fase da representação de conceitos que garante uma pesquisa consistente.

As linguagens alfabéticas estruturadas constituem sistemas de organização do conhecimento que correspondem às listas de cabeçalhos de assuntos e tesouros, construídos a partir de termos da linguagem especializada e da linguagem natural. A sua função é, por um lado, representar o conteúdo dos documentos num sistema de informação, e por outro, ajudar à sua recuperação por parte do usuário através da formulação de pesquisas, “função pelo conteúdo” e “função pelo uso” (Boccatto 2009). Neste contexto pode ainda falar-se da função de mediação, entre a linguagem usada pelo autor do documento e pelo usuário do sistema. O papel do profissional, ao realizar a atividade de indexação, deve ser o de estabelecer “pontes” entre a solicitação do usuário e a informação contida nos documentos. A utilização da linguagem controlada serve, pois, de apoio à prática da indexação na fase de representação de conceitos, e os termos dessa linguagem devem ser retirados de léxicos de assuntos ou de tesouros, uma vez que aí se encontram controlados. Estes são ferramentas de suporte essenciais para organizar e disseminar informação nos sistemas automatizados e o seu uso apropriado é fundamental, sobretudo em áreas científicas especializadas que exigem um grande controlo da terminologia, atribuindo qualidade ao processo da indexação e eficácia à recuperação da informação por parte dos usuários (Boccatto 2009; Boccatto e Fujita, 2011).

Os termos que resultam da fase do tratamento temático são visualizados por quem localiza e recupera a informação nos catálogos, pelo que a escolha de termos de uma linguagem documental deve ser condizente às necessidades dos que pesquisam, e o grau de especificidade dessa linguagem deve ser adequado ao perfil da comunidade de usuários. Essa linguagem serve, simultaneamente, o indexador na escolha dos termos a atribuir aos documentos, e o usuário no momento da pesquisa, que mesmo utilizando a linguagem natural, acede à informação através das relações semânticas estabelecidas na linguagem documental.

As linguagens “devem ser construídas com base na terminologia de uma área, com categorias e subcategorias delimitadas, a partir de relações lógico-semânticas de termos claros e bem definidos”, que permitam a representação adequada de conceitos. Neste contexto, a utilização da linguagem documental tem um papel fundamental e quando não se adequa aos conteúdos dos documentos, afeta tanto o trabalho do indexador, quanto as necessidades dos usuários que pesquisam (Boccatto e Fujita 2006; Boccatto 2009; Boccatto e Fujita 2011).

No processo de pesquisa por assuntos, a característica mais importante da linguagem de recuperação de informação é a capacidade de expressar os resultados da análise dos documentos. Um assunto pode ser expresso através de várias linguagens documentais como esquemas de classificação, ou termos controlados ou não controlados. Estas linguagens têm tipologias e características diferentes, com as quais os profissionais devem estar familiarizados (Hjørland 1997). Neste sentido, e face à importância da sua utilização, observamos de seguida as características gerais das linguagens de indexação.

1.2 Tipologia e características das linguagens de indexação

Linguagens de indexação⁶ constituem os instrumentos utilizados para representar a informação, de forma a recuperá-la posteriormente nos sistemas de pesquisa. As linguagens de indexação controladas são linguagens artificiais que foram construídas de acordo com regras explícitas cujo objetivo é possibilitar a recuperação de informações temáticas, reduzindo a diversidade da terminologia. As linguagens controladas mais usualmente conhecidas são os sistemas de classificação, as listas de cabeçalhos de assunto e os tesouros (Broughton 2006; Moreira 2005; Simões 2008).

Na organização do conhecimento (*Knowledge Organization*) a divisão considerada “tradicional” é entre sistemas de classificação e linguagens de indexação verbal (Hjørland 2012). De acordo com o esquema do autor (Fig.1.1) todos os tipos de linguagens de indexação representados (com a possível exceção dos sistemas de texto livre) são diferentes géneros de “sistemas de classificação”, na medida em que a ideia central é ‘agrupar por classes’, e que a diferença mais marcante entre as linguagens é, na realidade, entre os “sistemas de texto livre” e todos os outros, que são sistemas controlados.

⁶ Designações de ‘linguagem documental’ e ‘linguagem de indexação’ aparecem como sinónimos; a norma portuguesa NP 4285-4 define “linguagem documental” como a “linguagem formal utilizada para caracterizar os dados ou o conteúdo de documentos e permitir o seu armazenamento e recuperação”; a norma portuguesa NP 4036 define “linguagem de indexação” como o “conjunto controlado de termos escolhidos numa linguagem natural e utilizados para representar, sob forma coordenada, o conteúdo dos documentos.”

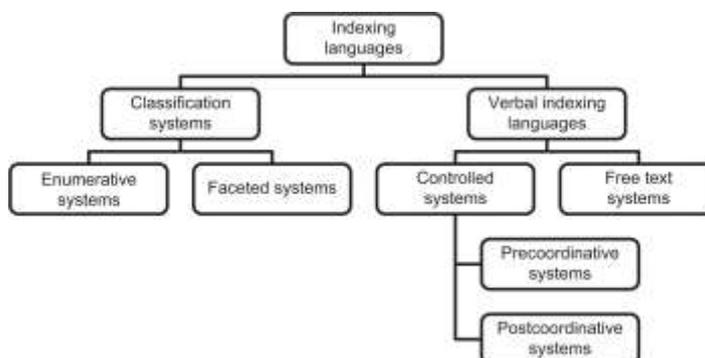


Fig. 1.1: Visão tradicional dos diferentes tipos de linguagens de indexação

Numa visão abrangente do esquema de Hjørland, diremos que a característica dominante das linguagens de indexação é a de serem linguagens controladas, pois só a linguagem verbal dos sistemas de ‘texto livre’ não é controlada; nestes é usada a linguagem natural em vez de termos selecionados de uma lista autorizada de cabeçalhos de assunto ou de descritores. Esta diferença define os dois grupos básicos de sistemas que permitem a recuperação de informação por assunto: os de linguagem natural que indexam através de palavras contidas no título ou resumo do documento, e os de linguagem controlada que indexam através da atribuição de termos prescritos de vocabulários estruturados (IFLA 2011).

De acordo com a sua tipologia, de sistemas para classificar ou para indexar, as linguagens controladas consistem num vocabulário (códigos ou termos para indexar ou representar classes), incluem usualmente uma estrutura de relações semânticas para ligar conceitos, e adotam regras para combinação de termos ou classes (sintaxe) (Broughton 2006).

Algumas das características próprias dos sistemas de classificação e dos sistemas de indexação verbal, são as seguintes:

- Os sistemas de classificação - usam a linguagem de tipo classificatório⁷, que se exprime por notações (símbolos de uma classificação, formados por números, letras ou combinações) e eventualmente expressões verbais, numa organização sistemática de conceitos e matérias mediante uma estrutura mono-hierárquica ou facetada. Os sistemas de classificação de tipo enumerativo⁸, listam as classes específicas de assuntos, e os de tipo facetado combinam classes, através da identificação de características comuns, organizando-se em facetas com classificação hierárquica dos termos no interior das facetas, e com uma sintaxe para combinar as notações das diversas facetas do código de classificação (ordenação sistemática por notação, e dentro desta, ordenação hierárquica por tópicos). Os sistemas de classificação mais conhecidos são a *Dewey Decimal Classification*⁹ (DDC), a *Universal Decimal Classification*¹⁰ (UDC), e a *Library of Congress*

⁷ Linguagem 'classificatória' e linguagem 'codificada' surgem como designações equivalentes.

⁸ Sistema de classificação enumerativo: sistema em que cada classe figura como entrada principal (NP 4285-4:2000).

⁹ Classificação de Dewey, <https://www.oclc.org/dewey.en.html> [Consulta: 25-04-2013]

¹⁰ Classificação Decimal Universal, <http://www.udcc.org/index.php/site/> [Consulta: 25-04-2013]

Classification (LCC)¹¹. Estas classificações são exemplos de sistemas enumerativos, embora a DDC e em especial a UDC tenham alguns elementos de classificações facetadas (Broughton 2006; Hjørland 2007; Svenonius 2000).

- As linguagens de indexação verbal - usam palavras em vez de códigos ou notações, num conjunto de termos que derivam da linguagem natural, com controlo do vocabulário no que respeita à sinonímia e à polissemia para reduzir a diversidade na terminologia e possibilitar a recuperação temática de forma consistente. Uma das maiores vantagens dos vocabulários controlados é a capacidade de síntese e de reunião de termos sinónimos (ou quasi-sinónimos)¹² num cabeçalho unívoco, ao qual se remetem os termos não autorizados. Este controlo é feito através de uma estrutura de relações, que serve para apoiar o usuário na pesquisa. Graças aos vocabulários controlados podem-se agrupar diversos recursos informativos em relação aos mesmos conceitos, reduzindo a dispersão de opções de pesquisa em catálogos e índices (Caro-Castro 2005; Fidel 1991b; Montalvo 2011; Svenonius 2003).

Nas linguagens de indexação verbal controladas, conforme o esquema de Hjørland (fig.1.1), aparece a divisão entre sistemas pré-coordenados e sistemas pós-coordenados. Trata-se de dois grandes grupos, os que usam termos pré-coordenados e os que dependem da pós-coordenação na pesquisa. Representam basicamente duas abordagens diferentes no processamento e recuperação da informação, que tiveram

¹¹ Classificação da Biblioteca do Congresso, <http://www.loc.gov/catdir/cpsolcc.html> [Consulta: 26-04-2013]

¹² Termo de forma diferente de outro, mas de significado afim, que pode ser tratado como sinónimo numa linguagem documental (NP 4285-4).

desenvolvimentos próprios, assim como as ferramentas que se aplicam a estas abordagens, os tesouros e os cabeçalhos de assunto (Miler e Teitelbaum 2002).

Estes dois tipos de sistemas, os tesouros e as listas de cabeçalhos de assunto, dispõem de uma estrutura semântica (relações hierárquicas e associativas) e de uma sintaxe (regras de combinação dos termos). A sintaxe nos tesouros é mais simples, limitando-se a algumas normas sobre os termos compostos, pois a combinação de descritores é feita no momento da pesquisa (pós-coordenada). Já no que respeita à estrutura e semântica os tesouros são os mais complexos (Caro-Castro 2005).

Os sistemas de indexação, para além do controlo do vocabulário ao nível da morfologia, formalizam também a sintaxe flexível da linguagem natural. Os cabeçalhos podem expressar a mesma ideia em mais do que uma maneira, (Psicologia da educação, ou Psicologia educacional), e daí a necessidade de formalizar a sintaxe da linguagem natural que permita só certas construções na criação de uma linguagem estruturada (Foskett 1997). Na construção dos cabeçalhos fala-se de sintaxe em relação à combinação das subdivisões que constituem um cabeçalho composto.

- Os tesouros são o tipo de linguagem de indexação verbal com a característica de identificar conceitos simples que mostram relações semânticas entre eles, numa apresentação alfabética. São instrumentos que têm servido principalmente para indexar materiais em domínios temáticos específicos, com o seu vocabulário próprio, como é o caso das bases de dados *online* produzidas por agências especializadas. São

sistemas desenhados para a indexação pós-coordenada¹³ em que o assunto de um documento é analisado por um indexador que atribui termos autorizados, os quais só são combinados na fase da pesquisa ao serem selecionados pelo usuário. Assim, para expressar um assunto complexo, a combinação de termos simples é feita na fase da pesquisa e não na fase da indexação (Broughton 2006; Foskett 1997; Hjørland 2007; IFLA 2011).

O objetivo tradicional de um tesouro é guiar o indexador e o usuário para escolherem o mesmo termo para o mesmo conceito. Para que isto seja possível, um tesouro deve listar todos os conceitos que possam ser úteis para recuperar informação num determinado domínio. Os conceitos são representados por termos, e para cada conceito, uma das possíveis representações é selecionada como ‘termo preferido’. Os termos preferidos são representados de forma a serem facilmente identificados, através das relações estabelecidas entre os termos (e/ou entre conceitos), numa apresentação estruturada (ISO/DIS 25964-1).

O controlo do vocabulário nos tesouros é feito com base em dois princípios fundamentais (IFLA 2011):

- A estrutura do tesouro, na sua apresentação de relações hierárquicas, indica o significado preciso de um termo, e no caso de não ser suficientemente explícito, é acompanhado de uma nota explicativa que indica o seu significado, podendo ainda

¹³ Indexação sem ordem pré-estabelecida de termos ou de índices (NP 4285-3).

indicar outros significados reconhecidos na linguagem natural, mas que foram deliberadamente excluídos para efeitos de indexação;

- Quando o mesmo conceito pode ser expresso por dois ou mais sinónimos ou quasi-sinónimos, um dos termos é selecionado como termo preferido, que deve ser usado consistentemente na indexação. Deve ser feita uma referência para o termo preferido a partir de qualquer sinónimo usado como ponto de acesso para o usuário.

- As listas de cabeçalhos de assunto são um tipo de vocabulário controlado usado para representar de forma sintetizada os assuntos contidos em documentos. Os cabeçalhos podem ser formados por um só elemento, ou por vários elementos encadeados (cabeçalho – subcabeçalho). Cada um destes elementos representa um conceito que pode expressar-se mediante uma ou várias palavras. Neste caso a sintaxe determina tanto a composição de cada um dos elementos, como a ordem e as ligações que permitem criar uma cadeia com elementos de diferentes categorias. Uma das vantagens reconhecidas neste tipo de linguagem é justamente a capacidade de expressar conceitos complexos mediante a construção de cadeias que constituem cabeçalhos previamente elaborados para descrever temas complexos (Caro-Castro 2005; Svenonius 2000).

Os cabeçalhos de assunto podem ser usados para indexar (cabeçalhos autorizados) ou não, no caso dos cabeçalhos não admitidos. Estes são comparáveis, respetivamente, aos termos preferidos e aos termos não preferidos num tesouro (ISO/DIS 25964-2). No contexto dos catálogos automatizados há maior idoneidade no uso de descritores em

relação a cabeçalhos de assunto. Houve uma progressiva “tesaurização” das listas de cabeçalhos de assuntos com o aparecimento das bases de dados, embora muitas bases de ciências sociais continuem a utilizar cabeçalhos de assuntos. Problemas sobre o uso de cabeçalhos para a recuperação de informação por assuntos foram demonstrados (Hunter 1991; Larson 1991), pois foram pensados para que o usuário os reconhecesse numa pesquisa alfabética lineal, mas não para que os formulasse, pois a sua linguagem não é a linguagem corrente; os subcabeçalhos nem sempre ajudam, e tanto a ordem da cadeia como as inversões carecem de sentido para a lógica de um usuário não profissional. Assim, os cabeçalhos de assuntos “pensados para um entorno manual em que a pesquisa se baseava no reconhecimento, não funciona num entorno automatizado, onde a pesquisa se baseia fundamentalmente na formulação”. Conclui-se que o uso de descritores é mais adequado na indexação das fontes eletrônicas, já que a sua relevância é maior e por conseguinte, garantem uma exaustividade maior nos resultados das pesquisas (Moscoso e Extremerño 1999).

Os cabeçalhos de assunto têm a sua origem nos catálogos de bibliotecas, e são projetados para a indexação pré-coordenada das suas coleções, usualmente de âmbito universal. A pré-coordenação é usualmente usada para combinar conceitos individuais em assuntos compostos ou complexos na fase da indexação, construindo-se cabeçalhos mais específicos mediante a atribuição de subdivisões. Ao contrário dos tesouros, os cabeçalhos de assunto incluem muitos componentes pré-combinados de termos. Num sistema pré-coordenado¹⁴ os termos autorizados atribuídos a um documento são combinados sintaticamente numa ou mais sequências representando as combinações

¹⁴ Indexação de documentos em que a ordem dos termos ou índices está estabelecida segundo uma linguagem documental ou um sistema de indexação (NP 4285-3).

disponíveis para fins de recuperação (Broughton 2006; IFLA 2011; Montalvo 2011). O exemplo de referência de uma lista de cabeçalhos de assunto, que tem sido adaptada em muitos países é a LCSH (*Library of Congress Subject Headings*).¹⁵

Pode-se dizer que estes sistemas partilham algumas características dos tesouros e dos esquemas de classificação. Como um tesouro, representam conceitos na forma de termos ou frases, e como um esquema de classificação sintético, têm regras sintáticas para combinação de termos em cadeias pré-coordenadas que representam conceitos e assuntos mais complexos. De facto a distinção entre sistemas de tesouros e cabeçalhos de assunto tem vindo a tornar-se um pouco confusa, muitos sistemas de cabeçalhos incluíram algumas características estruturais dos tesouros, como por exemplo as referências de *broader terms* (BT), *narrower terms* (NT), e *related terms* (RT). Os tesouros são atualmente muito usados em catálogos *online* e em bibliografias nacionais, pois é possível aplicar os seus princípios básicos a qualquer linguagem de indexação (IFLA 2011).

Na atualidade um sistema de indexação de assuntos pode abranger aspetos de ambos os tipos relacionados um com o outro, ou seja, incluir uma lista de cabeçalhos de assunto ou um tesouro para controlo do vocabulário e das relações semânticas, e regras sintáticas para a construção de cabeçalhos. É o caso do “*Nuovo Soggettario*”¹⁶, adotado pela Biblioteca Nacional de Florença (*Biblioteca Nazionale Centrale Firenze*), sobre o qual falamos adiante no capítulo 1.2.2.

¹⁵ <http://id.loc.gov/authorities/subjects.html> [Consulta: 02-03-2013].

¹⁶ http://thes.bncf.firenze.sbn.it/index_eng.html [Consulta: 02.05.2013].

Centramo-nos de seguida nas características dos sistemas de indexação verbal, que constituem o foco de interesse no nosso estudo.

1.2.1 Semântica e sintaxe dos vocabulários controlados

Nos vocabulários controlados é feita a normalização dos termos para limitar o seu significado, e essa regulação afeta a morfologia¹⁷, as relações semânticas e a sintaxe. O objetivo é evitar a polissemia, a homonímia e a homografia na linguagem natural, e conseguir uma linguagem em que cada termo represente um termo unívoco. O nível mais elementar de normalização é o que determina a categoria gramatical e a forma dos descritores, e o nível mais complexo é o que se refere à sintaxe (Caro-Castro 2005). Assim, na pesquisa de informação estão envolvidos dois tipos de relações (Foskett 1997):

- ✓ As relações semânticas, que partem da necessidade de pesquisar termos adicionais ou alternativos a partir do termo inicial de pesquisa;
- ✓ As relações sintáticas, que partem da necessidade de pesquisar a interseção de duas ou mais classes de diferentes conceitos.

¹⁷ Parte da gramática que trata da forma e dos processos de formação das palavras (Dicionário Priberam).

➤ As relações semânticas

Os termos de uma linguagem de indexação são relacionados por um conjunto de ligações, usualmente chamadas de “referências cruzadas” e que servem para chamar a atenção do usuário para as relações semânticas. A inclusão de uma rede de relações semânticas num sistema de indexação permite melhorar as formulações de indexação e pesquisa, ao sugerirem termos substitutos ou adicionais que se possam usar. Ao estabelecer essas relações na fase da indexação, estamos a tentar prever formas com as quais o usuário poderá encontrar um documento; quando as usamos na fase da pesquisa, tentamos atingir a melhor combinação possível entre a nossa formulação de pesquisa e os termos usados pelo indexador ou pelo autor. No fundo, as relações semânticas são usadas para permitir aos indexadores e usuários chegar a uma seleção de termos ótima, a usar numa determinada situação (Foskett 1997).

As relações entre conceitos, que constituem a estrutura semântica são as seguintes:

- ✓ Relações de equivalência – as relações deste tipo existem para garantir que cada conceito seja representado por um só descritor. Estas relações implicam a seleção de uma forma como termo preferido, fazendo uma referência cruzada entre este e o seu correspondente ‘não preferido’. Esta relação é aplicada entre termos, não entre conceitos, e a relação é recíproca, ou seja, é preciso mostrar o inverso dessas direções (tab.1.1). As relações são apresentadas através das expressões convencionadas:

USE, como prefixo do termo preferido

UF (*used for*), como prefixo do termo não preferido

Uma relação de equivalência é estabelecida nas seguintes situações (IFLA 2011):

- os termos são sinónimos
 - os termos são quasi-sinónimos
 - a especificidade de um termo é vista como desnecessária e é representado por outro termo de âmbito mais alargado
 - a especificidade de um termo é vista como desnecessária e é representado pela combinação de dois ou mais termos (equivalência composta)
- ✓ Relações hierárquicas – a principal função das relações hierárquicas é a de ajudar o indexador e o usuário a escolher o grau de especificidade apropriado às suas necessidades. Uma pesquisa pode ser alargada ou refinada através da escolha de termos nesta hierarquia. A relação hierárquica é baseada em graus ou níveis de superordenação e subordinação (tab.1.1), em que o conceito superordenado representa uma classe, e os conceitos subordinados referem os seus membros ou partes (IFLA 2011). As ‘etiquetas’ que devem ser usadas reciprocamente são:

BT (*broader term*) usado como prefixo do termo superordenado, que significa ‘termo genérico’,

NT (*narrower term*) usado como prefixo do termo subordinado, que significa ‘termo específico’

- ✓ Relações associativas – estas relações servem para associar conceitos que não estão relacionados hierarquicamente, mas estão associados semântica ou

conceptualmente, e devem portanto ter uma ligação explícita no tesauro, servindo como sugestão adicional ou alternativa de termos para indexação ou para pesquisa. A ‘etiqueta’ aplicada é RT (*related term*) que significa termo relacionado, e deve ser usada reciprocamente (tab.1.1).

Equivalência	Hierárquica	Associativa
artes plásticas USE belas-artes	artes NT arte popular	artes RT criação artística
belas-artes UF artes gráficas	arte popular BT artes	criação artística RT artes

Tabela 1.1: Exemplos da reciprocidade das relações semânticas

Enquanto as relações de equivalência são essencialmente estabelecidas num sentido, levando do termo não preferido para o autorizado, nas outras duas categorias de relações (hierárquicas e associativas) são necessárias referências cruzadas nos dois sentidos, ou seja, as relações são recíprocas, cada BT tem um correspondente NT, e vice-versa (tab.1.1).

Tal como nos tesauros, para estruturar os esquemas de cabeçalhos de assunto são usados os dois tipos de relações semânticas, entre termos (equivalência) e entre conceitos (hierárquicas e associativas). Inicialmente a forma como estas relações

estavam incorporadas nas listas de cabeçalhos de assunto foi criticada por ter algumas deficiências (como a estrutura hierárquica imperfeita, ou as relações associativas não diferenciadas das hierárquicas), que se foram normalizando mediante critérios similares aos da elaboração de tesouros. O sistema de reenvios foi melhorado através da conversão de referências anteriormente usadas, em siglas que identificam os termos (Caro-Castro 2005; Foskett 1997; Olson e Boll 2001; Svenonius 2000).

As relações nas listas de cabeçalhos de assunto podem, assim, aparecer na forma de ‘referência cruzada’ ou de ‘etiqueta’¹⁸:

- A relação de equivalência, indicada por referências cruzadas como *See, See from*, ou pelas etiquetas *USE, UF (used for)*
- A relação hierárquica, indicada pelas etiquetas *BT (broader term)*, *NT (narrower term)*, *RT (related term)*
- A relação associativa, indicada por referências cruzadas como *See also* ou pela etiqueta *RT (related term)*

Assunto	
1.	Psicologia infantil
2.	• Ver também: Crianças -- Psicologia
3.	Psicologia marxista
4.	Psicologia masculina
5.	• Ver: Homens -- Psicologia
6.	Psicologia organizacional
7.	• Ver: Psicologia das organizações

Fig. 1.2: Exemplo de referências cruzadas Ver e Ver também num índice de assuntos

¹⁸ Designações respetivamente correspondentes às de língua portuguesa Ver, Ver de, USE, UP (usado por), TG (termo genérico), TE (termo específico), Ver também, TR (termo relacionado).

1.2.1.1 Sintaxe dos cabeçalhos de assunto

As relações sintáticas têm o propósito de permitir melhorar a relevância na pesquisa, através da possibilidade de especificar assuntos com mais precisão; embora isso se consiga através das relações semânticas ao indicarem termos mais específicos, o método mais poderoso de aumentar a especificidade é através da coordenação de termos. Como já se falou anteriormente, nos sistemas pós-coordenados os termos são coordenados na altura da pesquisa através do uso do operador booleano ‘AND’¹⁹, podendo contudo conduzir a falhas na recuperação da informação. No entanto, num sistema de indexação em que os termos foram previamente coordenados, a pesquisa através desses termos, garante que a informação corresponde a essa expressão, (ex. Jovens -- Atitudes), e que a relevância na recuperação da informação é certamente diferente do que seria através da combinação dos termos de forma pós-coordenada. Quando se quer representar assuntos que consistem em mais do que um conceito usa-se a pré-coordenação de termos de uma linguagem de indexação em cabeçalhos de assunto (Foskett 1997).

Desde o surgimento dos sistemas pós-coordenados com o conceito de ‘unitermos’²⁰, que se constatou a conveniência de estabelecer uma certa pré-coordenação nos termos autorizados com vista a evitar as falsas coordenações. Por outro lado o progresso de diversos tesouros foi aproveitado convenientemente no aperfeiçoamento dos cabeçalhos de assunto usados nas bibliotecas. A introdução da automatização em quase todos os

¹⁹ Este operador booleano é o que permite combinar conceitos no mesmo sentido que a pré-coordenação.

²⁰ O mais pequeno elemento significativo de uma linguagem documental utilizado para representar um conceito específico num sistema de indexação coordenada; não é utilizado como sinónimo de palavra-chave ou descritor (NP 4285).

sistemas de recuperação de informação modificou o comportamento dos sistemas ‘puramente’ pré-coordenados, e foi produzindo gradualmente uma aproximação entre sistemas pós-coordenados e pré-coordenados (Bermello Crespo 2001).

Em relação à sintaxe os cabeçalhos têm as seguintes características (ISO/DIS 25964-2):

- ✓ Podem consistir numa ou mais palavras representando um só conceito, (ex. Aritmética / Geometria descritiva) ou de várias palavras representando a combinação de diferentes conceitos (ex. Minas e recursos minerais);
- ✓ Um cabeçalho pode ser simples, consistindo numa só entrada, ou pode ser composto, se forem adicionadas uma ou mais subdivisões ao cabeçalho inicial para formar uma cadeia pré-coordenada. (ex. Psicologia / Psicologia -- Métodos de investigação).

Os cabeçalhos compostos são cadeias pré-coordenadas, criadas através da adição ao cabeçalho inicial, de uma ou mais subdivisões para representar um assunto de forma mais completa. A ligação é feita geralmente com um traço duplo (ex. Automóveis -- Mecânica). A função das subdivisões é a de representar uma perspectiva, característica, forma, etc. sob a qual o tópico expressado pelo cabeçalho é apresentado num recurso de informação.

Os tipos comuns de subdivisões incluem:

- subdivisões do assunto - ex. *Artistas -- Bolsas de estudo*
- subdivisões geográficas (de lugar) - ex. *Artistas -- Porto (Portugal)*

- subdivisões cronológicas (de tempo) - ex. *Artistas chilenos -- 1984-1985*
- subdivisões de forma - ex. *Artistas brasileiros -- [Entrevistas]*

Quando mais do que uma subdivisão é adicionada a um cabeçalho, a sua ordem de atribuição deve ser a seguinte: -- Tópico -- Geográfica -- Cronológica -- Forma.

Ex. – *Artistas -- Condições de trabalho --1990 -- [Ensaio]*

Label	Vocabulary	Type
1. School sports--Fiction	LC Subject Headings	Complex Subject
2. Private flying--Accidents--United States--Periodicals	LC Subject Headings	Complex Subject
3. Costa Rica--History--1986-	LC Subject Headings	Complex Subject
4. Zimbabwe--Economic policy	LC Subject Headings	Complex Subject
5. Reinstatement	LC Subject Headings	Topic

Fig. 1.3: Cabeçalhos simples e complexos, com subdivisões (lista LCSH)

No caso da sintaxe de cabeçalhos, que através de subdivisões, permite especificar o assunto e contextualizar termos ambíguos, contribuindo para a representação da informação, embora envolvam alguma complexidade, não constituem um obstáculo para a recuperação em sistemas automatizados, pois qualquer termo é recuperado, seja qual for o seu lugar no cabeçalho (Caro-Castro 2005). Os cabeçalhos podem ser genéricos e abarcar vários aspetos de um tema, reduzindo as opções que terá o usuário na hora de pesquisar informação, mas também podem ser específicos, com o propósito

de localizar subtemas, aumentando tanto as opções de pesquisa, como as entradas da lista de cabeçalhos de assunto desenvolvida (Montalvo 2011).

1.2.2 Sistemas de indexação de referência

Uma linguagem de indexação verbal controlada de referência é a lista de cabeçalhos de assunto LCSH²¹ (*Library of Congress Subject Headings*), muito usada no Ocidente tanto na forma original, como em numerosas adaptações. Um estudo da IFLA (1994) com o objetivo de conhecer os cabeçalhos nos catálogos das bibliotecas nacionais, demonstrou, já então, um predomínio da utilização da LCSH em inglês ou traduzida (Caro-Castro 2005; Montalvo 2011). A automatização, o progresso das tecnologias de informação e o aparecimento dos OPACs na década de noventa proporcionaram que, à margem da LCSH se desenvolvessem catálogos de autoridade em bibliotecas nacionais, e sistemas tão importantes como RAMEAU (*Répertoire d'autorité matière encyclopedique et alphabetique unifié*) RSWK (*Regeln für den Schlagwortkatalog*) e SIPORbase (Sistema de Indexação em Português) desenvolveram-se durante essa década como um produto da crescente importância da indexação nos serviços nacionais (Caro-Castro 2005; Heiner-Freiling 2000).

A LCSH teve uma grande difusão nos Estados Unidos para o qual contribuiu a existência de um sistema de distribuição dos resultados da catalogação da Biblioteca do

²¹ <http://id.loc.gov/authorities/subjects.html> [Consulta: 14-02-2013]

Congresso; posteriormente as grandes agências de recursos bibliográficos como a OCLC (*Online Computer Library Center*) e a RLIN (*Research Libraries Information Network*), cujas bases de dados cobrem milhões de registos catalográficos, usaram este sistema para a atribuição de cabeçalhos de assunto aos seus registos; com isto implicaram um número cada vez maior de bibliotecas no uso desta linguagem para a representação dos conteúdos temáticos. O alcance das agências mencionadas e as facilidades de acesso remoto proporcionadas pela internet contribuíram para a divulgação da LCSH noutros países (Bermello Crespo 2001).

A estrutura da LCSH é essencialmente pré-coordenada. Quando os usuários tinham que consultar catálogos e índices impressos tinham que entender a lógica desta estrutura para aceder à informação (a ficha do catálogo era incompatível com a pós-coordenação, pelo que a pré-coordenação era a única opção). Muitos bibliotecários acolheram com muito agrado a opção dada pelos catálogos *online* de pesquisar por palavra-chave²², porque permitia usar uma linguagem mais natural e combinar termos de pesquisa de forma pós-coordenada combinando cabeçalhos simples. Apesar das vantagens da pré-coordenação, as pesquisas nesta estrutura exigem que os usuários estejam familiarizados com a sintaxe atribuída pelos catalogadores, de forma a interpretar a ordem hierárquica dos cabeçalhos e as subdivisões correspondentes. Questiona-se, então, se a estrutura pré-coordenada, por muito lógica que seja, responde à conveniência do usuário (Montalvo 2011).

²² Palavra ou grupo de palavras, eventualmente numa forma lexicográfica normalizada, escolhida do título ou do texto de um documento, que caracteriza o seu conteúdo e permite a sua recuperação (NP 4285).

Como já vimos anteriormente, todas as questões relacionadas com a indexação de documentos adquirem novas perspectivas à luz dos atuais contextos, em que proliferam os documentos digitais e as necessidades de pesquisa em rede. A adaptação de vocabulários controlados para indexação dos recursos eletrônicos tornou-se então uma questão emergente.

O enorme volume e o rápido crescimento de recursos disponíveis na *www* assim como a emergência de numerosos esquemas de metadados, provocaram o reexame da forma como os dados de assunto são atribuídos a recursos na *web*. Há um consenso alargado de que um esquema de metadados de assunto deve exibir simplicidade e interoperabilidade. A simplicidade refere-se à sua usabilidade por não catalogadores e a interoperabilidade permite aos usuários pesquisar através de vários sistemas de recuperação de informação. Em 1998 a OCLC (*Online Computer Library Center*) começou a explorar o desenvolvimento de um sistema de cabeçalhos de assunto que pudesse utilizar-se nos registos de metadados (*Dublin Core Metadata Initiative*)²³ para a descrição e indexação dos recursos eletrônicos. O desenvolvimento de um vocabulário controlado desta natureza podia conseguir-se usando uma lista já existente, e dos esquemas existentes a LCSH foi a escolha mais óbvia. A OCLC decidiu desenvolver o esquema FAST²⁴ (*Faceted Application of Subject Terminology*), que constitui um vocabulário normalizado, extraído da LCSH, que categoriza os cabeçalhos por facetas. Contém características de pré-coordenação, pois utiliza cabeçalhos com subdivisões, e de pós-coordenação porque, nos casos em que não se desenvolveram cabeçalhos complexos não se usam subdivisões, mas estabelecem-se cabeçalhos adicionais para

²³ <http://dublincore.org/> [Consulta: 06-03-2013]

²⁴ <http://www.oclc.org/research/activities/fast.html> [Consulta: 14-02-2013]

diferentes facetas. A maior alteração introduzida pelo esquema FAST²⁵ é a organização dos cabeçalhos em forma vertical, o que permite simplificar a sintaxe da LCSH e reduzir a quantidade de cabeçalhos únicos, sem prejuízo da sua riqueza léxica (Dean 2003; Montalvo 2011).

Os pontos fortes identificados nos vocabulários controlados, como é o caso da LCSH e de outras listas similares, fazem com que estes continuem a ser os temas centrais da investigação bibliográfica. Embora a pré-coordenação seja uma das grandes forças na hora de hierarquizar entradas de assuntos, é também um dos aspetos menos compreendidos pelos usuários, e esta dificuldade passou a ser superada pelo FAST ao permitir a construção de entradas pós-coordenadas. Além de simplificar a sintaxe agilizando a catalogação, conserva a riqueza léxica e o apoio logístico dos cabeçalhos LCSH. O projeto FAST veio demonstrar que é viável um novo esquema controlado de assuntos baseado na terminologia da LCSH, mas com a vantagem de uma sintaxe e regras de aplicação simplificadas (Dean 2003; Montalvo 2011).

Esta é uma tendência que se tem vindo a adotar em mais listas de cabeçalhos de assunto ou catálogos de autoridades acessíveis em linha, como é o exemplo do Nuovo Soggettario THESAURUS²⁶, que se publicou em linha numa versão facetada mais complexa que a adotada na LCSH. Editado pela BNCF (*Biblioteca Nazionale Centrale Firenze*), é a ferramenta italiana de indexação de assuntos para vários tipos de recursos. Foi desenvolvido em conformidade com as recomendações da IFLA, e outras normas

²⁵ Algo semelhante está em fase experimental no caso espanhol, com a publicação em *linked open data* da lista de cabeçalhos de assunto para as bibliotecas públicas, <http://id.sgcb.mcu.es/lem/>

²⁶ <http://thes.bncf.firenze.sbn.it/ricerca.php> [Consulta: 03-07-2013]

internacionais na área da indexação. Esta ferramenta foi criada para bibliotecas italianas gerais e especializadas, para museus, bibliotecas multimédia, arquivos e centros de documentação. A principal componente do *Nuovo Soggettario* é um tesouro geral disponível na *web* desde 2007. Dá apoio às novas práticas de indexação por assuntos e gere a terminologia que deriva da colaboração entre a BNCF (*Biblioteca Nazionale Centrale Firenze*), e outras bibliotecas. O projeto está a evoluir em várias direções e sustenta a interoperabilidade, promovendo o seu uso para além do tradicional ambiente das bibliotecas, no atual ambiente de *Linked data*²⁷ (Bergamin e Lucarelli 2013).

1.2.2.1 O Sistema de Indexação em Português - SIPORbase

O sistema SIPORbase²⁸ é um projeto que surgiu em 1988 com o propósito de desenvolver um catálogo alfabético de assuntos na Biblioteca Nacional de Portugal²⁹, para facilitar o acesso por assunto aos usuários sem um perfil comum, possibilitando a pesquisa através de uma linguagem terminológica e ultrapassar as insuficiências da CDU (Classificação Decimal Universal), anteriormente estabelecida como linguagem de recuperação de informação.

²⁷ É o conceito de usar a *web* para ligar dados relacionados que não estavam previamente ligados, ou usar a *web* para reduzir as barreiras à ligação de dados já ligados que usam outros métodos (<http://linkeddata.org>)

²⁸ A 1ª edição do manual SIPORbase é de 1988 (2ª ed. 1992, 3ª ed. 1998).

²⁹ Adiante designada BNP; <http://www.bnportugal.pt/>

Até 1988 não existia nenhum instrumento de trabalho senão as normas internacionais (ISO 5963 e 2788)³⁰ aplicáveis a qualquer língua, pelo que houve a necessidade de pensar em regras para a seleção e controlo da terminologia em língua portuguesa (Santos 2006). Sendo a BNP uma biblioteca geral com fundos bibliográficos heterogéneos, que não permite a adoção de uma linguagem terminológica pré-estabelecida, e que o nível de especificidade necessário só pode ser conseguido através da pré-coordenação, optou-se por uma linguagem de indexação de tipo pré-coordenada, construída de forma evolutiva.

O SIPORbase foi concebido com base na lista de cabeçalhos de assunto da LCSH, que serviu de base conceptual para a formulação das regras. O sistema foi assim desenvolvido a partir do estudo da aplicação à língua portuguesa das regras de construção e aplicação duma linguagem documental do tipo da LCSH. A implementação de uma linguagem de indexação pré-coordenada está assente sobre um corpo de regras que garantem a sua coerente construção, quanto ao controlo do vocabulário, à sintaxe e à semântica dos cabeçalhos. A opção da pré-coordenação da linguagem foi também justificada pelas condições do sistema automatizado, que já permitia o acesso multidirecional a cabeçalhos complexos, por palavras-chave ou termos componentes, assim como o registo da função semântica de cada elemento componente do cabeçalho. O sistema SIPORbase prevê nas suas regras a criação de registos de autoridade dos cabeçalhos pré-coordenados, constituindo assim um instrumento de normalização para toda a bibliografia portuguesa, que pode ser utilizado para construção dos ficheiros de autoridade de assuntos nas bibliotecas.

³⁰ ISO 5963 corresponde à norma portuguesa NP 3715:1989 (Método para a análise de documentos, determinação do seu conteúdo e selecção de termos de indexação); ISO 2788 corresponde à norma portuguesa NP 4036:1992 (Tesauros monolingues: directivas para a sua construção e desenvolvimento).

O SIPORbase é um sistema de normalização na construção da linguagem documental, que responde à necessidade de estabelecer princípios e critérios para a constituição de um catálogo de assuntos. Para além das regras que definem a filosofia do sistema, a política de indexação e a própria pré-coordenação, contempla um conjunto de instruções para a seleção e controlo da terminologia que podem ser usadas quer se tenha optado pela pré-coordenação quer pela pós-coordenação. Muitas bibliotecas portuguesas não têm recursos humanos suficientes para praticarem uma indexação pré-coordenada, mais complexa e morosa, decidindo-se pela pós-coordenação, e precisam de um instrumento que auxilie na seleção de terminologia a usar na indexação, quer esta seja retirada de fontes como tesouros ou outras, nomeadamente de obras de referência ou das próprias obras a indexar (Santos 2006).

O manual do sistema na sua 2ª edição (1992) passa a ser de autoria conjunta da BNP e do CLIP³¹ (Compatibilização de Linguagens de Indexação em Português), que é um projeto de cooperação da BNP com profissionais de bibliotecas especializadas, com vista à uniformização das linguagens documentais em várias áreas temáticas. A última edição (1998) é constituída por nove secções, sendo as três primeiras dedicadas à ‘organização e diretrizes de utilização’, aos ‘princípios e política de indexação’, e à ‘caracterização da linguagem documental’. As grandes secções seguintes referem-se a: Terminologia / Forma dos cabeçalhos / Estrutura de referências / Assuntos especiais / Documentos especiais. Os ‘assuntos especiais’ referem-se a literatura, cultura,

³¹Grupo criado em 1989 para o melhoramento das linguagens documentais utilizadas pelas bibliotecas cooperantes na Base Nacional de Dados Bibliográficos – PORBASE
http://www.bnportugal.pt/index.php?option=com_content&view=article&id=270&Itemid=194

civilização, história, filosofia e obras específicas; os ‘documentos especiais’ dizem respeito a obras de referência, obras para crianças e jovens, manuscritos e livro antigo, manuais de ensino, documentos visuais gráficos e publicações em série.

Para além do sistema SIPORbase, constituem instrumentos normativos de apoio à indexação as normas portuguesas (adaptações de normas internacionais) anteriormente referidas NP 3715:1989 e NP 4036:1992 e ainda a NP 4285-4:2000 ‘Documentação e informação. Vocabulário. Parte 4: Linguagens documentais’, que corresponde³² à norma ISO 5127-6:1983. As diversas partes desta norma têm como objetivo facilitar o intercâmbio no domínio da informação e documentação, contendo termos e definições dos conceitos na área das linguagens documentais.

A BNP tem desenvolvido e difundido as atividades normativas em Portugal através do seu Serviço de Normalização e Qualidade, que acompanha os desenvolvimentos internacionais de normas e práticas biblioteconómicas. Nos últimos anos foi dada prioridade à disponibilização de versões atualizadas em língua portuguesa de normas da IFLA, nomeadamente as relativas ao UNIMARC, aos princípios de catalogação e aos modelos conceptuais. A BNP é o Organismo de Normalização Sectorial para a Documentação e Informação através da sua Comissão Técnica (CT7) e o Instituto Português da Qualidade (IPQ). A nível internacional a CT7 acompanha os trabalhos da ISO/TC 46 e da ISO/TC 171 (*International Organization for Standardization - Information and documentation/Document management applications*) no sentido de

³² ISO 5127-6:1983 “*Documentation and information -- Vocabulary -- Part 6: Documentary languages.*”

harmonizar ou transpor para a língua portuguesa normas neste domínio de intervenção (BNP).

A normalização bibliográfica teve uma grande evolução nas últimas décadas, e a BNP adotou como uma das suas linhas estratégicas reforçar as atividades normativas no âmbito da biblioteconomia e no âmbito da informação e documentação em geral. Destacam-se as edições portuguesas de documentação da IFLA como os FRBR, as últimas edições de manuais UNIMARC e a Declaração dos Princípios Internacionais de Catalogação, as edições portuguesas de FRAD³³ e da ISBD consolidada³⁴, publicadas respetivamente em 2011 e 2012; Está também planeada a tradução de FRSAD, e depois de uma fase em que foi dada prioridade à atualização da literatura técnica da IFLA para disponibilizar em português, está prevista a condução de trabalhos para preparar as novas Regras Portuguesas de Catalogação de acordo com os novos Princípios Internacionais de Catalogação, os modelos conceptuais FRBR e FRAD e o código RDA (Galvão e Cordeiro 2010).

³³ <http://livrariaonline.bnportugal.pt/Issue.aspx?i=2151> [Consulta: 18-07-2013]

³⁴ <http://livrariaonline.bnportugal.pt/Issue.aspx?i=2163> [Consulta: 18-07-2013]

1.2.3 Evolução normativa dos vocabulários controlados

As ferramentas de indexação de assuntos devem obedecer a normas e diretrizes, para que sejam adotados critérios uniformes e conseguidos altos níveis de qualidade. A prática normativa permite em linhas gerais (IFLA 2011):

- ✓ Atingir um nível aceitável de qualidade,
- ✓ Minimizar a duplicação de trabalho, reduzindo custos,
- ✓ Garantir consistência e compatibilidade
- ✓ Facilitar dificuldades de pesquisa e recuperação da informação,
- ✓ Educar profissionais de informação
- ✓ Ajudar a ultrapassar barreiras linguísticas.

Ao nível dos organismos internacionais, na década de oitenta tomou relevo a *International Organization for Standardization* (ISO). Também a *International Federation of Library Association* (IFLA) se tem destacado pelo trabalho em estreita colaboração com outros grupos, atualizando as recomendações para tesouros multilingues (García Marco et al. 2007).

Em relação às listas de cabeçalhos de assunto, não há uma norma verdadeiramente internacional para a sua construção. A padronização de vocabulários controlados nos catálogos *online* é usualmente garantida pelo uso comum de uma lista de cabeçalhos de assunto *standard*, acompanhada por algum tipo de manual de aplicação. É típico que estes instrumentos sejam compilados e mantidos em bibliotecas nacionais responsáveis por indexar a bibliografia nacional. Muitas listas tornam-se “normas” através do uso

cooperativo e através da sua aplicação em registos legíveis por máquina que são distribuídos para outras bibliotecas, como é o caso da lista de cabeçalhos de assuntos LCSH (IFLA 2011).

A evolução das recomendações dos tesouros e sistemas de organização do conhecimento em geral gira em torno a algumas noções fundamentais como vocabulário controlado, descritores, relações, interoperabilidade, etc. A interoperabilidade é, aliás, a ideia mais inovadora desde que surgiram os tesouros, trazida do campo das TIC: define-se interoperabilidade a propriedade de um sistema cujas interfaces podem trabalhar com outros sistemas sem restrições de acesso ou implementação. Nos últimos cinco anos foram feitas revisões no seio dos organismos normalizadores, e um trabalho de consenso internacional culminou com a publicação da primeira parte da norma ISO para tesouros, que põe a ênfase na interoperabilidade (Sánchez, Colmenero, Moreiro 2012).

A primeira norma sobre tesouros foi a ANSI Z39.19-1974 “*Thesaurus Structure, Construction and Use*”, publicada pelo *American National Standards Institute*. Em 1993 a sua revisão deu origem às “*Guidelines for the Construction, Format, and Management of Monolingual Thesauri*”, directrizes que por sua vez foram revistas em 1998, tendo-se constatado a necessidade de estar a par das mudanças que ocorriam no âmbito da informação eletrónica. A 4ª edição desta norma ANSI/NISO Z39.19-2005, com o título “*Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*”³⁵ amplia o seu alcance com as diretrizes para o desenvolvimento de tesouros eletrónicos (García Marco et al. 2007).

³⁵ <http://www.niso.org/kst/reports/standards> [Consulta: 12-04-2013]

Com alguns avanços notáveis em relação à norma americana, no que diz respeito à interoperabilidade linguística, à análise de facetas e interoperabilidade entre linguagens documentais, foi publicada durante os anos 2005-2008 a norma britânica BS 8723 “*Structured vocabularies for information retrieval*”³⁶. As várias partes desta norma britânica foram anuladas depois da publicação da BS ISO 25964:2011, a versão como norma ISO com o título “*Information and documentation. Thesauri and interoperability with other vocabularies. Thesauri for information retrieval*” (García Marco 2007; Sánchez, Colmenero, Moreiro 2012).

Estas normas, a americana e a britânica para a construção de tesouros desenvolveram-se com base em duas normas de referência internacional: a ISO 5964 e a ISO 2788:

- A norma ISO 5964:1985 – “*Guidelines for the establishment and development of Multilingual thesauri*”, é a norma que instaurou os principais procedimentos para o tratamento de várias línguas num mesmo recurso, tendo sido adotada pelo padrão nacional em diferentes países.
- A norma ISO 2788:1986 – “*Guidelines for the establishment and development of monolingual thesauri*”, inicialmente publicada em 1974, foi revista e publicada em 1986. É a norma de nível internacional que se manteve até ao aparecimento da ISO 25964-1 em 2011. A norma foi traduzida e adaptada em cada país.

³⁶ <http://www.techstreet.com/products/1249195> [Consulta: 12-04-2013]

A norma ISO 25964:2011 (*Thesauri and interoperability with other vocabularies*), surgiu como revisão das normas ISO 2788 e ISO 5964, que depois de 25 anos já não respondem às práticas atuais de processos de indexação e recuperação de informação, nem cobrem a procura de aplicações de rede que utilizem recursos heterogéneos. Por um lado trata-se de uma mudança por obsolescência do formato (do papel para o suporte eletrónico), e por outro de uma mudança ou incremento da funcionalidade, e de uma adaptação a novos perfis de usuários (Sánchez, Colmenero, Moreiro 2012). A parte 1 desta norma (nos pontos de 1 a 13) corresponde em geral ao conteúdo da ISO 2788:1986 e da ISO 5964:1985. Os restantes pontos da parte 1, e toda a parte 2 cobrem um novo terreno que não estava em nenhuma norma internacional. (ISO/DIS 25964-2, 2011).

➤ A parte 1 (ISO 25964-1:2011) - “*Thesauri for information retrieval*” - dá recomendações para o desenvolvimento e manutenção de tesauros, tanto monolíngues como multilíngues, incluindo formatos e protocolos para troca de informação. Aplica-se a vocabulários usados para recuperar informação em todos os tipos de recursos, independentemente do suporte (texto, som, imagem, objeto físico, multimédia) incluindo bases de conhecimento e portais, bases de dados bibliográficas, textos, e *items* de coleções de museus ou multimédia. Os principais conteúdos da parte 1 são: os objetivos do tesouro e o propósito do controlo do vocabulário; forma dos termos, conceitos complexos e termos compostos, equivalência entre línguas, relações entre conceitos, análise de facetas, construção e

manutenção do tesauro, integração de tesouros com aplicações e interoperabilidade, formatos e protocolos.

- A parte 2 (ISO 25964-2: 2013) – “*Interoperability with other vocabularies*” - trata da interoperabilidade entre diferentes tesouros e com outro tipo de vocabulário estruturado, tais como taxonomias, listas de autoridade, ontologias, etc. Enquanto a parte 1 descreve características dos tesouros, esta parte da norma fornece alguma descrição elementar de outros vocabulários como esquemas de classificação, esquemas de cabeçalhos de assuntos, etc. Os pontos 1-12 e 14-16 desta parte tratam de princípios de interoperabilidade que se aplicam à maior parte dos vocabulários especialmente a tesouros (mapeamento). O ponto 13 fornece orientação sobre esquemas de classificação, e os pontos 17 a 24 são dedicados aos vocabulários usados para classificar e indexar recursos, esquemas de classificação, taxonomias, esquemas de cabeçalhos de assuntos e listas de autoridades de nomes. Cada ponto dá uma breve descrição das principais características do vocabulário, contrastando as suas componentes semânticas com as de um tesauro, de forma a contextualizar os requisitos de interoperabilidade (ISO/DIS 25964-2, 2013).

Quanto ao contributo da IFLA nestas matérias, destacam-se as “*Guidelines for Multilingual thesauri*” de 2009, cujas diretrizes expressam duas abordagens, a de construir um tesauro desde o início e a de combinar tesouros já existentes. A ideia da ligação de tesouros e de listas de cabeçalhos de assunto, é que os usuários, tanto os indexadores como os que pesquisam, possam continuar a usar o mesmo vocabulário de

assunto; através desta ligação, é dado acesso a recursos indexados através de diferentes tesouros ou listas de cabeçalhos de assunto, permitindo usar um vocabulário que seja familiar (IFLA 2011). Desde a construção das “*Guidelines for the Establishment and Development of Multilingual Thesauri*” nos anos setenta, dois desenvolvimentos tiveram um papel importante na maneira de pensar no acesso multilingue à informação: a construção de tesouros não simétricos e a ligação de dois ou mais tesouros e/ou vocabulários controlados. Na combinação de diferentes tesouros está igualmente contemplada a ligação com listas de cabeçalhos de assunto para uso na indexação e recuperação da informação. As ligações são feitas nas situações em que diferentes agências usam os seus próprios vocabulários de indexação, tornando assim possível aos usuários pesquisarem em todos os vocabulários ligados, usando qualquer dos tesouros ou listas de assuntos ligados. Um exemplo de um projeto deste tipo é o MACS³⁷ (*Multilingual Access to Subjects*), a que fazemos referência mais adiante; Identificar e localizar informação relevante noutros recursos e em vastas coleções é atualmente um desafio, daí a necessidade da interoperabilidade semântica. Para apoiar essa necessidade estão a ser ativamente desenvolvidos uma variedade de serviços *web* e outras ferramentas, como é o caso de repositórios de vocabulários e esquemas de metadados. (ISO/DIS 25964-2, 2011).

As diretrizes³⁸ da IFLA referentes à utilização dos vocabulários controlados nas bibliografias nacionais têm tido igualmente um papel muito importante, para a definição e apoio de políticas de indexação a nível internacional. Os objetivos definidos nos

³⁷ http://www.nb.admin.ch/nb_professionnel/projektarbeit/00729/00733/index.html?lang=en [Consulta: 03-05-2013]

³⁸ IFLA Classification and Indexing Section: <http://www.ifla.org/classification-and-indexing>

“*Principles Underlying Subject Heading Languages*” publicados em 1999, demonstram a sua abrangência:

- ✓ Facilitar o acesso por assunto à informação a um nível internacional
- ✓ Auxiliar no desenvolvimento de linguagens de cabeçalhos de assunto, estabelecendo o que é entendido por uma boa SHL (*Subject Heading Language*)³⁹, e quais são os princípios de construção e aplicação desejáveis para tais linguagens
- ✓ Promover o entendimento de diferentes SHLs identificando princípios comuns subjacentes e estabelecer uma estrutura para o seu estudo comparativo
- ✓ Estabelecer uma razão fundamental teórica para normas ou diretrizes particulares para a construção e aplicação de SHL.

São princípios básicos de construção de cabeçalhos de assuntos, que surgiram da necessidade de alcançar um consenso internacional acerca das bases conceptuais dos sistemas de recuperação temática alfabética (Bermello Crespo 2001)

A recente publicação das “*Guidelines for Subject Access in National Bibliographies*” de 2012, reafirma a importância destas diretrizes, na medida em que “no atual mundo da informação globalizada, as bibliografias nacionais continuam com a sua principal função inalterada: informar sobre a produção de publicações no país, a sua herança cultural e intelectual”. Estas diretrizes dão recomendações em relação a políticas de indexação para as agências bibliográficas nacionais, ilustrando através de exemplos de

³⁹ Adiante SHL

boas práticas, e do reconhecimento da vantagem do acesso à informação por assuntos que permite reunir as publicações dispersas em áreas específicas.

O processo de digitalização e colocação em rede de uma grande quantidade de recursos na internet tem sido um grande desafio para as linguagens documentais. Do ponto de vista dos usuários é reclamado um acesso integrado a todo o tipo de recursos. A consulta integrada exige que se assegure a compatibilidade entre as diferentes linguagens de modo a poder interligá-las (Garcia Marco et al. 2007).

Os tesouros e cabeçalhos de assuntos incluem-se nos chamados *Knowledge Organization Systems* (KOS) que no âmbito da *web* semântica⁴⁰ são apoiados por normas e recomendações desenvolvidas de forma evolutiva ao longo dos anos. A análise das normas de tesouros e de KOS em geral demonstra que, no seu processo evolutivo, se alteraram aspetos relacionados com o conteúdo, estrutura e função, requerendo uma representação formal mais lógica. Na última década a função dos KOS ampliou-se com a evolução das TIC, e a interoperabilidade, a vinculação entre sistemas e a reutilização de dados tornaram-se fatores essenciais. As normas e recomendações da ISO têm-se centrado mais em tesouros, que se têm adaptado ao entorno tecnológico e constituem uma peça fundamental como meio de organização do conhecimento para a *web* semântica (Sánchez, Colmenero, Moreiro 2012). Uma forma possível de criar tesouros é mediante a ligação de conceitos de diferentes vocabulários controlados, e a união de recursos distintos implica o desenvolvimento de mecanismos que permitam a

⁴⁰ A *web* semântica é designada, numa forma muito elementar, como uma extensão da *web* atual, que irá permitir que os computadores trabalhem em cooperação com os humanos. Tem como finalidade conseguir atribuir um significado aos conteúdos publicados na Internet de modo a ser perceptível tanto pelo humano como pelo computador, através da interligação de significados de palavras (Wikipédia). A ideia da *web* surgiu em 2001 com o trabalho de Tim Berners-Lee, James Hendler e Ora Lassila “*The Semantic Web : a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*”, publicado na revista norte-americana *Scientific American*.

interoperabilidade. Na evolução dos tesauros para a *web* semântica, uma das diferenças mais significativas que se encontram é a da estruturação terminológica numa dimensão de representação conceptual, onde as equivalências se criam entre conceitos e não entre palavras. O *software* de computadores necessita de um modelo explícito de dados que distinga entre termos e conceitos, e o desenvolvimento das normas ao longo dos anos deve ser observado, em particular, em torno dessa distinção, e em termos mais gerais, das alterações que facilitam a interoperabilidade dos dados nos tesauros. Este método caracterizara-se por uma gestão colaborativa do tesouro e a possibilidade de criar tesauros descentralizados de forma flexível. Sendo o tesouro uma ferramenta para apoio ao acesso por assunto à informação, a abordagem racional é a de que este pode guiar os usuários a utilizarem sempre os mesmos termos para os mesmos conceitos. Nesse sentido, as normas do tesouro são desenhadas de forma a torná-lo um guia fiável para escolher o termo certo para o conceito pesquisado (Clarke e Zeng 2012; Sánchez, Colmenero, Moreiro 2012).

Esta ideia está relacionada com os princípios de *Linked data*, de usar a *web* para ligar dados relacionados, como já referimos. Inicialmente, as tendências orientavam-se a representar KOS só para indexar e recuperar informação, e depois evoluíram em torno da *web* semântica para partilhar conhecimento. Embora se questione sobre a vigência dos tesauros, continuam a aparecer exemplos e iniciativas a favor dos vocabulários controlados, como foi recentemente o caso da Google, Microsoft e Yahoo que se uniram para por à disposição uma coleção de vocabulários partilhados: Schema.org⁴¹. O principal benefício para os três motores de pesquisa é partilhar os mesmos “mecanismos

⁴¹ <http://schema.org> [Consulta: 01-07-13]

de entendimento”, que permitem usar o vocabulário definido neste esquema. A recomendação SKOS⁴² ao apresentar uma formalização simples facilitou o uso e a eficiência dos vocabulários controlados na *web* semântica. (Sánchez, Colmenero, Moreiro 2012).

A recomendação SKOS (*Simple Knowledge Organization Systems*) desenvolvida pelo *World Wide Web Consortium* (W3C) entre 2004 e 2009, determina a forma de etiquetar as relações terminológicas e semânticas no contexto das tecnologias da *web* semântica, e especificamente de *Linked open data* (Sánchez, Colmenero, Moreiro 2012). Esta recomendação propõe um modelo de dados para definir e implementar estruturas básicas de conteúdo que possibilitem a sua reutilização e interoperabilidade. A literatura existente reflete uma ampla aceitação para formalizar os vocabulários controlados com SKOS, como é o caso dos tesouros Agrovoc e Eurovoc, e da LCSH.

⁴²*Simple Knowledge Organization Systems*, é uma área de desenvolvimento de especificações e normas de suporte ao uso de sistemas de organização do conhecimento (KOS) como tesouros, esquemas de classificação, sistemas de cabeçalhos de assunto e taxonomias no âmbito da *web* semântica; <http://www.w3.org/TR/skos-reference/> [Consulta: 18-02-2013]

1.2.4 Alternativas e tendências de indexação

Depois de observar as principais questões sobre o uso de vocabulários controlados na indexação, pensamos que é relevante referir a perspectiva do uso da linguagem livre e da contribuição da indexação automática na pesquisa, que constituem métodos alternativos nos sistemas de recuperação de informação.

Uma das recomendações da IFLA é no sentido de que deve ser disponibilizada aos usuários a indexação controlada e não controlada. Nos sistemas de pesquisa em texto livre, é usada a linguagem natural em vez de termos selecionados de vocabulários controlados como listas de cabeçalhos de assunto ou de descritores (IFLA 2011)¹.

A linguagem natural (LN)² é a linguagem do discurso comum, usada habitualmente na fala e na escrita. É também a linguagem do discurso técnico-científico usada pelos autores dos documentos, sendo por isso uma linguagem livre, sem qualquer tipo de controlo. No contexto de bases de dados bibliográficas e catálogos *online* esses termos encontram-se nos campos de título e de resumo, enquanto a linguagem controlada é registada nos campos de assunto onde ocorrem descritores ou termos de indexação, que são um conjunto de termos autorizados para uso na indexação e posteriormente na pesquisa (Lopes 2002; Ribeiro 1996).

¹ *Guidelines for Subject Access in National Bibliographies.*

² Adiante designada LN, a linguagem natural corresponde à linguagem humana, “cuja estrutura e regras evoluíram do seu uso ao longo de um extenso período de tempo, em oposição a uma linguagem artificial baseada em regras prescritas com vista ao seu desenvolvimento e utilização” (ODLIS - *Online Dictionary for Library and Information Science*).

Nas bases bibliográficas, a pesquisa nos campos do título e do resumo do documento através de termos da LN, conforme se ilustra na fig. 1.4, recupera registos em que a palavra “mulheres” ocorre. Esta forma de pesquisa apresenta vários problemas: por um lado recupera uma grande quantidade de informação, e por outro lado, não sendo feito o controlo da terminologia, pode haver muita informação relevante que não é recuperada.

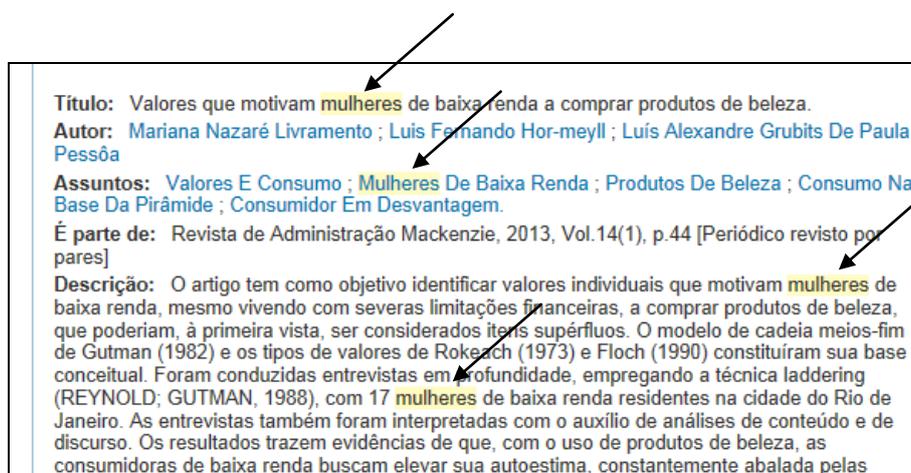


Fig. 1.4: Exemplo de linguagem natural na pesquisa de informação num resumo

Com esta opção de pesquisa obtém-se mais informação, no entanto poderá não corresponder na totalidade aos interesses do usuário, pois o termo pode ocorrer em vários contextos temáticos. As palavras significativas tiradas de títulos ou de resumos, podem constituir pontos de acesso nos índices de pesquisa, pois nesses campos cada palavra é candidata a ser pesquisada, excetuando-se aquelas designadas pelos sistemas como não-significativas. Assim as palavras são usadas para criar índices que podem ser pesquisados na linguagem natural. No momento da pesquisa, esse recurso pode ser utilizado para identificar termos e conceitos que ainda não foram incluídos na linguagem controlada como descritores, ou como cabeçalhos autorizados de assunto.

Foi defendido por vários autores que a pesquisa por LN nos resumos pode oferecer um melhor desempenho, pois está diretamente relacionada com o conteúdo dos documentos (Fidel 1986). Através da pesquisa em LN, a base de dados está efetivamente ‘auto indexada’, e cada palavra no documento torna-se um termo de pesquisa, e identifica ainda a unidade do texto no qual se encontra. Por esse motivo, a LN dos documentos é um potencial de termos de indexação ou pontos de acesso imediato, e os usuários podem interagir diretamente com os itens incluídos na base (Rothman 1983).

Ao contrário da LN a linguagem controlada (LC)³ é registada nos campos de assunto onde ocorrem descritores ou termos de indexação, constituindo assim, um conjunto de termos autorizados para uso na indexação e pesquisa⁴ (Ribeiro 1996; Lopes 2002). A utilização da LC permite ao usuário planejar uma estratégia de pesquisa, em que a recuperação da informação é feita em campos específicos de cabeçalhos controlados de assunto e/ou descritores, conforme se ilustra na fig. 1.5, em que o termo “mulheres” é recolhido do campo de assunto.

³ Adiante designada LC, a linguagem controlada significa “uma lista estabelecida de termos preferenciais da qual um catalogador deve selecionar os termos a atribuir a cabeçalhos de assunto ou os descritores para representar o conteúdo de uma obra num registo bibliográfico” (ODLIS).

⁴ No contexto da catalogação, a linguagem controlada significa “uma lista estabelecida de termos preferenciais da qual um catalogador deve selecionar os termos a atribuir a cabeçalhos de assunto ou os descritores para representar o conteúdo de uma obra num registo bibliográfico” (ODLIS).

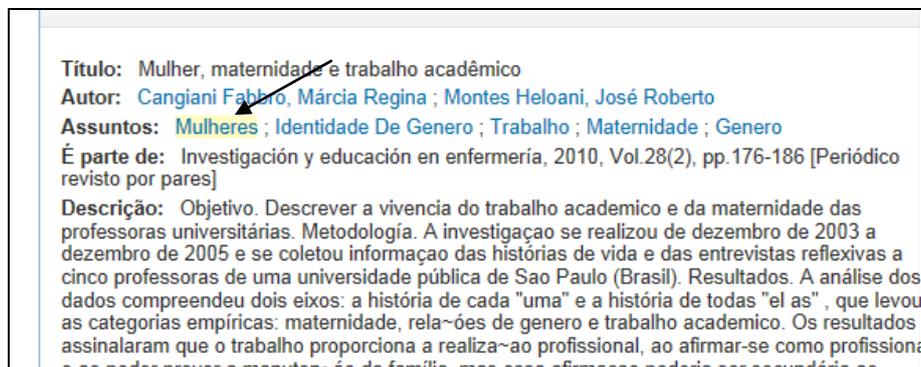


Fig. 1.5: Exemplo de linguagem controlada no campo de assunto

Nesta opção de pesquisa a informação recuperada garante ao usuário que os documentos tratam especificamente desse assunto, pois a atribuição do termo decorreu de uma análise prévia, e “pressupõe” um controle nesse campo do registo bibliográfico. Neste contexto, para melhorar a recuperação de informação o sistema de pesquisa deve também gerir as relações de equivalência (qualquer dos termos equivalentes deverá permitir ao usuário recuperar os documentos indexados com o descritor admitido), e as relações hierárquicas e/ou associativas (navegação e termos de pesquisa alternativos). Estas funcionalidades são possíveis através do controle de autoridade de que falamos adiante.

A LN é muito flexível e é muito vantajosa para os autores, que podem variar a sua terminologia para manter o interesse do leitor, mas para o indexador constitui um *handicap*, pois este está mais preocupado com as ideias expressas no documento (Foskett 1997). Na indexação feita pelos profissionais, há usualmente o recurso a

ferramentas de controlo do vocabulário, para uma maior qualidade no processo. A vantagem do uso de vocabulários controlados, através das suas listas de autoridade, é a escolha de descritores que uniformizam as variações ortográficas e estabelecem as relações semânticas, contribuindo para a consistência da linguagem entre o indexador e o pesquisador, de forma a facilitar a localização da informação (Savoy 2004).

A utilização da linguagem natural e da linguagem controlada nos sistemas de indexação é considerado um “eterno debate” desde há décadas. Estudos realizados desde a década de setenta sobre o uso dos dois tipos de linguagem aplicada à indexação em bases de dados, não revelaram especificamente qual o melhor tipo de linguagem para a recuperação de informação, pois uma série de fatores podem intervir no processo e afetar os resultados de pesquisa. Muitos desses estudos demonstraram que os dois tipos de linguagem apresentam vantagens e desvantagens, e que a sua utilização conjunta traz benefícios no desempenho dos sistemas favorecendo os resultados de pesquisa (Lopes 2002)⁵.

O desenvolvimento de *software* de computadores e de catálogos *online* tornou possível a combinação de diferentes linguagens de indexação, e essa funcionalidade vai de encontro ao que parece ser de consenso geral: sistemas controlados e sistemas não controlados (acesso por palavra-chave, pesquisa em texto livre) devem estar disponíveis ao usuário, porque tendem a complementar-se, adequando-se às diferentes capacidades e objetivos dos usuários (Foskett 1997).

⁵ Lopes, I. L. (2002). Uso das linguagens controlada e natural em bases de dados: revisão da literatura.

A combinação dos termos da LN e da LC são o ideal na elaboração de estratégias de pesquisa, apresentando melhor desempenho na recuperação do que as estratégias com cada tipo de linguagem individualmente (Henzler 1978; Muddamalle 1998). Na área de humanidades, em que há uma multiplicidade de abordagem dos mesmos conceitos, e que, não sendo controlada a sinonímia a pesquisa resulta em baixos níveis de recuperação e de precisão, a combinação de estratégias de pesquisa com LC e LN oferece melhores resultados na recuperação de informação (Knapp, Cohen e Juedes 1998).

A era mais recente sobre esta controvérsia surgiu com a pesquisa *online*. A popularidade da pesquisa em texto livre levou, contudo, a uma valorização dos vocabulários controlados. Várias experiências no contexto de bases de dados *online*, conduziram à descoberta de que os dois métodos podiam produzir a mesma precisão nas pesquisas, e que embora o vocabulário controlado produzisse uma melhor revocação do que a linguagem natural, os dois métodos se complementavam.

A construção e desenvolvimento de vocabulários controlados é um trabalho complexo que por um lado requer grandes investimentos de tempo para testar, sedimentar e manter atualizado, e por outro é uma tarefa que implica um trabalho intensivo, especialmente perante a alternativa da pesquisa em texto livre, onde o texto já está disponível, gerando índices automáticos (Fidel 1991b; Lopes 2002).

As necessidades de pesquisa dos usuários variam entre o *browsing* e a necessidade de uma pesquisa retrospectiva; um usuário pode estar interessado só num ou dois

documentos de um assunto e assim não precisar da ‘arrumação’ proporcionada pelos vocabulários controlados. Por outro lado há situações em que essa arrumação não deve ser sacrificada, como no caso de um investigador científico que pesquisa uma base de dados para uma verificação exaustiva de documentos sobre um assunto (Svenonius 2003).

Colocando o usuário no centro das pesquisas em bases de dados e catálogos *online*, é reconhecida a sua dificuldade em formular estratégias de pesquisa apropriadas, e portanto pode não beneficiar do uso das linguagens controladas associadas às bases. Muitos usuários empregam ambas as linguagens no momento de formulação da estratégia de pesquisa, principalmente porque, na maioria das bases de dados, é possível a pesquisa simultânea por campos que podem ser combinados entre si. Assim, os campos de resumo, de títulos, de descritores ou cabeçalhos de assunto podem ser amplamente utilizados visando à obtenção de um resultado mais satisfatório, independentemente da verificação, no momento de operação da pesquisa, de qual dessas linguagens terá melhor desempenho. O foco, portanto, está na obtenção de resultados satisfatórios, e não no instrumento utilizado para alcançar esses resultados (Foskett 1997; Markey 2007).

Na tab. 1.2 sintetizam-se as principais vantagens e desvantagens da utilização dos dois tipos de linguagem na elaboração de pesquisas no contexto dos catálogos bibliográficos, conforme identificadas por Lopes (2002).

Linguagem natural	
Vantagens	Desvantagens
1. Permite o registo imediato da informação numa base de dados, sem necessidade de consulta de uma linguagem controlada	Os usuários, no processo de pesquisa, precisam fazer um esforço intelectual maior para identificar sinónimos, grafias alternativas, homónimos, etc.
2. O processo de pesquisa não implica treinamento específico no uso de uma linguagem controlada	Alta incidência de respostas negativas ou de relações incorretas entre os termos usados na pesquisa, por ausência de padronização
3. Os termos são extraídos diretamente dos documentos que constituem a base de dados	Dificuldades de acesso tendem a aumentar com a entrada de termos de pesquisa aleatórios
4. Temas específicos citados nos documentos podem ser encontrados	Uma estratégia de pesquisa com os principais conceitos e seus sinónimos deve ser elaborada para cada base de dados
5. Elimina os conflitos de comunicação entre os indexadores e os usuários, pois ambos têm acesso aos mesmos termos	Perda de confiança do usuário perante uma possível resposta negativa

Tabela 1.2: Uso da linguagem natural na indexação (adapt. de Lopes, 2002)

Estreitamente ligado às questões anteriores está a contribuição da indexação automática. Nas diretrizes⁶ da IFLA 2011 para acesso por assuntos é reconhecido o potencial dos processos automatizados de indexação como alternativa aos métodos tradicionais, tendo em conta a crescente quantidade de documentos, relativamente aos quais pode ser apropriado usar procedimentos de indexação automáticos ou semiautomáticos.

A distinção entre indexação humana e indexação automática, corresponde genericamente às duas abordagens usadas para a análise de documentos, a análise feita pelos profissionais, e a análise automática algorítmica feita pelas máquinas, para extrair

⁶ *Guidelines for Subject Access in National Bibliographies.*

ou atribuir termos de indexação sem intervenção humana, uma vez que se tenham estabelecido programas ou normas relativas a esse procedimento. Tal como na indexação manual, o princípio de indexação automatizada é o de identificar um documento através de termos representativos do seu conteúdo, numa indexação livre por palavras que pertencem a um conjunto aberto de termos, ou numa indexação controlada, a um conjunto fechado de termos referenciados numa lista de autoridade ou num tesauro (Anderson e Pérez-Carballo 2001b; Méndez Rodríguez e Moreira González 1999).

A prática da indexação humana e da indexação automática, é uma questão que cada vez mais se coloca numa perspectiva de futuro. As razões que mais têm condicionado a visão destas duas práticas são, por um lado, a atual escassez de profissionais com formação específica nesta tarefa (que se nota na redução de currículos de formação de catalogadores nas escolas de biblioteconomia), e por outro lado, as dificuldades orçamentais que afetam as bibliotecas de uma forma geral. O maior condicionamento é, no entanto, a exigência dos usuários quanto a sistemas de recuperação de informação cada vez mais rápidos e disponíveis na internet, onde o crescimento dos recursos eletrónicos é exponencial (Montalvo 2011).

De acordo com Méndez Rodríguez e Moreira González (1999) um dos argumentos mais sólidos para o desenvolvimento de sistemas de indexação automatizada foi o custo elevado da indexação humana em termos de tempo, o que se constata cada vez mais. São tendências atuais, indexar os documentos digitais por meios eletrónicos, e aceder diretamente aos documentos em texto integral, pelo seu conteúdo, através de processamento linguístico automático. As investigações em torno da indexação

automática devem-se essencialmente a estes factos, conjugados com a informatização dos processos documentais e com a possibilidade de automatizar processos cognitivos.

A possibilidade da indexação automática é a análise de texto por meio de algoritmos para produzir cabeçalhos no índice. Em algumas bases de dados a indexação baseia-se simplesmente nas palavras do texto integral e em estratégias que extraem as palavras, frases ou chaves, pesquisadas pela combinação de métodos mais ou menos sofisticados, como a lógica booleana⁷, técnicas de lematização⁸ (truncatura de prefixos e sufixos), operadores de proximidade⁹, etc. (Svenonius 2003; Wellisch 1994). Informação sobre estas e outras técnicas de análise e indexação automática é detalhada no trabalho de Anderson e Pérez-Carballo (2001b).¹⁰

A alternativa da pesquisa em texto livre (quer nas partes do registo bibliográfico que contenham informação de assunto, ou nos próprios documentos), é designada de método de indexação derivativo, cujo início teve origem com os unitermos de Taube¹¹, e os chamados índices permutados,¹² que surgiram a partir dos anos sessenta, com o advento

⁷ O sistema que permite ao usuário combinar palavras ou frases que representam conceitos significativos, quando pesquisam num catálogo online ou numa base de dados bibliográfica, por palavras-chave. Os três operadores (comandos lógicos) estão disponíveis na maioria dos *software* de pesquisa.

⁸ Lematização (ou *stemming*) é a técnica que permite recuperar variantes de uma palavra através da sua raiz. A adição de símbolos de substituição de caracteres no início, meio e fim da palavra numa pesquisa por palavra-chave para recuperar as formas variantes. A truncatura é particularmente útil na recuperação das formas de singular e plural de uma palavra na mesma pesquisa (ODLIS); A lematização é um procedimento de redução ou de síntese parcial do léxico, que consiste na redução dos elementos lexicais a um único vetor centrado em torno de uma raiz temática ou de um vocabulário-chave (Zapparoli e Camlong 2002).

⁹ O *software* de pesquisa de algumas bases bibliográficas permite o uso de operadores de proximidade em expressões de pesquisa para especificar que um registo só será recuperado se a palavra-chave usada na pesquisa aparecer contida nos elementos de outras palavras, ou dentro da mesma frase ou parágrafo (ODLIS).

¹⁰ “*The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part II: Machine indexing, and the allocation of human versus machine effort.*”

¹¹ O ‘unitermo’ era fundamentado na hipótese de que cada ideia poderia ser representada por uma única palavra. O criador do ‘método unitermo’ foi Mortimer Taube, autor do livro *Computers and Common Sense* (1967), (DBpedia).

¹² A indexação automática mais simples é baseada no acesso às ocorrências de cada palavra. Este tipo de indexação é apresentada aos usuários através de índices nos formatos conhecidos como ‘permutados’, *keyword-in-context* (KWIC), *Keyword-out-of-context* (KWOC), *keyword-along-side-of-context* (KWAC). Em linguagem máquina isto equivale a indexação e recuperação de texto livre, pois irá encontrar cada ocorrência de cada carácter ou combinação de caracteres.

do computador, e a vantagem que este ofereceu da indexação com linguagem natural. (Anderson e Pérez-Carballo 2001b; Svenonius 2003).

Os trabalhos iniciais no uso de técnicas automatizadas são da autoria de Gerard Salton,¹³ que conduziu o desenvolvimento do sistema de recuperação de informação SMART (*System for the Mechanical Analysis and Retrieval of Text*), de que falamos adiante. Na defesa dos métodos automáticos, acerca da utilização de vocabulários controlados, este autor argumentou expressivamente (*apud* Hjørland 2012):

“Acting as if we were stuck in the nineteenth century with controlled vocabularies, thesaurus control, and all the attendant miseries, will surely not contribute to a proper understanding and appreciation of the modern information science field.”¹⁴

Estas questões foram efetivamente avaliadas nos projetos de Cranfield, que abordamos adiante no capítulo 1.4.1. O exemplo de um estudo em que se compara a eficácia da indexação humana baseada em vocabulários controlados com a indexação automática, demonstra que os descritores de uma lista de autoridades atribuídos manualmente, revelam melhor performance nos resultados de pesquisa, do que um esquema de indexação automática (Savoy 2004). Embora nem sempre seja encontrada uma diferença estatística significativa entre resultados de pesquisa decorrentes de cada uma das abordagens, a performance na recuperação com descritores atribuídos manualmente é melhor do que com a indexação automática baseada em títulos e *abstracts*. Quando é feita uma pesquisa exaustiva, dependendo da área de conhecimento, um ou outro sistema pode favorecer os resultados, no entanto é

¹³ Sobre os sucessos conseguidos nos métodos de indexação e recuperação automática, ver: Salton, G. (1986) *Recent trends in automatic information retrieval*; Salton e Buckley (1990) *Improving Retrieval Performance by Relevance Feedback*.

¹⁴ Salton (1996), *A new horizon for information science*.

demonstrada uma melhoria quando os descritores atribuídos na indexação manual são combinados com um esquema de indexação automática.

O debate sobre qual o melhor método, se o de indexação humana ou o de indexação automática continuou a decorrer na última década, em que persistem as experiências com a indexação automática, da qual decorrem melhoramentos. Os peritos nesta área reconhecem a importância de pesquisar eletronicamente, e os resultados da investigação, que têm vindo a ser testados nas TREC (Text REtrieval Conferences)¹⁵, continuam a ter impacto na indexação automática. Na perspetiva dos usuários, foi efetivamente demonstrado que estas duas abordagens são eficazes (Anderson e Pérez-Carballo 2001b; Savoy 2004; Shields 2005; Svenonius 2003).

Shields (2005) fez um interessante paralelo com a revolução industrial no sentido da forma como os humanos mudaram a maneira de criação dos *itens*, “os *itens* anteriormente criados à mão, de repente começaram a ser criados mais depressa e mais baratos usando humanos ajudados por máquinas”. Numa sociedade tecnologicamente desenvolvida é necessária a indexação automática para acompanhar a crescente quantidade de documentos eletrónicos a indexar.

As crescentes e sofisticadas técnicas de indexação automática têm grandes vantagens, no entanto, os métodos de análise de textos na indexação automática são só uma parte da prática da indexação, e a explosão da informação na internet é um exemplo de como é essencial desenvolver peritos humanos de forma seletiva, pelo que a

¹⁵ Conferências para fazer testes com grandes coleções de documentos no sentido de comparar vários sistemas de recuperação de informação de acordo com os mesmos parâmetros. A primeira conferência foi realizada em 1992; <http://trec.nist.gov/overview.html> [Consulta: 05-07-2013]

indexação humana, com o seu carácter intelectual, continua a ter um papel relevante e fundamental (Anderson e Pérez-Carballo 2001b; Cerbo 2011).

Numa perspetiva quantitativa das necessidades de indexação, um dos argumentos demolidores a favor da indexação automática tem sido o “impressionante auge da *web* e o êxito do Google” (Caro-Castro 2012). De acordo com a autora, a gestão da informação disponível em rede, em contínuo crescimento, constitui um desafio para os profissionais de documentação na aplicabilidade ao mundo digital das técnicas e ferramentas habitualmente usadas. As alternativas para enfrentar este desafio podem-se agrupar em duas tendências, uma que enfatiza o uso das técnicas automatizadas, e a outra a necessidade de informação “semântica”.

Em relação à primeira tendência, a solução está no aperfeiçoamento das técnicas de indexação automática sobre o texto completo, com o desenvolvimento de algoritmos sofisticados tanto para a seleção dos termos de indexação como para a organização e recuperação da informação. O objetivo é alcançar um nível mais alto de rendimento dos motores de pesquisa, porque à escala da *web*, as ferramentas e métodos bibliotecários tradicionais se têm vindo a tornar obsoletos. Muitos especialistas da organização do conhecimento são de opinião que os motores de pesquisa não tratam de maneira adequada os problemas lógico-linguísticos da representação e recuperação de informação, sendo necessárias as descrições “estruturadas” dos documentos que os permitam identificar com clareza; as insuficiências dos motores de pesquisa manifestaram as limitações de uma *web* baseada em documentos HTML e ligações hipertextuais, em que há excesso de informação desestruturada, opacidade do conteúdo

informativo disponível em bases de dados, e o não reconhecimento de conceitos, mas só de cadeias de caracteres (Caro-Castro 2012).

A segunda tendência, que se pode denominar de “semântica”, promove o desenvolvimento de modelos de metadados para estruturar a informação descritiva sobre os documentos digitais, e a implementação de vocabulários estruturados, ou listas de autoridade para controlar o conteúdo dos metadados e para organizar a informação, na qual se reconheça um significado dos dados e das relações entre eles. No desenvolvimento da *web* semântica “persegue-se a ideia de uma pesquisa inteligente”. A solução do problema de recuperação de informação na *web* exige o desenvolvimento da investigação em ambos os sentidos, tanto tecnológico como semântico, onde se combine o desenvolvimento de estruturas de meta informação com a potência do processamento informático (Caro-Castro 2012; Vatant 2010).

De acordo com o que observámos os dois tipos de indexação irão coexistir. A indexação humana é muito apreciada pelos seus contributos e vai permanecer disponível, embora com um número decrescente de pessoas a fazê-la, e em contextos limitados. Quanto à indexação automática, vão-se encontrar formas de a melhorar, através de investigações que tendem a ser mais numerosas, porque é inquestionável o seu contributo para resolver o problema da indexação ‘em massa’ necessária nos nossos dias. Em suma, a combinação da extração automática de descritores com vocabulários controlados é uma linha de trabalho muito importante atualmente, no sentido de

combinar as vantagens reconhecidas nos dois métodos. Um exemplo desta tendência é o projeto HIVE¹⁶ ou a aplicação PoolParty¹⁷.

A aplicação PoolParty é um sistema de gestão de tesouros e um editor SKOS para a *web* semântica, que inclui as funcionalidades de análise de texto. O *Poolparty Thesaurus Server* é uma das mais avançadas soluções para a criação de tesouros taxonomias, etc., combinando métodos baseados em *Linked data*, *data mining*.¹⁸ O sistema auxilia na construção e manutenção de tesouros multilingues, oferecendo ao usuário vários serviços semânticos através de um interface simples.

O HIVE (*Helping with Interdisciplinary Vocabulary Engineering*) é um sistema com tecnologia SKOS que pesquisa conceitos através de vários vocabulários de carácter interdisciplinar que podem ser seleccionados. Tem a funcionalidade de extrair conceitos de documentos ou de URLs, e assim indexar automaticamente. Esta operação é executada através de três passos muito intuitivos para quem usa o sistema: (1) Seleccionar um vocabulário, (2) Fazer *upload* de um documento ou indicar o URL, (3) Iniciar o processo. O resultado ilustra-se (fig. 1.6), através da nuvem de conceitos extraídos do tesouro Agrovoc, em que, seleccionando um dos conceitos “*Water management*”, temos acesso às relações semânticas estabelecidas (*Broader concepts*, *Narrower concepts*, etc.).

¹⁶ <http://hive.nescent.org:9090/home.html> [Consulta: 12-07-13]

¹⁷ <http://www.poolparty.biz/> [Consulta: 12-07-13]

¹⁸ Análise de dados contidos em texto de linguagem natural; equivale ao termo português ‘Prospecção de dados’, que é o processo de explorar grandes quantidades de dados para detetar relacionamentos sistemáticos entre variáveis, e novos subconjuntos de dados (Wikipédia)

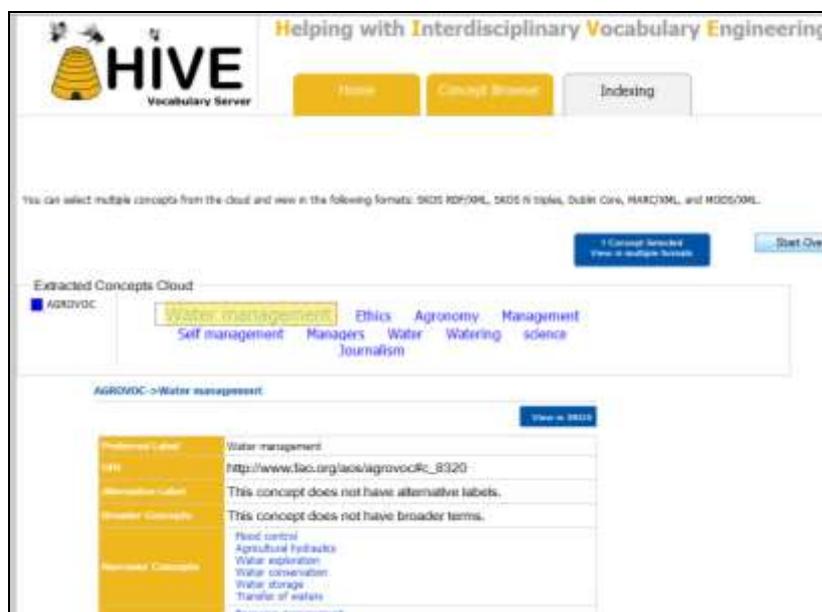


Fig. 1.6: Resultado da indexação de um documento no HIVE.

Por trás destes processos está um conjunto de boas práticas para publicar e ligar dados estruturados que é conhecido como *Linked data*¹⁹. Para além da interoperabilidade entre sistemas, o grande desafio reside em conseguir que os conteúdos estejam descritos por metadados dotados explicitamente de semântica, e é através da estruturação e normalização da informação, que deverá ser possível pesquisar por conceitos. O projeto da *web* semântica proporcionou o desenvolvimento de diversas normas (Dublin Core, XML, RDF, SKOS, OWL, etc.) que vêm contribuir para

¹⁹ <http://linkeddata.org/> [Consulta: 09-03-2013]; A designação de *Linked Data* significa o método de publicar dados na *web* e interligar dados de diferentes fontes. A *Linked Data* pode ser acedida usando *browsers* da *web* semântica, tal como os tradicionais documentos na *web* são acedidos usando *browsers* HTML. No entanto, em vez de seguirem *links* de documentos entre páginas HTML, os *browsers* da *web* semântica permitem navegar entre diferentes fontes de informação seguindo *links* RDF. Os *links* RDF podem ser seguidos por robots ou motores de pesquisa da *web* semântica (DBpedia).

melhorar os sistemas de pesquisa, especialmente se são aplicados a entornos controlados como repositórios, ou a bibliotecas digitais. Tradicionalmente as possibilidades de recuperação de informação numa biblioteca digital estão limitadas ao catálogo de objetos digitais através dos campos controlados de metadados atribuídos pelos profissionais. Uma combinação da biblioteca digital e do ambiente da *web* semântica parece óbvio, pois o nível de representação dos objetos ganha um valor acrescentado se estes forem contextualizados semanticamente, e isso pode ser conseguido pela tecnologia da *web* semântica que torna a informação compreensível para máquinas. Neste contexto a comunidade *Linked data* pretende ligar dados relacionados na *web* e a partir daí melhorar a interoperabilidade semântica. Por um lado novos recursos podem ser publicados e adicionados à nuvem de *Linked data*, por outro lado os recursos existentes podem ser reutilizados, contribuindo também para o crescimento da “nuvem”, que se ilustra no diagrama (fig.1.7).

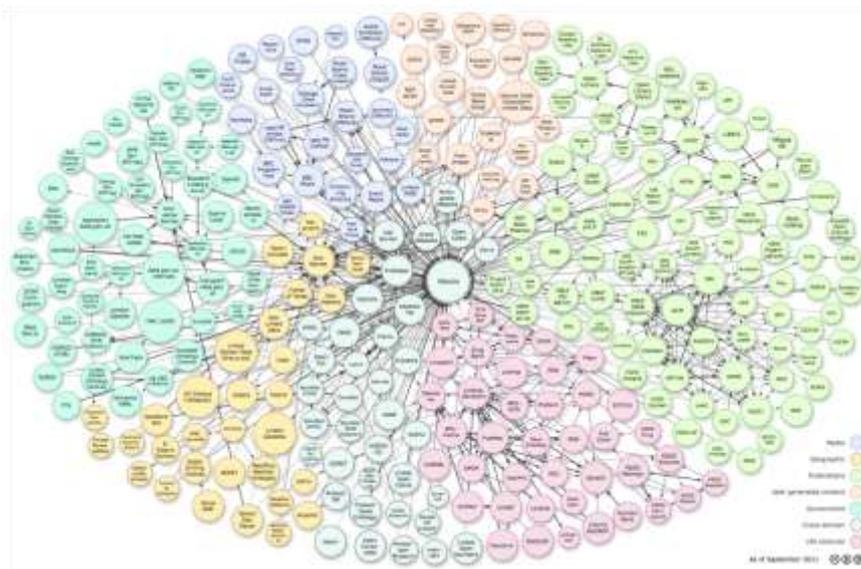


Fig. 1.7: “Linking Open Data cloud diagram”, (Richard Cyganiak e Anja Jentzsch)²⁰

Os formatos de intercâmbio de dados desenvolvidos no quadro da *web* semântica permitem a interoperabilidade entre sistemas, mas não resolvem o problema da compatibilidade entre vocabulários estruturados. Conseguir a interoperabilidade semântica continua a ser um desafio para os profissionais, especialmente num meio tão heterogêneo como a internet, e será necessário automatizar alguns dos processos que supõem a elaboração de grandes estruturas de conhecimento. Com o apoio da iniciativa *Linked data*, o modelo SKOS está a tornar-se comum para expressar sistemas de organização de conhecimento como tesauros, taxonomias, classificações ou listas de autoridade, sendo desde 2009 uma recomendação de W3C, que se passou a utilizar para levar a cabo a migração de vocabulários tradicionais como LCSH, EUROVOC, AGROVOC, e muitos outros, como se pode ver no relatório do *Linked Data Incubator*

²⁰ <http://lod-cloud.net/> [Consulta: 09-03-2013]

Group²¹ em que muitos vocabulários estruturados já estão publicados conforme as especificações desse modelo, adaptando-os aos padrões e às recomendações da *web* semântica e *Linked data*. Conforme se pode identificar nos “elementos mais importantes para a interoperabilidade semântica” (fig.1.8), SKOS estabelece um padrão para traduzir as classificações e tesouros existentes num formato legível por máquina (Caro-Castro 2012; Olensky 2010).

The skos:Concept Class	
<u>skos:Concept</u>	A unit of thought
Lexical Labels	
<u>skos:prefLabel</u>	To represent synonyms and define preferred terms analogue to a thesaurus (UF/USE)
<u>skos:altLabel</u>	
Semantic Relations	
<u>skos:broader</u>	To represent hierarchical structures
<u>skos:narrower</u>	
<u>skos:related</u>	To represent associative structures
Mapping Relations	
<u>skos:closeMatch</u>	To link two concepts that are sufficiently similar that they can be used interchangeably in some information retrieval applications (non-transitive)
<u>skos:exactMatch</u>	To link two concepts, indicating a high degree of confidence that the concepts can be used interchangeably across a wide range of information retrieval applications (transitive)
<u>skos:broadMatch</u>	To state a hierarchical mapping link between two concepts
<u>skos:narrowMatch</u>	
<u>skos:relatedMatch</u>	To state an associative mapping link between two concepts

Fig. 1.8: “Important elements of SKOS” (Olensky 2010)

Para melhor entendermos a relação dos valores do modelo SKOS com as relações semânticas tradicionais, apresentam-se os principais elementos para exemplificar o conceito de topo “Computer” (Sequeda 2012):

²¹ <http://www.w3.org/2005/Incubator/1ld/XGR-1ld-20111025/> [Consulta: 05-05-2013]

<p>Há dois tipos de etiqueta: Preferidas (skos:prefLabel) e Alternativas (skos:altLabel). Um conceito só pode ter uma etiqueta preferida por língua, mas pode ter várias etiquetas alternativas.</p>	<p>ex:Computer rdf:type skos:Concept; skos:prefLabel "Computer"@en; skos:prefLabel "Computador"@es; skos:altLabel "Ordenador"@es.</p>
<p>Nas relações semânticas, SKOS permite adicionar: relações hierárquicas, associativas e de mapeamento.</p>	
<p>As hierárquicas são (skos:broader) e (skos:narrower). Por exemplo o conceito “Computer” é ‘genérico’ do conceito “Laptop”.</p>	<p>ex:Computer rdf:type skos:Concept; skos:prefLabel "Computer"@en; skos:prefLabel "Computador"@es; skos:altLabel "Ordenador"@es ; skos:broader ex:Laptop.</p>
<p>Da mesma forma, o conceito “Laptop” é específico do conceito “Computer.”</p>	<p>ex:Laptop rdf:type skos:Concept; skos:prefLabel "Laptop"@en; skos:prefLabel "Portatil"@es; skos:narrower ex:Computer</p>
<p>A relação associativa (skos:related) é usada para apresentar uma relação entre dois conceitos. O conceito “Computer” está relacionado ao conceito “Software”.</p>	<p>ex:Computer rdf:type skos:Concept; skos:prefLabel "Computer"@en; skos:prefLabel "Computador"@es; skos:altLabel "Ordenador"@es ; skos:broader ex:Laptop ; skos:related ex:Software .</p>
<p>SKOS tem duas relações de mapeamento: skos:closeMatch e skos:exactMatch, que permitem representar um mapeamento entre diferentes conceitos.</p>	
<p>O skos:closeMatch indica que dois conceitos são suficientemente similares e ambos podem ser usados de forma alternada. Por exemplo, o conceito de ‘Laptop’ está estreitamente relacionado ao conceito ‘Netbook’.</p>	<p>ex:Laptop rdf:type skos:Concept; skos:prefLabel "Laptop"@en; skos:prefLabel "Portatil"@es; skos:narrower ex:Computer ; skos:closeMatch ex:Netbook.</p>
<p>O skos:exactMatch denota um maior grau de similaridade, nomeadamente, os dois conceitos têm o mesmo significado. Por exemplo, o conceito de ‘Laptop’ liga-se ao conceito de ‘Portatil’ existente noutro tesouro SKOS (ex2).</p>	<p>ex:Laptop rdf:type skos:Concept; skos:prefLabel "Laptop"@en; skos:prefLabel "Portatil"@es; skos:narrower ex:Computer ; skos:closeMatch ex:Netbook ; skos:exactMatch ex2:Portatil.</p>

1.3 Acesso por assuntos nos catálogos bibliográficos

Tendo como base o catálogo informatizado, a grande parte das bibliotecas acadêmicas dispõem do seu catálogo *online*, designado OPAC (*Online Public Access Catalog*), que através de uma interface gráfica na *web* permite acesso público e livre aos seus conteúdos. Voltamos a centrar-nos nos vocabulários controlados para acesso as bases bibliográficas.

A facilidade de pesquisa e rapidez de recuperação da informação na *web* são fatores que concorrem com a consulta de outro tipo de recursos disponíveis nas bibliotecas, como é o caso dos catálogos bibliográficos. A dificuldade em encontrar a informação desejada, juntamente com o desconhecimento das vantagens de uma pesquisa exaustiva nos mesmos, leva os usuários a preferir outras formas de acesso à informação. Apesar de muitos avanços a que temos assistido, continua atual a afirmação de Campos (2003) de que “muitos utilizadores que pesquisam na *web*, podem obter uma resposta que consideram satisfatória, mas ignoram que pode haver muito mais informação sobre o mesmo assunto que não foi recuperada”.

Embora o *software* dos catálogos *online* não seja estandardizado, a maioria dos catálogos são pesquisáveis pelos pontos de acesso de autor, título, assunto, palavras-chave, etc. O usuário faz as suas pesquisas por estes pontos, e recupera uma lista de registos que contêm todos os elementos necessários a uma descrição uniforme dos

documentos integrados nas bibliotecas, e constituem assim a ligação dos usuários aos documentos, permitindo a sua localização.

Em muitos sistemas o usuário tem disponíveis técnicas de pesquisa tais como a lógica booleana, operadores de proximidade, etc.¹ para refinar as expressões de pesquisa. Este conjunto de regras permitem afinar estratégias para tornar possível o encontro entre o desejo do usuário, expresso pela interrogação ao sistema, e a informação aí armazenada. No entanto, o usuário comum do catálogo não tem conhecimento dessas técnicas de pesquisa, e quase sempre interroga o sistema da forma mais intuitiva (Moreira 2005). O usuário comum, sem experiência ou conhecimentos das técnicas não sabe o que fazer se recupera poucos ou demasiados registros nos seus resultados de pesquisa (Caro-Castro 2005).

É muito frequente ouvir usuários da biblioteca que consultam o catálogo bibliográfico, dizerem que não encontram respostas para as suas necessidades, o que desde logo permite levantar várias questões relativas à pesquisa e recuperação de informação. Numa visão imediata diremos que a interrogação de pesquisa por parte do usuário não corresponde à terminologia que foi atribuída aos documentos no ato da indexação, ou que o sistema de pesquisa não redireciona o usuário desde os termos que utilizou na pesquisa até aos que existem no sistema, através de relações de equivalência (Caro-Castro 2005).

¹ Técnicas referidas no capítulo 1.2.4.

Para além de problemas conceptuais ligados ao processo de pesquisa no catálogo, diremos *a priori*, que outras situações podem dificultar o acesso à informação e o usuário não recupera o que deseja. Um dos motivos mais prováveis é ter sido feita uma indexação inadequada, sem bases normativas conforme observámos no capítulo 1.1, e não se dispor de um ficheiro de autoridades de assuntos que controle os termos e evite a sua dispersão nos índices. Outro problema frequente neste tipo de catálogos, é o facto do ficheiro de autoridades (caso exista) não ser utilizado para facilitar as pesquisas dos usuários, isto é, não serve para que o usuário encontre informação por qualquer termo considerado equivalente na representação de um conceito, e estas limitações são ainda bastante generalizadas. De acordo com Caro-Castro (2012), as normas relacionadas com a recuperação de informação põem ênfase no desenvolvimento da funcionalidade dos sistemas informáticos para converter estas ferramentas numa verdadeira ajuda (aspeto sobre o qual se insiste muito na norma ISO 25964), mas em concreto:

- ✓ Teriam que ajudar o usuário a expressar a sua necessidade de informação guiando-o para fazer uma análise conceptual de um tema mediante menus ou formulários de pesquisa facetados;
- ✓ Deveriam levar o usuário desde os seus termos de pesquisa aos descritores do tesauro e proporcionar-lhe a possibilidade de realizar pesquisas inclusivas, incorporando os temas específicos de um descritor determinado.

As primeiras investigações sobre o uso dos OPAC demonstram que estes incrementaram as pesquisas por assunto em relação aos catálogos manuais (Hancock-

Beaulieu 1990; Markey 1984), mas que o interesse inicial pareceu decair perante as dificuldades encontradas pelos usuários e a quantidade de pesquisas que não recuperavam registos (Larson 1991). Neste contexto surgiu a alternativa de recuperação de registos por palavras-chave, que resultou numa redução do uso do campo de assuntos para fazer pesquisas, tendo sido indicadas, em estudos posteriores, percentagens de utilização entre os 30 e os 57% desses campos de pesquisa (Yee e Layne 1998).

A maior parte das dificuldades identificadas na recuperação de informação em catálogos em linha estão relacionadas especificamente com as pesquisas por assuntos. O usuário deve ser capaz de expressar a sua necessidade de informação numa linguagem concreta, e a realização de pesquisas requer vários tipos de conhecimento, da área temática, das características dos campos que se vão utilizar na pesquisa, do vocabulário que se utilizou para a indexação etc. (Borgman 2000; Caro-Castro 2005; Olson e Boll 2001).

Apesar dos avanços tecnológicos e de muita investigação, os problemas nas pesquisas temáticas identificados na década de oitenta, continuam a fazer-se sentir em muitas bibliotecas, como pesquisas que não recuperam registos, sobrecarga de informação e pouca coincidência do vocabulário utilizado pelos usuários com o dos índices. Um dos grandes problemas do acesso temático nos catálogos é justamente a não coincidência entre os termos de pesquisa dos usuários, e os cabeçalhos de assunto existentes no sistema, sendo o motivo dessa disparidade atribuído à limitação do vocabulário, e a uma normalização e sintaxe estranhas ao uso “natural” da linguagem (Caro-Castro 2005; Foskett 1997; Rowley 1994).

Um dos aspetos que coloca maiores dificuldades na pesquisa por assuntos é o vocabulário, pois é difícil a coincidência na denominação dos conceitos de quem cria os documentos, de quem os indexa e de quem os pesquisa. Essa dificuldade é comprovada em estudos sobre as características dos termos e expressões de pesquisa dos usuários, e a sua coincidência com os índices de assuntos nos catálogos, sendo essa coincidência abaixo de 40% (Caro-Castro 2005; Collantes 1995; Drabenstott 1996b).

Na pesquisa de informação por assuntos, a interrogação ao sistema deve ser feita através dos campos de assunto ou índices. A evolução dos índices para o tipo permutado² permitiu que sejam criadas várias entradas de índice, em que cada subcampo de um cabeçalho constitui assim um ponto de acesso. A indexação mediante chaves de pesquisa dá origem aos índices de palavras-chave, que se incorporam nos catálogos para ampliar as possibilidades de pesquisa, introduzindo opções habituais nas bases de dados bibliográficas. A criação deste tipo de índices permite melhorar os efeitos da rigidez da pré-coordenação dos cabeçalhos nestes sistemas, reduzindo o risco de não recuperar registos na pesquisa, mas potenciando o volume de informação recuperada e o ruído documental.

A tendência dos atuais catálogos em linha é a de integração de recursos de informação, como é o caso do WorldCat³, que é a maior rede de conteúdos e serviços de bibliotecas. Embora essa situação não seja generalizada, pois em muitas bibliotecas o

² Índice automatizado que apresenta por ordem alfabética, as palavras-chaves, os descritores ou os elementos do título de um documento mantidos no seu contexto e combinados com uma referência que permite identificar esse documento (Dicionário de livro).

³ <https://www.worldcat.org/default.jsp> [Consulta: 15-07-2013]

catálogo é gerido de forma local, com os seus tradicionais registos bibliográficos, a tendência é no sentido de os sistemas de informação se irem modernizando através da aplicação de políticas nacionais, numa escala global.

1.3.1 O catálogo bibliográfico e de autoridades

Os catálogos bibliográficos são criados dentro dos padrões de qualidade exigíveis, e de acordo com as ferramentas informáticas disponíveis. Nestes sistemas automatizados, a influência do catálogo em fichas ainda é muito presente nos registos bibliográficos, e é enorme o volume de informação criada e armazenada de acordo com as regras tradicionais (Lopes e Galvão 2012).

Conforme evidenciado nos capítulos anteriores, um dos pontos fundamentais na tarefa de indexação é o controlo da terminologia, e nesse sentido, é imprescindível a existência de um catálogo de autoridades, que centralize a terminologia autorizada para ligação a um ou vários ficheiros bibliográficos. O ficheiro que contém registos de autoridade é a ferramenta utilizada pelos indexadores para organizar os catálogos das bibliotecas, no sentido de auxiliar os usuários a encontrar a informação que procuram, e a falta de um ficheiro deste tipo potencia uma indexação sem controlo dos termos e, em consequência, uma recuperação de informação inconsistente.

As atuais soluções de *software* bibliográfico dispõem de um ficheiro de autoridades, no qual se normaliza a terminologia, e que através da ligação ao catálogo bibliográfico, permite controlar o acesso aos termos nos índices de assunto. Em muitos casos só o catálogo bibliográfico está disponível para pesquisa dos usuários, no entanto, quer o catálogo de autoridades esteja visível ou não, é através deste que se controla a qualidade da terminologia.

O ficheiro de autoridades é constituído por listas estruturadas de nomes próprios ou de assuntos, processadas mediante uma série de operações que permitem elaborar os registos de autoridade, ou seja, estabelecer a forma dos termos que se vai utilizar para todas as ocorrências no catálogo na sua forma autorizada. Nestes registos são feitas as referências necessárias como as formas não autorizadas ou relacionadas, de maneira que tanto o profissional quando cataloga, como o usuário quando faz a pesquisa, utilizem sempre a mesma forma (Martinez et al. 1997). Um ficheiro de autoridades é o conjunto de registos de autoridade.

As funções básicas de um ficheiro de autoridade são (Galvão 2003):

- ✓ Criar registos de autoridades individualizados para entidades distintas, que estabeleçam e distingam formas do nome;
- ✓ Conter uma estrutura de referência que especifique formas autorizadas, não autorizadas e variantes;
- ✓ Estabelecer as bases para um sistema de autoridades informatizado, que permita comparar registos bibliográficos;
- ✓ Apoiar as necessidades de ‘arrumação’ do usuário.

As componentes dos registos de autoridade são (Campos 2003):

- ✓ Registo da forma normalizada de cada ponto de acesso;
- ✓ Possibilidade de agrupamento de todos os registos com o mesmo ponto de acesso;
- ✓ Garantia de normalização dos registos no catálogo;
- ✓ Documentação das fontes utilizadas para estabelecer os pontos de acesso;
- ✓ Registo de todas as formas do ponto de acesso (para além da escolhida como normativa), e das relações hierárquicas e associativas.

É necessário gerir adequadamente o catálogo de autoridades, no sentido de controlar a forma dos termos e, no caso de estar disponível aos usuários, rentabilizar as relações semânticas, para recuperar informação de forma eficaz nos OPAC. Para além de controlar os problemas terminológicos como a sinonímia, a homografia, a polissemia, etc., estabelecem-se relações hierárquicas e associativas entre conceitos, pessoas, entidades ou lugares. Os registos do catálogo de autoridades reproduzem todos os dados relativos aos cabeçalhos e subcabeçalhos, relacionando-os com os registos bibliográficos, o que do ponto de vista do acesso temático, permite dirigir o usuário dos seus próprios termos para os existentes no sistema, ou sugerir termos alternativos de pesquisa (Caro-Castro 2005).

Na criação de registos de autoridade, através dos quais se controlam os pontos de acesso na pesquisa, têm que ser observadas regras próprias e instrumentos normativos para seleção da terminologia⁴. A consistência do ficheiro de autoridade só é possível de manter recorrendo à utilização desses instrumentos que auxiliem ao controlo da

⁴ No caso das bibliotecas portuguesas as regras e instrumentos normativos para controlo do vocabulário são as Regras Portuguesas de Catalogação (reedição 2000); as “Recomendações para a construção de registos de autoridade de autor pessoa física”, (BNP 2005); o Sistema de Indexação em Português – SIPORbase.

terminologia. A criação de novos termos no catálogo bibliográfico é feita muitas vezes através do recurso a fontes de terminologia externa como léxicos de assunto ou tesouros, que apresentam os inconvenientes próprios de traduções e, na maioria dos casos, não respeitam os critérios de uniformidade em relação à língua portuguesa no uso de formas do singular e plural, controlo da sinonímia e homonímia, ou esclarecimento de termos obscuros. Também há por vezes a necessidade de introduzir novos termos, pois os tesouros não fazem a cobertura exaustiva de toda a terminologia em cada área do saber. A criação de nova terminologia, devidamente controlada na forma e no significado, com a respetiva estrutura de referências, deve ser feita nos registos de autoridade para posterior ligação aos registos bibliográficos. Essa metodologia envolve o levantamento das situações de controlo do vocabulário, de maneira a garantir o acesso aos conceitos, independentemente da forma pela qual possam ser pesquisados no catálogo (Santos e Almeida 2003; Santos 2006). A reconhecida vantagem do controlo do vocabulário através dos ficheiros de autoridade é a forma de garantir a coerência na representação de conceitos (IFLA, 2010).

O formato utilizado na maioria das bibliotecas portuguesas, tanto para registos bibliográficos como de autoridade, é o UNIMARC⁵. O objetivo que esteve subjacente à sua criação, inicialmente publicado em 1977, foi o de constituir um formato internacional de troca de dados bibliográficos entre diversos formatos MARC⁶. No entanto, a sua aplicação evoluiu para utilização como formato próprio em muitos países

⁵ O UNIMARC, como outros formatos MARC é uma implementação específica da ISO 2709, norma internacional que especifica a estrutura de registos bibliográficos. O UNIMARC veio dar resposta ao problema da incompatibilidade entre os vários formatos MARC, e foi adotado pela IFLA em 1985 como formato internacional para troca de registos entre agências bibliográficas nacionais.

⁶ MARC (*Machine Readable Cataloguing*) é o formato padrão de catálogos em muitas bibliotecas, para a descrição bibliográfica. Graças à sua utilização como estrutura para os dados bibliográficos há mais de três décadas, foi possível a automatização integrada das bibliotecas, e a troca de informação de forma normalizada. O conteúdo dos registos MARC é largamente influenciado por normas genericamente adotadas pela comunidade bibliotecária ao longo dos tempos, como é o caso das ISBD (*International Standard Bibliographic Description*), AACR2 (*Anglo-American Cataloguing Rules*), etc.

que estavam a iniciar a informatização das suas bibliotecas, e ainda não tinham definido um formato local de dados, como foi o caso em Portugal, onde o UNIMARC⁷ se tornou uma norma de facto. O formato tem evoluído de forma a adaptar-se a novas necessidades e a acompanhar a evolução das ISBD e dos novos modelos conceptuais FRBR (BNP).⁸

O formato UNIMARC bibliográfico e de autoridades constituem bases fundamentais para a produção e gestão de informação bibliográfica e de autoridades na generalidade das bibliotecas portuguesas. Nas bibliotecas que já produzem os seus registos, utilizando o formato bibliográfico, o formato de autoridades fornece as regras para que possam proceder também ao controlo adequado dos pontos de acesso (autor, título, assunto, etc.), permitindo-lhe estabelecer, de forma normalizada, a informação de autoridade que permite a partilha de dados e contribui para uma informação mais consistente nos catálogos. As bases de autoridades, obedecendo a regras e formatos uniformizados, desenvolveram-se no contexto da evolução dos catálogos coletivos como bases de dados bibliográficos de tipo cooperativo. Embora nem sempre seguida de forma homogénea pelas instituições bibliotecárias, essa evolução foi acompanhada da necessária normalização, no sentido de implementar critérios comuns, quer a nível da descrição bibliográfica, quer a nível das formas de entrada de nomes ou da representação de assuntos, para uma recuperação de informação eficaz (Santos 2007).

O UNIMARC autoridades tem como principal objetivo facilitar a troca internacional de dados legíveis por computador entre agências bibliográficas. Na descrição

⁷ Versão portuguesa do UNIMARC disponível *online*: http://www.ifla.org/files/uca/Unimarc_bib_abrev.pdf

⁸ Biblioteca Nacional de Portugal <http://www.bnportugal.pt>.

bibliográfica, um sistema de autoridades permite o controlo das formas de cabeçalho e facilita o trabalho do catalogador pois esses dados ficam ligados ao ficheiro bibliográfico. Ao ser criado um registo de autoridade o sistema informático deverá importar essa forma de cabeçalho escolhida para o registo bibliográfico. Quaisquer alterações posteriormente necessárias serão feitas sobre o registo de autoridade, alterando simultaneamente todos os registos bibliográficos a ele ligados (Galvão 2003).

Para melhor observar a relação entre os registos bibliográficos e os registos de autoridade, apresenta-se a correspondência entre os campos dos dois formatos: no formato UNIMARC Autoridades aparecem os campos relativos a cada tipo de cabeçalho, e no formato UNIMARC Bibliográfico aparecem os campos de assunto correspondentes (tab. 1.3).

UNIMARC autoridades		UNIMARC bibliográfico	
Bloco 2xx	Cabeçalho (campos)	Correspondência aos campos do Bloco 6xx (exemplos)	
200	Nome de pessoa	600	\$aVicente,\$bGil,\$f1465?-1537\$xDrama
210	Nome de colectividade	601	\$aFundação Manuel Cargaleiro.\$bMuseu
215	Nome geográfico	607	\$aSerra da Estrela (Portugal)
216	Marca registada	716	\$aKitekat
220	Nome de família	602	\$aBragança, Casa de\$xGenealogias
230	Título uniforme ⁹	605	\$aBíblia
235	Título colectivo uniforme ¹⁰	-	-
240	Autor / Título	604	\$aMozart, Wolfgang Amadeus (1756-1791)\$tDon Giovanni.
245	Autor/ Título colectivo uniforme	604	\$aWilde, Óscar\$tPeças de teatro. Excertos
250	Nome comum	606	\$aEducação
260	Acesso por lugar	620	\$aItália\$dVeneza
280	Forma, género ou características físicas	608	\$aChildren's stories\$jPictorial works

Tabela 1.3: Correspondência de campos UNIMARC (Autoridades e Bibliográfico)

⁹ O Cabeçalho de Título uniforme destina-se a reunir os registos de edições de uma obra publicada sob diferentes títulos, ou para reunir itens de uma série.

¹⁰ O Cabeçalho de Título colectivo uniforme destina-se a reunir os registos das várias obras de um autor com extensa produção.

Na fig. 1.9 podemos ver o exemplo de um registo da base de autoridades de autores¹¹ da Biblioteca Nacional, com o nome de autor e a relação de equivalência, respetivamente nos campos 200 e 400 do formato de autoridades.



Fig. 1.9: Registo de autoridade de nome de autor (visualização UNIMARC)

Na visualização do mesmo registo em formato ISBD, como é comum para os usuários (fig.1.10), observamos o cabeçalho de nome autorizado, e a relação de equivalência UP (usado por) a preceder as formas do nome não autorizadas.

¹¹ PACWEB – Base de Autoridades – Pessoa Física - <http://pacweb.bn.pt/autores.htm> [Consulta: 16-09-2012]

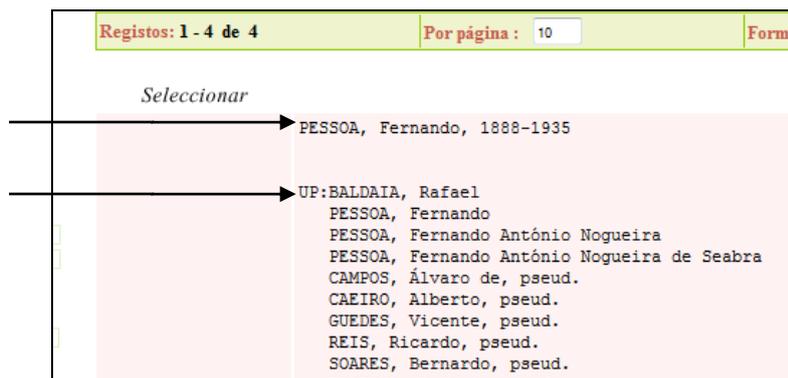


Fig.1.10: Registo de autoridade de nome de autor (visualização ISBD)

Numa pesquisa por assunto de “Fernando Pessoa”, podemos ver no registo recuperado (fig.1.11) que o nome de autor utilizado como assunto corresponde ao campo 600 do formato bibliográfico, ou seja, no formato de autoridade são controladas as variantes do nome no campo 200, que para ser usado como assunto deve corresponder ao campo 600 do formato bibliográfico.

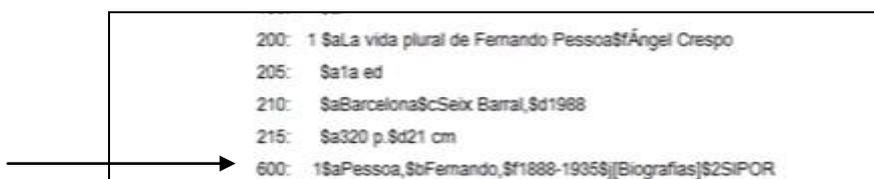


Fig. 1.11: Nome de autor como assunto num registo bibliográfico

Conforme já referido, os registos de autoridade permitem controlar a terminologia e estabelecer relações semânticas através de pistas de referência (fig.1.12); no bloco 4xx do UNIMARC autoridades, são estabelecidas as pistas “Ver”, que contêm cabeçalhos

variantes, e no bloco 5xx são estabelecidas as pistas “Ver também”, que contêm cabeçalhos uniformes relacionados. Nas relações semânticas da estrutura dos tesouros, a pista ‘Ver’ corresponde à relação de equivalência ‘USE’, que é a relação formal entre termos que devem ser representados pelo mesmo descritor ou índice, e a pista ‘Ver também’ corresponde, quer à relação hierárquica, quer associativa ‘TR’ de termo relacionado, que não sendo sinónimo é associado a outro por uma relação coordenada (NP4285-4).

Assunto	
1.	Psicologia infantil
2.	• Ver também: Crianças – Psicologia
3.	Psicologia marxista
4.	Psicologia masculina
5.	• Ver: Homens – Psicologia
6.	Psicologia organizacional
7.	• Ver: Psicologia das organizações
8.	Psicologia patológica
9.	• Ver: Psicopatologia

Fig. 1.12: Pesquisa por assunto com remissivas ‘Ver’ e ‘Ver Também’

Como observámos, podem-se considerar duas classes de registos de autoridade (Caro-Castro 2005):

- Registos de autoridade propriamente ditos, que contêm a informação relativa a um cabeçalho autorizado e as suas referências associadas;

- Registos de referências, que servem para dirigir o usuário, de um termo não autorizado para um autorizado ou para o conduzir a termos relacionados.

A atual expansão da informação em rede contribuiu para a diversificação de serviços de pesquisa e acesso à informação, e essa diversificação é o fator que tem vindo a reforçar a necessidade da construção de vocabulários controlados representados em formato legível por computador, que contribuam para o desenvolvimento da chamada *web* semântica.

As bases de dados de autoridade das bibliotecas são atualmente muito valorizadas pela sua disponibilização na *web* como recursos de informação próprios e não apenas como elementos internos ligados às suas bases bibliográficas. Este aspeto é de extrema importância, na medida em que para além da troca de registos de autoridade entre sistemas bibliográficos, torna possível a sua reutilização por outros sistemas de informação (Manual UNIMARC, 2008).

1.3.2 Controlo de autoridades

Subjacente ao recurso técnico que é o catálogo de autoridades, está a questão mais abrangente do ‘controlo de autoridades’.

Na descrição bibliográfica, conforme já observámos, o controlo de autoridades é utilizado para manter a coerência da forma usada na representação dos pontos de acesso e das relações entre conceitos. As formas normalizadas desses pontos são feitas ao nível dos nomes de pessoas, entidades, títulos uniformes e assuntos, com a finalidade de facilitar a identificação e a recuperação dos documentos, evitando as confusões geradas por homónimos, sinónimos ou a variedade de nomes com que uma pessoa ou uma entidade podem ser denominadas (Martínez et al. 1997).

De acordo com Gorman (2003), a catalogação não pode existir sem pontos de acesso normalizados, e o controlo bibliográfico é impossível sem controlo de autoridade, mecanismo pelo qual se consegue o necessário grau de uniformização. Os pontos de acesso têm as funções básicas de permitir ao usuário do catálogo encontrar o registo e agrupar registos que partilhem uma característica comum. Nesse sentido, o controlo do vocabulário é o aspeto mais importante do trabalho de autoridade. Não se consegue ter um verdadeiro serviço de biblioteca sem uma arquitetura bibliográfica e não se pode ter essa arquitetura sem controlo de autoridades.

O controlo de autoridades é considerado como um recurso inerente às funções do catálogo bibliográfico e foi desenvolvido essencialmente em relação aos cabeçalhos de autores. A maior parte dos estudos sobre esta matéria têm-se dirigido aos pontos de acesso na perspetiva da catalogação, relativamente a autores e títulos, mas nos casos de linguagens de indexação, os fundamentos teóricos são os mesmos, “desenvolver uma linguagem controlada, é aplicar os princípios do controlo de autoridade às rubricas de assuntos”, uma vez que esses fundamentos são os de controlar pontos de acesso, tal

como acontece nos da catalogação. O controlo é concebido de forma a estabelecer uma forma única que constitui o cabeçalho autorizado, com a criação de remissivas para as formas não autorizadas. O principal objetivo é, em todos os casos, o de manter a consistência na representação dos pontos de acesso num catálogo e o processo subsequente de mostrar as relações entre nomes, obras e assuntos (Ribeiro 1996; Wolverton 2006).

A linguagem documental usada na indexação é uma componente muito importante dos catálogos bibliográficos; no seu duplo papel, primeiro no processo de indexação e depois no processo de recuperação de informação, é imprescindível que seja compatível com a linguagem de pesquisa do usuário. Quando isso não se verifica, as pesquisas por assunto resultam insatisfatórias e é posta em causa a credibilidade do sistema, que leva o usuário a abandonar esse recurso para obter informação. A observação destes factos permite-nos afirmar que se entra num círculo vicioso, em que não é dada muita atenção à indexação porque o usuário não faz pesquisas por assuntos, e os usuários não fazem pesquisas por assuntos porque não recuperam a informação específica que desejam (Bocato, Fujita, Leiva 2011).

A prática do controlo de autoridades foi reforçada a partir do alto índice de fracassos constatado pelos usuários na pesquisa dos catálogos *online* nas bibliotecas. É então apontada como principal vantagem da prática desse controlo, a melhoria significativa na precisão e na revocação da informação. A atribuição de termos controlados melhora a precisão (capacidade de limitar uma pesquisa através do controlo da homonímia e da polissemia), e melhora a revocação (capacidade de ampliar uma pesquisa através do

controle de sinónimos e do estabelecimento de relações associativas). O maior requisito de qualquer sistema de recuperação de informação é conseguir juntar o material de um dado assunto, e esse requisito não pode ser alcançado por uma linguagem de recuperação cheia de termos não controlados. O sistema que não vai ao encontro desse requisito é considerado como tendo uma má performance no que respeita à revocação, e no que respeita à precisão se permitir recuperar material irrelevante. A atribuição de termos controlados proporciona orientação ao usuário durante a pesquisa, e por outro lado a garantia de recuperação de resultados, uma vez que esses pontos de acesso nos índices de assunto, garantem a existência de documentos indexados na base. Por outro lado, a falta de controlo dos pontos de acesso origina que numa pesquisa o usuário recupere informação que não vai ao encontro da sua necessidade, como por exemplo numa pesquisa por palavra-chave, como já referimos acerca da linguagem natural, que pode recuperar registos não relacionados com o assunto da pesquisa; nesse caso fala-se de ‘ruído’ que significa uma ‘falsa combinação’, conforme designa a expressão em inglês *false drop*, usada na área da recuperação de informação (Martínez *et al.* 1997; Svenonius 2003). O ruído pode dever-se quer a uma falsa combinação na pesquisa, quer a ambiguidade (polissemia, homografia, sinonímia) ou falta de precisão na indexação.

O trabalho de autoridade tem vindo a ser reclamado, desde a década de setenta, como a parte mais dispendiosa da catalogação, e continuam a ser procurados caminhos para automatizar e simplificar essa tarefa, no sentido de reduzir os seus custos. Foi dado um grande passo nessa direção com o desenvolvimento de meios para partilhar ficheiros de autoridades entre bibliotecas. A tecnologia atual dá oportunidades para ligar esses ficheiros, e abre assim novos serviços para os usuários. O controlo de autoridade

aplicado no atual ambiente da *web* traz precisão às pesquisas, e permite a navegação ao usuário através da estrutura sindética¹² de referências, para variações e formas controladas de títulos e assuntos. A *web* permite aumentar os objetivos de aplicação dos registos de autoridade, passando-se da situação de ficheiros de autoridade individuais de uma instituição, ou de ficheiros partilhados *online*, para uma situação de partilhar ficheiros entre todas as comunidades. A informação de autoridade partilhada traz, para além do benefício de reduzir os custos, a possibilidade de um acesso controlado com todas as suas vantagens. Poder aceder aos registos de autoridade num âmbito internacional, simplifica extraordinariamente a sua criação e manutenção, possibilitando aos usuários a grande vantagem de aceder à informação na própria língua. As facilidades tradicionais do controlo, como criar e atualizar registos, ou apresentar referências nos sistemas de bibliotecas integradas, têm tido maior visibilidade para os catalogadores. No entanto, a abertura desses ficheiros para acesso pela internet, torna-os uma ferramenta útil tanto para os profissionais como para os usuários finais. Estes ficheiros podem ser usados tanto como ferramenta de referência para variantes de nome ou identificação de entidades, mas também como canal para aceder a registos bibliográficos, e a partir daí aceder diretamente a recursos digitalizados (Tillet 2003).

A superioridade de um sistema de informação com controlo de autoridade é óbvia, pois a recuperação de milhares de registos sem ordem nem diferenciação numa pesquisa, não terá grande utilidade para o usuário. É também óbvio que fornecer o vocabulário e controlo de autoridade necessários para fazer uma pesquisa com sucesso com o Google, seria muito moroso e caro. São estes dois fatores que estão no cerne do

¹² Estrutura sindética num catálogo ou índice é a rede de referências do tipo “Ver” e “Ver também”, mostrando o relacionamento entre os cabeçalhos ou descritores (Glossário de biblioteconomia).

dilema da “catalogação da *web*”, ou seja, de ter o mundo da internet e da *web* sob controlo bibliográfico. Para se assegurar precisão nas pesquisas, precisamos de vocabulários controlados, no entanto, não temos capacidade para estender esse controlo à vasta massa marginal de documentos *web*. A solução parece ser não abandonar a questão dos metadados como algo que será sempre de utilidade em grandes bases de dados usadas por bibliotecários e usuários, e investir nos esquemas de metadados com os atributos dos tradicionais registos bibliográficos. O conceito de metadados (*data about data*)¹³ surgiu do desejo dos “não bibliotecários” de melhorar a capacidade de recuperação de páginas *web* e outros documentos na internet. A base deste conceito é de que se consegue um grau de revocação e precisão ao pesquisar em bases de dados, sem os processos morosos e caros da catalogação padronizada, isto é, algo entre a pesquisa em texto livre de motores de pesquisa (que é rápida, barata e ineficiente) e a catalogação completa (que pode ser lenta, de trabalho intensivo, cara, e altamente eficaz) (Gorman 2003).

Os desafios para realizar o controlo de autoridades, ou partilhar esses registos numa escala global, são muitos, pois envolvem questões relacionadas com diferentes sistemas, regras de catalogação, línguas, escritas, desafios técnicos de acesso, formatos de comunicação, etc.; é neste contexto que no mundo das bibliotecas digitais e das bibliotecas em geral, há um crescente foco na necessidade de interoperabilidade (Tillet 2003; Gorman 2003). Estes desafios são repensados no atual contexto da *web* semântica e de *Linked open data*, de que são exemplos alguns catálogos de autoridades como

¹³ Os metadados costumam ser definidos como dados estruturados sobre outros dados; um registo de metadados constitui um conjunto de atributos ou elementos necessários para descrever um recurso (Ríos Hilario 2006).

VIAF (*The Virtual International Authority File*)¹⁴, ou de listas de cabeçalhos de assunto e tesouros em formato RDF/XML com SKOS, como LCSH, RAMEAU, EUROVOC, etc.

A implementação da *web* semântica origina a passagem de um espaço global de informação, com documentos interligados, para outro em que estão ligados documentos e dados, numa “rede estruturada e organizada” onde o elemento principal são os objetos de informação. Na *web* semântica os dados são interligados, expressos em RDF¹⁵ para que os sistemas os possam explorar de forma automática (recompilando ou publicando); os dados convertem-se nos objetos de informação que podem ser documentos, páginas *web*, imagens, vídeos, pessoas, instituições, lugares ou conceitos. Neste contexto o conceito de documento passa a ser entendido como um conjunto de “átomos informativos autónomos”, que por sua vez estão interligados ou interrelacionados. No mundo das bibliotecas e arquivos os índices de assuntos ligados aos tesouros, e especialmente os catálogos de autoridades fundamentam-se num princípio similar: diferentes elementos do registo bibliográfico, que descrevem o documento no seu conjunto, ligam-se com os registos dos catálogos de autoridades (Caro-Castro 2012; Olensky 2010).

Os novos modelos comuns de dados para partilha e ligação de sistemas na *web*, como SKOS (*Simple Knowledge Organization Systems*), e os princípios de *Linked data*¹⁶ vêm facilitar a ligação de dados, de informação e de conhecimento, constituindo uma evolução que vem ao encontro das necessidades atuais de pesquisa de informação. É o

¹⁴ <http://viaf.org/>. [Consulta: 11-06-2013]

¹⁵ RDF (*Resource Description Framework*) é uma linguagem para representar informação na internet.

¹⁶ <http://linkeddata.org/> [Consulta: 19-02-2013]

contexto da *web* semântica, que surge como uma “grande base de dados coletiva”, com a necessidade de agilizar o uso de vocabulários controlados (Caro-Castro 2005).

1.3.2.1 Referências de controlo de autoridades

O aparecimento dos catálogos em linha generalizou a realização do controlo de autoridades¹⁷. Conseguida a normalização da descrição bibliográfica através das ISBD, e a possibilidade do seu processamento informático através do formato MARC, o intercâmbio internacional de registos bibliográficos contribuíram para o acesso universal à informação, o que supunha que um mesmo documento pode apresentar pontos de acesso distintos segundo o país ou centro de catalogação. Além disso, os sistemas de catalogação centralizada e cooperativa necessitavam de uma normalização de todo o processo que passava pelas formas dos cabeçalhos (Wolverton 2006).

A IFLA iniciou trabalhos neste sentido em 1978 com a criação de um grupo de trabalho sobre um Sistema Internacional de Autoridades, que tinha dois objetivos: definir os elementos constitutivos de um registo de autoridade e preparar um formato para o intercâmbio de dados. Os resultados deste grupo de trabalho surgiram em 1983, na apresentação das Normas para ‘autoridades e referências’ “*Guidelines for Authority*

¹⁷ A necessidade de unificar os pontos de acesso nos catálogos foi levantada por Charles Cutter (1961), no entanto dois fatores contribuíram para tornar essa necessidade imprescindível, a expansão literária depois da Primeira Guerra Mundial, e a crescente utilização de nomes de entidades como pontos de acesso nos registos bibliográficos.

and Reference Entries” (GARE),¹⁸ que foram publicadas no ano seguinte. Estas diretrizes referem-se a cabeçalhos relativos a nomes de pessoas ou entidades e a títulos uniformes e, portanto não contemplavam assuntos.

Em relação aos cabeçalhos de assuntos, os sistemas de classificação como a CDU e as listas publicadas para bibliotecas em muitos países, serviam de referência para a normalização desses cabeçalhos. Não sendo considerados suficientes estes instrumentos para normalização, a IFLA formou em 1988 um grupo de trabalho sobre ‘diretrizes para ficheiros de autoridade de assunto’, que elaboraram as normas “*Guidelines for Subject Authority and Reference Entries* (GSARE), publicadas em 1993.

Nas últimas décadas emergiram investigações dedicadas ao controlo de autoridades, assim como a forte tendência da sua internacionalização. Numa perspetiva de catalogação colaborativa, o desenvolvimento de estudos e recursos permite acompanhar o avanço desta área. Constituem referências no panorama internacional os seguintes programas e projetos:

Os programas da *Library of Congress*, NACO e SACO ambos integrados no PCC (*Program for Cooperative Cataloging*) que é uma organização colaborativa dedicada a permitir aos catalogadores criarem e tornarem acessíveis registos bibliográficos e de autoridade estruturados de acordo com normas aceites mutuamente:

¹⁸ Desde a edição das GARE (1984) houve muitos avanços na capacidade tecnológica das bibliotecas, foi estabelecido o formato UNIMARC para autoridades, nasceu a internet, o protocolo Z39.50 foi afinado permitindo uma nova perspetiva para a partilha internacional de ficheiros de autoridade. A sua edição revista e atualizada (2001) deu origem às GARR (*Guidelines for Authority Records and References*).

- NACO¹⁹ (*Name Authority Cooperative Program*), iniciado em 1976, é um modelo cooperativo para construir e manter autoridades de nomes numa base de dados partilhada. O programa permitiu à *Library of Congress* reduzir consideravelmente o investimento numa das mais consumidoras e dispendiosas partes da catalogação, o trabalho de autoridade. É um programa em tempo real com experiência comprovada e de rápido crescimento, pois cada vez mais bibliotecas reconhecem o valor do controlo de autoridade e querem participar nesta construção de uma base partilhada para benefício de todos os membros. O número de parceiros a nível internacional é crescente, pois é um modelo de parceria bem-sucedido entre essa biblioteca e uma comunidade internacional de catalogadores, que tem vindo a demonstrar o poder da colaboração e a importância das normas comuns (Byrum 2003).

- SACO²⁰ (*Subject Authority Cooperative Program*) é descrito como a componente de assunto do programa PCC que permite aos catalogadores submeterem propostas de cabeçalhos de assunto para possível inclusão na lista dos LCSH ou modificação dos existentes, sendo a sua génese em 1981. Com o avanço da tecnologia, em 1992 o programa SACO foi desenvolvido para permitir que os cabeçalhos de assunto desenvolvidos de acordo com as diretrizes da linguagem LCSH pudessem ser adicionados ao ficheiro de autoridades da *Library of Congress*. Como consequência o programa contou com a participação de mais de oitenta instituições, tais como a *British Library*, o *Trinity College Dublin*, as universidades de Oxford e Cambridge e algumas Bibliotecas Nacionais. Um dos grandes desafios do programa SACO é permitir à linguagem LCSH lidar com cabeçalhos de assunto em língua não inglesa,

¹⁹ <http://www.loc.gov/aba/pcc/naco/index.html> [Consulta: 28-01-2013]

²⁰ <http://www.loc.gov/aba/pcc/saco/index.html> [Consulta: 28-01-2013]

uma vez que alguns portais de assunto estão atualmente a ter capacidade de pesquisas multilingues. Este desafio está a ser conseguido através do uso de um “ficheiro de autoridade internacional virtual” (VIAF), orientado para pontos de acesso descritivos (de pessoas, entidades, títulos uniformes, etc.) apresentado mais adiante. O programa SACO continua a fornecer um mecanismo para o desenvolvimento de cabeçalhos do tipo LCSH, que promove a estrutura de um ficheiro padronizado, para poder ser usado de forma uniforme, ajudando assim a prevenir a duplicação de esforços dos que usam esses cabeçalhos (Cristán 2003; Wolverton 2006).

- O projeto AUTHOR, desenvolvido entre 1995 e 1997 com o apoio da Comunidade Europeia, em que participaram as Bibliotecas Nacionais do Reino Unido, França, Bélgica, Espanha e Portugal, demonstrou que as regras referentes à criação de registos de autoridade eram diferentes nos vários países envolvidos. Este projeto limitava-se a nomes de pessoas e entidades, e pretendeu unificar os formatos nacionais em autoridades UNIMARC, e estabelecer uma rede com acesso à lista de autoridades de cada país. Embora fossem seguidos os Princípios de Paris, houve a necessidade de revisão dos princípios estabelecidos e foram desenvolvidas novas diretrizes que resultaram nas “*Guidelines for Authority Records and References*” (GARR), publicadas pela IFLA em 2001 (Santos 2007; Wolverton 2006).
- O projeto MACS²¹ (*Multilingual Access to Subjects*) iniciado em 1997, desenvolveu um estudo para acesso multilingue à informação temática, e foi realizado entre a *Swiss National Library*, a *Deutsche National Bibliothek*, a *Bibliothèque Nationale de France* e a *British Library*, com a finalidade de estabelecer ligações entre as listas de

²¹ http://www.nb.admin.ch/nb_professionnel/projektarbeit/00729/00733/index.html?lang=en [Consulta: 22-11-2012]

cabeçalhos de assunto das quatro instituições. Constitui um ficheiro de autoridade virtual multilingue que foi criado em resposta a um pedido da CENL (*Conference of European National Librarians*) para encontrar uma solução para o problema do acesso por assunto multilingue às bases de dados europeias, através do estabelecimento de ligações entre diferentes linguagens. O projeto MACS foi estabelecido com o objetivo de “providenciar os meios pelos quais as bases de dados das bibliotecas podem ser acedidas de forma multilingue, através do uso de cabeçalhos equivalentes de ficheiros de autoridade de assunto.” A incorporação de uma ligação para os elementos dos registos de autoridade dentro dos ficheiros de autoridade nacionais poderá criar um ficheiro de autoridade multilingue virtual (Clavel-Merrin 2004).

- O projeto VIAF²² (*The Virtual International Authority File*), concebido em 1998 pela *Deutsche Nationalbibliothek* e a *Library of Congress*, teve como primeiro objetivo reduzir os custos de catalogação, tornando acessíveis os registos de autoridade criados por diferentes bibliotecas a nível mundial. Em 2003 a OCLC (*Online Computer Library Center*) juntou-se aos parceiros iniciais com o objetivo de desenvolver o projeto combinando os ficheiros de autoridade de nome de pessoa da Biblioteca do Congresso e da Biblioteca Alemã num único serviço de autoridade. O VIAF tornou-se um projeto conjunto, a que se juntou a *Bibliothèque Nationale de France*, com o objetivo de ligar os ficheiros de autoridade destas bibliotecas nacionais num serviço global acessível aos usuários a nível mundial através da *web*. Este projeto pode ser encarado como um dos blocos básico da construção da ‘*web*

²² <http://viaf.org/> [Consulta: 15-07-2013]

semântica'. A aplicação do controlo de autoridade neste ambiente estende a esse domínio as vantagens da precisão nas pesquisas, para além de permitir a navegação através da estrutura de referências, e uma capacidade alargada para estabelecer ligações. Os modelos VIAF são usados para maximizar os benefícios do trabalho de autoridade partilhado (Tillet 2003; Cristán 2003), constituindo uma ferramenta acessível cujo uso é recomendado pela Google, Microsoft e Yahoo, que se uniram na iniciativa Schema.org²³, para por à disposição uma coleção de vocabulários controlados (Sánchez, Colmenero, Moreiro 2012).

- O projeto LEAF²⁴ (*Linking and Exploring Authority Files*), decorreu entre 2001 e 2004, e teve como objetivo conceber um modelo de arquitetura para estabelecer ligações entre registos de autoridade. Essa arquitetura permite recolher, ligar e dar acesso a informação existente de autoridades de nomes, independentemente da sua criação (em bibliotecas, arquivos, museus) ou de diferenças nacionais. O modelo LEAF é baseado numa ideia simples, em que dados de autoridade de diferentes “fornecedores” serão armazenados num servidor central, que será carregado regularmente para acrescentar novos dados, de cada vez que um usuário fizer uma pesquisa. O usuário é assim integrado no estabelecimento de um ficheiro de autoridades central europeu (Weber 2003).

Da abordagem que fizemos ao controlo de autoridade, verificamos que o interesse em redor desta questão continua a crescer, e que os atuais desenvolvimentos tecnológicos de sistemas permitem torná-lo uma realidade para muitas instituições

²³ <http://schema.org/> [Consulta: 15-07-2013]

²⁴ <http://www.leaf-eu.org/index.html> [Consulta: 28-01-2013]

através de uma partilha à escala internacional. Como já referimos anteriormente, a questão do controlo de autoridades coloca-se atualmente na perspetiva da *web* semântica, como um espaço global de informação em que estão ligados documentos e dados numa “rede estruturada e organizada” (Caro-Castro 2012).

No ambiente em que atualmente se inserem os catálogos, a catalogação não se deve limitar ao conceito de descrição dos documentos pensada para produzir catálogos constituídos por registos autónomos, ignorando a riqueza de relações que se podem estabelecer entre entidades bibliográficas e as possibilidades de visualização de dados que elas permitem (Lopes e Galvão 2012).

No processamento catalográfico ainda existente nas bibliotecas há duas grandes estruturas, a descritiva, que engloba os princípios, códigos e normas que estão na base da descrição bibliográfica, e a estrutura de comunicação que se reflete no modo como os registos bibliográficos são apresentados ao utilizador (Lopes e Galvão 2012). Nestes dois planos têm sido reconhecidos por um lado a desadequação das funcionalidades dos catálogos bibliográficos em relação às exigências atuais, e por outro, o fraco aproveitamento do potencial informativo das bases de dados (Galvão e Cordeiro 2010). Foi com base nestas perspetivas que se deu o desenvolvimento dos chamados “modelos conceptuais”.

Numa época em que o usuário de serviços de informação dispõe de tantas opções e possibilidades de pesquisa, faz-se sentir a necessidade de repensar o papel dos catálogos nos dias de hoje. O desenvolvimento e a utilização de vocabulários controlados

encontram-se atualmente enquadrados em alterações nos princípios internacionais de catalogação, nas regras de catalogação nacionais e nos requisitos funcionais dos registos bibliográficos e de autoridade, e os seus desenvolvimentos continuam a demonstrar o valor atribuído ao controlo de autoridades da informação nos catálogos das bibliotecas. A proposta dos Requisitos Funcionais (FRBR), pensados especificamente como alternativa às ISBD²⁵, e da nova Declaração Internacional de Princípios de Catalogação vão ao encontro desse cenário (Montalvo 2011; Moreno 2009).

Nos contextos das funcionalidades de pesquisa, um dos maiores inconvenientes para melhorar o rendimento dos catálogos, segundo Caro-Castro (2005), é a rigidez dos registos bibliográficos em formato MARC. Este formato, cuja estrutura se baseia nas ISBD e em Regras de Catalogação, foi desenhado para aceder à informação de um registo completo e não para estabelecer relações lógicas entre os elementos informativos que o formam, ou entre registos de diferentes bases de dados. O formato MARC reproduz estas características, adaptando-se bem à comunicação da informação bibliográfica, mas não à manipulação automatizada dos dados. As alternativas têm sido orientadas no sentido de encontrar formatos mais flexíveis que facilitem a criação e a apresentação das relações entre os campos para utilização da estrutura sindética do catálogo. Estas alternativas são particularmente importantes para a pesquisa por assuntos, porque sem fazer uso das relações semânticas que existem nos vocabulários controlados, estes não deixarão de ser o obstáculo até agora existente no acesso temático. As limitações nas possibilidades de pesquisa provocadas pela rigidez deste formato de descrição bibliográfica, impulsionaram o desenvolvimento de um modelo de

²⁵ Sobre a integração das ISBD no ambiente da *web* semântica ver “*Interoperability of ISBD within Linked Data Environment*”.²⁵

descrição baseado nos requisitos funcionais FRBR (*Functional Requirements for Bibliographic Records*) (Caro-Castro 2005).

1.3.3 Requisitos funcionais dos registos bibliográficos e de autoridade

As ISBD foram as normas de descrição de recursos bibliográficos de excelência a nível internacional durante mais de duas décadas, mas a proliferação de recursos e a evolução tecnológica decorridos na década de 90, deram lugar a uma multiplicação de normas, modelos e esquemas para dar resposta aos problemas da descrição bibliográfica (Ríos Hilario 2003, 2006).

Nos anos que se seguiram à publicação da primeira das ISBD, a ISBD(M)²⁶ em 1971, ocorreram uma série de desenvolvimentos no ambiente em que operam os princípios e as normas de catalogação, tais como a expansão de sistemas automatizados, a criação em grande escala de bases de dados bibliográficos pelas agências nacionais de catalogação, a emergência do acesso em rede à informação eletrónica e a novas formas de publicação, alterações essas que originaram a necessidade de uma reavaliação da teoria da catalogação (ODLIS).²⁷

²⁶ *International Bibliographic Description for Monographic Publications*

²⁷ *Online Dictionary for Library and Information Science*

A partir da necessidade de repensar os conceitos subjacentes ao controlo bibliográfico, a IFLA desenvolveu estudos para reavaliar os custos de catalogação no sentido de atualizar a função dos catálogos à luz da diversidade de suportes e das necessidades do usuário. Embora os objetivos do catálogo bibliográfico de Charles Cutter continuem a ser reconhecidos, pode dizer-se que a tentativa mais recente de descrever as suas funções se deu com o estudo dos FRBR. O Primeiro Encontro Internacional da IFLA de “Especialistas em Código de Catalogação Internacional” (IME ICC), realizado em Frankfurt em 2003, foi considerado uma oportunidade para a reavaliação dos Princípios de Paris (1961), à luz dos FRBR e dos atuais ambientes *online* (Ríos Hilario 2003; Tillet 2003).

A iniciativa do estudo dos FRBR (*Functional Requirements for Bibliographic Records*) enquadra-se no reconhecimento da crescente diversificação dos recursos e das alterações tecnológicas dos ambientes de informação. A chamada “FeRBeRerização” dos catálogos, significa a necessidade de se extraírem dos catálogos tradicionais as funcionalidades e redes de relações dos requisitos funcionais (Galvão e Cordeiro 2010).

A introdução do modelo Entidade – Relação trazida pelos FRBR possibilita a reanálise dos objetivos dos catálogos, para que deixem de ser vistos como réplicas dos catálogos em fichas e passem a apresentar a informação de forma mais contextualizada, dando ao utilizador possibilidades de ‘navegar’ pelo universo das entidades no catálogo (Galvão e Cordeiro 2010).

Publicados pela IFLA em 1998 os FRBR²⁸ têm o objetivo de delinear as funções a considerar para o registo bibliográfico, no pressuposto de ir ao encontro das quatro tarefas essenciais para o usuário, de encontrar, identificar, selecionar e obter. Este foi o primeiro de uma série de estudos sobre modelos conceptuais do universo bibliográfico, ao qual se seguiram o estudo para os registos de autoridade, FRAD (*Functional Requirements for Authority Data*), publicado em 2009, e o estudo para os registos de autoridade de assuntos FRSAD (*Functional Requirements for Subject Authority Data*), publicado *online* em 2010.

Numa sucessão de aprofundamentos teóricos, foi assim desenvolvido um conjunto de estudos inter-relacionados que assentam num modelo conceptual (baseado no modelo computacional) do tipo ‘entidade-relacionamento’, e que constituem uma proposta de registo bibliográfico para todos os tipos de materiais e tarefas de usuários, associados aos recursos descritos nos catálogos *online* de bibliotecas (Tillet 2003; Moreno 2007).

Este modelo conceptual de relação entre ‘entidades’ relaciona as tarefas do usuário nos catálogos de bibliotecas e em bases de dados bibliográficas, e constitui uma “abordagem holística” ao acesso e recuperação da informação, no sentido em que as relações entre as entidades contêm *links* que permitem navegar pela hierarquia das relações estabelecidas. É um modelo separado de normas de catalogação específicas como as AACR²⁹ ou ISBD (*International Standard Bibliographic Description*). Nos FRBR as ‘entidades’ podem representar os diferentes aspetos do interesse do usuário

²⁸ Requisitos funcionais dos registos bibliográficos (BNP, 2008).

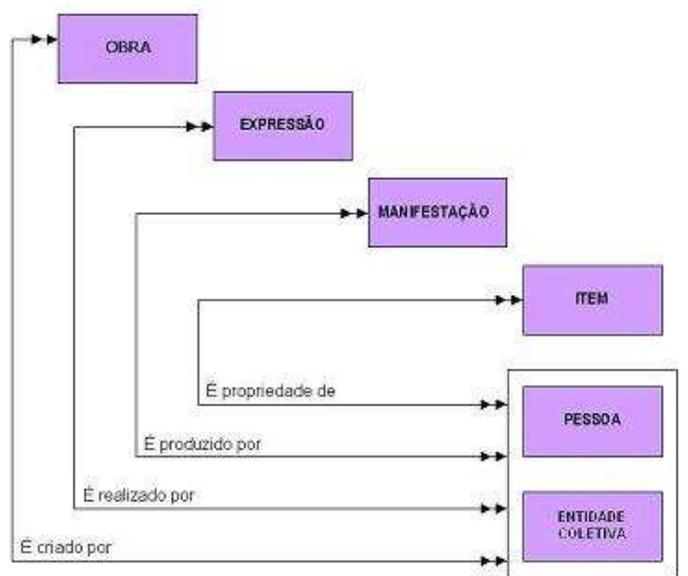
²⁹ *Anglo-American Cataloguing Rules*, com a última revisão em 2005, foram sucedidas pelas RDA (*Resource Description and Access*) em 2010, que constitui o novo código com base nos FRBR, concebido para ter uma estrutura mais flexível adaptada ao ambiente digital.

nos produtos de criação intelectual ou artística, e as ‘relações’ servem para determinar as ligações entre essas ‘entidades’, o que permite ao usuário navegar num catálogo ou base de dados bibliográfica, explorando assim o universo aí representado. O usuário faz a pesquisa utilizando um ou mais atributos da ‘entidade’, e as relações existentes no registo bibliográfico fornecem informação adicional, ajudando a encadear a entidade com outras relacionadas. Os relacionamentos que ocorrem entre entidades proporcionam informação adicional que ajudam o usuário a fazer novas ligações (permitindo navegar pelas entidades que pesquisa e por outras relacionadas). Através desses relacionamentos, o registo bibliográfico deve permitir descobrir e obter a entidade descrita, e esse é o nível básico de funcionalidade que o catálogo deve preencher. Esta característica fez com que o modelo sirva de base para estruturar ontologias ou modelos de metadados na *web*. Os relacionamentos (representados adiante na fig.1.13) podem ser entre obras, entre expressões e entre manifestações de diferentes obras (Moreno 2007; Pacheco 2003; Ríos Hilario 2003).

Quando o modelo FRBR foi elaborado, foi reconhecido que era necessária a realização de outros estudos sobre registos de autoridades, bem como um maior detalhe sobre assuntos, pouco tratados no modelo original. Surgiram então os Requisitos Funcionais para Dados de Autoridade - FRAD (*Functional Requirements for Authority Data*), o modelo conceptual de relação de entidades, para relacionamento de dados contidos nos registos de autoridade, e assim facilitar a partilha de informação de acordo com as necessidades do usuário (Moreno 2009).

No FRAD, tal como no modelo original FRBR, as entidades, atributos e relacionamentos foram estabelecidos em relação às quatro tarefas do usuário, no sentido de explicar como os dados de autoridade apoiam essas mesmas tarefas. Os usuários dos dados de autoridade são: os profissionais de informação que criam e mantêm dados de autoridade, os intermediários, e os usuários finais que pesquisam para satisfazer as suas necessidades de informação, seja através do acesso direto aos dados de autoridade, ou indiretamente, através dos pontos de acesso controlados (formas autorizadas do nome, formas variantes do nome, referências, etc.) presentes em catálogos, bibliografias nacionais, ou outras bases de dados similares, etc. (FRAD, 2009).

Para melhor observação dos relacionamentos, identificamos as entidades dos dois grupos base (fig.1.13):



Fonte: IFLA (1998)

Fig. 1.13: FRBR – Modelo Entidade-Relacionamento Grupo 1 e 2

Grupo 1 - *Contém as entidades que são produto de trabalho intelectual ou artístico:*

Obra - é a criação intelectual ou artística de um determinado autor, independentemente do seu formato físico e da forma do seu conteúdo;

Expressão - é a realização intelectual ou artística específica que assume uma obra (som, imagem, objeto, etc.);

Manifestação - é a materialização de uma expressão de uma obra (quando há alteração na forma física do suporte, o produto resultante é considerado uma nova manifestação);

Item - é um exemplar único de uma manifestação.

Grupo 2 - *Contém as entidades responsáveis pelo conteúdo intelectual ou artístico representado no grupo 1 (autores, compositores, artistas, editores, tradutores):*

Pessoa - pessoas que refletem o seu papel no que respeita à obra, à expressão, à manifestação ou ao item;

Entidade colectiva - organizações que refletem o seu papel no que respeita à obra, à expressão, à manifestação ou ao item.

O modelo FRAD e o que se lhe seguiu, de Requisitos Funcionais para Dados de Autoridade de Assunto (FRSAD) (*Functional Requirements for Subject Authority Data*), constituem uma extensão dos modelos anteriores para classificação e indexação. Os FRSAD são o modelo de relacionamento de entidades que surge na continuação de trabalho efetuado no modelo FRBR, e explica como as entidades do grupo 3 (que servem como assuntos da criação intelectual ou artística) podem ser relacionadas e

controladas a nível bibliográfico). O Grupo 3 inclui os elementos adicionais de representação de conteúdo dos registos bibliográficos, que são as informações relacionadas aos pontos de acesso por assunto. Os FRSAD constituem uma extensão dos estudos iniciados nos FRBR. Este e o anterior FRAD complementam e desenvolvem alguns aspetos do modelo conceptual original FRBR. O conjunto dos três modelos é designado a família FRBR, no sentido em que são considerados partes de um modelo geral maior (IFLA, 2010). A chamada “FRBR *family*”³⁰ consiste, assim, em três modelos conceptuais, cada um cobrindo um aspeto da informação registada nos registos bibliográficos e de autoridade.

As entidades previstas, no seu conjunto, aparecem assim reunidas em três grupos, conforme representado na fig. 1.14: as do grupo 1 (*obra, expressão, manifestação e item*) representam os produtos da criação intelectual ou artística; as do grupo 2 (*pessoa, instituição, família*) representam os responsáveis pelo conteúdo intelectual ou artístico, a produção física e disseminação de tais produtos; e as do grupo 3 (*conceito, objeto, evento, lugar*) representam os assuntos relacionados com as criações intelectuais ou artísticas dos grupos 1 e 2 (Galvão e Cordeiro 2012; Hatsek 2012; Pacheco 2003; Ríos Hilario 2006; Zavalina 2012).

³⁰ <http://www.ifla.org/node/2016> [Consulta: 21-11-2012]

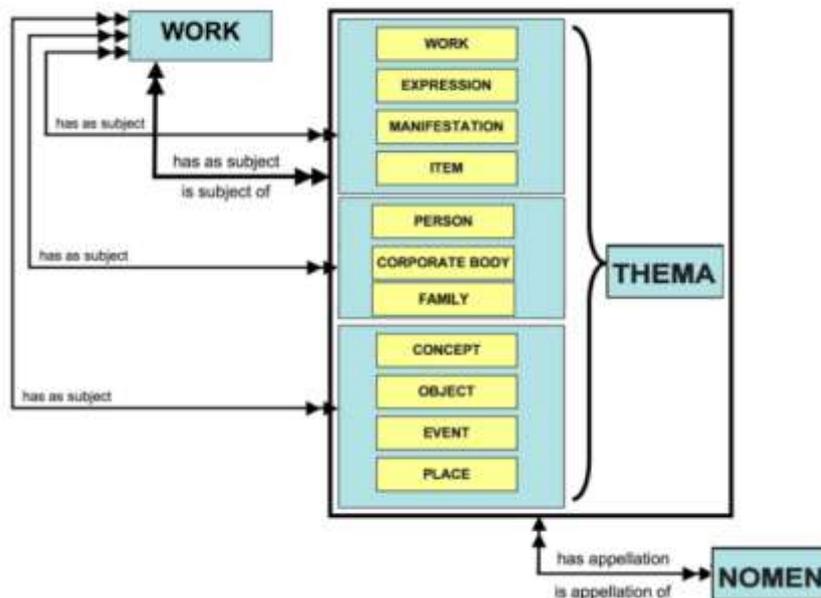


Fig. 1.14: Relação FRSAD com FRBR (com adição da família de entidades FRAD)³¹

Numa representação geral dos FRBR como se observa na fig.1.14, o diagrama representa as relações de “assunto” entre a obra e as entidades do grupo 1, grupo 2 e grupo 3. As entidades no grupo 3 representam um conjunto adicional de entidades que servem como os assuntos das obras.

Neste diagrama foram introduzidas duas entidades:

Thema – qualquer entidade utilizada como matéria (assunto) de uma obra;

Nomen – qualquer signo ou sequência de signos (caracteres alfanuméricos, símbolos, sons, etc.) mediante os quais se conhece, se refere e se aborda um *thema*.

³¹ Adaptado de: “Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model” IFLA 2010.

Foram eleitos os termos latinos *Thema* (plural *Themas*) e *Nomen* (plural *Nomens*) por serem culturalmente neutros e não requerem tradução. Qualquer Obra pode ter mais de um *Thema* e qualquer *Thema* pode ser matéria (assunto) de mais que uma Obra.

Na conclusão do relatório FRASAD é referido que este modelo:

- Constitui um modelo conceptual no quadro dos FRBR que se refere ao “*aboutness*”³² das obras;
- Define um marco estruturado de referência para relacionar os dados dos registos de autoridade de assunto com as necessidades dos usuários;
- É desenvolvido com o objetivo de colaborar numa avaliação do potencial de intercâmbio internacional e da utilização de dados de autoridade de assunto;
- Melhora as considerações para os requisitos funcionais de dados de matéria a um nível independente de qualquer aplicação, sistema ou contexto específico.

³² A totalidade de matérias explícitas ou implícitas contidas no texto de um documento, incluindo mas não limitado ao(s) significado(s) do título, às intenções declaradas ou não declaradas do autor, e as formas em que a informação pode ser usada por leitores (ODLIS. *Online Dictionary for Library and Information Science*).

As tarefas realizadas pelos usuários, no conjunto dos modelos são as seguintes:

FRBR	FRAD (autoridade)	FRSAD (autoridade assunto)
Encontrar (materiais)	Encontrar (autoridade)	Encontrar (matéria)
Identificar (entidade)	Identificar (autoridade)	Identificar (matéria)
Selecionar (entidade)	Contextualizar (situar autoridade)	Selecionar (matéria)
Obter (acesso)	Justificar (autoridade)	Explorar (relações matéria)

Tabela 1.4: Definição das tarefas dos usuários nos três modelos

Os FRBR refletem as necessidades dos usuários ao realizarem pesquisas:

Encontrar - materiais (que correspondam aos critérios de pesquisa)

Identificar - uma entidade (nos dados recuperados)

Selecionar - uma entidade (adequada às necessidades do usuário)

Obter - acesso (à entidade selecionada)

No contexto dos FRAD, as tarefas do usuário são as seguintes:

Encontrar – entidades correspondentes ao critério de pesquisa; ou explorar o universo de entidades bibliográficas;

Identificar – uma entidade (distinguindo entidades com características similares) ou validar a forma do nome a ser usada como ponto de acesso controlado;

Contextualizar – situar (pessoa, coletividade, obra etc.) num contexto pelo qual é reconhecida

Justificar – documentar a razão da escolha, pelo criador da autoridade, de um nome ou forma do nome pelo qual o ponto de acesso foi controlado.

No contexto dos FRSAD, as tarefas dos usuários são as seguintes:

Encontrar – matérias que correspondam aos critérios estabelecidos pelo usuário

Identificar – uma matéria distinguindo-a entre outras com características similares

Selecionar – uma matéria adequada às necessidades do usuário

Explorar – as relações entre as matérias

A opinião generalizada é de grande confiança no modelo FRBR para melhorar a capacidade representativa do registo bibliográfico e, simultaneamente, o desempenho dos catálogos (Pacheco, 2003). De acordo com Tillet (2003) os FRBR têm um grande impacto na catalogação, e a aplicação deste modelo é encorajada pelos especialistas, no sentido de facilitar a padronização internacional e reduzir os custos com a catalogação à escala global, prevendo-se que tenham uma profunda influência no projeto de sistemas futuros.

O modelo FRBR representa a base para a evolução dos princípios e regras de catalogação a nível internacional, e a par dos modelos conceptuais surgiu a necessidade de reavaliar os princípios subjacentes à elaboração de catálogos, que não eram ‘revisitados’ desde os “Princípios de Paris 1961” (Galvão e Cordeiro 2010).

1.3.4 Os novos princípios de catalogação e o código RDA

A partir das alterações e desenvolvimentos que originaram a reavaliação da teoria da catalogação, surge a nova Declaração de Princípios Internacionais de Catalogação³³, desenvolvida pela IFLA e publicada em 2009. Esta Declaração vem substituir a Declaração de Princípios de Paris (1961), e surge com um âmbito ampliado ao incluir todos os tipos de materiais para além dos textuais, e todos os aspetos dos dados bibliográficos e de autoridade utilizados em catálogos OPAC. A nova declaração tem por base os anteriores princípios catalográficos e o recente modelo conceptual FRBR (*Functional Requirements for Bibliographic Records*), e pressupõe essencialmente facilitar o intercâmbio internacional de dados bibliográficos e de autoridade, através de orientações para o desenvolvimento de códigos de catalogação (IFLA 2009).

Na introdução da Declaração é expressa a sua atualidade:

“[...] Passados quarenta anos, ter um conjunto comum de princípios internacionais de catalogação tornou-se ainda mais desejável uma vez que catalogadores e respectivos clientes, em todo o mundo, usam OPAC (Online Public Access Catalogues). Neste momento, ao alvorecer do século XXI, a IFLA desenvolveu um esforço para produzir uma nova declaração de princípios que se destinam aos catálogos de bibliotecas em linha e outros. O primeiro princípio é servir a conveniência dos usuários do catálogo.”

³³Disponível online http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-pt.pdf; corresponde aos “*Statement of International Cataloguing Principles*”(ICP).

As orientações principais da Declaração dizem respeito aos pontos: ‘Objetivos e funções do catálogo’, ‘Descrição bibliográfica’, ‘Pontos de acesso’, e ‘Fundamentos para a funcionalidade de pesquisa’.³⁴

O princípio da “conveniência do usuário” é considerado o mais determinante para orientar a construção de códigos de catalogação, expressando que “as decisões relativas à descrição e a formas controladas dos nomes para acesso devem ocorrer tendo em conta o usuário.”

Em relação a “entidades, atributos e relações” são referidos os mesmos elementos definidos nos modelos conceituais:

- Entidades que podem ser representadas por dados bibliográficos e de autoridade (obra, expressão, manifestação, item, pessoa, família, coletividade, conceito, objeto, evento, lugar);
- Atributos (que identificam cada entidade);
- Relações (entre entidades, que sejam bibliograficamente significativas).

Os conceitos trazidos pelo modelo conceptual FRBR e pelos Princípios Internacionais de catalogação correspondem à consciencialização de novas exigências e também à reformulação de necessidades antigas (Lopes e Galvão 2012).

³⁴ Está também incluído um glossário, onde é referido que “alguns termos usados nos Princípios de Paris ou em regras de catalogação antigas” já não são usados nos Princípios Internacionais de Catalogação, como por exemplo os que usamos frequentemente neste trabalho: Cabeçalho (*Ver* Ponto de acesso autorizado, Ponto de acesso controlado); Referência (*Ver* Forma variante de nome).

A reanálise dos objetivos dos catálogos através dos FRBR influenciou a preparação de revisões dos códigos de catalogação, e contextualizou o surgimento do novo código de catalogação RDA³⁵ (*Resource Description and Access*) publicado em 2010.

Os Princípios Internacionais de Catalogação foram indispensáveis para a normalização internacional dos códigos de catalogação existentes e para a evolução da catalogação, porque trouxe uma nova visão sobre o tratamento de recursos bibliográficos. O estabelecimento do RDA mudou a forma de se pensar da catalogação, representando uma mudança nessa prática. Este código apresenta uma estrutura para descrever todos os tipos de recursos, dando uma nova abordagem à descrição nos registros bibliográficos, tendo como base os modelos FRBR e FRAD. O RDA garante melhor navegação no catálogo e uma melhor visualização dos registros bibliográficos, define os elementos dos dados com precisão, e mantém a ligação com outras comunidades de metadados.

Por ser uma norma de conteúdo de dados, o RDA é aplicável a diversos recursos de informação, tais como documentos bibliográficos, arquivísticos ou museológicos, em formato físico ou eletrônico, e traz maior coerência na realização de pesquisas (Hatsek 2012). O novo código é uma realidade a nível internacional, que tem vindo a gerar profundas mudanças no processo de catalogação e informatização dos registos nas bibliotecas, e estas mudanças precisam de ser assimiladas pelos profissionais e paulatinamente colocadas em prática (Hatsek 2012).

³⁵ <http://www.rdatoolkit.org/> [Consulta: 15-07-2013]

As alterações ao nível da informatização, da crescente complexidade dos conteúdos, e da publicação dos modelos relacionais, foram o contexto em que surgiu o código RDA. Um dos seus aspetos mais relevantes é a independência relativamente a qualquer estrutura ou formato específico para visualizar ou armazenar metadados. O código apresenta-se como uma norma de conteúdo para descrever e proporcionar o acesso a recursos. A forma de registar a informação permite que esta possa ser codificada independentemente do esquema de metadados. Podem ser usados esquemas já existentes, como MARC21, UNIMARC, Dublin Core, etc., ou outros futuros. Torna-se assim possível assegurar o aspeto fundamental da integração dos dados criados com base neste novo código, com os dados já armazenados, que foram criados de acordo com regras de catalogação pré-existentes, sejam elas as AACR (*Anglo-American Cataloguing Rules*) ou outras. Embora na perspetiva dos catalogadores as regras do novo código devam fazer o que as AACR têm feito nos últimos 40 anos, o ambiente tecnológico em rápida mudança cria novas exigências. É um dos primeiros códigos gerais de catalogação a ser criado desde que o suporte digital começou a suplantar o impresso, tornando-se dominante. Assim, o RDA pode ser aplicado usando a tecnologia e as bases de dados atuais, por exemplo, em registos MARC armazenados nos catálogos “tradicionais”, e pode também ser usado no ambiente *web* e em novos tipos de estruturas de bases de dados (Danskin 2013; Lopes e Galvão 2012).

O futuro do controlo bibliográfico está muitas vezes associado com o RDA, cuja estrutura é largamente baseada nos requisitos funcionais, pois este código é de facto uma implementação de dois modelos, o FRBR e o FRAD, que definem as entidades ou objetos de interesse especificados no RDA, e o conjunto dos seus elementos

corresponde aos atributos e relações definidas para essas entidades (Danskin 2013; Zavalina 2012). O RDA está disponível *online* através de uma ferramenta chamada *RDA Toolkit*.³⁶

1.4 Avaliação da qualidade da indexação

Avaliar a qualidade da indexação é de extrema importância, uma vez que esta constitui a tarefa central dos sistemas documentais, e está diretamente ligada à eficiência da recuperação de informação. O conceito de qualidade deve ser tido em conta na pesquisa dos catálogos bibliográficos, e a manutenção dessa qualidade deve ser feita através dos pilares do controlo de autoridades, e da respetiva avaliação (Albás 2005; Extremeño 1999).

É de salientar que a questão da avaliação de bibliotecas implica algum discernimento, e sobretudo uma sustentação sólida. Embora seja um processo fundamental para conhecimento e aproximação às expectativas dos usuários, deve-se ter em conta a realidade única de cada instituição resultante de uma série de fatores, que dificultam a utilização de normas gerais para referências comparativas (Gómez Hernandez 1995).

³⁶ <http://www.rdatoolkit.org/> [Consulta: 15-07-2013]

No seguimento de reflexões teóricas ligadas ao nosso estudo, é importante fazer uma breve referência aos principais pontos da avaliação da qualidade em sistemas de informação, pois é a forma que permite reorientar o desenvolvimento de trabalhos nessa área. Assim, depois de abordar os antecedentes de avaliação de Sistemas de Recuperação de Informação, observam-se em linhas gerais os enfoques metodológicos e as medidas usadas para avaliar a eficácia desses sistemas.

1.4.1. Antecedentes de avaliação de Sistemas de Recuperação de Informação

A avaliação da qualidade da indexação em bases de dados documentais enquadra-se na perspectiva mais abrangente da avaliação de Sistemas de Recuperação de Informação (SRI). Muitos dos conceitos envolvidos que vamos expor têm origem em estudos determinantes nesta temática, iniciados nos anos 50. Os principais projetos de avaliação de SRI que apresentamos são referidos por vários autores em estudos que se foram sucedendo (Bocato e Fujita 2006, 2010; Fidel 1991b; Melo 1994; Muddamalle 1998; Ribeiro 1996; White e Griffith 1987).

- Os Projectos Cranfield I e II:

O Projecto Cranfield I, desenvolvido de 1957 a 1962 e orientado por Cleverdon, teve o objetivo de conhecer a eficiência de sistemas de indexação, através de um método objetivo de avaliação comparada. Foram comparados quatro sistemas, uma classificação facetada, a CDU, cabeçalhos alfabéticos de assuntos e um sistema *Uniterm*. As principais conclusões retiradas deste projeto foram: - que a linguagem de indexação usada não alterava substancialmente a taxa de revocação “*versus*” taxa de relevância, e que alguma variação significativa na curva desses valores se devia a inadequadas decisões intelectuais sobre os conceitos do assunto na indexação ou na pesquisa; - que se verificava uma melhoria na indexação à medida que os indexadores adquiriam conhecimentos e experiência; - que o fator isolado que mais contribuiu para a presença de falhas nos sistemas foi o erro humano, tanto na indexação como na pesquisa. A principal contribuição deste projeto foi a nível teórico, na definição dos conceitos de revocação e relevância e na formulação dos critérios de especificidade e exaustividade (Melo 1994).

O Projecto Cranfield II, que surgiu em 1963 na sequência dos resultados obtidos no primeiro, foi de carácter puramente experimental e comparou a eficácia de recuperação de informação, medida em termos de precisão (*precision*) e revocação (*recall*) de 33 linguagens de indexação, incorporando vários graus de controlo. As principais conclusões retiradas deste projeto foram: - que a utilização de termos simples retirados dos documentos deu bons resultados, piorando, no entanto, quando a linguagem natural foi utilizada para exprimir

conceitos com termos compostos; - que os vocabulários controlados originaram piores resultados que os termos simples, exceto na expressão de termos compostos em que se revelaram muito melhores. O seu resultado mais citado foi que um controlo mínimo do vocabulário, em que só os sinónimos e a terminação das palavras eram normalizados, dava resultados tão bons, e por vezes ainda melhores que um total controlo de vocabulário (Melo 1994; Svenonius 2003).

- O Sistema MEDLARS (*Medical Literature Analysis and Retrieval System*) é um sistema informatizado de recuperação de referências criado pela *National Library of Medicine* nos Estados Unidos. Num projeto iniciado em 1965, conduzido por Lancaster, pretendeu-se avaliar o comportamento do sistema e analisar os fatores que afetam a recuperação da informação e as necessidades para uma maior eficiência e menor custo. Foi o maior teste de avaliação realizado até hoje, no que se refere à quantidade de documentação envolvida (cerca de 800.000 documentos). Teve a importante característica de ter sido efetuado num sistema de informação em funcionamento, e não num ambiente laboratorial. A análise dos resultados permitiu calcular a média de valores característicos da eficácia do sistema, bem como identificar falhas, no que respeita a documentos incorretamente recuperados e documentos relevantes não recuperados. O teste funcionou como indicador do desempenho do sistema, mas também como meio de corrigir deficiências detetadas. As principais conclusões retiradas da avaliação do sistema foram as seguintes: - em relação à linguagem de indexação, os fracassos deveram-se à falta de especificidade e a relações incorretas entre os termos, e à inadequada exaustividade na indexação, quer

insuficiente ou em demasia; - em relação ao processo de indexação os erros humanos deram origem a muitas falhas, quer por atribuição inadequada de termos, quer por omissões; - em relação ao processo de pesquisa revelaram-se muitas falhas, e o uso da linguagem natural proporcionava melhores resultados nas pesquisas heurísticas (em que o usuário modificava a sua estratégia de acordo com o resultado que ia obtendo) do que nas pesquisas interativas. Os principais melhoramentos propostos a partir deste projeto foram no sentido de melhorar a especificidade da linguagem de indexação (Melo 1994; Ribeiro 1996).

- Projecto Aberystwyth, realizado em 1968 e levado a cabo por E. Michael Keen, comparou três linguagens pós-coordenadas de indexação, nomeadamente de termo condensado (vocabulário mínimo de uma lista), linguagem natural não controlada (palavras simples derivadas dos documentos) e linguagem estruturada hierarquicamente (baseada numa classificação). As principais conclusões retiradas foram sobre o desempenho dos sistemas em função da revocação, que indicaram ser o termo condensado o melhor sistema em função de tempo de duração da pesquisa, e que o desempenho do sistema numa linguagem pré-coordenada era pior. Este estudo revelou que não são significativos os ganhos decorrentes do uso de linguagens complexas de indexação (Melo 1994).

- Projecto SMART (*System for the Mechanical Analysis and Retrieval of Text*), orientado por Gerard Salton em 1965 nos Estados Unidos, no qual se estudou a

avaliação de sistemas informatizados de recuperação de informação, em que as pesquisas eram realizadas em texto livre, sem uso de qualquer controle no vocabulário de indexação e de pesquisa. As principais conclusões retiradas foram que a análise automática da linguagem era tão eficaz na recuperação quanto a obtida por indexação intelectual (não automática), e que a utilização do procedimento com o thesaurus para reconhecer sinónimos e termos relacionados melhorava a eficácia na recuperação da informação. Apesar de algumas críticas, o projeto SMART permitiu demonstrar limitações inerentes aos SRI, e definir ações para melhorar o seu desempenho. A principal conclusão foi a de que não se justificava realizar a indexação manual, com uso do vocabulário controlado, pois a linguagem natural e a pesquisa em texto livre revelavam a mesma eficácia na recuperação da informação (Melo 1994; Ribeiro 1996).

As contribuições destes projetos têm servido de base a todas as investigações relacionadas. A partir dos anos setenta é difícil generalizar conclusões nesta matéria, tendo sido divulgados outros testes de avaliação de desempenho a par dos avanços tecnológicos e do aparecimento de numerosos sistemas de recuperação de informação. Deve-se salientar que todas essas experiências se basearam nos conceitos e nos métodos definidos nos projetos pioneiros de Cranfield (Melo 1994; Ribeiro 2005), e a linha de trabalho da avaliação da indexação em função dos resultados da recuperação em entornos controlados tem tido continuidade nas conferências TREC (*Text REtrieval Conferences*).³⁷

³⁷ <http://trec.nist.gov/> [Consulta: 17-06-2013]

1.4.2. Metodologias de avaliação

A avaliação de Sistemas de Recuperação de Informação (SRI), de um modo geral, tem sido realizada em torno do desempenho de diferentes sistemas, para testar a sua eficácia e avaliar resultados de pesquisa, no sentido de determinar a qualidade dos instrumentos e das linguagens de indexação. No entanto, a tendência da avaliação evoluiu no sentido de enfatizar mais as necessidades dos usuários e ir ao encontro dos seus comportamentos na pesquisa de informação (Ribeiro 2005).

Desde finais da década de oitenta que a avaliação de SRI colocou os usuários dos sistemas no centro das investigações. Foi reconhecido que, para determinar a eficácia dos sistemas, as variáveis relacionadas com os usuários são tão importantes como as relacionadas com os métodos de indexação. A complexidade dessas variáveis tem dificultado determinar, se os resultados da avaliação dos sistemas decorrem diretamente do tipo de indexação utilizada (humana ou automática), ou de questões particulares ligadas às próprias variáveis, como diferentes níveis de exaustividade na indexação, diferentes níveis de especificidade de vocabulário, de diferentes opções de pesquisa, ou da combinação de algumas dessas características (Anderson e Pérez-Carballo 2001b).

Neste enquadramento há duas linhas de investigação fundamentais na área da recuperação de informação: uma com o enfoque centrado no sistema, designada tradicional, e outra com o enfoque centrado no usuário, designada cognitiva e sociológica. Para o enfoque tradicional, o sistema é o mecanismo que permite armazenar e recuperar a informação, tendo a perspetiva cognitiva ampliado os limites

do “sistema”, com a introdução do “fator humano” e do ambiente em que se desenvolve a pesquisa (Caro-Castro 2005).

Estas linhas de investigação estão relacionadas com os paradigmas epistemológicos da Ciência da Informação, que se inter-relacionam: o paradigma físico (centrado no sistema), o paradigma cognitivo (centrado no usuário) e o paradigma social (interação do usuário individual e do ambiente organizacional) no processo de recuperação de informação (Bocato e Fujita 2010). Foi a partir da década de 1990 que alguns investigadores inseriram o elemento ‘contexto’ ao processo cognitivo individual, evidenciando assim uma visão sociocognitiva no desenvolvimento desse processo.

Os métodos de avaliação de linguagens documentais em SRI inserem-se, assim, nesses paradigmas, com os seus respetivos “objetos”:

- no paradigma físico, o catálogo *online* que permite a comunicação com o usuário, através dos registos bibliográficos com termos de indexação;
- no paradigma cognitivo, o usuário que se coloca numa posição ativa, em contexto individual, ao elaborar conceitos de pesquisa segundo as suas necessidades de informação;
- no paradigma sociocognitivo que acrescenta o ‘contexto’ na realização desse processo individual, ou seja, o usuário num contexto socio cultural.

Na linha de investigação centrada no usuário, as metodologias e variáveis utilizadas são heterogéneas, pois há uma grande variedade na definição de variáveis e na sua

utilização. A categoria do ‘processo de pesquisa’ é a menos uniforme e a do ‘entorno de pesquisa’ é a mais difícil de caracterizar. Embora não seja clara a vinculação desta linha de investigação com uma metodologia concreta, há uma certa homogeneidade nas técnicas de recolha de dados e nos métodos de análise. Há um predomínio da análise transaccional que dá informação suscetível de ser analisada quantitativa e qualitativamente. A utilização de técnicas e métodos qualitativos é imprescindível quando o objetivo não é só descrever o processo, mas também entender as causas e consequências das ações realizadas durante o mesmo (Caro-Castro, Cedeira, Travieso, 2003).

Embora na avaliação de SRI tenham sido feitos estudos³⁸ com aplicação de abordagens quantitativas, a investigação qualitativa tem-se vindo a desenvolver em detrimento dos enfoques quantitativos que vigoraram durante cerca de meio século. A mudança de perspectiva contextualiza-se na alteração de paradigma na própria Ciência da Informação, em que o enfoque nas problemáticas do âmbito psicológico ou psicocognitivo e sociológico é inevitável, pois a informação é encarada como um fenómeno humano e social (Ribeiro 2005).

A metodologia qualitativa, ligada ao paradigma cognitivo, diz respeito aos processos de tratamento temático e de recuperação da informação, relevando em muitos deles a participação do indexador, para intervenção eficaz nos sistemas de informação. Os estudos que utilizam a metodologia qualitativa cognitiva evidenciam a subjetividade da atividade intelectual dos indexadores e dos usuários no desenvolvimento desses

³⁸ Primeiros trabalhos com utilização da abordagem quantitativa: Cleverdon e Lancaster (1964), testes da ASTIA, projectos Cranfield I e II, sistema Medlars, e teste de Aberystwyth.

processos. O contexto sociocognitivo liga-se à representação e à recuperação da informação, na medida em que o indivíduo está inserido no seu contexto sócio cultural e histórico, pressupondo que a necessidade de informação do usuário é influenciada pelo seu meio, e que será representada de acordo com os seus modelos mentais, associados à sua concepção do mundo através da sua cultura e história de vida pessoal, refletindo-se nos seus interesses concretos de informação (Dal'Evedove e Fujita 2008).

A abordagem qualitativa cognitiva trouxe inúmeras contribuições através do modelo de avaliação que considera o usuário como um elemento fundamental. Este modelo utiliza instrumentos de recolha de dados como o *sense-making* e o protocolo verbal, que permitem refletir as suas necessidades reais. Esta abordagem está presente em estudos sobre o tratamento temático e a recuperação da informação, evidenciando as práticas dos indexadores e dos usuários no desenvolvimento destes processos, e o fator de subjetividade relacionado com ambos (Fujita 2003; Ingwersen 1982, 2003; Vargas-Quesada, Moya Anegón, Olvera Lobo 2002). Embora a pesquisa de informação e respetiva recuperação seja a conjugação integrada do sistema, do próprio usuário e da linguagem documental adotada, a participação do indexador para intervenção eficaz no bom desempenho desta, é um fator essencial na obtenção de resultados de pesquisa satisfatórios (Boccatto e Fujita 2006). Esta perspetiva é muito interessante para o nosso estudo, uma vez que na qualidade dos índices de assunto que observamos, está inerente a participação do indexador na construção dessa linguagem, cujos termos serão utilizados nas interrogações ao sistema de pesquisa.

1.4.3. Medidas de avaliação

As novas possibilidades criadas pela informática permitiram o desenvolvimento de projetos de avaliação de Sistemas de Recuperação de Informação, no sentido de determinar a eficácia do seu desempenho, e também para detetar falhas e estudar formas de melhorar a sua qualidade. A avaliação de qualidade de bases de dados é de grande utilidade, e “não deve ter por objetivo realçar o bom funcionamento de um produto de informação, mas antes centrar-se nos aspetos que precisam de ser melhorados” (Rodríguez Yunta 1998).

Como vimos anteriormente, a indexação é um processo que tem vindo a ser avaliado de diferentes modos desde a década de 60, e um deles tem a ver com as propriedades, estrutura e características das linguagens. No âmbito das bases bibliográficas a avaliação da linguagem documental utilizada na indexação é de extrema importância, quer do ponto de vista do indexador, quer do usuário pois permite verificar a eficácia no desempenho do sistema. A qualidade da indexação determinará tanto a pertinência como a exaustividade na pesquisa, e estas medidas são determinantes para a satisfação do usuário. Estas questões têm uma importância contínua, pois o aumento da informação eletrónica trouxe maior exigência no nível de qualidade, e essa exigência significa uma “maior preocupação em relação à fiabilidade da informação” (Boccatto e Fujita 2006; Extremeño 1999; Soler Monreal 2009).

As medidas usadas na avaliação dos sistemas (*recall* e *precision*) foram definidas a partir dos projetos de Cranfield. Avalia-se a eficácia de um sistema pela taxa de revocação¹ (*recall ratio*) e pela taxa de precisão (*precision ratio*). A revocação é a percentagem dos documentos relevantes existentes na base de dados que são recuperados numa pesquisa, e num sistema eficaz todos os que são relevantes deverão ser recuperados. A precisão é a percentagem de material recuperado relevante em resposta a uma pesquisa, isto é, todos os registos recuperados devem ter uma relação exata com os termos de pesquisa. Todos os sistemas de recuperação de informação dependem destas duas medições cruciais (Gorman 2003; Svenonius 2003).

A determinação dessas taxas tem inerente o conceito de relevância, isto é, a utilidade de um documento pela informação de interesse em função de uma questão; a relevância significa, pois, a exatidão com que um conceito é representado por um descritor, não devendo este ser nem mais específico nem mais genérico do que o conceito (Svenonius 2003). O princípio da relevância diz respeito à necessidade de utilizar descritores que sejam um reflexo do conteúdo do documento (Moscoso e Extreño 1999). De um ponto de vista subjetivo, a utilidade de um documento em relação a uma necessidade de informação concreta por parte do usuário, corresponde ao conceito de pertinência, ou seja, um documento pode ser ‘relevante’ face a uma determinada questão, mas não ser ‘pertinente’ para um determinado usuário na medida em que poderá não acrescentar informação à sua necessidade de informação. É referido a este respeito que “a medida da eficácia de um sistema deverá pressupor o conceito de relevância, baseado numa opinião consensual e não no juízo de valor subjetivo de um utilizador” (Ribeiro 1996).

¹ Uma das principais dificuldades em usar a revocação como uma medida de eficácia de pesquisa é que poderá ser praticamente impossível determinar o número total de registos relevantes a não ser em bases de dados muito pequenas (ODLIS – *Online Dictionary of Library and Information Science*).

No projeto Cranfield II (1963) a designação de “precisão” veio substituir a de “relevância” por se considerar mais objetiva (Svenonius 2003).

Nestas medidas de avaliação influem diretamente a exaustividade e a especificidade na indexação; a exaustividade está relacionada com o número de conceitos identificados no conteúdo do documento e o número de descritores atribuídos para o descrever; assim, quanto mais exaustiva for a indexação, maior será a potencialidade do sistema recuperar documentos relevantes (taxa de revocação); quanto mais específica for a operação, maior será a taxa de precisão, na medida em que haverá mais termos de relação exata com a pesquisa (Moscoso e Extreño 1999). Segundo Gil Leiva (1997) a exaustividade e a especificidade são critérios aplicados durante a indexação, posteriormente aplicados na avaliação da qualidade dessa operação, através dos resultados da resposta documental. Estas medidas devem ser usadas, avaliadas e reavaliadas, na perspectiva de um trabalho de reindexação, que permite ir afinando todo o processo.

Complementares às taxas de revocação e precisão são, respetivamente, as taxas de silêncio e ruído, que correspondem aos documentos relevantes que não são recuperados, e aos que são recuperados e não são relevantes.

Para observarmos melhor os parâmetros de avaliação, sistematizamos os vários conceitos envolvidos (tab. 1.5):

Medidas de avaliação da eficácia de um SRI		
Na pesquisa		
Taxas	Revocação	Número de registos relevantes recuperados numa pesquisa, em relação ao número total de registos relevantes na base de dados.
	Precisão	Número de registos relevantes recuperados, em relação ao número total recuperado na pesquisa.
	Silêncio	Número de documentos não recuperados, face ao número de documentos relevantes existentes, ou seja, documentos relevantes não recuperados.
	Ruído	Número de documentos não relevantes que são recuperados na pesquisa, ou seja, documentos recuperados não relevantes.
Conceitos inerentes	Relevância	Utilidade de um documento recuperado face a um tema de pesquisa.
	Pertinência	Utilidade para o usuário de um documento recuperado face às suas necessidades de informação.
Na indexação		
Influem diretamente	Exaustividade	Está relacionada com o número de noções que caracterizam o conteúdo do documento e o número de descritores usados para descrever os conceitos.
	Especificidade	Exatidão com que um determinado conceito é representado por um termo de indexação.
	Consistência	Grau de coerência na atribuição de descritores a um mesmo documento, por um ou por vários indexadores.

Tabela 1.5: Medidas de avaliação da indexação

A ‘consistência’ é um dos parâmetros fundamentais para avaliar a qualidade da indexação. O princípio da consistência estabelece que um mesmo conceito deve expressar-se sempre através do mesmo descritor (Moscoso e Extremerño 1999).

O indicador de ‘consistência’ interessa-nos em particularmente, pois está diretamente ligado ao nosso estudo, de um ponto de vista da representação formal, uma vez que é observada a terminologia atribuída por vários profissionais.

A consistência representa o grau de coerência no tratamento de diferentes documentos em relação a uma mesma matéria. O grau de consistência é um indicador da análise do tratamento documental e pode ser usado quer a nível interno de uma base, ou entre sistemas diferentes. O princípio de consistência estabelece que a relação entre conceito e descritor deve ser biunívoca, ou seja, a cada conceito deve corresponder um descritor e vice-versa, e essa relação pode medir-se pela proporção de descritores coincidentes atribuídos a documentos com a mesma temática. Num sistema de informação o controlo da linguagem documental deve servir para garantir ao usuário a existência de um termo de indexação que defina corretamente a sua pesquisa, sem que seja necessário ampliar a estratégia de pesquisa para recuperar documentos sobre o tema definido. Neste sentido, a análise da consistência é de grande utilidade não só para o controlo da qualidade da indexação, como também para a orientação de indexadores na coerência da atribuição de termos. (Extremeño 1999; Rodríguez Yunta 1998).

A utilização deste indicador para comparar a coincidência na atribuição de termos, deve ter em conta a presença de diversas variáveis e o seu controlo. Embora seja difícil retirar conclusões definitivas dos resultados das avaliações se não forem controladas todas essas variáveis, a sua observação contribui para acrescentar conhecimentos acerca da complexidade do processo de indexação. As variáveis envolvidas são, a formação e experiência do indexador, o domínio que este tem das ferramentas que utiliza, o conhecimento do âmbito temático em que se enquadra o documento, diferentes diretrizes de trabalho, diferentes unidades de informação, e sobretudo a amostra sobre a qual é realizada a avaliação (Gil Leiva 2002, 2008).

Neste contexto, há uma série de fatores que influem no grau de consistência da indexação (Soler Monreal 2009):

- ✓ O tipo de vocabulário usado; uma linguagem controlada proporciona maior coerência que um vocabulário livre;
- ✓ A abrangência e a especificidade do vocabulário; quanto mais amplo for o vocabulário, a sua especificidade tende a ser maior, o que diminui a consistência;
- ✓ A matéria e a terminologia; as matérias mais concretas favorecem a consistência, enquanto as mais abstratas a diminuem;
- ✓ Características dos indexadores; quando a sua formação e experiência nesta tarefa é similar, revelam maior grau de concordância;
- ✓ As ferramentas de indexação; quando são usadas as mesmas ferramentas, a consistência é maior;
- ✓ O número de termos atribuídos (exaustividade); quanto mais descritores se atribuem, mais difícil é a concordância entre indexadores, não tanto ao nível de termos principais, mas mais nos termos secundários;

Ao caracterizar a consistência pelo grau de semelhança na representação de termos selecionados pode-se falar de ‘consistência interindexador’ que é realizada por vários indexadores, através da análise de um mesmo documento, e a ‘consistência intra-indexador’, que é realizada pelo mesmo indexador, através da análise do mesmo documento em períodos diferentes. Quanto à utilização de parâmetros de análise, estes podem ser muito diferentes em cada investigação, como por exemplo, comparação da

indexação em bibliotecas ou em bases de dados, indexação com palavras-chave (linguagem natural) ou com descritores (linguagem controlada), tipos de suporte, profissionais com ou sem experiência, etc., sendo recomendado que as comparações sejam estabelecidas entre estudos que coincidam nos parâmetros. Independentemente dos parâmetros utilizados, o grau de consistência na indexação costuma variar entre os 20% de mínima e os 60% de máxima (Gil Leiva 2001, 2008).

Uma fórmula aplicada para medir os índices de consistência, já usada anteriormente em estudos de comparação de sistemas, ou entre indexadores (Hooper 1965; Lancaster 1991; Salton e McGill 1983; Tonta 1991), (*apud* Gil-Leiva 2001), apresenta-se da seguinte forma:

$$C_i = \frac{T_{co}}{(A + B) T_{co}}$$

(Gil Leiva 1997)

em que,

C_i = a consistência entre os dois sistemas ou dois indexadores.

T_{co} = o número de termos comuns atribuídos pelos sistemas ou indexadores.

A = o número de termos atribuídos pelo sistema 1 ou indexador 1.

B = o número de termos propostos pelo sistema 2 ou indexador 2.

T_{co} = o número de termos comuns atribuídos por ambos os sistemas ou indexadores.

O estudo dos índices de consistência como indicador de uma indexação correta poderá ser problemático, uma vez que se pode encontrar uma indexação consistente mas incorreta, que se verifica quando os profissionais indexam um documento e cometem o mesmo erro (Gil Leiva 1997). Por esse motivo, embora a utilização desta medida seja de extrema utilidade, devem ser igualmente desenvolvidos estudos com aplicação de outras medidas, para uma verdadeira avaliação dos sistemas.

Outras duas formas de avaliar a qualidade da indexação, que se enquadram na linha de investigação para avaliar bases de dados e catálogos bibliográficos, são, por um lado, comprovar a coincidência dos termos utilizados pelos usuários na pesquisa e os termos disponíveis nos índices² (Caro-Castro 2005) e por outro lado comprovar a adequação dos termos dos índices às normas sobre terminologia dos vocabulários controlados, que nos interessa em particular pela proximidade à nossa investigação.

² Ver a este respeito “El acceso por materias en los catálogos en línea: índices y terminología de los usuarios en el catálogo CISNE”.

2. Contexto das bibliotecas do ensino superior politécnico

2.1 A função da biblioteca no ensino superior

Algumas considerações teóricas da relação entre a biblioteconomia e a educação superior referidas num estudo de Gómez Hernández (1995) continuam atuais no enquadramento das bibliotecas de ensino superior. É referido que a função da biblioteca é recolher e proporcionar acesso aos recursos de informação necessários para produzir conhecimento, quer através do estudo, do ensino, da investigação ou da transmissão cultural. A missão da biblioteca é fazer com que os estudantes e professores conheçam as metodologias documentais, que lhes permitam selecionar, localizar, usar e avaliar a informação de que necessitam. Fala-se de formação documental como um meio para tornar possível a autonomia na aprendizagem e a capacidade para a investigação. Se considerarmos que as funções da universidade são essencialmente comunicativas e informativas (recolher, transmitir, produzir e ensinar a produzir informação), será evidente a importância da biblioteca como instituição mediadora nesse processo informativo. O autor justifica assim a metáfora da biblioteca como coração da universidade.

"La biblioteca es el corazón de la Universidad. No hay otro factor material más estrechamente ligado a la calidad de la Educación Superior. Unas pocas universidades con escasos medios bibliotecarios han logrado un desarrollo importante en algunas áreas, porque los recursos de laboratorio pueden ser más importantes que la biblioteca en algún campo particular, o en otros casos porque las universidades estén situadas junto a otras colecciones de grandes bibliotecas (...) Pero las instituciones que se destacan en todas las disciplinas tienen invariablemente las mejores bibliotecas de investigación del país".

Nunca é demais reforçar as seguintes considerações teóricas do autor, que continuam, infelizmente, por aplicar em algumas instituições:

- ✓ Para que a biblioteca constitua um verdadeiro apoio na transformação da informação em conhecimento, deve-se ter em conta o atual contexto da ciência e da sociedade, caracterizado pela grande quantidade da informação, e da sua diversidade, especificidade e ritmo de transformação;
- ✓ A biblioteca deve ajustar os seus serviços a estas circunstâncias, tendo como missão facilitar o acesso a recursos de informação, e também proporcionar formação documental aos seus usuários, no sentido de estimular a capacidade da própria aprendizagem;
- ✓ Estando a atividade de investigação implícita no ensino de nível superior, é indiscutível a necessidade da existência de boas bibliotecas especializadas como base de apoio para o conhecimento científico.

Pode dizer-se que “educar é sobretudo ensinar a investigar, num contexto em que o professor deve criar uma atmosfera de investigação, e a biblioteca deve constituir um local eficaz de aproximação à informação documental”. Os alunos deveriam ter algumas destrezas documentais ao chegar ao ensino superior, saber que instrumentos existem nas bibliotecas para obter informação, como no caso dos catálogos, mas na realidade a maioria chega sem qualquer experiência de usar as bibliotecas.

As bibliotecas têm-se vindo a converter num ponto de acesso a qualquer tipo de documento, através das redes, do correio eletrónico, etc. As novas bibliotecas são cada vez mais fruto de um conjunto de utilização dos meios eletrónicos para o armazenamento e transmissão da informação, juntamente com as técnicas tradicionais, que devem tornar possível aos usuários o acesso simultâneo aos recursos de informação tanto locais como remotos. A biblioteca tornou-se num serviço de informação, em que o bibliotecário é um informador e um especialista nos canais de acesso à informação.

Estas reflexões permitem-nos verificar a importância da informação no contexto da sociedade em que vivemos, e a relevância do papel das bibliotecas como ‘laboratórios’ para o desenvolvimento dos vários processos envolvidos.

Numa reflexão sobre a sociedade do conhecimento e o ensino superior Veiga Simão (2003) refere:

“A transição da sociedade atual para uma sociedade baseada no conhecimento, [...] onde a informação assume papel determinante e as tecnologias de informação e comunicação (TIC) emergem como suporte crucial do desenvolvimento sustentável e da competitividade entre nações, só pode ser levada a cabo com sucesso se os poderes políticos procederem a profundas reformas estruturais de natureza social e económica, orientadas pelo fomento da cultura, da qualidade, da produtividade e da inovação.”

No enquadramento do desafio estratégico definido pela União Europeia entre 2002-2020, considerou-se básica a necessidade de desenvolver “um programa mobilizador de educação e formação ao longo da vida” que passa pela eliminação do analfabetismo informático e da iliteracia cultural e económica. No atual contexto da sociedade do conhecimento deve-se privilegiar a criação do saber, e as instituições do ensino superior

são determinantes para o desafio dessa construção. Segundo o autor, a evolução desta nova sociedade deve assentar nos pilares essenciais: da cidadania, da cultura, da ciência e da inovação. No pilar da cidadania as instituições de ensino superior, devem ter na sua missão a criação de capital social (inteligência), pois é o que gera sucesso na sociedade do conhecimento, e o “agente determinante do progresso das instituições”. Trata-se de um ‘capital’ que deve ser procurado e fomentado nas comunidades envolventes, a nível regional, nacional, europeu e internacional.

De igual forma, o pilar cultural está ligado ao conhecimento, pois é claro que “um cidadão mais culto é um cidadão mais livre”, e com maior capacidade de afirmação na sua participação social. Neste contexto, as TIC têm contribuído para a consistência do conceito de educação e formação ao longo da vida “fomentando a interdisciplinaridade e abrindo caminhos para novas profissões”.

No pilar da ciência é imprescindível tomar consciência do papel que as universidades terão a desempenhar na construção da sociedade do conhecimento, que se transforma e internacionaliza permanentemente. Em Portugal este papel é repartido com as instituições politécnicas, que tal como as instituições universitárias devem ter consciência de que o futuro depende em grande medida do desenvolvimento cultural, científico e técnico que se cria em centros de conhecimento e de investigação.

Tanto nas universidades como nos institutos politécnicos, as atividades didáticas devem ser indissociáveis da investigação, embora esta tenha características diferentes nos dois tipos de instituição. O equilíbrio entre as funções de ensino e de investigação é

muito relevante quer a nível individual, quer institucional, e deverá existir uma racionalização de recursos humanos e materiais, por meio do incentivo às redes de investigação que possibilitem maximizar e complementar potencialidades. No caso dos institutos politécnicos deve ser cultivada a investigação aplicada e a integração em redes regionais de I&D em parceria com universidades e/ou empresas.

No pilar da inovação o ensino superior constitui um instrumento decisivo da inovação e competitividade, pois são as instituições de ensino superior que qualificam os recursos humanos, pelo que é imprescindível, no caso português, investir na sua qualidade, numa dimensão europeia. A qualidade está intimamente associada a este pilar da inovação, e deve ser criado um relacionamento universidade-sociedade e instituto politécnico-sociedade perante a evolução da ciência e da tecnologia numa ligação global à economia do conhecimento. Estas tendências de evolução caracterizam-se pela abertura a novos mecanismos de cooperação entre universidades e entre estas e institutos politécnicos. Nas palavras de Veiga Simão,

“O equilíbrio nas conceções de universidade e de instituto politécnico irá residir essencialmente no peso que a cada instituição será atribuído pelos seus programas estratégicos e na clarificação de objetivos fundamentais, como sejam: visão estratégica, dimensão cultural, internacionalização, impacto regional, I&D, inovação, especialização e interdisciplinaridade”.

Na organização do ensino superior português, o ensino politécnico é ministrado em instituições de ensino superior não universitárias. Os dois tipos de instituição conferem os graus académicos¹ de licenciatura e mestrado (graus de Bolonha), sendo o grau de

¹ O ciclo de estudos de licenciatura no ensino politécnico tem uma duração de seis semestres curriculares, no ensino universitário tem uma duração entre seis e oito semestres curriculares. O grau de mestre é igualmente conferido por instituições universitárias e politécnicas, com um ciclo de estudos de duração compreendida entre três e quatro semestres curriculares.

doutor somente conferido pelas instituições universitárias. Têm, no entanto um carácter diferenciado mais evidenciado ao nível do 2º ciclo de estudos, em que, de acordo com informação da Direção Geral do Ensino Superior², o grau de mestre no ensino politécnico “deve assegurar, predominantemente, a aquisição de uma especialização de natureza profissional”, e no ensino universitário “deve assegurar, predominantemente, a aquisição de uma especialização de natureza académica com recurso à atividade de investigação ou que aprofunde competências profissionais.”

A importância do papel que as bibliotecas desempenham na sociedade em geral, e ao nível do ensino superior em especial, tem vindo a ser cada vez mais reconhecido. Em instituições de ensino superior exige-se que estas unidades de informação sejam devidamente apetrechadas e com altos níveis de qualidade.

O novo modelo de Educação Superior (decorrente da Declaração de Bolonha)³ teve grande repercussão nas bibliotecas universitárias em toda a Europa. A nova situação estabelecida no sistema de ensino superior europeu criou um novo paradigma para a comunidade académica, onde se encontra também a biblioteca universitária. A necessidade de adaptar a biblioteca ao novo contexto docente e às atuais exigências de informação levou a que se estabelecesse o modelo CRAI (Centro de Recursos de Aprendizagem e Investigação) que vem substituir o modelo tradicional de biblioteca. As primeiras correntes bibliotecárias deste modelo foram implementadas no Reino Unido, Estados Unidos, Canadá e Holanda, com a designação anglo-saxónica de “*Learning Resources Centre*” ou “*Learning Commons*”, seguindo-se em Espanha através da

²<http://www.dges.mctes.pt/DGES/pt/Reconhecimento/NARICENIC/Ensino+Superior/Sistema+de+Ensino+Superior+Português/>

³ http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11088_pt.htm [Consulta: 03-01-2013]

REBIUN⁴ (Red de Bibliotecas Universitárias). A nova expressão do conceito de biblioteca implica a adaptação à sociedade do conhecimento, deixando de ser entendida como depósito de livros, e passando a ser um centro de recursos para a aprendizagem e a investigação, ao serviço das novas metodologias docentes. O tradicional modelo de ensino centrado na docência, evoluiu para um baseado na aprendizagem interativa, que a biblioteca apoia como complemento da aula presencial. Neste novo modelo educativo, a aplicação das tecnologias de informação acrescenta às funções tradicionais da biblioteca a possibilidade de formar e instruir sobre o uso de novos recursos de informação de forma eficaz. O bibliotecário vai deixar de ser um técnico especialista no tratamento da documentação para passar a ser um novo agente educativo da universidade que apoia o professor e o estudante neste novo contexto (Delandés Palomares 2012; Taladriz 2007).

No atual ambiente globalizado, e no pressuposto da aprendizagem ao longo da vida, as universidades são um ponto central da informação, educação e inovação, como agentes desses processos. Nesta perspetiva, a anterior biblioteca de ensino superior deve realizar importantes mudanças na sua organização e gestão, e os profissionais de documentação terão um papel fundamental como gestores do conhecimento. A biblioteca que tem sido um serviço de apoio à universidade, deverá transformar-se num serviço estratégico que facilite o acesso a estudantes e professores na gestão da informação. A imagem conservadora da biblioteca deverá tornar-se num ponto aglutinador de uma série de serviços de ponta na nova conjuntura de ensino superior. Através de uma forte componente tecnológica a nova biblioteca converte-se numa

⁴ <http://www.rebiun.org/index.html> [Consulta: 03 -01-2013]

biblioteca digital que reúne os sistemas de informação, e que reforça igualmente a ideia social de encontro e comunicação da comunidade académica, constituindo um verdadeiro centro de recursos integrados. Para as direções das instituições o modelo CRAI representa uma nova forma de distribuir recursos e equipamentos, solucionando alguns desequilíbrios tecnológicos dentro do mesmo campus de ensino superior. Há igualmente um melhor aproveitamento das políticas de inovação, e uma nova imagem de prestígio que atrai novos estudantes. A biblioteca, anteriormente vista como um serviço de ‘segunda ordem’, está-se a converter através do modelo CRAI, num elemento estratégico das próprias universidades e institutos superiores (Delandés Palomares 2012; Martínez 2004).

As razões para a mudança do conceito de biblioteca são essencialmente a necessidade de eficiência na gestão universitária no atual contexto económico, como centralizar serviços, criar modelos mais eficientes que evitem duplicação de estruturas e aumento de pessoal, racionalização orçamental, ou seja, oferecer maiores e melhores prestações a menor custo. As instituições devem convergir para este novo modelo, embora esse processo, como sabemos, esteja dependente da sua história e características. As vantagens da convergência são essencialmente o fortalecimento de liderança para a planificação estratégica e visão integrada, que potenciam o desenvolvimento de projetos transversais. É fundamental um maior aproveitamento de recursos humanos e materiais, um maior grau de visibilidade e qualidade de serviços, que globalmente contribuem para a vantagem competitiva das instituições de ensino superior (Celestino Angulo 2007).

Como é sabido, a natureza do tipo de ensino, o carácter dos cursos, o enfoque da docência e da investigação e outros fatores determinam o crescimento das bibliotecas nas instituições, e sob este ponto de vista pensamos que as bibliotecas inseridas no sistema politécnico e universitário têm características diferentes sobretudo ao nível de recursos humanos e materiais. Neste contexto apresenta-se de seguida informação de carácter geral referente aos dois tipos de bibliotecas.

2.2 Bibliotecas do ensino universitário e do ensino politécnico

No contexto das bibliotecas de ensino superior politécnico, a Associação Portuguesa de Bibliotecários Arquivistas e Documentalistas (BAD) tem tido ao longo dos tempos um papel fundamental na área da biblioteconomia em Portugal através de todas as suas atividades. O Grupo de Trabalho das Bibliotecas do Ensino Superior (GTBES), tem como objetivo central ‘conhecer e refletir sobre as questões mais importantes na esfera das bibliotecas do ensino superior⁵’. Anteriormente designado por Grupo de Trabalho das Bibliotecas Universitárias, tomou a nova designação a partir de 1999. Esse ponto foi um marco importante, pois veio integrar as Bibliotecas do Ensino Superior Politécnico. Esta decisão de alargar a esfera de ação a todo o ensino superior permitiu, nas palavras do coordenador do GTBES⁶, “uma agregação e nova gestão de recursos, que se tem

⁵ As primeiras jornadas dedicadas às bibliotecas universitárias foram realizadas em 1980, assim como vários encontros dedicados à temática ao longo dessa década.

⁶ Professor Carlos Lopes, em entrevista informal concedida a 11 de Maio de 2009.

refletido nas bibliotecas na última década [...] foi um fator de inclusão muito forte que resultou numa boa dinâmica”.

Surgiram então diversos trabalhos resultantes dessa vontade de cooperação, que contribuíram para uma efetiva tomada de consciência e melhoria de todo o sistema de bibliotecas de ensino superior, para além da alteração de visão estratégica para o futuro. Os trabalhos e projetos realizados culminaram no projeto da Biblioteca Científica Digital, designada Biblioteca do Conhecimento Online (B-on)⁷, dinamizada pela Agência para a Sociedade do Conhecimento (UMIC)⁸, um instituto público que tem por missão o desenvolvimento de projetos na área da sociedade da informação, sob tutela do Ministério da Ciência, Tecnologia e Ensino Superior.

Com este projeto foi criado um consórcio nacional constituído por Universidades, Politécnicos e instituições ligadas à investigação científica, que disponibiliza o acesso a textos integrais de milhares de publicações científicas das principais editoras internacionais. O projeto, para além de ter servido de alavanca ao desenvolvimento das bibliotecas politécnicas, contribuiu para a sua integração no panorama da informação universitária. Segundo as palavras do Coordenador do GT BES da BAD, podemos dizer que “existe uma situação antes e pós B-on” nestas bibliotecas.

A biblioteca B-on começou a ser planeada em 1999, e em 2004 foram disponibilizados os primeiros textos integrais. A infraestrutura técnica é assegurada pela

⁷ <http://www.b-on.pt>

⁸ http://www.unic.pt/index.php?option=com_frontpage&Itemid=307 [Consulta: 13-03-2013]

Fundação para a Computação Científica Nacional (FCCN)⁹. Na sua missão a B-on refere o objetivo de “garantir o acesso a publicações de natureza científica e serviços eletrónicos à comunidade académica e científica nacional.”

Alguns dos objetivos que estão na base da sua criação são:

- Desempenhar um papel ativo na construção da Sociedade do Conhecimento;
- Estimular a cooperação entre as entidades académicas e científicas nacionais;
- Racionalizar custos através da negociação centralizada com editoras e fornecedores de conteúdos;
- Ser referência no fornecimento de serviços eletrónicos e produtos para dar resposta às necessidades dos usuários;
- Desenvolver uma política de preservação dos conteúdos.

No sector das bibliotecas universitárias portuguesas devemos referir também, que o trabalho desenvolvido na área da avaliação de qualidade na última década, tem vindo a servir de orientação para desencadear processos de avaliação de serviços nas instituições. São conhecidos alguns estudos que focam a necessidade de normalização e avaliação no contexto da qualidade dos serviços e das boas práticas em bibliotecas do ensino superior em Portugal (Galvão e Machado 2004; Lopes 2006; Serrano 2003). Embora a avaliação nas bibliotecas universitárias tenha começado no início da década

⁹ <http://www.fccn.pt/> [Consulta: 13-03-2013]

de noventa, foi reconhecida a necessidade de publicar mais estudos para que a realidade nacional dos serviços de informação se pudesse conhecer melhor, e pensamos que o aumento da investigação nesta área é uma tendência que se tem vindo a notar.

Em linhas gerais consideramos que o contexto em que as bibliotecas do ensino politécnico se encontram atualmente, de aproximação às bibliotecas universitárias, é favorável ao desenvolvimento sustentado nas boas práticas, e que a par de outras estratégias de intervenção possíveis, se poderão fazer verdadeiros avanços para uma melhor prestação de serviços no acesso à informação.

2.3 O caso das Bibliotecas do Instituto Politécnico de Portalegre

A parte prática da nossa investigação inclui um estudo de caso que importa contextualizar neste último capítulo teórico. Trata-se de uma rede de bibliotecas que fazem parte da mesma instituição de ensino, e fazem esforços conjuntos para cumprir os pressupostos da sua missão enquanto sistemas de informação. Estas bibliotecas foram criadas separadamente, e cresceram de forma isolada ao longo do tempo, desenvolvendo fundos documentais para dar resposta aos públicos que servem. Estão integradas em Escolas Superiores com carácter disciplinar diferente, e os fundos documentais refletem não só características temáticas muito diversas, como também estruturas de desenvolvimento singular.

Desde meados da década anterior que os profissionais das bibliotecas iniciaram a intenção de cooperar, no sentido de rentabilizar recursos humanos e materiais, até que se chegou ao projeto de um catálogo coletivo que, de forma centralizada, pudesse servir melhor uma comunidade académica mais alargada. Todos os esforços foram conjugados neste sentido, e as bibliotecas agora com um catálogo bibliográfico em rede, podem de facto dar visibilidade aos seus fundos documentais a um maior número de usuários.

Até chegar a este ponto das bibliotecas se encontrarem ligadas num catálogo coletivo virtual, foram seguidas as regras e normas de descrição bibliográfica, assim como a adoção do mesmo *software* de gestão bibliográfica, o que permitiu uma melhor uniformização de procedimentos biblioteconómicos. No entanto, ao nível da indexação, pela complexidade que a prática envolve, não foi possível desenvolver essa tarefa de forma uniforme em todas as bibliotecas.

Este grupo de bibliotecas está integrado no projeto PORBASE¹⁰ – Base Nacional de Dados Bibliográficos – que é o catálogo coletivo das bibliotecas portuguesas, coordenado pela Biblioteca Nacional de Portugal, e que tem contribuído de forma decisiva para a normalização das práticas profissionais. O sistema de indexação SIPORbase, que surgiu no contexto deste projeto é uma referência nacional, e neste sentido constitui uma base de orientação para o trabalho das bibliotecas nesta área, embora, pela sua difícil implementação, nem sempre seja possível a sua aplicação.

¹⁰<http://porbase.bnportugal.pt/#focus> [Consulta: 14-03-2013]

Apesar destas unidades de informação terem fundos documentais com dimensões e temáticas diferentes, partilham atualmente de um espírito de cooperação que permite direcionar estratégias para objetivos comuns. Neste contexto têm sido desenvolvidos, a nível interno, trabalhos relacionados com a gestão das Bibliotecas, no sentido de implementar melhorias contínuas. Um dos pontos centrais em que se observou a necessidade de intervenção foi ao nível da pesquisa nos catálogos bibliográficos.

Constatou-se através de um estudo sobre estas bibliotecas¹¹ (Martins 2009), e do contacto direto com os responsáveis das mesmas, que os procedimentos de indexação não eram uniformes, e que se revelavam dificuldades na recuperação de informação por assuntos nos catálogos bibliográficos. Ligando essa dificuldade à indexação, um ponto referido pelos responsáveis foi a necessidade de uma política de indexação.

A intervenção especializada dos profissionais a este nível implica um esforço, mas é da qualidade dessa intervenção que dependerá a consistência dos catálogos. Nesse sentido devem ser definidas estratégias no planeamento de serviços de informação dentro das instituições.

¹¹ Este estudo partiu da necessidade de fazer um levantamento de procedimentos de indexação, com vista à uniformização da prática, de acordo com os princípios e normas adequadas.

2.3.1. Fontes de terminologia preferenciais

Como referimos, os fundos documentais destas Bibliotecas têm características temáticas diferentes que abrangem as grandes áreas de ciências da educação, ciências da saúde, ciências empresariais, engenharia e tecnologias, ciências agrárias. Devido a esta diversidade temática são usadas diferentes fontes de informação para recolha de termos, e este facto coloca atualmente alguns problemas de consistência nos índices do catálogo coletivo virtual, de que falamos adiante.

Referem-se as fontes de informação preferencial que têm sido usadas, que revelam as diferenças temáticas:

- **AGROVOC Thesaurus**¹² – é o tesouro utilizado como fonte preferencial na Biblioteca da Escola Superior Agrária; é um tesouro multilingue para áreas das ciências agrárias, que cobre temáticas relacionadas com alimentos, nutrição, agricultura, pescas, florestas, ambiente, etc.; está inserido na AIMS (*Agricultural Information Management Standards*) que é um portal da FAO (*Food and Agriculture Organization*) direcionado para a disseminação de normas e boas práticas na gestão de informação para as grandes áreas da agricultura e desenvolvimento rural.

¹² <http://aims.fao.org/standards/agrovoc/about>

● **Catálogo da BNP¹³** - Biblioteca Nacional de Portugal - O catálogo da BNP é a fonte preferencial do fundo documental de carácter mais generalista, servindo também de referência para as outras bibliotecas, sobretudo para a construção de cabeçalhos pré-coordenados em temáticas gerais. A pesquisa nos índices da BNP permite aceder a autoridades de autores, e índices controlados de assuntos. Estes pontos de acesso são especialmente utilizados para localização e importação de registos bibliográficos para as bases locais, sendo uma mais-valia nas tarefas de catalogação.

● **DeCS¹⁴ - Lista de Descritores em Ciências da saúde** – Este recurso é utilizado como fonte preferencial na Biblioteca da Escola Superior de Saúde. O sistema DeCS é um vocabulário estruturado criado pela BIREME¹⁵ (Biblioteca Regional de Medicina - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde). Esta terminologia foi desenvolvida como linguagem de indexação para as áreas das ciências da saúde, a partir do MeSH (Medical Subject Headings) da *U.S. National Library of Medicine*. A terminologia DeCS é usada na pesquisa de assuntos nas fontes de informação LILACS¹⁶ e MEDLINE¹⁷ da Biblioteca Virtual em Saúde (BVS)¹⁸. Este recurso possibilita a pesquisa *online* em três idiomas (inglês, português, espanhol).

¹³ <http://catalogo.bnportugal.pt/#focus>

¹⁴ <http://decs.bvs.br/>

¹⁵ A Bireme é um Centro Especializado da OPAS/OMS – Organização Pan-Americana da Saúde/Organização Mundial da Saúde.

¹⁶ LILACS – Literatura Latino-Americana e do Caribe em Ciências da Saúde.

¹⁷ MEDLINE – Literatura Internacional em Ciências da Saúde.

¹⁸ BVS – Rede de bibliotecas do BIREME.

● **EuroVoc**¹⁹ Thesaurus - O tesouro multilingue da União Europeia é utilizado como fonte preferencial para as áreas da educação, da gestão e tecnologias. É um tesouro com características multidisciplinares, que pode ser consultado em qualquer língua da União Europeia. A abrangência de domínios e a facilidade de utilização contribuíram para que se tornasse um instrumento de apoio à indexação muito útil nestas bibliotecas. O tesouro é estruturado por grandes domínios especialmente direcionados para as temáticas das instituições da União Europeia.

¹⁹ <http://eurovoc.europa.eu>

2.3.2 O sistema bibliográfico e de autoridades

O sistema de gestão bibliográfica usado nas Bibliotecas é o PORBASE5, promovido pela Biblioteca Nacional, entidade que assumiu o papel de Centro Distribuidor Nacional do CDS/ISIS¹ em 1987. As várias parametrizações PORBASE do Mini-Micro CDS/ISIS concebidas para MS-DOS, foram produzidas e distribuídas até 2000, sendo então substituídas pelo novo sistema, concebido para funcionar em ambiente Microsoft Windows. A utilização do mesmo sistema de gestão bibliográfica permitiu que as atualizações de *software* fossem decorrendo em períodos de tempo razoável.

Verificou-se no entanto nos últimos anos, a impossibilidade de realizar estas atualizações de forma uniforme, o que originou algumas dificuldades no desenvolvimento do trabalho em cooperação. Assim, foi considerada a hipótese de ligar as bibliotecas através da última versão disponível do sistema, atualmente designado Mind PRISMA Porbase5 (SQL-Server), na versão transacional (1.6) para redes de bibliotecas.



Fig. 2.1: Interface de pesquisa do catálogo coletivo virtual

¹ *Software* de gestão bibliográfica desenvolvido e disponibilizado pela UNESCO.

A adoção deste sistema, que tem um interface de pesquisa multibases (fig.2.1) constitui uma vantagem fundamental, pois para que um catálogo virtual seja funcional, as bases de dados participantes devem oferecer um conjunto uniforme de opções de pesquisa que permitam recuperar documentos comparáveis de cada catálogo (Caro-Castro 2005). Quanto à uniformidade dos índices, é uma questão igualmente desejável mas de implementação mais complexa, conforme veremos mais adiante.

Analisadas as vantagens do novo sistema, a nova versão foi implementada em 2008. Esta decisão resultou da proposta de ligação em rede das bibliotecas no âmbito do sistema de gestão de qualidade a decorrer no Instituto Politécnico de Portalegre, fundamentada pela redução de custos, melhoria de desempenho do *software* de gestão, rentabilização de recursos de informação e de tarefas profissionais.

Na parametrização do sistema, a indexação da base de dados é a operação de criação dos índices contendo os termos de pesquisa, e só será possível o acesso aos termos no índice após o registo ter sido indexado. O módulo de catalogação possui um modo de funcionamento de indexação automática, isto é, após o registo ter sido criado ou modificado o sistema atualiza automaticamente o índice gerando os termos de pesquisa.

Este módulo permite que no processo de catalogação se possa trabalhar com uma base de autoridades ligando-a à base bibliográfica, trazendo assim as vantagens que o uso do controlo de autoridades possibilita de uniformizar os termos nos diferentes registos bibliográficos. No caso destas bibliotecas, as bases bibliográficas só podem estar associadas a uma base de autoridades (fig.2.2), e por sua vez, a base de autoridades

pode ter associadas as várias bases bibliográficas. Este facto colocou-nos a questão de existirem vários registos de autoridade não controlados, uma vez que este ficheiro foi criado de forma automática a partir dos termos de indexação existentes em cada uma das bases bibliográficas, aquando da transação de registos para o sistema.

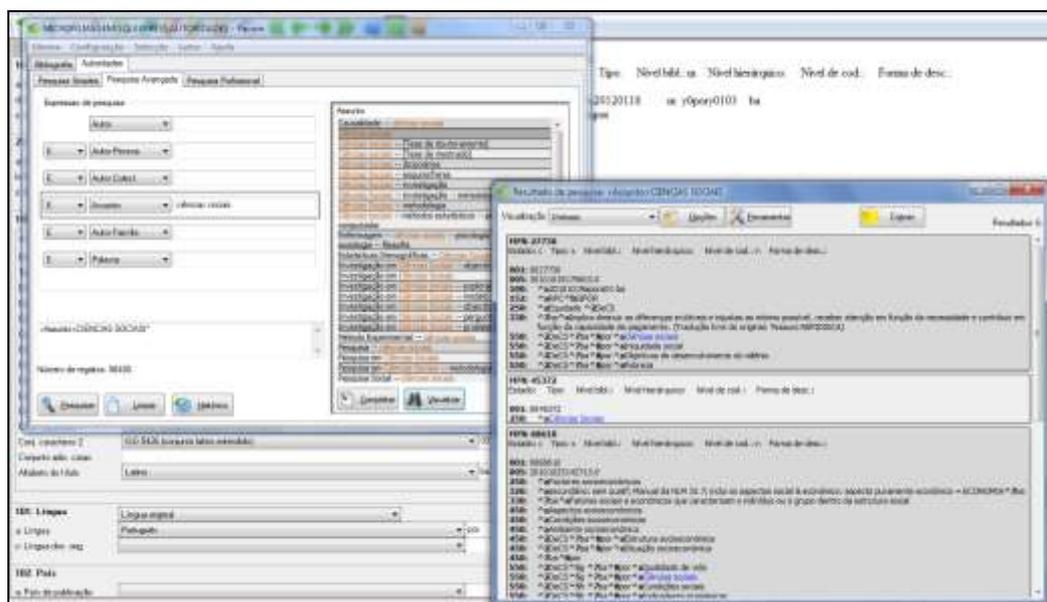


Fig. 2.2: Ficheiro de autoridades, com visualização de registos selecionados

A situação das bases bibliográficas em relação a esta questão coloca-se da seguinte forma: as cinco bases estão ligadas a um ficheiro de autoridades de assunto comum a todas; em cada uma das bases há a possibilidade de trabalhar na opção ‘com controlo de autoridades’ ou ‘sem controlo de autoridades’ (fig.2.3).

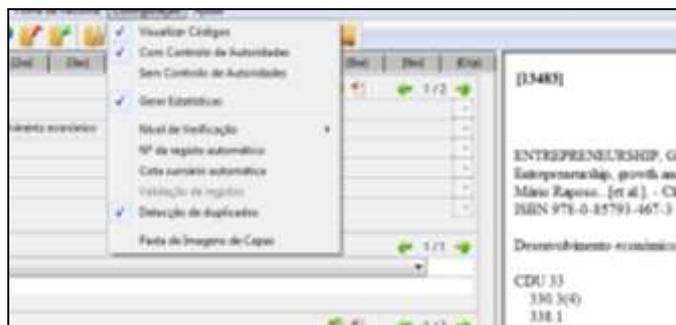


Fig. 2.3: Registo bibliográfico sem e com controlo de autoridades

Considera-se que a melhor opção é trabalhar com ‘controlo de autoridades’, através da qual se faz a ligação do registo bibliográfico a um registo de autoridade já existente (fig.2.4).

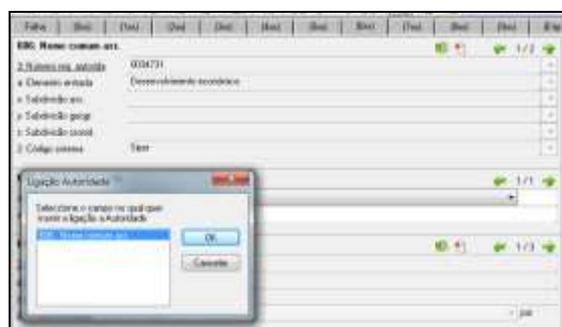


Fig. 2.4: Ligação do registo bibliográfico ao registo de autoridade

Na prática verifica-se que, em muitos casos as ligações de autoridade não se encontram controladas, o que resulta do facto de termos que foram atribuídos nesse campo sem a opção de controlo, terem criado automaticamente um registo de

autoridade, que de facto o não é. Para evitar esse perpetuar de registos não controlados, são criados registos de autoridade (se for o caso), em que são preenchidos os respetivos campos do UNIMARC. No entanto, apesar de esse módulo comum de autoridades servir às cinco bases bibliográficas, na realidade os procedimentos e controle de indexação não são feitos de acordo com uma política de indexação comum.

Relativamente ao catálogo de autoridades devemos realçar o seguinte: no sistema informático utilizado, ao ser feita a ligação do catálogo bibliográfico ao catálogo de autoridades, este foi preenchido automaticamente nos campos de assunto, o que desde logo coloca a dificuldade de terem sido criados milhares de registos de autoridade com os termos não controlados existentes no catálogo bibliográfico.

A criação da base de dados de autoridades foi feita através do processo automático de importação dos dados presentes no bloco 6xx do UNIMARC bibliográfico para o ficheiro de autoridades. De acordo com os especialistas da aplicação informática, “é a forma mais automática de fazer esta criação em massa de registos de autoridade; a única maneira de fazer esta operação de forma diferente é criar cada autoridade manualmente para cada um dos registos bibliográficos.

No caso da migração de bases bibliográficas com milhares de registos para este sistema, não era viável fazer a ligação do catálogo bibliográfico ao catálogo de autoridades de outra forma que não a automática, pois aumentava a tendência de criar mais termos para além dos já existentes, ainda de forma não controlada. Para ilustrar esta questão, apresentam-se alguns exemplos de termos (tab.2.1) retirados do catálogo

de autoridades, e que requerem uma intervenção urgente no controlo da terminologia, uma vez que não estando controlada a autoridade, o processo de ligação dos registos bibliográficos ao ficheiro de autoridades fica comprometido. Na tarefa de catalogação, ao atribuir um termo no campo do assunto deparamo-nos com duas hipóteses: se no sistema seleccionarmos a funcionalidade ‘sem controlo de autoridades’, um termo atribuído cria automaticamente um registo de autoridade; trabalhando ‘com controlo de autoridades’, teremos que escolher um termo já existente para associar o registo; ora se não estão controlados no ficheiro de autoridade, iremos alimentar as situações de dispersão de termos para os mesmos conceitos. Embora o sistema tenha recentemente introduzido a funcionalidade de ‘fusão de registos de autoridade’, essa opção coloca a questão de reavaliar os registos para comprovar a sua validação adequada.

Termos de indexação não controlados

EDUCAÇÃO DA CRIANÇA	GESTÃO DE PESSOAL
EDUCAÇÃO DAS CRIANÇAS	GESTÃO DE RECURSOS HUMANOS
EDUCAÇÃO EM SAÚDE	MÁQUINA AGRÍCOLA
EDUCAÇÃO PARA A SAÚDE	MAQUINARIA AGRÍCOLA
GESTÃO DE ÁGUA	MÁQUINAS AGRÍCOLAS
GESTÃO DA ÁGUA	MÁQUINAS PARA AGRICULTURA
GESTÃO AMBIENTAL	MERCADO AGRICOLA
GESTÃO DO AMBIENTE	MERCADOS AGRICOLAS
GESTÃO DA INFORMAÇÃO	SISTEMA DE AGRICULTURA
GESTÃO DE INFORMAÇÃO	SISTEMAS AGRÍCOLAS
GESTÃO DA QUALIDADE	SISTEMAS DE AGRICULTURA
GESTÃO DE QUALIDADE	SAÚDE COMUNITÁRIA
GESTÃO DA ORGANIZAÇÃO	SAÚDE DA COMUNIDADE
GESTÃO DAS ORGANIZAÇÕES	SAÚDE DA FAMÍLIA
GESTÃO ORGANIZACIONAL	SAÚDE FAMILIAR

Tabela 2.1: Termos existentes no ficheiro de autoridades de assuntos

Estes termos de indexação resultam da utilização de linguagem natural em vez de uma linguagem documental, que evite a representação de conceitos iguais com termos diferentes. É de notar que nalguns casos, os termos aqui agrupados por conceito, ao serem visualizados no índice de assuntos do catálogo, aparecem de forma ainda mais dispersa, permitindo a recolha de informação de forma inadequada.

No contexto da cooperação entre bibliotecas do ensino superior tem sido bastante evidenciada a importância do controlo de autoridades para o processo de pesquisa nos catálogos informatizados. São realçadas as vantagens das pesquisas exaustivas nos catálogos bibliográficos (Campos, 2003) em que “a estrutura habitual de um catálogo controlado é a que melhor facilita a procura e o agrupamento de informação [...] as remissivas asseguram que os materiais relevantes serão encontrados independentemente da forma usada na pesquisa”.

Embora a construção de um ficheiro de autoridades implique etapas exigentes, e a disponibilidade de profissionais especializados, só por esse meio é possível melhorar a consistência dos índices de pesquisa, reduzindo as probabilidades de “ruído” ou “silêncio” na recuperação da informação.

2.3.3 O projeto do catálogo coletivo: necessidade de uniformização

Há que estabelecer a diferença entre os catálogos coletivos tal como se entendiam em redes de cooperação bibliotecária, desde os anos 90 ao início do séc. XXI, e a ideia de catálogos coletivos na atualidade. A ideia do catálogo coletivo deu um salto qualitativo e quantitativo na situação atual de integração de recursos. O problema de acesso por assuntos continua a ser o mesmo ou maior, por exemplo nos catálogos virtuais como EUROPEANA¹ porque se integram recursos de diferentes países em diferentes línguas. O mesmo tipo de problema está-se a reproduzir em relação aos repositórios dentro da iniciativa OAI (*Open Archives Initiative*)², onde a informação é recompilada sem que que exista uma normalização dos descritores temáticos de diferentes vocabulários.

A perspetiva que aqui observamos é a do catálogo coletivo nas redes de cooperação bibliotecária em sistemas à escala institucional.

As questões que se colocam perante a construção e manutenção de um catálogo coletivo são de várias ordens, e implicam a observação de procedimentos técnicos e metodológicos das partes envolvidas, especialmente no que diz respeito à indexação de assuntos.

Nas características funcionais do catálogo coletivo, as funções que lhe estão genericamente associadas são a possibilidade de criação de serviços de empréstimo de

¹ <http://www.europeana.eu/portal/>

² <http://www.openarchives.org/> [Consulta: 31-07-2013]

documentos entre bibliotecas, e a partilha de sistemas de catalogação. São de evidenciar as funções informativas/identificativas, de localização e empréstimo de documentos, e da partilha de registos catalográficos. Nestes catálogos é possível recuperar um conjunto de registos provenientes das coleções que integram as bases de dados envolvidas, pelo que estas funções têm uma grande utilidade (Pacheco 2003).

No entanto, um dos problemas que se coloca no acesso à informação em bases de dados de tipo cooperativo ou em rede, seja a informação de carácter multidisciplinar ou especializado, tem a ver com a recuperação por assuntos, em que se colocam as questões da heterogeneidade, ou seja, para além de existirem inúmeros sistemas de representação de assuntos utilizados nas bibliotecas, e da diversidade de critérios subjacentes à sua aplicação, coloca-se ainda a incontornável questão do nível de profundidade da análise e representação de assuntos, dependente das políticas de indexação adotadas pelos serviços (Santos 2007).

As consequências mais evidentes desta variedade no uso de linguagens são a falta de consistência na representação de assuntos, e a possível inexistência de uma estrutura semântica coerente, dois fatores que dificultam as pesquisas nos catálogos coletivos. Para além de todas as possibilidades de pesquisa, com maior ou menor grau de sofisticação tecnológica, atualmente proporcionadas pelos catálogos, os índices de palavras-chave e os índices de cabeçalhos de assuntos constituem elementos de indexação, cuja consistência da terminologia é imprescindível, o que implica a coerência na indexação por parte dos profissionais das várias instituições, no caso de catálogos coletivos (Caro-Castro 2005).

A perspetiva de utilização de uma linguagem comum, que parece ser a forma de ultrapassar em grande parte o problema da heterogeneidade da terminologia, coloca outras dificuldades na gestão de catálogos coletivos, como o equilíbrio entre exaustividade e especificidade num cenário multidisciplinar de temáticas, para além das dificuldades resultantes de uma política de indexação que também se deseja equilibrada entre a flexibilidade e a rigidez.

Embora as novas formas de recolha e sistematização de conteúdos viabilizem a recuperação da informação de forma mais consistente nos catálogos coletivos, no contexto das bibliotecas académicas, com áreas temáticas científicas especializadas, a utilização de uma linguagem documental comum não irá corresponder igualmente às necessidades de todos os usuários do catálogo (Bocato 2009). Esta autora concluiu na sua investigação que é necessário verificar o uso da linguagem documental de forma a contribuir para a sua adequação à comunidade académica, e que essa verificação deve ser feita avaliando a sua atualização, especificidade e compatibilidade com as necessidades de pesquisa.

Tecnicamente pode-se fazer a distinção entre o catálogo coletivo composto por uma única base de dados e o catálogo coletivo virtual, composto por várias bases de dados pesquisáveis em simultâneo. Neste contexto, a definição de catálogo coletivo, baseia-se no facto de este “receber”, ou “conter” dados de diferentes fontes. Independentemente das características técnicas do catálogo, é o processo de indexação que está em causa, pois para além dos problemas que se colocam com a terminologia recolhida de várias

fontes de informação, a adoção de diferentes metodologias e a falta de uma política de indexação comum, são as principais causas ligadas à inconsistência na realização dessa tarefa.

Os catálogos coletivos, apesar de não serem uma inovação tecnológica, e serem ainda a aspiração de muitas bibliotecas que não dispõem deles, apresentam uma série de problemas derivados da sua própria natureza e conceção. Podem ser consultados em linha mas em muitos casos não se atualizam de forma simultânea. O processo habitual é que cada centro envie periodicamente, para uma unidade central, dados atualizados da descrição dos fundos, e aí coloca-se o problema da atualização (Castillo 2003). No caso de cada catálogo ser gerido de forma independente e se atualizar em linha, a informação encontra-se permanentemente atualizada, mas coloca-se o problema da falta de controlo centralizado dos registos.

O modelo físico do catálogo coletivo inicialmente adotado nas bibliotecas em estudo, foi o que recebe registos exportados pelos catálogos locais. Considerou-se este tipo de catálogo mais consistente por haver um controlo centralizado no cumprimento de regras estabelecidas, sendo no entanto de difícil manutenção, quer a nível de recursos, quer pela dependência do envio de registos por parte das bibliotecas envolvidas. Mais tarde tornou-se evidente a dificuldade de manter este tipo de manutenção, e optou-se pela aquisição de um interface de pesquisa comum que percorresse as bases de dados, e que permitia a atualização permanente destas, em detrimento do controlo de registos.

Aquando da observação para este estudo, os fundos documentais das bibliotecas representavam um universo de cerca de 72.200 documentos de *vária* tipologia numa diversidade de áreas temáticas. A nova versão do sistema informático trouxe uma série de novas funcionalidades, entre as quais, e conforme referido anteriormente, a possibilidade de ligação em rede, com a pesquisa conjunta das diferentes bases de dados, e a ligação automática a uma base de autoridades, o que desde logo prevê a necessidade de controlo do vocabulário de indexação, e a construção sistemática de uma lista de autoridades de assuntos.

Conforme observámos no capítulo anterior, é notória a falta de controlo da terminologia que provoca a dispersão de termos nos índices de assunto. Ao nível das opções de pesquisa para os usuários, esses índices podem contribuir mais para complicar do que para facilitar a tarefa.

Antes de realizar uma pesquisa os usuários não conhecem o cabeçalho que necessitam, e frequentemente não entendem os critérios de ordenação nem a função dos cabeçalhos e subcabeçalhos (Drabenstott, Simcox e Fenton 1999; Caro-Castro 2005). No entanto, conforme ilustra a fig. 2.5, a facilidade dos índices permutados já existentes em muitos sistemas, permitem que os subcampos sejam pontos de acesso, o que “permite atenuar a rigidez dos cabeçalhos pré-coordenados”, uma vez que pesquisa os termos em qualquer ponto do cabeçalho. De acordo com Caro-Castro (2005) é comum nas pesquisas por assunto, que os usuários expressem os temas através de uma ou duas palavras, e que comecem as pesquisas com termos mais gerais do que as suas necessidades de informação, especificando-as posteriormente nos resultados obtidos.

Ora num índice de assuntos não controlado é pouco provável que o usuário, por um lado, tenha facilidade em optar pelos índices disponíveis, e que por outro, verifique resultados consistentes que decorrem das várias escolhas (fig.2.5).

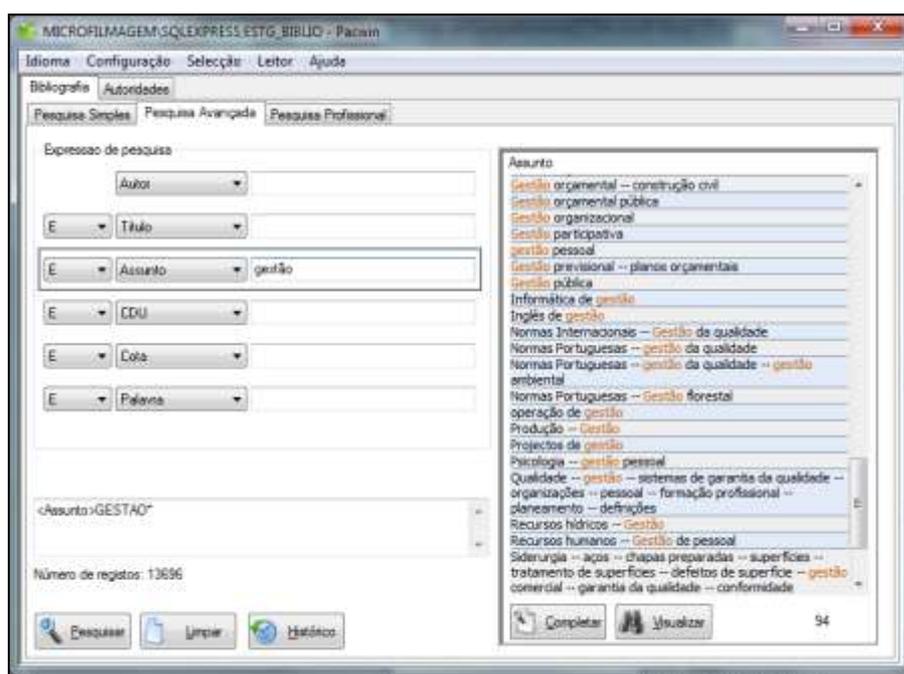


Fig. 2.5: Exemplo de pesquisa nos índices

A manutenção do catálogo coletivo, para ser crescente e real a sua utilidade, pressupõe a necessidade de consolidar procedimentos metodológicos na indexação. Ao serem realizadas pesquisas simultaneamente em todas as bases bibliográficas através do interface Multibases, constatou-se essa necessidade, pois deparámo-nos com uma grande inconsistência ao nível da terminologia de pesquisa por assuntos.

Podemos referir o exemplo de vários registos do mesmo documento, a que são atribuídos termos de indexação conceptualmente diferentes, e que se juntam no catálogo

coletivo, onde a inconsistência dos índices se torna mais evidente. Esta evidência prejudica tanto a recuperação da informação, como a fiabilidade do catálogo.

(Ex.) Título: Manual de investigação em ciências sociais

➤ Termos atribuídos para pesquisa por assunto:

- Investigação em educação
- Ciências sociais
- Sociologia
- Investigação social
- Método de investigação
- Investigação em ciências sociais

Esta inconsistência na indexação dá origem às seguintes situações:

Há muitas entradas de termos não controlados que ficam dispersos no índice; a atribuição de termos muito genéricos dá origem a resultados de pesquisas não pertinentes. Estas situações tornam-se muito evidentes no catálogo coletivo, que corre o risco de perder as vantagens de centralizar a informação, caso não se proceda a uma intervenção sistemática.

➤ Pesquisas dos usuários

Depois de identificadas as inconsistências de indexação nas bases, tivemos interesse em verificar alguns dados referentes à pesquisa dos usuários dos catálogos. Através de

estatísticas de pesquisa fornecidas pelo *software* bibliográfico, referentes a 2009 e 2010, observamos no conjunto das bases, que os pontos de acesso mais utilizados pelo público foram ‘Palavra’ (53%), ‘Assunto’ (18%), ‘Título’ (15%), e ‘Autor’ (14%). A maioria das pesquisas é feita por palavra, o que pode constituir uma intenção de procurar assuntos, com os inconvenientes da pesquisa por palavras, pois não se trata de ‘palavra em assunto’, mas em qualquer campo do registo bibliográfico. Numa primeira análise pensamos tratar-se de uma opção do usuário pela possibilidade de pesquisa que produz mais resultados. Verifica-se que a pesquisa por índices de assunto é inferior à de índices de palavras, o que pressupõe que o usuário pesquisa por palavra, pela fraca recuperação de informação através de índices de assunto.

De acordo com Larson (1991) o declínio no uso do índice de assuntos num catálogo deve-se à substituição da pesquisa por palavras-chave, e que está relacionada com a frustração do usuário ao utilizar o índice de assuntos, que por um lado tem dificuldades em entender, e por outro origina falhas na pesquisa.

Individualmente em cada biblioteca, podem ser de várias ordens as razões que exercem influência nas opções de pesquisa dos usuários, para além da questão da indexação que afeta as pesquisas por assunto. O tipo de sala de leitura, que pode proporcionar uma maior proximidade dos profissionais e dos usuários, os acessos informáticos disponíveis numa sala de leitura que potenciam outro tipo de pesquisa para além do catálogo, e outro tipo de fatores ligados a comportamentos de literacia informacional. Consideramos, no entanto, que a qualidade da catalogação está diretamente ligada com o controlo de autoridades, e que no caso destas bibliotecas é

imprescindível tanto o controlo da nova terminologia, como a correção retrospectiva dos registos de autoridade. Foi a constatação de inconsistências nos índices de assunto que deu origem ao presente estudo empírico.

A prática da indexação é uma das tarefas realizadas nas bibliotecas que mais dificuldade coloca aos profissionais. É uma área de trabalho que exige conhecimentos especializados e muita dedicação, e não há um equilíbrio entre os recursos humanos disponíveis e a atual exigência de serviços por parte dos usuários. Pensamos, contudo, que é possível melhorar esta prática, através de uma nova metodologia de trabalho baseada nas recomendações normativas e nas ferramentas de apoio atualmente disponíveis.

3. Objetivos e Metodologia

Nos capítulos anteriores estudamos os contextos teóricos que enquadram a nossa investigação, dos quais se destacam a importância dos instrumentos normativos que orientam a indexação, a utilização de uma linguagem documental para uma maior coerência, e a avaliação da qualidade dos sistemas de recuperação de informação.

Perante a complexidade que envolve uma indexação de qualidade, quisemos saber qual era a perceção do problema por parte dos profissionais de instituições congéneres (bibliotecas do ensino politécnico), nomeadamente no tipo de ferramentas de apoio que utilizam na execução dessa tarefa.

Por outro lado, tivemos interesse em observar as características da indexação ao nível do estudo de caso, no sentido de avaliar os índices de assunto e verificar as suas debilidades. A partir dessas constatações pretende-se delinear uma política de indexação adequada, de forma a melhorar todo o processo. O estudo de caso, contextualizado no capítulo 2.3, envolve cinco bases bibliográficas que se ligaram em rede para utilização de um catálogo coletivo. Na utilização do catálogo verificámos que os índices de assunto não se encontravam controlados, e que a indexação era heterogénea. Os índices apresentavam uma grande dispersão de termos referentes aos mesmos conceitos, e por outro lado, uma grande falta de consistência na atribuição de termos a documentos iguais. A avaliação da qualidade dos índices produzidos nessas bases revelou-se uma necessidade estratégica para fomentar a consulta dos catálogos, e neste sentido, é feita uma análise formal dos índices de assunto. Esta análise constitui uma abordagem de

grande importância à problemática da indexação, na medida em que uma intervenção para melhoria dos índices irá potenciar bons resultados na pesquisa por assuntos.

3.1 Objetivos

Os pontos centrais da nossa investigação são, em primeiro lugar, comprovar a prática da tarefa de indexação de acordo com a normativa adequada nas bibliotecas politécnicas portuguesas, e em segundo lugar, verificar a qualidade dos índices de assunto nas bases bibliográficas de um grupo de bibliotecas, através da análise de uma amostra da terminologia, de modo a comprovar que há debilidades no desenvolvimento do processo de indexação, e que não há um controlo sistematizado da terminologia. Através da confirmação destas hipóteses, pretende-se demonstrar que a utilização de instrumentos normativos na indexação é pouco consistente, e que esse facto pode ter influência na criação de índices de qualidade.

Neste estudo pretendemos, numa primeira fase, identificar a utilização de instrumentos normativos no processo de indexação por parte dos profissionais das bibliotecas do ensino politécnico. A nossa hipótese de partida em relação a este ponto é que na maioria destas bibliotecas se faz a tarefa de indexação em condições pouco favoráveis, no que diz respeito à utilização de instrumentos normativos e ao controlo da linguagem documental. Como o instrumento de referência nacional é o sistema de

indexação em português SIPORbase, há especial interesse em saber se serve de apoio à construção da linguagem controlada.

Numa segunda fase é analisada a terminologia das bases bibliográficas num grupo de bibliotecas para conhecer as características da indexação, e assim comprovar dificuldades decorrentes de uma indexação não adequada. Perante o conhecimento dessas características, pretende-se justificar uma proposta de melhoria metodológica no processo de indexação nesse grupo específico. Se no primeiro nível de investigação podemos saber sobre a utilização do SIPORbase, neste segundo nível, das características da indexação, verifica-se se a terminologia está representada de acordo com esse sistema.

Para tornar possível o conhecimento destes fatores, foram estabelecidos objetivos, o primeiro de carácter mais geral, e o segundo definindo vários objetivos específicos:

Ob1 – Conhecer a situação das bibliotecas do ensino superior politécnico¹ em Portugal, no que diz respeito à utilização de instrumentos normativos e ferramentas de apoio ao processo de indexação.

Ob2 – Identificar a situação de um grupo específico de Bibliotecas (Instituto Politécnico de Portalegre) quanto às características dos índices de assunto nos catálogos bibliográficos, para:

¹ Adiante designadas de BESP.

- a) Conhecer as características quanto ao uso do formato de intercâmbio UNIMARC;
- b) Conhecer as características quanto ao controlo de terminologia;
- c) Conhecer as características da sintaxe;
- d) Comprovar a consistência dos índices (campo 6xx) com as normas estabelecidas no SIPORbase;
- e) Comprovar a coerência da terminologia em relação as outras fontes de controlo.

3.2 Metodologia

A metodologia para o presente estudo foi delineada de acordo com os objetivos estabelecidos e desenvolve-se a dois níveis, conforme já referido: o primeiro nas Bibliotecas do Ensino Superior Politécnico, e o segundo na rede de Bibliotecas do Instituto Politécnico de Portalegre.

3.2.1 Nível 1: Instrumentos de apoio à indexação nas Bibliotecas do Ensino Superior Politécnico

A rede do ensino superior politécnico em Portugal² é constituída por 98 Escolas Superiores estabelecidas em 15 Institutos Politécnicos e em algumas Universidades (Anexo II). As unidades de informação (bibliotecas, centros documentais, etc.) que pertencem a estas instituições encontram-se, nalguns casos, fundidas numa biblioteca geral, ou num sistema de catalogação centralizada, pelo que se verifica uma ligeira diferença entre o número de instituições existentes e o número de unidades apuradas para recolha dos dados.

3.2.1.1 Recolha de dados

Neste nível de investigação de carácter exploratório, foi feito o levantamento das bibliotecas do ensino politécnico, cujos responsáveis foram contactados por *e-mail* com apresentação do estudo e pedido de colaboração. Foram contactadas 56 instituições em Fevereiro e Março de 2009. Em Novembro de 2010 foi reforçado o pedido de colaboração junto das instituições que não participaram na primeira fase através de uma nova abordagem. Foi apurado um total de 44 respostas válidas, o que corresponde a 78,5% do universo inquirido.

² <http://www.dges.mctes.pt/DGES/pt/Instituicoes>

Foi usado o método de inquérito por questionário por se considerar o mais adequado, uma vez que dá resposta às questões pretendidas de forma prática, e é possível contactar todo o universo em causa.

O questionário foi desenhado na tecnologia *Google Docs*, e contém questões sobre a utilização de instrumentos normativos, e sobre as ferramentas de apoio na recolha e controlo da terminologia. As questões são colocadas com possibilidade de resposta Não/Sim, sendo as respostas afirmativas direccionadas para a designação de normas e ferramentas, em resposta aberta³ (Anexo IV).

3.2.1.2 Seleção da amostra

Conforme referido anteriormente, neste nível de observação foi possível seleccionar todo o universo de estudo, uma vez que foram contactadas todas as instituições do ensino politécnico em Portugal.

³ Pensamos que o facto de não ser dada hipótese de escolha entre designações poderá corresponder melhor à realidade.

3.2.1.3 Variáveis

Conforme foi evidenciado na parte teórica (1.1) do trabalho, é fundamental a utilização das normas para desenvolver a indexação, pois o seu desconhecimento pressupõe que não são observados os princípios básicos para o desenvolvimento adequado do processo. Evidenciou-se também a importância das linguagens controladas na fase de representação dos conceitos, e nesse sentido é relevante obter informação sobre a utilização de fontes de terminologia como auxiliares na recolha de termos, e também sobre o controlo de autoridades para fixação da terminologia e estabelecimento de relações. Para ir ao encontro destes pressupostos teóricos, foram estabelecidas variáveis que nos permitem obter essa informação.

II - Utilização de normas no processo de indexação

(deve estar presente uma norma como a NP 3715, a NP 4036⁴ ou um sistema de controlo terminológico como o Siporbase);

(a referência a instrumentos não considerados como tal, ou respostas que não permitam identificar o instrumento utilizado, são consideradas numa categoria de 'não especificado').

⁴ NP 3715:1989 Documentação. Método para análise de documentos, determinação do seu conteúdo e selecção de termos de indexação. NP 4036:1992 Documentação. Tesouros monolíngues: directivas para a sua construção e desenvolvimento.

Exemplos de respostas não especificadas⁵:

- CDU
- CDU e Thesaurus temáticos
- Eurovoc - Lince (Educação)
- Lista cabeçalho de assuntos
- Listas de assuntos da Universidade de...
- Manual da Biblioteca Nacional
- Manual interno
- Normas
- NP's
- Política de indexação (interna)
- Regras Portuguesas de Catalogação
- Thesaurus e listas de termos
- Thésaurus, Lista de descritores

I2 - Utilização de léxico(s) de assuntos ou tesouro(s) para apoio à indexação

(presença de léxico(s) de assunto ou tesouro(s) para recolha de termos);

(a referência a instrumentos não considerados como tal, ou respostas que não permitam identificar o instrumento utilizado, são consideradas numa categoria de 'não especificado').

Exemplos de respostas não especificadas⁶:

- as duas maneiras
- CDU
- CDU, Listagens elaboradas pela própria biblioteca
- Ciências Económicas, Direito e outros
- Léxico de assuntos (não existe thesaurus em língua portuguesa na área temática em causa)
- Sendo especializada foi criado um léxico de assuntos pelo bibliotecário
- SIPORbase, CDU
- Thesaurus
- Vários disponíveis na internet, de acordo com as várias áreas científicas
- vários thesaurus e listas de assuntos

⁵ Considerou-se favorável exemplificar esta categoria para melhor compreensão; os exemplos foram recolhidos de respostas obtidas nos inquéritos, as quais demonstram desde logo alguma confusão existente entre normas e ferramentas, como se verá nos resultados mais adiante.

⁶ Idem.

I3 - Utilização de ficheiro de autoridades de assunto ligado à base bibliográfica

(*sim ou não*)

3.2.1.4 Análise

As respostas ao inquérito sobre “utilização de instrumentos de apoio à indexação” foram recolhidas em folha Excel; os dados foram inseridos e tratados no *software* SPSS *Statistics* 17.0, a partir do qual resultaram tabelas com a distribuição de frequências das respostas.

Sintetizam-se as variáveis estabelecidas para o Nível 1 – Instrumentos de apoio à indexação nas BESP (tab.3.1).

Nível 1 – Instrumentos de apoio à indexação nas BESP

I 1	Utilização de instrumentos normativos
I 2	Utilização de léxico(s) de assuntos ou tesauro(s)
I 3	Utilização de ficheiro de autoridades

Tabela 3.1: Grupo de variáveis do nível 1

3.2.2 Nível 2: Características da indexação nas Bibliotecas do Instituto Politécnico de Portalegre

O Instituto Politécnico de Portalegre dispõe de quatro unidades de informação inseridas nas suas Escolas Superiores⁷ e de um fundo documental⁸ em acesso reservado. Para o nosso estudo são observadas as cinco bases bibliográficas correspondentes, que se encontram disponíveis ao público através dos catálogos *online*.

3.2.2.1 Recolha e depuração de dados

Neste nível é estudada a informação contida nas bases bibliográficas adiante designadas de B1EDU, B2SAU, B3TGE, B4AGR, B5GER, e que representam áreas do conhecimento em vários domínios, inseridos num contexto de ensino superior, conforme representados de seguida (tab.3.2).

	Bibliotecas / Fundos	Áreas temáticas
B1	Esc. Superior de Educação	Ciências da educação e comunicação
B2	Esc. Superior de Saúde	Ciências da Saúde
B3	Esc. Superior Tecnologia e Gestão	Gestão; Engenharia e Tecnologias
B4	Esc. Superior Agrária de Elvas	Ciências agrárias; Ciência animal
B5	Fundo Prof. Fortunato Queirós	Geral; História; Educação

Tabela 3.2: Bibliotecas e respetivas áreas temáticas

⁷ Escola Superior de Educação, Escola Superior de Saúde, Escola Superior de Tecnologia e Gestão, Escola Superior Agrária de Elvas.

⁸ Fundo documental Professor Fortunato Queirós.

A recolha de dados é feita a partir de registos do UNIMARC bibliográfico⁹, dos quais se analisa a terminologia dos índices de assunto apresentada no bloco 6xx.

O sistema de gestão bibliográfica utilizado nestas bibliotecas (Porbase5 – SQL) permite uma intervenção em ‘modo profissional’ em todas as bases para recuperar registos de acordo com os campos especificados do UNIMARC. Assim, foram percorridas as bases bibliográficas, para verificação de preenchimento dos campos (tab.3.3) em todos os registos.

Campos do UNIMARC bibliográfico – Bloco 6xx (Assuntos)
600 – Nome de Pessoa
601 – Nome de Colectividade
602 – Nome de Família
604 – Autor-Título
605 – Título
606 – Nome Comum (<i>análise de registos em amostra</i>)
607 – Nome Geográfico

Tabela 3.3: Campos pesquisados nas bases para verificação de conteúdo

Começámos por identificar o preenchimento de todos os campos acima mencionados, tendo usado a expressão de pesquisa P(v6xx) que verifica a presença de um determinado campo ou subcampo do bloco 6xx em que P é Presente, e a sintaxe é P(<campo>).

⁹ Formato utilizado na PORBASE (Base Nacional de Dados Bibliográficos), que adotou o UNIMARC como formato normalizado desde a sua criação em 1986, para troca de dados bibliográficos com as bibliotecas cooperantes.

Foram usadas as seguintes expressões de pesquisa:

P(<v600>) P(<v601>) P(<v602>) P(<v604>) P(<v605>) P(<v606>) P(<v607>).

Assim, a partir do número de registos nas bases à data¹⁰ da pesquisa, é identificado o preenchimento de campos de assunto, o que permitiu desde logo reconhecer, por um lado, os registos sem assunto, e por outro, os registos com preenchimento no campo 606 e nos restantes campos do bloco 6xx, (tab.3.4 e 3.5).

Observando a totalidade de registos nas bases, verificamos que dos 86813 registos existentes, 38169 não têm campos de assunto preenchidos, o que equivale a cerca de 44%, que consideramos uma percentagem muito elevada, pelo que representa de documentos não acessíveis através de pesquisas por assunto. Os registos que contêm assuntos (48644) representam cerca de 56%.

Bases bibliográficas	Total registos	Registos sem assunto		Registos com assunto	
B1EDU	47233	28291	59,8%	18942	40,1%
B2SAU	12230	2544	20,8%	9686	79,1%
B3TGE	13020	4831	37,1%	8189	62,8%
B4AGR	13662	2059	15,0%	11603	84,9%
B5GER	668	444	66,4%	224	33,5%
Total	86813	38169		48644	

Tabela 3.4: Número de registos sem assunto e com assunto

¹⁰ Estas pesquisas foram realizadas nos dias 2 e 3 de Junho 2011.

Tendo presente a ideia de catálogo coletivo, é relevante observar a situação heterogénea das bases bibliográficas, quanto à distribuição de registos sem assunto (tab.3.4). Estes dados serão confrontados mais adiante, no sentido de observar a eventual relação de erros com a quantidade de cabeçalhos preenchidos, ou seja, um menor número de registos indexados, poderá equivaler a uma indexação mais cuidada e portanto mais demorada?

Dos registos com assunto verificamos que 47433 (97,5%) pertencem ao campo 606, referentes aos nomes comuns, e os restantes 1211 (2,5%) pertencem aos outros campos¹¹ do bloco 6xx, referentes a nomes próprios (tab.3.5). O preenchimento do campo 606 concentra quase a totalidade da indexação.

	<i>Campo 606</i>		<i>Outros campos 6xx</i>	
Total	47433	97,5%	1211	2,5%
B1EDU	18938	99,9%	4	0,0%
B2SAU	9644	99,5%	42	0,4%
B3TGE	7575	92,5%	614	7,4%
B4AGR	11159	96,1%	444	3,8%
B5GER	117	52,2%	107	47,7%

Tabela 3.5: Registos com assunto (campo 606 e outros campos do bloco 6xx).

Na observação dos dados individualmente por bases, os campos preenchidos do bloco 6xx representam, nalguns casos, uma percentagem praticamente nula (tab.3.5). Só

¹¹ Nesta contagem encontram-se 16 registos que contêm também o campo 606 preenchido.

no caso da base *B5GER* se observa um equilíbrio entre o preenchimento de campos relativos a nomes comuns e a nomes próprios.

3.2.2.2 Seleção da amostra

Foi possível analisar a totalidade de registos (1211) referentes a campos do bloco 6xx. No entanto, para análise da terminologia no campo 606, houve necessidade de fazer a observação por amostra. A amostragem estatística é um dos métodos de avaliação usados com mais frequência para análise da qualidade das bases de dados, e o tamanho da amostra pode oscilar entre 200 e mais de 2000 registos. O procedimento mais objetivo para seleção de registos é o de extração ao acaso. No entanto, dependendo das características do estudo, é possível utilizar critérios de seleção para criação de amostras representativas, tais como, a data de criação dos registos, o assunto dos documentos, as regras ou normas de descrição, etc. (Ortego e Bonal 2000).

Embora os métodos aleatórios de amostragem sejam os que à partida melhor asseguram a representatividade da amostra, a sua validade não coloca de parte os métodos não aleatórios. A dificuldade em obter dados de alguns elementos no universo de estudo, como restrições relacionadas com o *software*, etc., levaram a optar por uma das técnicas não aleatórias. Um dos tipos de amostra não aleatória é a amostra intencional, que é “composta de elementos selecionados deliberadamente, geralmente

porque o investigador considera que possuem características que são típicas ou representativas da população” (Vicente *et al.* 1996).

Para seleção da nossa amostra selecionámos um critério cronológico da criação de registos. Recolheram-se os registos criados nos anos 2008, 2009 e 2010, e esta decisão foi tomada com base, por um lado numa melhor representação das tendências de indexação, e por outro, na adequação do *software* bibliográfico para essa recolha.

Podemos, então, dizer que em relação ao campo 606 de Nomes Comuns, se estudou uma amostra ‘intencional’, pois foram selecionados os registos de anos recentes, que melhor poderão indicar as características da indexação.

Foram selecionados os registos referentes a monografias, pois é a este nível bibliográfico que está localmente instituído proceder à indexação por assunto. Como nalgumas bibliotecas também não se indexam documentos eletrónicos, foi, nesses casos, usada uma expressão de pesquisa que conjuga, através de operadores booleanos, a data de criação dos registos com o nível bibliográfico monografias e exclui os documentos eletrónicos. Foi usada a expressão de pesquisa das datas conjugadas com os outros elementos:

(<Data de Criação>20080101 até <Data de Criação>20101231)

E (<Nível Bibliográfico>M)

E NÃO(<Tipo de Documento>L)

Em relação aos outros campos do bloco 6xx não houve necessidade de seleccionar uma amostra, uma vez que o número de ocorrências permitia a observação de todos os registos.

Assim, foi apurada uma amostra de 1573 registos, dos quais resultaram 2218 cabeçalhos de assunto referentes ao campo 606 UNIMARC (Nomes comuns). Em relação aos outros campos do bloco 6xx foram recolhidos todos os registos (1211), dos quais resultaram 1282 cabeçalhos de Nomes próprios. Foram analisados 3500 cabeçalhos no total.

Os registos foram recolhidos em formato de visualização UNIMARC, de onde foram retirados os campos não necessários, permanecendo aqueles cuja informação é relevante para a análise, como o campo de título e de assunto, e em alguns casos o campo de notas (fig.3.1). Foram organizados por bases para observação individual em cada conjunto, e os dados foram posteriormente inseridos em folhas Excel para filtragem e análise.

200:1 ^aA revolução francesa e as suas consequências^fAntónio de Sèves
210: ^aLisboa^cPro Domo^d1944
225:2 ^aCultura política^v2
300: ^aConferência lida na Sala Algarve da Sociedade de Geografia, em Maio de 1944.
606: ^384159^aRevolução francesa, 1789-1799^2SIPOR
606: ^384738^aRevolução francesa, 1789-1799^xConsequências^2SIPOR

Fig. 3.1: Registo bibliográfico em visualização UNIMARC

Na tabela 3.6 pode-se observar o peso das bases na amostra de dados. Nos nomes comuns é mais ou menos equilibrado o peso de três bases, sendo bastante mais pequeno noutras duas (B4AGR e B5GER). Nos nomes próprios verifica-se uma distribuição de registos menos uniforme, sendo quase nula no caso de uma base (B1EDU).

Bases/áreas temáticas	Nomes comuns	Nomes próprios
B1 Educação	35%	0%
B2 Saúde	27%	5%
B3 Tecnol. Gestão	23%	48%
B4 Agrária	7%	37%
B5 Geral	8%	10%

Tabela 3.6: Peso das bases na amostra de cabeçalhos

Relativamente ao nível de significação da amostra, consultámos a tabela¹² de Arkin e Colton (*apud* Bugada 1974), para determinação de uma amostra¹³ de um universo finito, e verificámos que para um universo de 50000 elementos, com um nível de confiança de 99,7%, e com uma margem de erro de 4%, a tabela apresenta o valor de 1368 elementos. A nossa amostra tem o valor de 1573 registos num universo de 48644, pelo que consideramos a sua significação dentro destes parâmetros.

¹² “*Tables for Statisticians*“ dos autores Herbert Arkin e Raymond Colton (1963); As tabelas apresentam a amplitude da população (universo) e os valores de acordo com o nível de confiança e a margem de erro correspondente.

¹³ Considera-se um universo finito menos de 100.000 elementos; Um nível de confiança de 99,7% (que corresponde a uma probabilidade de 997 por 1000), representa uma segurança quase absoluta, e a margem de erro não deve ultrapassar os 5% (Bugada 1974).

3.2.2.3 Variáveis

A sistematização de variáveis para as características da terminologia das cinco bases, foi estabelecida a partir de evidências identificadas de um modo geral em todas as bases, como:

- ✘ Preenchimento incorreto de campos e subcampos UNIMARC, em que o termo atribuído pertence a outro campo, como por exemplo, um ‘Nome comum’ no campo de um ‘Nome próprio’;
- ✘ Existência de termos ambíguos, como é o caso de termos relacionados com saúde humana e saúde animal, que sem qualificadores, num catálogo coletivo não permitem selecionar esta diferenciação no índice;
- ✘ Existência de termos obscuros, cujo significado depende do contexto e não têm limites definidos. São termos que, sendo criados em cabeçalhos isolados não têm significado explícito, e fora do contexto¹⁴ de um assunto principal, têm significado ambíguo, como é o caso do termo ‘Previsão’;
- ✘ Problemas de carácter conceptual, que afetam a recuperação de informação¹⁵, verificando-se nalguns casos a atribuição de termos de indexação muito genéricos. Ao realizar pesquisas na base com esses termos, recupera-se uma grande quantidade de registos em que se identificam conceitos mais específicos, ou até conceitos diferentes.

¹⁴ Referido por Svenonius (2003) como “*context independence*”, no sentido de que os termos de indexação devem ser significativos fora do contexto.

¹⁵ Na análise de registos é tida em conta a recomendação da norma NP3715, ‘ponto 5. Análise do documento’, sendo os termos de indexação analisados no contexto do título ou outros elementos do registo bibliográfico, que permitem considerar explícita as características observadas.

A natureza das situações foi identificada de acordo com o sistema SIPORbase, que nos encaminhou para a análise da forma dos cabeçalhos, do controlo da terminologia e da representação de conceitos, revelando-se pontos essenciais a ser observados.

Com base no referencial teórico do Sistema SIPORbase, e para uma análise mais consistente, foram definidas variáveis para três eixos principais de análise que são o formato UNIMARC, a sintaxe de cabeçalhos e a terminologia. Considerámos que a observação destes três eixos é fundamental para conhecer as características da indexação, ao nível da inserção de dados, ao nível da sintaxe dos cabeçalhos pela complexidade da sua construção, e ao nível da terminologia pela dificuldade que implica a normalização do vocabulário.

Embora estes eixos sejam transversais à análise dos cabeçalhos de nomes comuns e de nomes próprios, as características de cada um destes grupos exigem a sua observação de forma separada. No eixo da terminologia revelou-se adequado observar algumas características particulares dos nomes comuns no que respeita à forma gramatical dos termos (simples ou compostos) nos cabeçalhos não subdivididos (*ex.* ‘Economia’ ou ‘Condições económicas’).

Para além dos três eixos centrais houve ainda a necessidade de estabelecer duas outras variáveis: uma que diz respeito ao princípio do cabeçalho específico que constitui um dos princípios básicos da indexação do sistema SIPORbase, e outra que diz respeito

à cobertura terminológica das bases, uma vez que em alguns casos a terminologia é recolhida de fontes como tesouros ou listas de descritores de temáticas específicas.

Nos três eixos principais as variáveis e respetivas categorias foram definidas de acordo com a tipologia dos cabeçalhos:

Cabeçalhos de **Nomes Comuns (NC)**

São nomes comuns os substantivos (ou locução nominal) que correspondem ao campo 606 UNIMARC (Nome comum usado como assunto). Quanto à forma podem ser de dois tipos:

Não subdividido (NC1) - quando contém apenas o subcampo ^a

(*ex.* Economia)

Subdividido (NC2) - quando contém uma ou mais subdivisões

(*ex.* Economia -- [Relatórios] -- Portugal – 2000)

As subdivisões dos cabeçalhos dos nomes comuns são representadas no UNIMARC pelos seguintes códigos:

\$a Elemento de entrada (assunto principal)

\$j Subdivisão de forma

\$x Subdivisão de assunto

\$y Subdivisão geográfica

\$z Subdivisão cronológica

Cabeçalhos de **Nomes Próprios (NP)**

São nomes próprios os nomes de pessoa, entidade, e lugar,¹⁶ que correspondem respetivamente aos campos 600, 601, 607 UNIMARC. Estes cabeçalhos apresentam subcampos para representação dos cabeçalhos segundo as Regras Portuguesas de Catalogação; não se trata pois de subdivisões de assunto como no caso dos nomes comuns. Os cabeçalhos dos nomes próprios são representados no UNIMARC pelos seguintes códigos:

600 Nome de pessoa (subcampos):

- \$a Palavra de ordem
- \$b Outra parte do nome
- \$c Elementos de identificação ou distinção
- \$d Numerais romanos
- \$f Datas

Ex.: Pessoa, Fernando, 1888-1935

601 Nome de colectividade (subcampos):

- \$a Palavra de ordem
- \$b Subdivisão (ou nome da colectividade)
- \$c Elementos de identificação ou distinção

Ex.: Instituto Politécnico de Portalegre. Escola Superior de Saúde

607 Nome geográfico (subcampos):

- \$a Elemento de entrada

Ex.: Alentejo (Portugal)

¹⁶ É igualmente considerado como nome próprio o nome de família, que corresponde ao campo 602 UNIMARC, que não ocorreu nos registos.

Erros UNIMARC

Verifica-se se os termos são inseridos nos devidos campos e subcampos UNIMARC, dos cabeçalhos de Nomes Comuns e de Nomes Próprios.

U1

Erro na inserção de termos não correspondentes à etiqueta de campo do bloco 6xx.

600:ªAnálise de dados

(nome comum no campo de nome de pessoa)

606:ªInstituto Politécnico de Portalegre

(nome de coletividade no campo de nome comum)

606:ªPortugal

(nome geográfico no campo de nome comum)

U2

Erro na inserção de termos em qualquer subcampo nos cabeçalhos de Nomes Comuns (^j ^x ^y ^z), conforme os exemplos que se seguem:

U2^j

606:ªCuidados veterinários^jAnimais de companhia

(erro em ^j que é um subcampo referente à forma)

U2^x

606:ªDesenvolvimento regional^xAlentejo

(erro no subcampo ^x Alentejo que deve ser representado pelo subcampo ^y, de aspeto geográfico)

U3

Erro na inserção de termos em qualquer subcampo nos cabeçalhos de Nomes próprios (^a, ^f) conforme os exemplos que se seguem:

U3^a

600:^aFernando Pessoa

(erro no subcampo ^a que representa a entrada pelo apelido do autor, que deve ser seguido do ^b para o nome)

U3^f

600:^aPessoa^bFernando

(erro no subcampo ^f que representa a(s) data(s) que deve(m) fazer parte do cabeçalho de autor - ^aPessoa^bFernando^f1888-1935)

Erros de Sintaxe

É analisada a ordem dos campos e subcampos na representação dos cabeçalhos.

S1

Nos nomes comuns verifica-se se os cabeçalhos subdivididos são representados ao nível conceptual pelo assunto principal no ^a e no(s) subsequente(s) ^j, ^x, ^y, ^z, correspondendo respetivamente a faceta de forma, subdivisão de assunto, faceta geográfica ou cronológica.

Se a ordem dos subcampos é incorreta ou se os conceitos representados em subcampos não se consideram como tal, conforme os exemplos:

606: ^aEscolas^xCentros de formação

(Centros de formação não se consideram subdivisão de assunto de Escolas)

606: ^aProjecto de instalação^xSistema de rega

(o conceito 'Sistema de rega' deve ser considerado como assunto principal e o conceito 'Projecto de instalação' deve ser considerado subdivisão de assunto)

606: ^aEstatísticas^xPoder de compra

(o conceito 'Poder de compra' deve ser considerado assunto principal e 'Estatísticas' neste caso, representa um qualificador de forma que deve ser inserido em ^j).

S2

Nos nomes próprios esta variável é observada da mesma forma, tendo em conta as subdivisões de assunto existentes depois do cabeçalho de nome próprio, construído de acordo com as Regras Portuguesas de Catalogação, e cuja forma é analisada na variável de terminologia.

600: ^aPessoa^bFernando^f1886-1935^jCatálogos^xExposições

(a subdivisão de forma deve ser apresentada depois da subdivisão de assunto)

601: ^aSociedade Portuguesa de Autores^z1985^xAniversários

(a subdivisão cronológica deve ser apresentada no final do cabeçalho)

Terminologia

É analisada a terminologia de acordo com os índices da BNP, os quais obedecem às regras do sistema de indexação SIPORbase.

T1

Não coincidência na designação de Nomes Comuns em qualquer campo ou subcampo.

Alguns dos exemplos considerados nesta categoria são exemplificados de seguida. É o caso da terminologia referente à saúde humana e à saúde animal, que tem um significado ambíguo se não apresentar um qualificador. Outro erro comum é o uso de um termo qualificador de forma como assunto principal (Bibliografia).

	<i>(não conforme)</i>	<i>(mas sim)</i>
T1^a	606: ^a Obesidade	^aObesidade (Animais)
T1^a	606: ^aBibliografia	^aAssunto principal^jBibliografias
T1^j	606: ^aDesign^jDicionário	^aDesign^jDicionários
T1^x	606: ^aEconomia^xEercícios de economia	^aEconomia^jExercícios
T1^y	606: ^aSociologia rural^yAlentejo	^aSociologia rural^yAlentejo (Portugal)

T2

Utilização de termos compostos nos cabeçalhos não subdivididos dos Nomes Comuns.

Na observação dos cabeçalhos¹⁷ de nomes comuns, constituídos por um só campo, ou seja, sem subdivisões, é relevante analisar a utilização de termos compostos. De acordo com as referências SIPORbase, este tipo de cabeçalhos (não subdivididos) devem corresponder à representação de um assunto simples, constituído por um só conceito e representado por um termo de indexação; gramaticalmente esse termo pode ser expresso em formas de termo simples ou termo composto. São analisados os termos quanto à forma gramatical, no sentido de identificar se correspondem a este princípio. São exemplos de termos compostos ‘não conforme’:

	<i>(não conforme)</i>	<i>(mas sim)</i>
T2 ^a	606: ^aEducação de adultos	606: ^aAdultos^xEducação
T2 ^a	606: ^aCondição feminina	606: ^aMulheres^xCondições sociais
T2 ^a	606: ^aDesenvolvimento curricular	606: ^aCurrículos^xDesenvolvimento

- Os termos compostos ‘Educação de adultos’, ‘Condição feminina’ e ‘Desenvolvimento curricular’ representam mais do que um conceito, e devem ser decompostos para dar origem a cabeçalhos subdivididos;

¹⁷ Cabeçalho (de termo simples ou termo composto) - é o elemento da linguagem documental que representa um assunto e pode ser constituído por um só termo de indexação, se se tratar de um assunto simples (um só conceito), ou por dois ou mais termos de indexação pré-coordenados, se se tratar de um assunto complexo (mais do que um conceito).

T3, T4, T5

Não coincidência na designação de Nomes Próprios (de pessoa, entidade, lugar) em qualquer campo ou subcampo, conforme os exemplos:

	<i>(não conforme)</i>	<i>(mas sim)</i>
T3 ^a	600: ^aMaria Ulrich	600: ^aUlrich^bMaria^f1908-1988
T4 ^a	601: ^aEscola Superior de Educação de Portalegre	601: ^aInstituto Politécnico de Portalegre^bEscola Superior de Educação
T5 ^a	607: ^aExtremadura	607: ^aExtremadura (Espanha)
T5 ^a	607: ^aAlqueva	607: ^aBarragem do Alqueva (Conselho de Portel, Portugal)

Nos nomes próprios o controlo da terminologia obedece às Regras Portuguesas de Catalogação, conforme a sequência dos exemplos, para nome de pessoa nome de instituição, nome geográfico.

Cabeçalho específico

Na observação dos nomes comuns é ainda estabelecida a variável do Cabeçalho específico, em que se analisa a terminologia dos cabeçalhos ao nível da representação de conceitos, através da identificação de termos muito genéricos.

O princípio do cabeçalho específico é um princípio básico de indexação consignado no sistema SIPORbase, e estabelece que “cada assunto deve ser representado por um cabeçalho cujo sentido não seja nem mais lato nem mais restrito que esse assunto”. Assim, a partir da análise de registos em que é explícito no título a atribuição de um cabeçalho mais específico do que o apresentado, são marcados os casos que não correspondem a esse princípio, conforme os exemplos que se seguem:

- A um documento com o título “O Estuário do Tejo”, e com o termo de indexação – Ecologia
(poderia ser atribuído um termo mais específico)
 - Estuário do rio Tejo (Portugal)

- A um documento com o título “Anuário estatístico da Região Alentejo 2009”, e com o cabeçalho – Estatística -- Alentejo
(poderia ser atribuído um cabeçalho mais específico)
 - Alentejo (Portugal) -- 2009 -- [Dados estatísticos]

- A um documento com o título “Carta geológica de Portugal: notícia explicativa da folha 33-C Campo Maior”, e com o termo de indexação – Geologia
(poderia ser atribuído um cabeçalho mais específico)
 - Cartas topográficas -- Campo Maior (Portugal)
 - Informação cartográfica -- Campo Maior (Portugal)

- A um documento com o título “Estudo da caracterização da qualidade ambiental na indústria corticeira em Portugal”, e com o termo de indexação – Agronomia
(poderia ser atribuído um cabeçalho mais específico)
 - Cortiça – Portugal
 - Indústria da cortiça – Portugal
 - Ambiente – Qualidade -- Portugal

- A um documento com o título “Lei n.º 62/2007, de 10 de Setembro: regime jurídico das instituições de ensino superior”, e com o termo de indexação – Regime jurídico
(poderia ser atribuído um cabeçalho mais específico)
 - Ensino superior – Portugal – 2007 – [Legislação]

- A um documento com o título “English grammar in use: a self-study and practice book...”, e com o termo de indexação – Língua inglesa
(poderia ser atribuído um cabeçalho mais específico)
 - Língua inglesa – Gramática – [Exercícios]

Fontes de terminologia

Verifica-se a cobertura das bases em relação às fontes de informação preferenciais utilizadas (tesauros e índices da BNP). São confrontados os termos de indexação dos registos com os descritores dos tesauros e índices autorizados da BNP. Só é considerada a equivalência exata nos tesauros e índices, conforme os exemplos (tab.3.7).

<i>termos</i>	<i>(presença)</i>	<i>coincidência / (não coincidência)</i>
Educação cívica	Eurovoc	(BNP) Educação para a cidadania
Educação das crianças	-	BNP
Política de saúde	Decs	BNP
Nutrição animal	Agrovoc	(BNP) Animais -- Alimentação

Tabela 3.7: Exemplos de cobertura das fontes de terminologia

- O termo ‘Educação cívica’ aparece como descritor no tesauro Eurovoc, mas nos índices BNP o termo autorizado é ‘Educação para a cidadania’. Neste caso o termo exato foi identificado no tesauro.
- O termo ‘Educação das crianças’ aparece como índice autorizado na BNP, e não consta como descritor no tesauro Eurovoc. Neste caso o termo exato foi identificado nos índices da BNP.
- O termo ‘Política de saúde’ aparece identificado no DeCs e nos índices da BNP. Neste caso conta como tendo cobertura das duas fontes.

O mesmo procedimento é feito de acordo com as fontes preferenciais usadas em cada base, ou seja, são consultados o tesauro Eurovoc no caso das bases de ciências de educação, de gestão e geral, o Agrovoc para a base de ciências agrárias, a lista de descritores DeCs para a base de ciências da saúde, e os índices BNP para todos os casos (Anexos V e VI).

Para melhor observação, sintetizam-se as variáveis estabelecidas para o Nível 2 – Características da indexação nas BIPP (tab.3.8).

Nível 2 – Características da indexação nas BIPP

TIPO DE CABEÇALHOS		UNIMARC (erros)	SINTAXE (ordem)	TERMINOLOGIA (forma)		Cabeçalho específico	Fontes de terminologia
NC Nomes Comuns	NC1 não subdividido	U1 (campo)	<i>(não se aplica)</i>	T1 ^a	T2^a Utilização dos termos compostos	Utilização do princípio do cabeçalho específico	Tesauros Lista descritores
	NC2 subdividido	U1 (campo) U2^ (subcampos) ^a, ^j, ^x, ^y, ^z	S1 (ordem) ^j, ^x, ^y, ^z	T1 ^ a, ^j, ^x, ^y, ^z	<i>(não se aplica)</i>		
NP Nomes Próprios		U1 (campo) U3^ (subcampos) ^a, ^b, ^c^, ^f	S2 (ordem) ^j, ^x, ^y, ^z	T3 (pessoa) T4 (entidade) T5 (geográfico) ^a, ^b, ^c, ^f	<i>(não se aplica)</i>	<i>(não se aplica)</i>	Índices BNP

Tabela 3.8: Grupo de variáveis do nível 2

3.2.2.4 Análise

O método de análise para observação das variáveis das características da indexação foi de acordo com os procedimentos que a seguir se apresentam:

- Análise da tipologia dos cabeçalhos dos Nomes Comuns, para distinção entre cabeçalhos não subdivididos e cabeçalhos subdivididos. Esta observação permite saber qual o tipo de cabeçalhos mais usado.
- Nos cabeçalhos não subdivididos houve interesse em observar a utilização de termos compostos¹, pois devem corresponder à expressão de um só conceito. Esta foi uma categoria estabelecida na variável de terminologia, que nos permite saber se expressam essa característica, ou se, tratando-se de mais que um conceito, deveriam dar origem a um cabeçalho subdividido, representado através da combinação dos termos correspondentes aos vários conceitos, conforme o “princípio da pré-coordenação”.
- Relativamente ao controlo da terminologia, tanto nos Nomes Comuns como nos Nomes Próprios foram elaboradas listas de ocorrência de termos (Anexos V e VI), para verificação do vocabulário nos índices de assunto da BNP, que são controlados de acordo com o sistema SIPORbase.², e nos tesouros e lista de descritores usados como fonte de terminologia.

¹ Termo composto - locução nominal, na forma gramatical de expressão adjetiva ou prepositiva, designativa de um único conceito; termo de indexação que pode ser decomposto morfológicamente em componentes distintos, podendo cada um destes ser expresso ou reexpresso por uma palavra suscetível de servir, por si própria, de termo de indexação.

² O SIPORbase é um sistema pré-coordenado e todos os termos são criados à medida que se vai fazendo a indexação dos documentos. A terminologia é controlada no ficheiro de autoridade, seguindo todas as regras desse sistema. Os termos nos índices

- Para análise da variável ‘Cabeçalho Específico’ são observados os registos completos (título, notas, etc.), e marcados os cabeçalhos com termos em que se verificou a ausência do estipulado nesse princípio. Para a identificação de termos muito genéricos relativamente aos conceitos expressos, apoiamo-nos nos índices de assunto da BNP e no tesouro Eurovoc. Este tesouro serve de apoio a todas as bibliotecas, e verificámos a ocorrência de termos que podemos considerar ‘muito genéricos’; isto é, o resultado de pesquisa de um termo no tesouro apresenta o número de conceitos em que ocorre esse termo. Alguns exemplos desses termos são: Gestão tem 37 conceitos, Economia 35, Educação 36, Energia 41, Ensino 43. Nestes casos pressupõe-se a possibilidade de poder especificar um conceito a partir das sugestões apresentadas no tesouro.
- Na variável de ‘Fontes de terminologia’ verificou-se se os termos correspondiam aos descritores dos tesouros ou listas que são usados em cada caso, pois a não concordância com os índices da BNP poderá ser justificada por um maior recurso aos tesouros, que podem apresentar terminologia mais específica.

Toda a informação foi tratada manualmente através do registo de ocorrências em folhas Excel, a partir das quais foram feitas as respetivas marcações, filtragens e contagens, para elaboração de tabelas e gráficos.

BNP são criados no ficheiro de autoridade para controlo da terminologia. Assim, o índice de assuntos é o resultado da indexação, isto é, dos cabeçalhos atribuídos aos registos bibliográficos, conjuntamente com os termos criados no ficheiro de autoridade para controlo da terminologia” (informação cedida pela Dra. Manuela Santos, responsável da Área de Classificação e Indexação na BNP).

Nas técnicas de análise foram usadas estatísticas descritivas e para contrastar a dependência de variáveis foi usado o coeficiente de correlação de *Pearson*³, que mede o grau de relação entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1, e os resultados podem ser interpretados da seguinte forma:

0.70 para mais ou para menos indica uma forte correlação

0.30 a 0.7 positivo ou negativo indica correlação moderada

0 a 0.30 indica uma fraca correlação.

Na aplicação deste coeficiente às variáveis da nossa investigação, o cálculo foi feito em Excel recorrendo à função CORREL.

³ Em estatística descritiva, o coeficiente de correlação de Pearson mede o grau de correlação (e a direção dessa correlação – se positiva ou negativa) entre duas variáveis. Este coeficiente, normalmente representado por r assume apenas valores entre -1 e 1. $r = 1$ significa uma correlação perfeita positiva entre as duas variáveis, $r = -1$ significa uma correlação negativa perfeita entre as duas variáveis, isto é, se uma aumenta, a outra sempre diminui; $r = 0$ significa que as duas variáveis não dependem linearmente uma da outra, no entanto, pode existir uma dependência não linear (Wikipédia) <http://pt.wikipedia.org>.

4. Resultados

Os resultados da nossa investigação são apresentados conforme a definição das variáveis para os dois níveis de observação, e organizadas de acordo com as respetivas categorias.

4.1. Nível 1 - Instrumentos de apoio à indexação nas BESP

Conforme definido no primeiro objetivo da nossa investigação, pretendemos saber como se processa a indexação nas bibliotecas do ensino politécnico em Portugal, nomeadamente no que diz respeito a três pontos essenciais, como a utilização de normas para desenvolver a tarefa, o recurso a ferramentas para recolha de terminologia controlada, e a utilização de um ficheiro para controlo de autoridades. De acordo com as variáveis estabelecidas neste nível de observação, apresentamos os resultados obtidos nos inquéritos:

Na variável **II**, sobre a utilização de normas no processo de indexação, a resposta foi afirmativa na maioria dos casos, sendo, no entanto, quase equiparada a percentagem dos que não usam qualquer instrumento normativo (tab.4.1).

Valid	Frequency	Percent
não	19	43,2
sim	25	56,8
Total	44	100,0

Tabela 4.1: Uso de instrumentos normativos na prática da indexação

Das respostas afirmativas à questão anterior, pode-se verificar (tab.4.2) que ao pedir a designação desse instrumento normativo, 50% não responde, e 38,6% deu respostas não especificadas, ou seja, não enquadráveis nos instrumentos considerados normativos como a NP 3715, NP 4036, ou SIPORbase. A soma destas duas categorias de resposta (88,6%) é muito significativa, pois revela que na grande maioria dos casos se faz indexação sem a observação de instrumentos normativos.

Valid	Frequency	Percent
não especifica	17	38,6
NP 3715	2	4,5
Siporbase	2	4,5
NP 4036	1	2,3
não responde	22	50,0
Total	44	100,0

Tabela 4.2: Designação do instrumento normativo utilizado

Na variável **I2**, sobre a utilização de léxico(s) de assuntos ou tesouro(s) para apoio à indexação, na maioria dos casos são usados léxicos de assuntos ou thesauri (tab.4.3), não sendo usados esses instrumentos em cerca de 18% dos casos.

Valid	Frequency	Percent
não	8	18,2
sim	36	81,8
Total	44	100,0

Tabela 4.3: Uso de léxico ou tesouros no auxílio à indexação

Como se trata de uma questão aberta, agruparam-se as respostas de acordo com a tipologia possível (tab.4.4). Nalguns casos foram referidos mais do que dois instrumentos, que, por não corresponderem a respostas uniformes, categorizámos em ‘vários’. Na categoria ‘não específica’ foram considerados os casos em que as respostas não referem qualquer instrumento, ou referem um que não se enquadra nas respostas válidas. Constatamos que esta categoria é a que revela a maior percentagem de respostas (25%). A questão não é respondida em 20,5% dos casos.

As bibliotecas que utilizam um ou dois tesouros/léxicos têm pouca representação, ao contrário das que utilizam mais do que dois (categoria ‘vários’), que representam 22,7% (tab.4.4).

Valid	Frequency	Percent
não especifica	11	25,0
não responde	9	20,5
Eurovoc e TESE	1	2,3
Agrovoc	5	11,4
Lince	2	4,5
Decs	2	4,5
Unesco educação	1	2,3
Unesco e Lince	2	4,5
Mesh e Lince	1	2,3
vários	10	22,7
Total	44	100,0

Tabela 4.4: Designação de léxicos ou tesouros usados

Na variável **I3**, sobre a utilização de ficheiro de autoridades de assunto ligado à base bibliográfica, podemos observar (tab.4.5) que a resposta foi equitativa para Sim e Não. Assim, em metade dos casos não há ligação a um ficheiro de autoridades, o que pode dificultar consideravelmente o controlo da terminologia.

valid	Frequency	Percent
não	22	50,0
sim	22	50,0
Total	44	100,0

Tabela 4.5: Utilização de ficheiro de autoridades de assunto

Sintetizando os resultados, é interessante observar graficamente a representação das percentagens na utilização de instrumentos normativos e na utilização de tesouros/léxicos no apoio à indexação (fig.4.1). Em ambos os casos é muito elevada a percentagem de “não responde” ou de respostas não especificadas, revelando-se sobretudo nos instrumentos normativos.

Em relação a estes, verificamos que são muito baixas as percentagens de respostas que referem as normas ou o sistema SIPORbase, que são na realidade os três instrumentos que dão orientações para a prática de uma indexação controlada. As altas percentagens de ‘não específica’ e de ‘não responde’, tratando-se de uma resposta aberta, pressupõem a dificuldade de identificação destes instrumentos por parte dos profissionais.

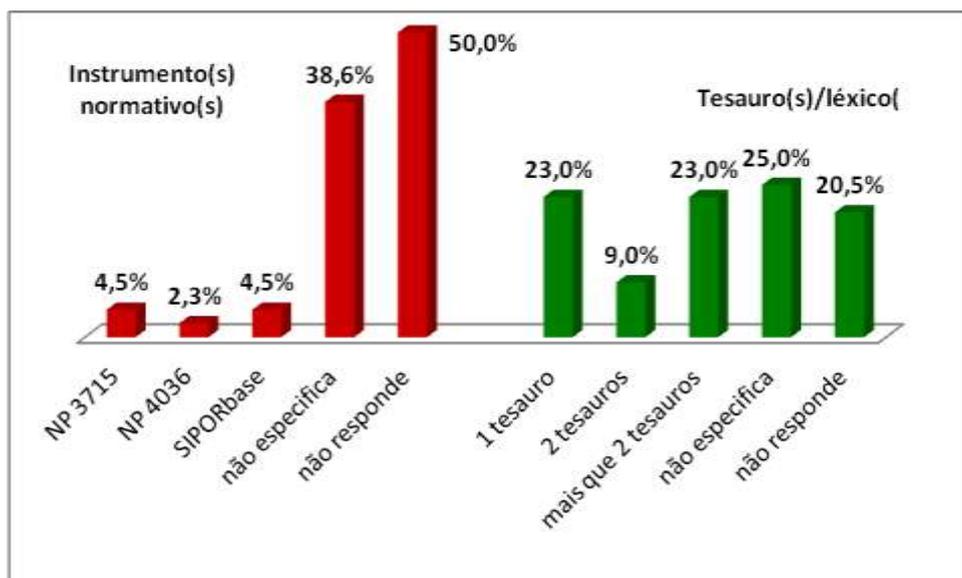


Fig. 4.1: Utilização de instrumentos normativos e de Tesouros ou léxicos

Quanto à utilização de tesouros ou léxicos (fig.4.1), consideramos que a percentagem de 23% referente ao uso de ‘mais que dois tesouros’, é significativa, no sentido em que tal situação potencia dificuldades, tanto ligadas às particularidades de cada tesouro, mas sobretudo ao nível da uniformização na construção de registos de autoridade.

4.2. Nível 2 - Características da indexação nas BIPP

Antes da apresentação dos resultados do nível 2, conforme a ordem estabelecida nas variáveis, observamos a tipologia dos cabeçalhos existentes:

Cabeçalhos de **Nomes Comuns (NC)**

Nos cabeçalhos do tipo NC1 (não subdividido) e NC2 (subdividido) verificámos que em três bases bibliográficas (B1EDU, B2SAU, B3TGE) há mais cabeçalhos do tipo ‘Não subdividido’, e em duas outras bases (B4AGR e B5GER) há mais do tipo ‘Subdividido’, ou seja, constituídos por subdivisões do assunto principal (fig.4.2). Não era previsível este resultado, pois contraria o pressuposto de que as bases com temáticas mais específicas como no caso das B1EDU, B2SAU, e B3TGE apresentem mais cabeçalhos do tipo subdividido, enquanto a de temática geral (B5GER) tenha mais

cabeçalhos não subdivididos. Estes resultados podem ter relação com os da variável ‘Cabeçalho específico’, que iremos comprovar mais adiante.

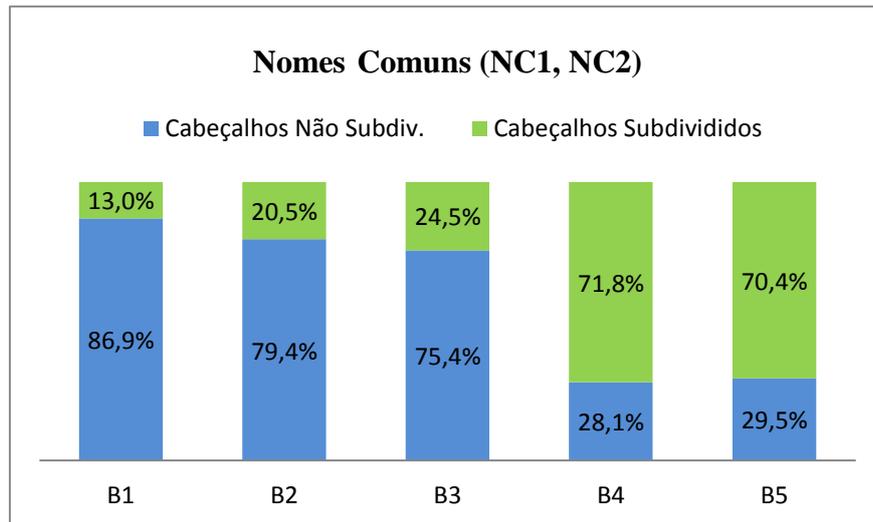


Fig. 4.2: Tipo de cabeçalhos, subdivididos e não subdivididos, por base bibliográfica

Erros UNIMARC

Nomes Comuns (U1,U2)

As categorias U1 (erros de campo) e U2 (erros de subcampos) desta variável foram observadas nos cabeçalhos ‘não subdivididos’ e ‘subdivididos’, podendo a U2 ocorrer só nos cabeçalhos do tipo ‘subdividido’. A U1 diz respeito ao ^a do campo 606, e a U2

aos subcampos $\wedge j$, $\wedge x$, $\wedge y$, $\wedge z$. Conforme definido nas categorias, consideram-se erros quando algum termo é inserido na etiqueta errada, como por exemplo, um nome próprio no campo 606, um qualificador de forma no $\wedge x$, uma faceta de assunto no $\wedge j$, etc.

Os erros U1 ocorreram em todas as bases exceto na *B3TGE* (tab.4.6), embora apresentando valores pouco significativos. Relativamente a erros de subcampo U2, que só podem ocorrer em cabeçalhos subdivididos (NC2), encontram-se alguns nas bases *B1EDU*, *B3TGE* e *B4AGR*, sendo a percentagem inferior a 1% na *B2SAU* e nula na *B5GER*.

Nesta categoria os indicadores de subcampo com erros que mais ocorrem são no $\wedge x$ que corresponde a subdivisão de assunto e no $\wedge j$ que diz respeito a subdivisão de forma. São exemplos destes erros, $\wedge a$ Estatística $\wedge j$ Congresso, ou $\wedge a$ Equipamentos de rega $\wedge x$ Catálogo, em que o termo ‘Congresso’ é atribuído como forma e o termo ‘Catálogo’ como subdivisão de assunto.

A não existência de erros em subcampos $\wedge y$, $\wedge z$, corresponde à quase inexistência desses subcampos nos registos, o que não é um bom indicador, pois em muitos casos seria eventualmente possível representar um assunto mais específico com a ajuda dessas subdivisões. Por outro lado, a inexistência de erros de subcampos na *B5GER* é um indicador de boas práticas, uma vez que é a base em que há mais cabeçalhos subdivididos.

Cabeçalhos de Nomes Comuns					
UNIMARC	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
U1 (campo 606)	0,1%	2,3%	-	1,8%	1,2%
U2^j (subcampo)	-	-	0,7%	5,0%	-
U2^x (subcampo)	9,4%	0,8%	1,7%	2,5%	-

Tabela 4.6: Resultados nas variáveis UNIMARC

Nomes Próprios (U1,U3)

Conforme referido na recolha de dados, são observados todos os cabeçalhos de Nomes Próprios existentes nas bases. Individualmente em cada base é muito heterogénea a sua ocorrência, o que deve ser tomado em conta ao observar os resultados na representação em percentagens (tab.4.7).

Na variável de erros UNIMARC, foram encontrados mais erros nas bases B2SAU e B4AGR, tratando-se de nomes comuns que foram incorretamente inseridos nos campos de nomes de pessoa e de nomes geográficos, e de campos sem preenchimento.

Na categoria U3 os erros encontrados dizem respeito ao ^f (subdivisão de datas no nome de pessoa), representando valores irrelevantes.

Cabeçalhos de Nomes Próprios					
UNIMARC	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
U1 (campos)	50%	51,7%	4,1%	74,1%	-
U3 (subcampos)	-	1,7%	2,7%	0,6, %	-

Tabela 4.7: Resultados nas categorias da variável UNIMARC

Erros de Sintaxe

Nomes Comuns (S1)

Os erros de sintaxe (ordem incorreta dos subcampos) dos nomes comuns só podem ocorrer nos cabeçalhos subdivididos. Um exemplo destes erros é o cabeçalho ‘Desenvolvimento regional--Alentejo’, que deveria ser representado ‘Alentejo (Portugal)--Desenvolvimento regional’. Para melhor observar os resultados desta variável, tem interesse ver o número de cabeçalhos subdivididos no total dos registos (26,3%), e respetivamente em cada base (tab.4.8).

	Nº de Cabeçalhos		Subdivididos	
	Total		Total	%
Total	2218		584	26,3%
B1EDU	771		101	13,0%
B2SAU	607		125	20,5%
B3TGE	514		126	24,5%
B4AGR	160		115	71,8%
B5GER	166		117	70,4%

Tabela 4.8: Distribuição dos cabeçalhos subdivididos

Os resultados desta variável apresentam uma percentagem de erros mais elevada na base *B1EDU* (9,4%) e na *B4AGR* (6,2%). Nas bases *B2SAU*, *B3Tecn* e *B5GER* é onde ocorrem menos erros (tab.4.9). A percentagem mais elevada na *B1EDU* deve-se a um procedimento de cabeçalhos criados de forma repetida, ou seja é usado o mesmo cabeçalho para uma grande quantidade de registos. Estas percentagens representam os erros de sintaxe em relação ao total dos registos da amostra; pensamos, contudo, que é relevante ver o que representam dentro dos cabeçalhos subdivididos, que é onde ocorrem. No caso da *B1EDU*, por exemplo, representa uma percentagem de 72%, o que significa que poucos cabeçalhos são corretos na sua sintaxe.

Cabeçalhos de Nomes Comuns					
SINTAXE (S1)	<i>B1EDU</i>	<i>B2SAU</i>	<i>B3TGE</i>	<i>B4AGR</i>	<i>B5GER</i>
% no total de registos	9,4%	0,6%	1,9%	6,2%	1,8%
% nos cabeçalhos subdivididos	72,2%	3,2%	7,9%	8,6%	2,5%

Tabela 4.9: Erros de sintaxe nos cabeçalhos de nomes comuns

Para saber se é significativa a relação entre o número de cabeçalhos subdivididos e o número de erros, ou seja, se ocorrem mais erros de sintaxe nos casos em que há mais cabeçalhos subdivididos, aplicou-se o teste do coeficiente de correlação de *Pearson*¹, que resultou no seguinte:

¹ Coeficiente positivo: implica variação das variáveis no mesmo sentido; Coeficiente negativo: implica a variação em sentidos opostos; Coeficiente de correlação superior a 0,7 indica relação (positiva ou negativa) forte.

Na situação A, em que se consideram todas as bases de dados, a relação entre o número de cabeçalhos e os erros de sintaxe é negativa, ou seja variam inversamente. Há mais erros quando o número de cabeçalhos subdivididos decresce e vice-versa. Contudo, este resultado pode ser influenciado pelos valores da base B1EDU que apresenta um número elevado de erros, face ao menor número de cabeçalhos subdivididos.

A

-0,87	Correlação entre o nº de cabeçalhos Subdivididos e os erros de Sintaxe
-------	--

Na situação B, em que se exclui a base B1EDU, a relação entre o número de cabeçalhos e os erros de sintaxe é positiva, mas fraca, o que indica que o número de erros não depende linearmente do número de cabeçalhos. É possível que exista uma relação não linear entre as variáveis, que esteja relacionada com outros fatores.

B

0,01	Correlação entre o nº de cabeçalhos Subdivididos e os erros de Sintaxe
------	--

Nomes Próprios (S2)

Não foram encontrados erros de sintaxe nos cabeçalhos de Nomes Próprios, sendo escassa a ocorrência de subdivisões de assunto.

Terminologia

Nomes Comuns (T1, T2)

Para a observação da terminologia dos cabeçalhos de Nomes Comuns, foram verificados os termos nos índices de assunto da BNP, que estão controlados segundo o sistema de indexação SIPORbase. Os termos que não coincidem com esses índices, são designados de ‘não conforme’; não se trata de erros, pois em muitos casos são termos recolhidos em tesouros ou listas de descritores mais adequados às temáticas específicas.

A primeira observação da variável foi feita nos cabeçalhos subdivididos, onde se verifica a terminologia no assunto principal (^a), e em todos os subcampos (tab.4.10).

Como podemos verificar, não é significativa a percentagem de termos ‘não conforme’ nos subcampos ^j, ^y, ou ^z (tab.4.10), os quais ocorrem com pouca frequência, como já se referiu na variável UNIMARC.

Cabeçalhos subdivididos	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
T1 ^a (assunto principal)	12,9%	6,5%	7,3%	53,1%	8,4%
T1 ^j (subdiv. forma)	0,5%	-	4,0%	4,3%	-
T1 ^x (subdiv. assunto)	3,7%	11,3%	8,3%	51,8%	7,8%
T1 ^y (subdiv. lugar)	-	-	2,1%	4,3%	3,6%
T1 ^z (subdiv. cronológica)	-	-	-	-	1,2%

Tabela 4.10: Terminologia ‘não conforme’ os índices BNP (cabeçalhos subdivididos)

A observação seguinte foi nos cabeçalhos não subdivididos (que só apresentam o ^a), onde se verifica a terminologia contabilizando os termos simples e os termos compostos (tab.4.11). Estas categorias T1 e T2 correspondem respetivamente à forma gramatical de ‘termos simples’ e de ‘termos compostos’ nos cabeçalhos Não Subdivididos. A sua diferenciação, que observamos adiante, é interessante no sentido de prever que erros de termos compostos (mais do que um conceito), poderiam dar origem à subdivisão de cabeçalhos. Um exemplo de termo na categoria T2^a é ‘Avaliação de currículos’ num cabeçalho Não Subdividido, que poderia dar origem ao cabeçalho Subdividido ‘Currículos -- Avaliação’. Nos cabeçalhos subdivididos, os termos, quer sejam simples ou compostos, são contabilizados na categoria T1.

Cabeçalhos Não subdivididos	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
T1 ^a (termos simples)	5,1%	14,1%	7,7%	8,7%	0,6%
T2 ^a (termos compostos)	25,0%	30,3%	32,1%	15,6%	4,2%

Tabela 4.11: Terminologia ‘não conforme’ os índices BNP

A soma de ocorrências no assunto principal (^a) dos cabeçalhos não subdivididos e subdivididos, juntamente com a ocorrência na subdivisão de assunto (^x) nestes últimos, dá-nos a ideia global de terminologia ‘não conforme’, como se ilustra na fig. 4.3.

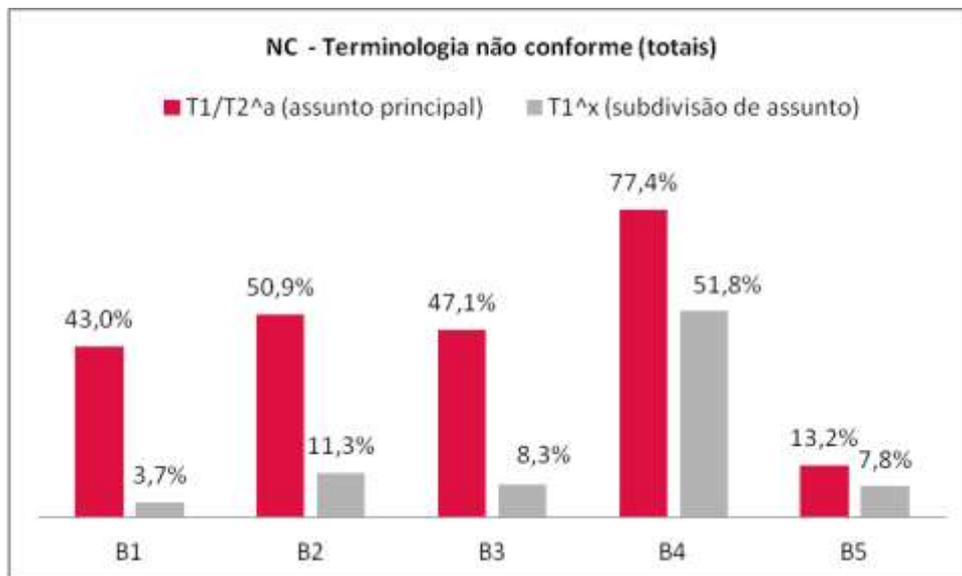


Fig. 4.3: Termos ‘não conforme’ os índices da BNP

Estas duas categorias relativas ao assunto principal e à subdivisão de assunto, ocorrem em todas as bases. É na base B4AGR que se encontram mais termos ‘não conforme’. Embora esta base cubra temáticas específicas, seria de esperar que outras como a B1EDU, B2SAU e B3TGE, com a mesma característica, apresentassem valores semelhantes, sendo contudo mais baixos.

Genericamente (fig.4.3), é nos termos de assunto principal que há percentagens mais elevadas de termos ‘não conforme’, o que, no caso das três primeiras bases, está de acordo com o facto de terem mais cabeçalhos não subdivididos (fig.4.2 - tipo de cabeçalhos, pág. 245). Já no caso da *B4AGR*, que tem mais cabeçalhos subdivididos, não se poderá fazer essa relação. A base *B5GER* é a que apresenta maior conformidade com os índices BNP, tanto no assunto principal como na subdivisão de assunto.

Conhecida a dificuldade na representação de conceitos, é importante para a nossa investigação identificar a utilização de termos simples e de termos compostos nos cabeçalhos não subdivididos, conforme referimos anteriormente.

Nestas categorias (T1, T2) da variável de terminologia, os resultados indicam o predomínio de termos compostos neste tipo de cabeçalhos. Conforme se verifica na fig.4.4, o único caso em que tal não acontece é na *B1EDU*.

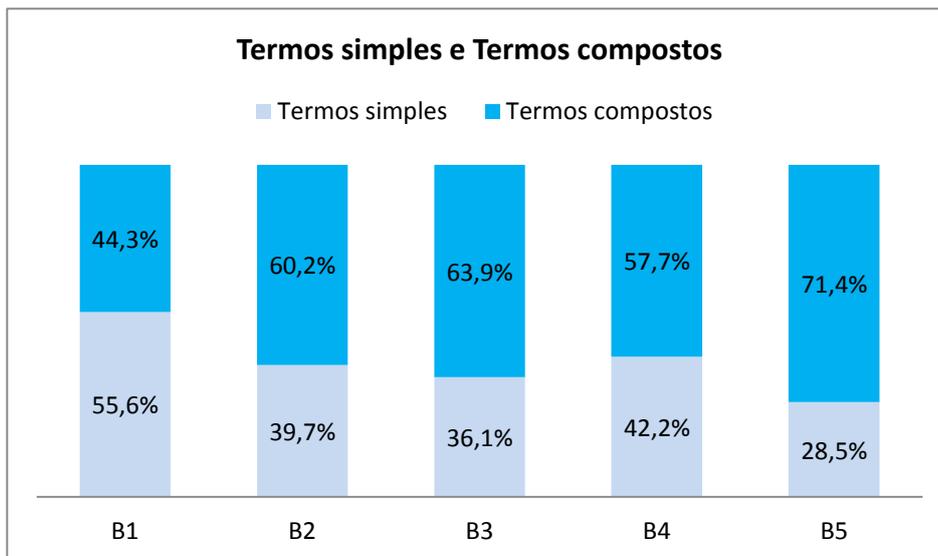


Fig.4.4: Termos simples e termos compostos nos cabeçalhos 'Não subdivididos'

Observámos anteriormente (tab.4.11) que é na base *B5GER* que há mais conformidade com os índices da BNP tanto nos termos simples como nos compostos, e que é comum às restantes bases a situação de 'não conforme' nos termos compostos. Curiosamente a base *B5GER* é a que apresenta maior percentagem de termos compostos (fig.4.4); pelo contrário, a *B1EDU*, que tem menos termos compostos, apresenta uma percentagem elevada na categoria de termos compostos 'não conforme' (25%), (tab.4.11).

Tivemos interesse em verificar se há uma relação significativa entre o número de cabeçalhos e o número de erros que ocorrem nos termos simples e compostos, pelo que se aplicou o teste do coeficiente de relação de *Pearson*, que resultou no seguinte:

0,70	Correlação entre o nº de cabeçalhos e erros nos Termos Simples
0,95	Correlação entre o nº de cabeçalhos e erros nos Termos Compostos

Este teste mostra que há uma correlação significativa entre o número de cabeçalhos e o número de erros que ocorrem nos termos simples e compostos. Embora não exista correlação perfeita entre as variáveis, há maior relação entre o nº de cabeçalhos e os erros nos termos compostos, do que entre os primeiros e os erros nos termos simples. A correlação linear perfeita implicaria um coeficiente de correlação de *Pearson* = 1 (correlação positiva) ou = -1 (correlação negativa).

Nomes Próprios (T3, T4, T5)

Na variável de terminologia (tab.4.12) a distribuição de termos ‘não conforme’ é muito heterogénea. Na categoria de nomes de pessoa (T3) só se encontram erros na *B1EDU*, *B3TGE* e *B5GER*, sendo nas duas últimas um resultado pouco significativo (0,2% e 0,7%), e na *B1EDU* só referente a 1 registo em 4 (25%).

Na categoria de nomes de entidade (T4) são encontrados erros na *B3TGE* (35,4%), na *B4AGR* e na *B5GER* com resultados quase nulos de 0,2% e 0,7%.

Na categoria de nomes geográficos (T5) só se encontram erros nas *B2SAU*, *B3TGE* e *B4AGR*, apresentando a última um valor significativo (20,7%).

Nomes Próprios		B1	B2	B3	B4	B5
<i>(Número de cabeçalhos)</i>		<i>(4)</i>	<i>(58)</i>	<i>(620)</i>	<i>(472)</i>	<i>(128)</i>
<i>UNIMARC</i>	<i>U1 (campo)</i>	50%	51,7%	4,1%	74,1%	-
	<i>U2 (subcampos)</i>	-	1,7%	2,7%	0,6%	-
<i>Terminologia</i>	<i>T3 (pessoa)</i>	25%	-	0,2%	-	0,7%
	<i>T4 (entidade)</i>	-	-	35,4%	0,2%	0,7%
	<i>T5 (geográfico)</i>	-	3,4%	0,3%	20,7%	-

Tabela 4.12: Distribuição de erros nas variáveis dos Nomes Próprios

Sintetizando os resultados para uma observação geral, verificamos que nos cabeçalhos de Nomes Comuns (tab.4.13) as percentagens mais elevadas se registaram na variável de Terminologia, que no total das categorias apresenta a mais alta na B4AGR (77,4%) e a mais baixa na B5GER (13,2%), com valores relativamente equiparados nas outras três bases. Na variável UNIMARC, o conjunto das categorias U1 e U2, verifica-se uma ocorrência de erros na base B4AGR (12,9%), seguida da B3TGE (8,7%), sendo os valores pouco significativos nas restantes bases.

<i>Nomes Comuns</i>	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
UNIMARC (U1,U2)	0,1%	2,1%	8,7%	12,9%	1,2%
Sintaxe	9,4%	0,6%	1,9%	6,2%	1,8%
Terminologia (T1,T2)	43,0%	50,9%	47,1%	77,4%	13,2%
<i>Nomes Próprios</i>					
UNIMARC (U1, U3)	50,0%	53,4%	6,8%	74,7%	-
Terminologia (T3,T4,T5)	25,0%	3,4%	35,9%	20,9%	1,4%

Tabela 4.13: Soma de resultados por variável nos cabeçalhos de NC e NP

Nos cabeçalhos de Nomes Próprios (tab.4.13) foi na variável UNIMARC que foram encontrados mais erros; no total das categorias U1 e U3, verifica-se que a base B4AGR apresenta maior percentagem de erros (74,7%), seguida da B2SAU (53,4%) e da B1EDU (50%), sendo os valores pouco significativos nas bases B3TGE e B5GER (6,8% e 0%). Refira-se que o valor apresentado na B1EDU diz respeito só a 4 cabeçalhos existentes na base.

Embora os erros encontrados ao nível do UNIMARC sejam significativos, uma vez que poderão por em causa a correta ligação aos pontos de acesso, é na terminologia que se verificam mais situações de ‘Não conforme’ os índices da BNP, que se vinculam à utilização do sistema de indexação SIPORbase. Este facto é bastante relevante, pois demonstram a utilização de outras fontes de terminologia, ou eventualmente uma falta de controlo de termos, conforme poderemos analisar adiante na variável ‘Fontes de terminologia’.

Na sequência das variáveis nos três eixos UNIMARC, Sintaxe e Terminologia dos cabeçalhos de Nomes Comuns e de Nomes Próprios, passamos a apresentar os resultados das variantes do ‘Cabeçalho específico’ e de ‘Fontes de terminologia’.

Cabeçalho Específico

Nesta variável observamos que é considerável o número de registos em que não é cumprido o princípio do cabeçalho específico, com menos relevância no caso da base *B5GER*. É nas bases *B1EDU* e *B4AGR* que se observam mais casos de cabeçalhos muito genéricos, o que é curioso, pois tratando-se de áreas temáticas como ciências da educação e ciências agrárias, deveriam necessitar de cabeçalhos de assunto mais específicos (fig.4.5).

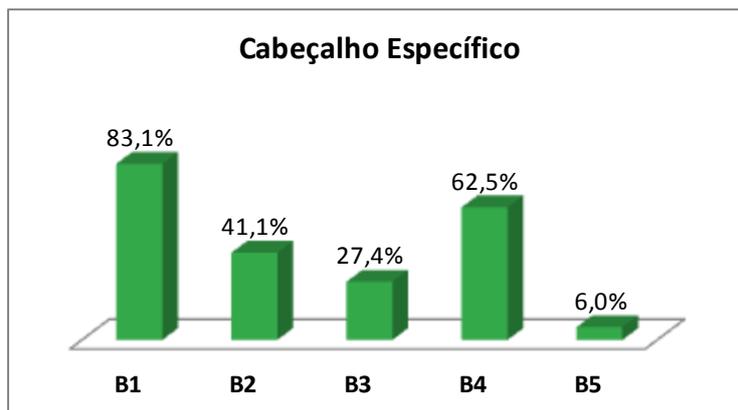


Fig. 4.5: Não conformidade com o princípio do cabeçalho específico

Tivemos interesse em verificar se existe uma relação entre o tipo de cabeçalho Não Subdividido (NC1) e Subdividido (NC2) e a variável Cabeçalho Específico, que verifica esse princípio. Aplicou-se o teste do coeficiente de relação de *Pearson*, que resultou no seguinte:

0,87	Correlação entre o número de cabeçalhos Não Subdivididos e os erros CE
------	--

Existe uma relação positiva entre os cabeçalhos Não Subdivididos e os erros de Cabeçalho Específico, ou seja, a ocorrência de erros nos CE tende a crescer com o aumento dos cabeçalhos não subdivididos.

-0,70	Correlação entre o número de cabeçalhos Subdivididos e os erros CE
-------	--

Existe uma relação negativa entre os cabeçalhos Subdivididos e os erros do Cabeçalho Específico, ou seja, a ocorrência de erros CE tende a decrescer com o aumento dos cabeçalhos subdivididos.

Fontes de terminologia

Verificámos na análise da terminologia que eram evidentes duas situações, por um lado, termos específicos de áreas temáticas retirados de tesouros, e termos mais genéricos, que se encontravam controlados nos índices de assuntos do catálogo em linha da BNP.

Consideramos que os termos de cada base devem ser analisados não só na perspetiva do catálogo coletivo, mas também à luz da especificidade temática a que correspondem. Neste sentido, a terminologia é verificada nos tesouros que servem de apoio, o que permite perceber a utilização dos mesmos na adoção de termos.

Como em todos os casos se recorre a tesouros e listas de descritores para recolha de termos, e para melhor percebermos as características da indexação praticada, verificamos a cobertura desses instrumentos em cada base bibliográfica. A partir das listas de ocorrência de termos, foram identificados os descritores nos tesouros e os termos nos índices da BNP.

Assim, de acordo com a prática em cada caso, foram obtidos os seguintes resultados:

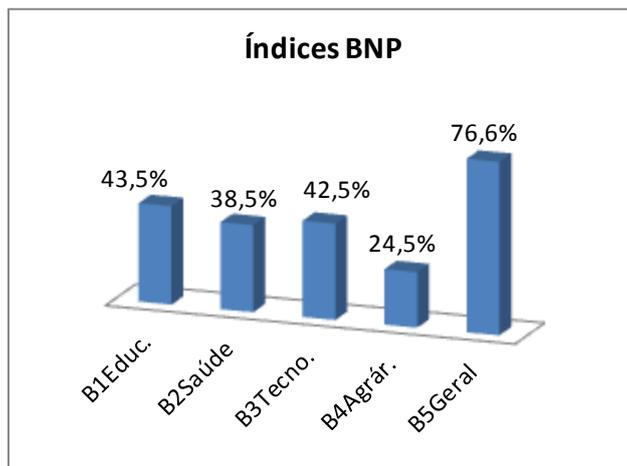


Fig. 4.6: Termos dos índices da BNP

Como podemos observar (fig.4.6) a maior cobertura dos índices da BNP verifica-se na *B5GER* que se trata de um fundo documental de carácter mais geral em que as temáticas dominantes se encontram nestes índices de assuntos. É nas bases *B2SAU* e *B4AGR* que se verifica menor cobertura pois são bases com temáticas específicas nas áreas das ciências da saúde e das ciências agrárias, em que há um maior recurso aos tesouros.

Em todos os outros casos se verifica que é maior a percentagem de termos retirados dos tesouros e lista de descritores, concretamente do Eurovoc nos casos da *B1EDU* e *B3TGE*, do Agrovoc na *B4AGR* e da lista de descritores DeCs na *B2SAU*, (fig.4.7).

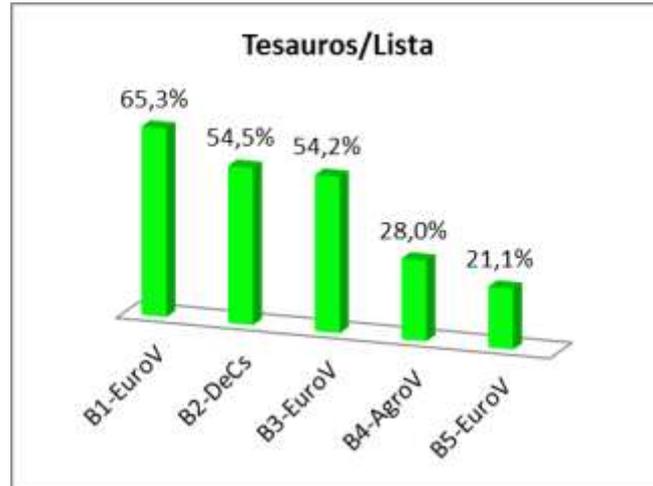


Fig.4.7: Termos dos tesouros / lista de descritores

Sintetizando os resultados das últimas variáveis (tab.4.14), é curioso na B1EDU que apresenta uma elevada percentagem de erros no cabeçalho específico, e portanto tendo mais cabeçalhos genéricos, sejam utilizados termos do tesouro. Em princípio os tesouros deveriam ser mais específicos. No caso do Eurovoc pode-se explicar pois não é um tesouro específico de educação. Já no caso da B4AGR, apesar de ser usado o tesouro Agrovoc como fonte de terminologia, há muitos termos que não coincidem com os descritores do tesouro. As percentagens assinaladas a cinzento não perfazem 100%, porque, neste caso, foram identificados termos que se repetem nas duas fontes de terminologia.

Cabeçalho específico		<i>B1EDU</i>	<i>B2SAU</i>	<i>B3TGE</i>	<i>B4AGR</i>	<i>B5GER</i>
		83,1%	41,1%	27,4%	62,5%	6,0%
Fontes de Terminologia	Índices BNP	43,5%	38,5%	42,5%	24,5%	76,6%
	Tesouro/Lista	65,3%	54,5%	54,2%	28,0%	21,1%

Tabela 4.14: Variáveis ‘Cabeçalho específico’ e ‘Fontes de Terminologia’

5. Discussão

Neste capítulo comentamos aspetos dos resultados do nosso estudo destacando alguns pontos de convergência dos dois níveis de investigação. A informação de carácter geral obtida no primeiro nível, como a fraca utilização normativa, a recolha de termos de várias fontes de informação e o controlo de autoridades, é comprovada no segundo nível na observação de características dos índices de assuntos. Estas constatações reforçam as linhas centrais da nossa investigação que são a necessidade de orientar o processo de indexação pela sua inerente subjetividade, e o benefício da utilização de um vocabulário controlado para consistência dos índices.

5.1 Instrumentos normativos

A utilização de instrumentos normativos na indexação foi a questão colocada para saber se nas bibliotecas politécnicas em geral são seguidas as normas de apoio ao desenvolvimento do processo. Os resultados obtidos apoiam a hipótese de que os procedimentos não são adequados à boa prática desta tarefa, comprometendo-a logo de início na fase da análise dos documentos.

- ✓ Identificámos que é reduzida a utilização de normas, e que podem não ser do conhecimento comum, pois existe alguma confusão em relação à sua designação, como é o exemplo de respostas ‘CDU, Tesouros e listas de termos, Regras Portuguesas de Catalogação’, etc., e só cerca de 11% das respostas se enquadram nas categorias definidas na variável.

- ✓ A observação das características dos índices de assunto revela que há sinais de uma indexação inconsistente. Esta constatação relaciona-se com o resultado do nível 1, na medida em que a não utilização das normas conduz a uma indexação inadequada. A fraca utilização normativa ocorre sobretudo na variável do cabeçalho específico, em que se encontram muitos registos cujos termos de indexação atribuídos não espelham uma análise dos documentos. Embora a observação desta variável possa ser afetada por se basear no título¹ ou outra informação do registo bibliográfico, como o campo de notas, identifica-se que era possível especificar facetas, ou adicionar qualificadores cronológicos, geográficos, formais, ou seja, na maioria dos casos verifica-se a presença de conceitos mais específicos.

Esta questão relacionada com os cabeçalhos específicos é de extrema importância, uma vez que constitui um dos princípios básicos de indexação, e a atribuição de termos muito genéricos pode estar ligada a vários fatores.

¹ O ponto 5.2 da NP 3715, refere: “Não é recomendado uma indexação feita com base unicamente no título [...] alguns títulos podem induzir em erro”; como os documentos em análise podem conter conceitos relevantes que não estão expressos no título, só são considerados para determinar a ocorrência de erro os casos em que é possível, através de expressão explícita, determinar a presença de um conceito mais específico

Pode tratar-se de um procedimento na atribuição de assuntos através da ficha catalográfica da publicação, como é referido num estudo de Boccato e Fujita (2010)² onde se verificou a falta de sistematização na análise documental, ao serem atribuídos assuntos dessa forma, em detrimento de uma análise focada no conteúdo e no interesse dos usuários. Outro fator que pode estar na base dessa decisão é ser o modelo de indexação usualmente seguido em bases de dados multidisciplinares no qual predomina a atribuição de termos genéricos, o que por vezes resulta na incapacidade de satisfazer necessidades de informação mais específicas (Rodríguez 1998).

No caso que se estuda podem estar presentes estes dois fatores, pois não havendo uma política de indexação, recorre-se com frequência à prática referida. Em relação à característica de base multidisciplinar, trata-se de facto de um modelo composto por várias bases em que o usuário pode pesquisar individualmente em cada uma, de acordo com as temáticas, ou em todas através dum único ato de pesquisa. Nesta perspetiva podemos dizer que se trata de uma consulta multidisciplinar, com as vantagens e desvantagens dessa característica. É comum a coexistência de bases de diferentes tipos num contexto de ‘catálogo coletivo’, sendo contudo desejável a consistência da terminologia nos índices de pesquisa, em especial na de carácter geral comum a todas as bases.

Esta reflexão sobre um equilíbrio entre a multidisciplinaridade e diferentes níveis de especificidade, leva-nos para o ponto seguinte sobre a necessidade e possibilidade de

² O uso de linguagem documentária em catálogos colectivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal.

controlo de autoridades num catálogo coletivo que envolve diferentes temáticas, mais gerais ou mais específicas.

5.2 Ferramentas de apoio: diversidade e uniformidade

- ✓ Na questão sobre as ferramentas de apoio, os resultados indicam que nas bibliotecas politécnicas se utilizam várias fontes de informação; em 23 % dos casos são usados vários tesouros ou léxicos para recolha de termos.

Ainda neste nível, na questão sobre o ficheiro de autoridades, os resultados indicam que em metade das bibliotecas politécnicas não é feito controlo de autoridades. Além disso, deve-se ter em conta que nos casos em que há ligação ao ficheiro de autoridades, não fica excluída a dificuldade do controlo da terminologia, ou seja, os registos de autoridade podem estar criados mas não estarem controlados; isto passa-se no estudo de caso, conforme referimos no capítulo 2.3.2, em que existe um ficheiro de autoridades que foi criado automaticamente e não foi feita a correção retrospectiva dos registos, que implica a confirmação do termo autorizado e o estabelecimento de relações semânticas. Para comprovar o efetivo controlo de autoridades nas bibliotecas politécnicas, teria que se verificar cada caso.

Embora seja já conhecida a utilidade do ficheiro de autoridade por parte dos profissionais, na maioria dos casos não há disponibilidade para o desenvolver e manter de acordo com as normas recomendadas. Embora a implementação e manutenção de listas de autoridade implique grande esforço, consideramos que o controlo centralizado de autoridades constitui uma grande vantagem no caso de um catálogo coletivo. Devido à complexidade na efetiva aplicação das normas, é de extrema importância a cooperação entre sistemas ao nível da implementação de léxicos controlados, de acordo com as necessidades específicas, aproveitando assim muito do trabalho já desenvolvido. O controlo de autoridades tem sido durante muito tempo uma parte muito importante do processo de catalogação, e o desenvolvimento tecnológico decorrido nas últimas décadas, tornou-o mais fácil de implementar em muitas bibliotecas. O desenvolvimento do controlo de autoridades e a evolução dos OPAC permitem a partilha de recursos e um ambiente favorável ao contexto profissional, no qual muitos bibliotecários aprendem de facto sobre o controlo de autoridades, e esta perspetiva deve ser potenciada (Martínez *et al.* 1997; Wolverton 2006).

- ✓ Ao nível das características dos índices que observámos no nível 2, os resultados mais relacionados com esta questão das ferramentas de apoio foram os da variável de terminologia, onde se identificaram práticas muito diferentes na recolha de terminologia e na formulação de cabeçalhos. Duas situações evidentes são: a existência de termos em linguagem livre, e de termos controlados recolhidos de tesouros e outras listas de descritores.

Como já referimos, as Bibliotecas do estudo de caso constituem um catálogo coletivo virtual, sendo na realidade uma rede bibliotecária que quer partilhar os seus fundos bibliográficos, e centralizar essa informação para uma comunidade académica. Contudo tem alguns dos problemas reconhecidos nos catálogos coletivos, como a duplicação de catalogações, ou a falta de consistência derivada do uso de diversas linguagens de indexação.

Muitos dos casos tratados como catálogos coletivos são na realidade redes bibliotecárias, e a gestão interbibliotecária é da maior importância, uma vez que poderá trazer benefícios para os profissionais e para os usuários. O trabalho desenvolvido isoladamente em bibliotecas resulta numa diversidade de procedimentos na atribuição de cabeçalhos de assunto. Na hora de partilhar recursos e criar catálogos coletivos revelam-se dificuldades decorrentes de situações muito heterogêneas. Nestes contextos as redes de cooperação bibliotecária são uma colaboração muito importante para a eficácia da recuperação da informação, que passa pela criação de pontos de acesso homogêneos, baseados numa normativa e diretrizes comuns, que regulem o controlo de autoridades. (Caro-Castro e Travieso Rodríguez 2003; Castillo 2003).

Uma das principais barreiras existentes na recuperação da informação é a diversidade de vocabulários empregues pelas diversas fontes e usuários. A necessidade de facilitar o acesso à informação incluída em sistemas com linguagens de pesquisa diferentes, levou ao desenvolvimento de vários projetos de compatibilização de linguagens. O facto de conviverem na mesma rede ou catálogo, bibliotecas com diferentes necessidades no que diz respeito à especificidade dos termos de indexação é um dos “eternos debates” na

área da recuperação de informação. Sistemas como o *Broad System of Ordering*³ (BSO) foram criados com a intenção de dar resposta a esse problema. O BSO é um sistema de classificação para pesquisa de assuntos na internet ou em qualquer grande sistema de informação, cobrindo todo o conhecimento.

O formato SKOS (*Simple Knowledge Organization Systems*) ao qual nos referimos no capítulo 1, é também um exemplo de resposta a estas questões; formaliza os vocabulários controlados possibilitando a sua interoperabilidade, e prevê a ligação de conceitos de várias estruturas de conhecimento, em que as equiparações podem ser exatas, mais genéricas ou mais específicas, ajudando a resolver o problema de diferentes níveis de especificidade (Sánchez-Cuadrado, Colmenero-Ruiz, Moreira 2012).

Um dos projetos mais avançados neste sentido é o Sistema Unificado de Linguagens em Medicina (*Unified Medical Language System – ULMS*), que é um repositório de vocabulários biomédicos desenvolvido na *US National Library of Medicine*. Este projeto representa o esforço de desenvolvimento na recuperação de informação proveniente de múltiplas fontes em sistemas automatizados. O UMLS constitui um conjunto de ferramentas que permitem a interoperabilidade entre sistemas e junta vários vocabulários e normas na área da saúde e biomedicina (Bermello Crespo 2001; Bodenreider 2004).

Ao nível das bases bibliográficas, em sistemas informáticos em que as linguagens se compatibilizam através do controlo de autoridades, essa compatibilização poderá ser

³ <http://www.ucl.ac.uk/fatks/bs0/about.htm> [Consulta: 13-06-2013]

implementada através do estabelecimento de uma política de indexação bem definida, cujos resultados devem ser avaliados periodicamente para monitorização da eficácia junto das pesquisas dos usuários.

É sabido que uma absoluta precisão na indexação é pouco provável até com vocabulários controlados, e que os índices de assunto nos catálogos não são tão consistentes como seria desejável. Também é difícil de alcançar um grande nível de uniformidade dos vocabulários, mas contudo é indispensável o controlo da terminologia de acordo com critérios orientadores, para uma maior consistência dos índices (Sievert e Andrews 1991; Montalvo 2011).

6. Conclusões

Neste capítulo apresentam-se as conclusões da investigação de acordo com os resultados obtidos. Conforme definimos nos objetivos a investigação foi feita em dois níveis que se complementam. As questões de carácter geral colocadas no primeiro nível, dão respostas que caracterizam uma tipologia de bibliotecas quanto à utilização de normas e de ferramentas no apoio à indexação. A análise dos índices de um pequeno grupo de bibliotecas no segundo nível, permitiu relacionar a qualidade e características destes com a utilização das normas e ferramentas, conforme já se comentou na discussão. Em última análise os resultados recolhidos permitem deduzir que o tipo de problemas detetados nos índices que observámos, poderão ocorrer igualmente nas bibliotecas que indagámos quanto ao processo de indexação.

6.1 Conclusões gerais

Neste estudo foram concentradas as atenções no processo de indexação, de modo a refletir a utilidade desta tarefa na qualidade dos índices de assunto, para maior eficácia na recuperação da informação.

Na parte teórica da nossa investigação vimos como é fundamental o conhecimento das normas orientadoras do processo de indexação. O seu desconhecimento potencia as

marcas de subjetividade do indexador e o conseqüente prejuízo na pesquisa da informação. Para ir ao encontro das necessidades de pesquisa dos usuários, é basicamente fundamental a análise adequada do conteúdo dos documentos conjugada com a estrutura de relações semânticas nos vocabulários controlados (Bertrand, Cellier, Giroux 1996; Hjørland 1997). Tentámos demonstrar a importância da observação das fases prescritas pelas normas e do controlo de autoridades, como forma de atingir melhores resultados ao nível da consistência dos índices.

Os resultados da nossa investigação evidenciaram a dificuldade do processo de indexação ao nível dos seus princípios essenciais, como a identificação de conceitos e a representação dos mesmos de forma controlada. Embora os dados obtidos não permitam identificar o tipo de dificuldades sentidas nas fases do processo, podemos afirmar que certo tipo de erros não ocorria se as normas tivessem sido observadas. Por exemplo, a atribuição dos termos genéricos, que já comentámos na discussão, pode significar, por um lado, a análise¹ pouco exaustiva do documento, e por outro, a dificuldade na identificação e seleção dos conceitos. Neste ponto em concreto são encontradas debilidades, pois é evidente que não são tidos em conta os princípios básicos de indexação prescritos nos instrumentos normativos, como o ‘princípio do cabeçalho específico’, e o ‘princípio do cabeçalho uniforme’.

Sintetizamos as principais conclusões (tab.6.1) que podem relacionar de imediato os dois níveis de investigação, em que a informação recolhida nas bibliotecas politécnicas pode ser comprovada na prática de um pequeno grupo de bibliotecas, através de

¹ O ponto 5.2 da NP 3715 refere: “o indexador deve certificar-se de que nenhuma informação útil foi descurada.

características dos índices. A fraca utilização de instrumentos normativos manifesta-se na falta de princípios de indexação, e a variedade de fontes de terminologia sem o controlo de autoridades manifesta-se na utilização inconsistente de palavras-chave, descritores, e cabeçalhos de assunto.

Indícios nas BESP (nível 1)		Manifestações nas BIPP (nível 2)
Utilização de Instrumentos normativos	43% não usam 11% normativos 4,5% SIPORbase	- Não observação de princípios de indexação, (cabeçalho específico e uniforme) - Terminologia não conforme o SIPORbase
Várias fontes de terminologia	32% até 2 tesauros 23% mais que 2	- Utilização indiscriminada de descritores, palavras-chave, cabeçalhos pré-coordenados
Ficheiro de autoridades	50% não utilizam	- Terminologia não controlada

Tabela 6.1 Ligação dos níveis de investigação:os indícios e as manifestações

Como se verifica na tab. 6.1, nas bibliotecas politécnicas há um número considerável de casos em que não se usam instrumentos normativos (43%), e nos casos em que são usados só 11% das respostas se podem considerar como tal (NP3715, NP 4036, SIPORbase). Em relação à utilização do SIPORbase, só em 4,5% dos casos foi referida.

No que diz respeito às fontes de terminologia, sendo muitos os casos em que são usados dois ou mais tesauros/léxicos, e em 50% dos casos não é utilizado um ficheiro

de autoridade, a manifestação no nível dois é a utilização indiscriminada de termos nos cabeçalhos e a sua falta de controlo.

6.2 Conclusões específicas (BESP)

- ✓ Ficámos a conhecer a situação das bibliotecas politécnicas em Portugal (BESP) no que respeita à utilização de instrumentos normativos e ferramentas de apoio à indexação.

Estes resultados permitem-nos concluir que numa grande maioria destas bibliotecas se faz indexação sem orientação normativa, o que consideramos um pouco surpreendente, uma vez que o conhecimento normativo deve ser a base de todo o processo de indexação

Em relação à utilização de ferramentas de apoio, a maioria destas bibliotecas (82%) usa tesouros ou léxicos de assuntos, mas só em metade dos casos as respostas se consideram de acordo com as categorias definidas, ou seja, não referiram tesouros ou léxicos em concreto², verificando-se alguma confusão quanto à designação de ‘ferramentas’, como são exemplos as respostas ‘Thesaurus’, ‘Vários disponíveis na internet’, ‘Listas de assuntos’, etc. Curiosamente o mesmo tipo de confusão se verificou na designação de instrumentos normativos.

² O tipo de respostas pode ser visto no capítulo 3.2.1.3.

Em última análise concluímos que há um desconhecimento do que são instrumentos normativos e ferramentas de apoio à indexação por parte dos profissionais, o que pode ser indício de que a indexação é praticada de forma inadequada por profissionais sem formação específica nessa área.

Quanto à utilização do ficheiro de autoridades, é equitativa a percentagem de respostas dos que utilizam e dos que não utilizam. Conforme comentámos na discussão a este respeito, o efetivo controlo de autoridades deve representar uma baixa percentagem, pelo que concluímos que sendo utilizadas várias fontes de terminologia, torna-se difícil o seu controlo, ponto em causa a consistência dos índices.

6.3 Conclusões específicas (BIPP)

Ficámos a conhecer algumas características da indexação do grupo de bibliotecas BIPP através da análise dos índices de assunto dos seus catálogos bibliográficos. Através das variáveis que estudámos identificou-se de forma empírica que os diferentes resultados entre as bases estão relacionados com fatores como a recolha de termos de diversas fontes temáticas, mas também com os procedimentos de indexação ao nível da sintaxe dos cabeçalhos, ou da aplicação de princípios básicos de indexação.

✓ Da observação dos três eixos centrais (formato UNIMARC, sintaxe de cabeçalhos e controlo da terminologia), retirámos as seguintes conclusões:

○ UNIMARC

Não é significativa a percentagem de erros que ocorrem no UNIMARC ao nível dos ‘nomes comuns’, no entanto seria desejável uma maior utilização das potencialidades do formato relativamente à representação de facetas de assunto. Embora os resultados desta variável apresentem uma baixa percentagem de erros (média não superior a 3%), devem ser considerados relevantes, pois a vinculação entre o catálogo bibliográfico e o catálogo de autoridades é feita através da ligação das etiquetas UNIMARC, que devem estar corretamente atribuídas. Foi nos subcampos de subdivisão de assunto que ocorreram mais erros, tratando-se, na maioria, de termos que indicam subdivisão de forma, o que revela desconhecimento de princípios básicos de indexação. Ao nível do preenchimento de ‘nomes próprios’, embora seja reduzida a sua utilização, encontraram-se bastantes erros em proporção aos dos ‘nomes comuns’, (uma média de 36% em erros de campo nos nomes próprios, e de 1% nos nomes comuns), o que confirma o desconhecimento ou dificuldade no uso do formato.

É possível relacionar que a insignificante ou mesmo inexistente presença de erros em subcampos, se verifica nas bases que têm mais cabeçalhos simples, e portanto sem utilização de subcampos. Os erros que encontramos revelam sobretudo algum desconhecimento da utilidade do formato, e a conseqüente dificuldade dos profissionais em explorar possibilidades de uma indexação mais adequada. Estes resultados

coincidem com as conclusões de um estudo de análise da frequência de utilização dos elementos do UNIMARC que foi levado a cabo sobre os registos da base de dados bibliográficos PORBASE, no qual se constatou que 35% dos campos do bloco 6-- não são utilizados, e que há uma baixa frequência de utilização do bloco 6xx. Uma das conclusões retiradas foi que “praticamente metade dos subcampos disponíveis nunca foi utilizada, e que “uma percentagem considerável de campos e subcampos disponíveis para descrição bibliográfica têm uma utilização praticamente nula”. Outra questão interessante é a observação de que o profissional de informação, ao fazer a descrição dos recursos, ainda encara o UNIMARC simplesmente como um “hospedeiro” dos elementos das ISBD e das regras de catalogação, não explorando as potencialidades que o formato oferece ao nível da recuperação da informação nos sistemas”³ (Lopes *et al.* 2007).

A qualidade dos registos UNIMARC pode ser vista pelo seu preenchimento sem erros, e também pelo nível da utilização de subcampos. A qualidade dos registos é muito importante para o usuário, embora o critério de qualidade possa variar com o tipo de usuário e o uso específico do registo. Seja qual for o critério, para ser considerado de qualidade aceitável, um registo deve conter a informação necessária, pois informação insuficiente, dados incorretos e códigos inadequados irão colocar problemas. É fundamental que um registo seja preciso, consistente com outros registos, e funcional no sistema no qual é usado (Chapman 1994).

³ Relativamente ao preenchimento do ‘Unimarc essencial’ e o ‘nível mínimo de preenchimento’ de registos definido em 2004 para integração de registos das bibliotecas cooperantes na Porbase.

○ SINTAXE

No que respeita à sintaxe dos cabeçalhos não é significativa a percentagem de erros encontrados, e o domínio de cabeçalhos não subdivididos influi nesse resultado. No entanto é de salientar que o contraste das variáveis de erros de sintaxe com os cabeçalhos não subdivididos e subdivididos indicou uma fraca relação, ou seja, o número de erros não depende linearmente do número de cabeçalhos.

Em relação a esta matéria, as principais tendências de desenvolvimento das linguagens de indexação no contexto internacional são a convergência de sistemas pré-coordenados e pós-coordenados (Bermello Crespo 2001). Uma das tendências é “decompor” os cabeçalhos para evitar os problemas que decorrem da sintaxe, como já referimos no capítulo 1.2.2 acerca do sistema FAST (*Faceted Application of Subject Terminology*) e do *Nuovo Soggettario*, ferramentas que se destacam no apoio às novas práticas de indexação por assuntos, suportada na interoperabilidade entre diferentes sistemas.

○ TERMINOLOGIA

Quanto à terminologia verifica-se que estamos perante procedimentos de indexação diferentes, com linguagem natural, com termos controlados de tesouros/listas de descritores, e com a construção de cabeçalhos pré-coordenados. Predominam os cabeçalhos não subdivididos, o que resulta na atribuição de cabeçalhos muito genéricos que recuperam demasiados registos e colocam problemas de especificidade, conforme já foi comentado na discussão. É evidente que as práticas de tratamento documental nestas bibliotecas têm sido seguidas de forma heterogénea. Deve ser tido em conta o facto de

estarmos a observar estes termos à luz de uma nova perspectiva, no âmbito de um ‘catálogo coletivo’. Refira-se que as bases bibliográficas evoluíram isoladamente, e que os termos de indexação foram atribuídos em função de um público específico numa determinada área, como por exemplo nas áreas de saúde, ou de zootecnia; perante essa perspectiva poderia parecer ao indexador que nalguns casos não era fundamental delimitar a área temática, pois quem os pesquisava, procurava-os nesse contexto temático.

A questão da diferença encontrada nos modelos de indexação é muito significativa no nosso estudo, pois consistem em opções muito diferentes. Embora o uso dos tesouros permita um bom nível de controlo do vocabulário, também nestes produtos se produzem numerosas inconsistências na forma dos descritores, o que resulta na dispersão de entradas nos índices. Essa é mesmo a maior debilidade, é que não há um controlo total do vocabulário, o que dá lugar ao aparecimento de sinónimos ou problemas de polissemia. É possível conseguir um grau de convergência nos modelos de indexação na medida em que forem seguidas normas básicas sobre a atribuição de termos ou na limitação de cabeçalhos pré-coordenados (Rodríguez e Alonso 2011).

O uso de diferentes cabeçalhos de assuntos dificulta a recuperação de informação, sobretudo nos catálogos coletivos, e converte a pesquisa por assuntos numa das opções menos utilizada. De uma forma geral, embora tenha havido grandes mudanças tecnológicas nas últimas décadas, não se deu um desenvolvimento equivalente nas políticas e na qualidade da indexação, e são encontradas muitas deficiências a este nível. É comum a situação da utilização de várias linguagens de indexação nas bibliotecas, ou

de se defenderem as listas de cabeçalhos de assunto, ou a utilização de tesouros e outros vocabulários livres, levando-as a desenvolver caminhos diferentes na implementação dos seus ficheiros de autoridades de assuntos.

No caso das bibliotecas universitárias espanholas é referido que essa problemática se deve à carência de uma política de indexação homogénea (Caro-Castro e Travieso Rodríguez 2003; Villén-Rueda 2006). Já no caso português dispomos do sistema de indexação SIPORbase, que define uma política de indexação e um conjunto de regras que podem ser adotados por qualquer biblioteca. A sua utilização é, no entanto, dificultada em grande parte pela sua complexidade e manuseamento, cuja versão impressa não possibilita uma interatividade adequada.

Por último, coincidindo com Strehl (1998) constatou-se que neste grupo de bibliotecas não há manuais ou diretrizes que estabeleçam critérios comuns a adotar por todas, apesar da inconsistência dos índices; a falta de um padrão a ser seguido prejudica não apenas o bibliotecário no momento da definição dos termos mas principalmente, coloca em risco a qualidade da recuperação de informação. Atualmente, no contexto ‘coletivo’ há uma forte necessidade de rever procedimentos no sentido de uniformizar metodologias e cooperar na melhoria dos catálogos.

O uso de uma linguagem documental deve facilitar o processo de pesquisa de informação, e é por isso de extrema importância o seu papel no processo de atribuição de cabeçalhos para produção de índices de qualidade, que servem de pontos de acesso para os usuários. A linguagem documental poderá ajudar no processo cognitivo do

indexador que atribui os termos, e do usuário que os escolhe para definir a sua expressão de pesquisa (Boccatto 2009; Fujita 2003; Hjørland 1997).

6.3.1 Conclusões de utilização do SIPORbase

- ✓ Verificámos o controlo da terminologia com base nas orientações do sistema de indexação SIPORbase, através da conformidade com os índices de assunto da BNP (Biblioteca Nacional de Portugal).

Verifica-se que o sistema SIPORbase tem pouca utilização por parte das bibliotecas politécnicas, conforme os dados recolhidos no nível 1 (indicado só em 4,5% dos casos), e só como instrumento normativo; os índices da BNP, que são controlados de acordo com esse sistema, não foram mencionados como instrumento de apoio em nenhum caso. Embora seja um sistema de referência nacional que contempla as diretrizes para uma prática correta, a sua difícil utilização revela-se nos resultados do segundo nível em que se observam as características da indexação. Ai, na comparação da terminologia com os índices de assunto da BNP (que estão de acordo com o sistema) é onde se encontra a situação mais heterogénea de todas as variáveis. Nos índices observados há elevadas percentagens de termos ‘não conforme’ o sistema, e revelam-se mais nas bases de áreas específicas como ciências da saúde e ciências agrárias. Apenas uma das bases apresenta maior conformidade, o que parece estar ligado ao facto de ser uma base com temáticas de carácter mais geral que permitem adotar termos dos índices da BNP. O mesmo não

se verificou quando analisámos a cobertura de tesouros temáticos nas bases que não usam o SIPORbase, por exemplo, uma base que tem uma percentagem de ‘não coincidência’ de 24,5% tem uma percentagem de 28% de cobertura do tesouro da sua temática. Trata-se, então, da utilização de termos de linguagem livre em vez de descritores do tesouro ou dos índices BNP.

6.4 Sugestões para investigações futuras

Esperamos que este estudo possa contribuir para uma melhoria no processo de indexação e na eficácia da recuperação de informação por parte dos usuários. Pensamos que os resultados obtidos são válidos para o caso concreto que analisámos, de um grupo de bibliotecas com áreas temáticas particulares; neste contexto levantam-se outras questões que deverão ser objeto de estudo mais aprofundado, como a verificação da utilidade da política de indexação que se irá propor, e a efetiva vantagem introduzida pela uniformização da linguagem controlada, ao nível dos resultados de pesquisa para os usuários do catálogo.

Ao longo do trabalho foram surgindo algumas questões que gostaríamos de deixar como sugestões para futuras investigações, e que poderão complementar os resultados da nossa investigação.

- Verificar como é utilizado o ficheiro de autoridades nas bibliotecas, de acordo com a tipologia existente de *software* bibliográfico;
- Comprovar se a linguagem documental é compatível com a linguagem de pesquisa do usuário do catálogo;
- Os profissionais têm atualmente que atender a uma multiplicidade de tarefas que condicionam o desenvolvimento de trabalho técnico; teria interesse saber se os profissionais que fazem indexação têm formação específica, e qual a disponibilidade para desenvolver essa tarefa;
- As dificuldades que se revelaram ao nível da qualidade dos índices, poderiam ser analisadas através de um estudo qualitativo que implicasse os indexadores, de forma perceber o tipo de dificuldades que se colocam no desenrolar do processo;
- Haveria interesse em alargar o estudo às bibliotecas universitárias, uma vez que devem ter tido um desenvolvimento diferente e mais consolidado do que as bibliotecas politécnicas.

7. Proposta de “Metodologia de Indexação”

Um dos objetivos que nos propusemos alcançar foi fazer uma proposta de melhoria metodológica no processo de indexação para apoio dos profissionais. Nas Bibliotecas do nosso estudo, a falta de uma política orientadora foi o fator que mais contribuiu para o desenvolvimento heterogêneo das bases bibliográficas no que respeita à terminologia de assunto. A observação dos índices nos catálogos demonstram a diferença de procedimentos adotados, e a necessidade de uma política que defina a utilização de linguagens controladas compatíveis com as necessidades dos usuários, tendo em conta a especificidade de certas áreas temáticas.

A necessidade de políticas de indexação é uma referência constante em torno das temáticas que estudámos. Elas definem os princípios básicos de acordo com as normas gerais, e a maioria das grandes instituições documentais conta com uma política explícita. É o caso do sistema de indexação em português SIPORbase, que foi concebido como um sistema adequado para um catálogo alfabético de assuntos na Biblioteca Nacional, e que constitui um instrumento de referência para todas as bibliotecas portuguesas.

Trata-se, no entanto, de um sistema complexo e de difícil utilização por parte dos profissionais que não tenham formação específica nessa área, ou disponibilidade de dedicação exclusiva à tarefa, conforme referimos na apresentação do nosso estudo. No sentido de dar a conhecer de forma mais generalizada este instrumento de apoio,

elaborámos umas tabelas adaptadas do SIPORbase, que se apresentam em anexo, e às quais se recorre durante o processo proposto na metodologia apresentada neste capítulo.

Na definição de uma política de indexação para um grupo concreto de bibliotecas, é necessário ter em conta as suas especificidades para que a implementação seja adequada. A definição de política requer a análise de vários fatores como o tipo de linguagem a utilizar, que deve refletir os interesses da unidade de informação, as características dos usuários, e os meios técnicos e humanos disponíveis (Ribeiro 1996). Nesta perspetiva pensamos que é necessário delinear uma política de indexação interna para as Bibliotecas do IPP, que para além dos princípios básicos, defina alguns pontos específicos de acordo com as necessidades particulares deste grupo.

A política de indexação serve de base ao processo, guiando sobre os procedimentos a ser seguidos no uso de vocabulários controlados e no controle de autoridades; Contudo, essa política deve ser vista como uma filosofia que reflita os interesses e os objetivos da unidade de informação, mais do que como uma lista de procedimentos a seguir (Boccatto, Fujita, Leiva 2011; Fujita; Leiva, 2009).

7.1 Modelo de indexação

Uma das principais diferenças que identificámos nas bibliotecas do nosso estudo foi o modelo de indexação usado, ou a falta de um em concreto. Verificámos o uso

indiscriminado de termos retirados da linguagem natural (palavras-chave sem controlo de vocabulário) e de linguagem controlada em tesouros (descritores), assim como de cabeçalhos pré-coordenados. Trata-se de opções aparentemente muito diferentes, que devem, no entanto, manter um grau de convergência no seguimento de normas básicas sobre a construção e normalização dos termos, controlo dos nomes próprios, ou a limitação da pré-coordenação aos conceitos que só podem ser expressos sem ambiguidade de forma composta, etc. (Rodríguez e Alonso 2011).

Na sequência da reflexão sobre a implementação de uma política, e simplificando uma questão que é complexa, propomos um modelo para o desenvolvimento do processo de indexação. As orientações que se propõem consideram-se adequadas para o grupo de Bibliotecas a que se destinam, pelo que, numa fase experimental, se recomenda que sejam observadas com o máximo rigor. No entanto, as instruções apresentadas deverão ser consolidadas posteriormente numa política de indexação interna, a ser construída de forma evolutiva e amadurecida.

Propõe-se um modelo que representa as três etapas básicas da indexação, ao qual atribuímos o nome IRNA, acrónimo de **I**dentificação, **R**ecolha, **N**ormalização e **a**utoridades:

- 1- Identificação e seleção de conceitos através da análise conceptual de documentos, de acordo com instrumentos normativos.
- 2 - Recolha de termos em fontes de terminologia.
- 3 - Normalização de termos e de cabeçalhos, e controlo de autoridades.

1 Identificação e seleção de conceitos

a) Análise do documento para definição do seu conteúdo (NP 3715)¹

As partes do texto que devem ser analisadas com especial atenção são o título, o sumário e o resumo, caso exista. Não é recomendado uma indexação feita com base unicamente no título, pois alguns podem induzir em erro, e em relação aos resumos, alguns poderão ser insuficientes para representar o conteúdo dos documentos, pelo que a informação necessária para o indexador deve ser observada noutras partes como a introdução, início de capítulos, conclusão, ilustrações, quadros etc.

b) Identificação e seleção dos conceitos representativos do conteúdo²

O critério principal na seleção dos conceitos deve ser sempre o seu valor potencial, como um elemento na expressão do conteúdo de um documento e na sua recuperação. Ao escolher os conceitos, o indexador deve ter presentes as perguntas que podem ser feitas ao sistema de informação, ou “quem é o meu usuário?”

c) Para melhor identificação dos assuntos e posterior localização em fontes de terminologia, devem-se relacionar genericamente os conceitos com:

¹ Primeira das três fases consideradas na norma portuguesa NP 3715, onde são recomendados os procedimentos para a análise e descritos os métodos gerais a aplicar, que se destinam aos sistemas de indexação em que se exprime o conteúdo de forma resumida, através de termos de uma linguagem de indexação controlada.

² O sistema SIPORbase estabelece que sejam identificados os conceitos que expressam cada assunto, e sejam selecionados os que forem considerados pertinentes. São considerados pertinentes “os que constituem limitações intencionais ou evidentes do autor na sua abordagem ao assunto, e desde que tenham, na obra em causa, expressão significativa relativamente ao assunto básico”, e são considerados não pertinentes “os que não constituem parâmetros de limitação evidentes, surjam como informações pontuais, com funções ilustrativas ou acessórias relativamente ao assunto básico”.

- a. os três grupos de entidades dos requisitos funcionais FRBR; uma obra tem uma expressão, e pode ter várias manifestações que dão origem a itens únicos (grupo 1); por outro lado, a responsabilidade pelo conteúdo da obra pode ser de pessoa, entidade ou família (grupo 2); por último, o que está relacionado com a obra, pode ser um conceito, um objeto, um evento ou lugar (grupo 3), ou seja, um assunto, concreto ou abstrato.
- b. as categorias da norma ISO 25964-1 (de Nomes Comuns e Nomes Próprios), em que se identificam ‘coisas, materiais, atividades, eventos, propriedades, disciplinas, pessoas, entidades, lugares, etc.

Da fusão desses instrumentos, e na linha de Van Slype (1991), resulta o modelo de categorias que a seguir se apresenta, e que deverá ser operacional a nível de uma divisão conceptual de assuntos.

	Categorias	Gerais (com Nome Comum)	Específicas (com Nome Próprio)
Concretos	Seres vivos	Animais, pessoas, plantas	Cervantes Saavedra, Miguel de
	Objetos e suas partes	Edifícios, automóveis	Torre Eiffel (Paris) Audi A3
	Obras e criações artísticas	Cinema, Livros de caballería	Casablanca (Película cinematográfica) Amadís de Gaula
	Materiais	Madeira, plástico, titânio	
	Lugares	Rios, cidades, países, bairros, comarcas	Ebro (Espanha, Rio), Sri Lanka
	Entidades	Associações, Congressos, Museus, Universidades Famílias	Museo do Prado Universidade de Salamanca Borgia (Familia)
Abstractos	Atividades ou processos	Costura, fertilização, automatização,	
	Eventos	Aniversários, guerras, revoluções	Guerra civil de Espanha
	Propriedades	Flexibilidade, resistência	
	Disciplinas	Arqueologia, química, teologia	
	Unidades de medida	Ampere, metro, litro	
	Delimitações cronológicas	Idade Média, S. XVI	
	Estilos artísticos	Pintura barroca	

Tabela 7.1: Categorias para identificação de conceitos

2

Recolha de termos em fontes de terminologia

Nesta segunda etapa do processo estão subjacentes os princípios gerais de indexação que devem ser observados na fase de representação dos conceitos (SIPORBase): para além do princípio terminológico, que prevê o controlo da terminologia conforme as instruções do sistema de indexação (presente nos índices de assunto do catálogo da BNP), devem ser observados:

- ✓ O princípio da entrada única - cada assunto deve ser representado por um só cabeçalho co extensivo com o conteúdo temático a que se aplica;
- ✓ O princípio da entrada direta - a entrada de assunto deve ser constituída por um cabeçalho em que o assunto é dado diretamente pelo termo que o representa, não recorrendo à interposição de termo(s) mais genérico(s);
- ✓ O princípio da entrada específica – a entrada de assunto deve ser constituída por um cabeçalho específico, isto é, cujo sentido não seja nem mais lato nem mais restrito que o conteúdo temático a que se aplica;
- ✓ O princípio da pré-coordenação – um assunto complexo, isto é, não suscetível de ser representado apenas por um termo de indexação, deve ser representado através da combinação dos termos correspondentes aos vários conceitos (ver formas de pré-coordenação nas tabelas SIPORbase).

- a) O termo ou expressão a atribuir (caso não exista já na base de autoridades local), deve ser localizado em primeiro lugar no catálogo da Biblioteca Nacional (BNP)¹.
- b) Termos de carácter geral, comuns a todas as áreas, devem ser recolhidos do catálogo da BNP. Se estes termos não existirem no catálogo da BNP, devem ser procurados nos tesouros, e se também aí não se encontrarem, devem ser controlados de acordo com as diretrizes do SIPORbase (ver tabelas em anexo I).
- c) No caso de se tratar de um termo específico de área temática, devem ser consultados os respetivos tesouros, devendo sempre ser selecionado o descritor autorizado.

Os recursos que constituem as fontes de terminologia, apresentam-se de acordo com a prioridade de utilização e áreas temáticas:



Os índices de assunto do catálogo da Biblioteca Nacional são a fonte preferencial, uma vez que cumprem os princípios de indexação do SIPORbase, tanto ao nível do controlo da terminologia, como na construção de cabeçalhos pré-coordenados.

¹ http://www.bnportugal.pt/index.php?option=com_content&view=article&id=226&Itemid=54&lang=pt

Embora seja um catálogo de carácter ‘generalista’, inclui terminologia específica que em muitos casos serve as necessidades encontradas.

No caso de não se encontrar a terminologia desejada no catálogo da BNP, é recomendada a consulta dos seguintes recursos de temáticas mais específicas.



O tesouro Eurovoc é recomendado para todas as áreas. Constitui um bom apoio nas temáticas ligadas a atividades do âmbito da Comunidade Europeia.



Para vocabulário específico nas áreas da educação, cultura, ciências naturais, ciências sociais e humanas, comunicação e informação, pode-se consultar o Tesouro da UNESCO. Embora não esteja disponível uma versão em português, é muito útil a identificação abrangente dessas temáticas. Recomenda-se a versão SKOS do ‘Tesouro de la UNESCO’.



Para vocabulário na área das ciências da saúde deve-se consultar a listas de descritores DeCS. O vocabulário DeCS é uma linguagem para a indexação, pesquisa e recuperação de assuntos desenvolvido a partir do MeSH – *Medical Subject Headings* da *U.S. National Library of Medicine*. Para além dos termos médicos originais do MeSH foram desenvolvidas as áreas de Saúde pública, Homeopatia, Ciência e saúde e Vigilância sanitária. O DeCS participa no projeto da rede semântica em saúde, UMLS - *Unified Medical Language System* da *NLM* com a responsabilidade da atualização e envio dos termos em português e espanhol.



Para vocabulário na área das ciências agrárias recomenda-se o uso do tesauro Agrovoc. É um vocabulário controlado, multilingue, que cobre as áreas de interesse da FAO (*Food and Agriculture Organization of the United Nations*), como alimentação, nutrição, agricultura, pescas, floresta, ambiente, etc.



Para vocabulário específico nas áreas de ciência e tecnologia, consultar a “Nomenclatura de Ciencia y Tecnología de la UNESCO”. É um sistema de classificação do conhecimento largamente usado em projetos de investigação e teses de doutoramento.



Para vocabulário específico na área do ambiente, pode-se consultar o tesouro GEMET. É um tesouro alojado na EIONET (*European Environment Information and Observation Network*), uma rede parceira da *European Environment Agency* (EEA).



Para recolha de autoridades de nomes próprios é recomendada a consulta do VIAF. O VIAF combina vários ficheiros de autoridades de nomes próprios num único serviço de autoridade alojado na OCLC. Este serviço tem como objetivo a redução de custos e o incremento da utilidade de ficheiros de autoridade de bibliotecas através de ligações e disponibilização da informação na *web*. É um projeto de várias bibliotecas nacionais e agências regionais e transnacionais para partilha multilingue de autoridades de nomes próprios.

Sintetizando as ligações:

Consulta preferencial

[Catálogo BNP](#) (Biblioteca Nacional)

Temáticas da União Europeia [EUROVOC](#) Thesaurus

Educação, cultura, comunicação [UNESCO](#) Thesaurus

Ciências da saúde [DeCs](#) - Lista de descritores

Ciências agrárias [AGROVOC](#) Agricultural Thesaurus

Ciência e tecnologia [UNESCO](#) nomenclature

Ambiente [GEMET](#) General Multilingual Environmental Thesaurus

Autoridades de Nomes Próprios

[VIAF](#) – Virtual International Authority File

3

Normalização de termos e de cabeçalhos, e controlo da linguagem no ficheiro de autoridades.

A 3ª etapa do modelo prevê a normalização de termos e cabeçalhos para posterior controlo de autoridades.

Uma das características da indexação num catálogo coletivo é a falta de coerência nos índices, devido ao uso de vários tipos de vocabulários. Será útil fazer a ligação entre terminologia especializada de tesouros, e terminologia mais geral dos índices da BNP, ou seja, estabelecer relações entre os diferentes vocabulários controlados. Sempre que

se verifique possível essa vinculação, os índices do catálogo coletivo resultarão mais uniformes para os usuários. Podem ser estabelecidas relações semânticas, de equivalência, hierárquicas ou associativas, utilizando os blocos 4xx e 5xx do UNIMARC Autoridades.

Assim, propõe-se observar o seguinte:

- a) Os índices da BNP devem ser sempre consultados, especialmente quando se pretende estabelecer cabeçalhos pré-coordenados, com subdivisões de assunto, de forma, de lugar ou cronológicas; os termos compostos, no caso de conterem mais do que um conceito devem dar origem a subdivisão do cabeçalho. A sintaxe dos cabeçalhos deve ser rigorosamente respeitada, de acordo com as subdivisões estabelecidas no UNIMARC (ver ‘Cabeçalhos’ nas tabelas SIPORbase, anexo I). No caso de consulta posterior de outra fonte, na qual se encontrem várias opções de termos, essas expressões poderão ser usadas para estabelecer as relações semânticas nos registos de autoridade.

Por exemplo: o termo ‘Proteção ambiental’ nos índices BNP remete para o cabeçalho autorizado ‘Ambiente--Proteção’ que cumpre as regras do sistema SIPORbase de divisão de termos compostos. No entanto, ao pesquisar o termo no tesouro Eurovoc, podemos beneficiar de uma série de termos não autorizados, como ‘conservação da natureza, proteção da natureza, salvaguarda do ambiente’, que são de grande utilidade para o controlo do vocabulário.

Termo nos índices BNP	Proteção ambiental Ver: Ambiente -- Proteção
Termo no tesouro Eurovoc	Proteção do ambiente UP Conservação da natureza Proteção da natureza Salvuarda do ambiente

Neste caso poderiam ser estabelecidas as relações de equivalência:

Ambiente--Proteção

UP

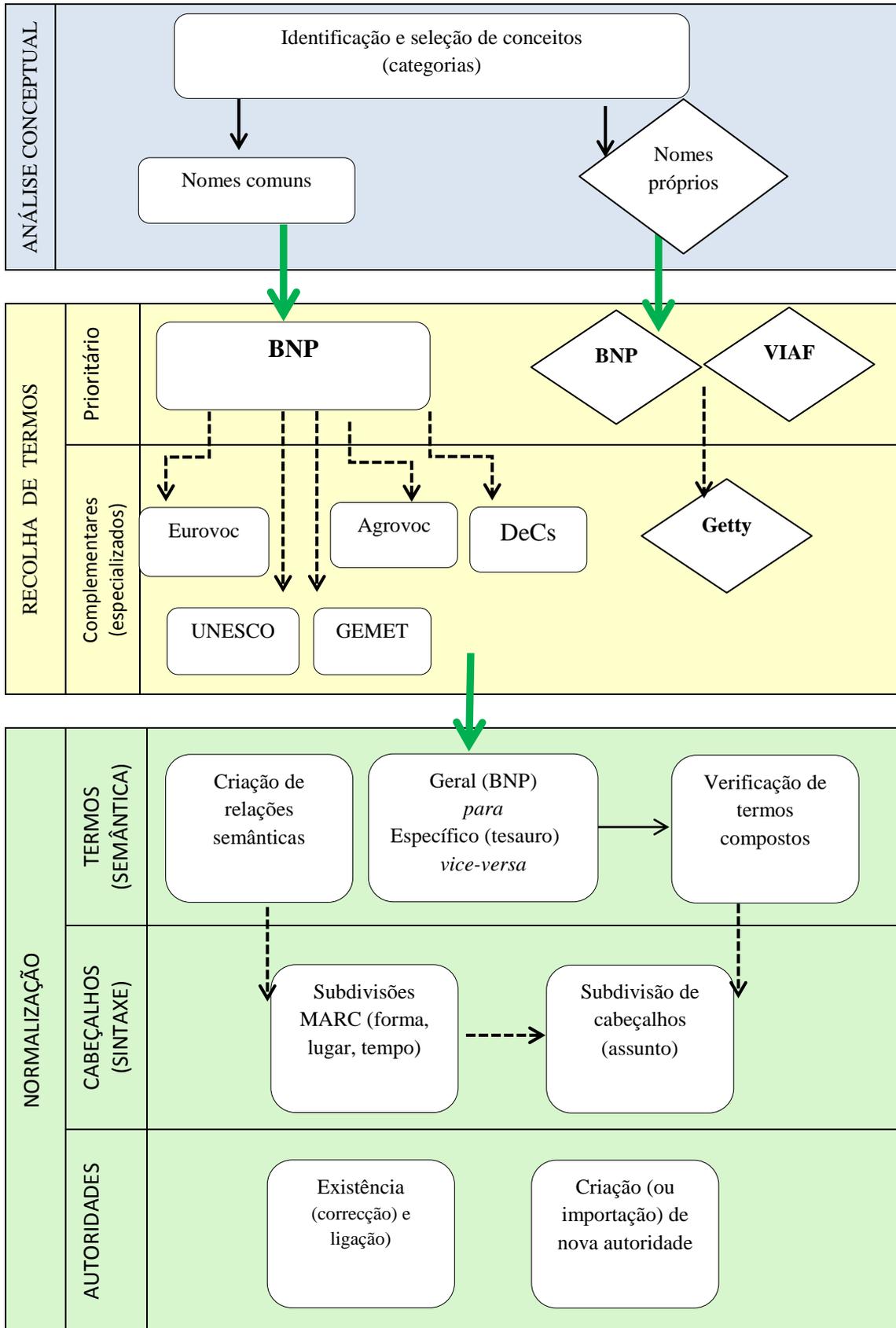
Proteção ambiental
Proteção do ambiente
Conservação da natureza
Proteção da natureza
Salvuarda do ambiente

- b) Sempre que seja possível devem ser igualmente estabelecidas relações ascendente e descendentes entre os termos, ou seja, de um mais específico para um mais geral e vice-versa;
- c) O controlo da terminologia no ficheiro de autoridade é imprescindível. Caso se encontrem termos não controlados no ficheiro de autoridade, selecionar os índices existentes e visualizá-los no formato UNIMARC para verificar se estão preenchidos, ou se contêm apenas termos resultantes de ‘criação automática’ (a opção de ‘controlo de autoridade’ deve estar sempre ativada). Selecionar o registo de autoridade que se encontra completo, para o associar ao registo bibliográfico. No caso de não existir um registo de autoridade devidamente preenchido, mas que contenha o termo que se decidiu atribuir, esse registo deve

ser preenchido adequadamente, de acordo com o nível mínimo exigido, que no UNIMARC Autoridades correspondem às etiquetas: 001 Identificador do Registo / 100 Dados Gerais de Processamento / 2xx Cabeçalho / 4xx Pistas ‘ver’ (caso existam) / 5xx Pistas ‘ver também’ (caso existam) / 801 Fonte de origem.

- d) Os termos que representam conceitos novos devem ser verificados quanto ao conteúdo e à forma, com a ajuda de instrumentos de referência (dicionários, enciclopédias); os novos termos devem ser recolhidos nas fontes de terminologia recomendadas; no caso de não se encontrarem aí os termos desejados, devem ser controlados de acordo com as diretrizes do SIPORbase (anexo I).

Fluxograma do modelo de indexação



8. Bibliografia

- Albás, L. 2005. *Control de autoridades y calidad de puntos de acceso del catálogo automatizado de la biblioteca de la Universidad de Salamanca* [Em linha]. Disponível em: <http://web.usal.es/~alar/Bibweb/Temario/Lorenzo.PDF>. [Consult. 12 Jun. 2013].
- Anderson, J. D., Pérez-Carballo, J. 2001a. The nature of indexing: how humans and machines analyse messages and texts for retrieval. Part I: Research, and the nature of human indexing. *Information Processing and Management*, 37: 231-254.
- Anderson, J. D., Pérez-Carballo, J. 2001b. The nature of indexing: how humans and machines analyse messages and texts for retrieval. Part II: Machine indexing, and the allocation of human versus machine effort. *Information Processing and Management*, 37: 255-277.
- Bergamin, G., Lucarelli, A. 2013. The Nuovo soggettario as a service for the linked data world. *Italian Journal of Library and Information Science*, v4 (1). Disponível em: <http://leo.cilea.it/index.php/jlis/article/view/5474> [Consult. 10 Mai. 2013].
- Bermello Crespo, L. 2001. Tendencias de desarrollo de los lenguajes de encabezamientos de materia en las bibliotecas. *Acimed*, v9 (2). Disponível em: <http://eprints.rclis.org/5190/> [Consult. 10 Ago. 2013].
- Bertrand, A., Cellier, J. M., Giroux, L. 1996. Expertise and strategies for the identification of the main ideas in document indexing. *Applied cognitive psychology*, v10: 419-433.
- Bocato, V. R. C., Fujita, M. S. L. 2006. Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica. *Perspectivas em Ciência da Informação*, v11, n2: 267-281.

- Bocato, V. R. C. 2009. *Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo sociocognitivo com protocolo verbal*. Universidade Estadual Paulista, Campus de Marília.
- Bocato, V. R. C., Fujita, M. S. L. 2010. O uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal. *Perspectivas em Ciência da Informação*, v15, n3: 23-51.
- Bocato, V. R. C., Fujita, M.S.L. 2011. Evaluation of indexing language used in collective catalogues of university libraries: a socio-cognitive study using verbal protocol. *International Journal of Library and Information Science*, v3(8): 162-178.
- Bocato, V. R. C., Fujita, M. S. L., Gil Leiva, I. 2011. Avaliação comparada do uso de linguagens de indexação em catálogos de bibliotecas universitárias para recuperação por assunto. *Representación y Organización del Conocimiento*, v17 n1. Disponível em: <http://www.iberid.eu/ojs/index.php/scire> [Consult. 20 Set. 2012].
- Bodenreider, O. 2004. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acid Research*, v32: 267-270.
- Bonal Zazo, J. L., Ortego de Lorenzo-Cáceres, M. P. 2000. Métodos de evaluación de calidad de catálogos automatizados: análisis comparativo. Jornadas Españolas de Documentación (7ª). *La gestión del conocimiento: retos y soluciones de los profesionales de la información*. Bilbao. Disponível em: <http://dialnet.unirioja.es/servlet/articulo?codigo=568773> [Consult. 6 Mar. 2013].

- Borgman, C. L. 2000. Why are digital libraries hard to use? In *From Gutenberg to the global information infrastructure: access to information in the networked world*. Cambridge: MIT, 2000. p117-141.
- Bourdon, F., Landry, P. 2008. Best Practices for Subject Access to National Bibliographies: interim report by the Working Group on Guidelines for Subject Access by National Bibliographic Agencies. *International Cataloguing and Bibliographic Control (ICBC)*, v37, n4: 63-65.
- Broughton, V. 2006. *Essential thesaurus construction*, London, Facet publishing.
- Brundage, C. A. 1989. Teaching controlled vocabulary and natural language to end-users. *Science & Technology Libraries*, v10, n1: 3-13.
- Bugeda, J. 1974. *Manual de tecnicas de investigacion social*. Madrid: Instituto de Estudios Politicos.
- Byrum, J. D. 2003. Naco: a cooperative model for building and maintaining a shared name authority database. *International Conference on Authority Control: Definition and international experiences*. Florence, Italy. Disponível em: <http://www.sba.unifi.it/ac/en/program.htm> [Consult. 30-12-2012]
- Campos, F. M. 2003. Controlo de autoridade: novos contextos e soluções. Jornadas Porbase (9^{as}). *Novos contextos para a cooperação*. Lisboa. Disponível em: <http://purl.pt/331/1/programa.html> [Consult. 12-08-2012]
- Caro-Castro, C. 2005. *El acceso por materias en los catálogos en línea: índices y terminología de los usuarios en el catálogo CISNE*. Universidad de Salamanca.
- Caro-Castro, C. 2012. Vocabulários estruturados, Web Semântica y Linked Data: oportunidades y retos para los profesionales de la Documentación. In *Arquivologia, Biblioteconomia e Ciência da Informação: identidades, contrastes*

e perspectivas de interlocução. Eduardo I. Murgia, Mara E. F. Rodrigues (eds). Niterói, RJ: Editora da UFF: 139-155.

Caro-Castro, C., Travieso Rodríguez, C. 2003. Encabezamientos de materia en las bibliotecas españolas: perspectiva histórica y situación actual. In *Tendencias de investigación en organización del conocimiento*. Crispulo Travieso Rodríguez, José António Frías Montoya (coord.). Salamanca. p. 327-334. Disponible em: <http://dialnet.unirioja.es/servlet/articulo?codigo=1225339> [Consult. 15-06-2013].

Caro-Castro, C., Cedeira Serantes, L., Travieso Rodríguez, C. 2003. La investigación sobre recuperación de información desde la perspectiva centrada en el usuario: métodos y variables. *Revista española de documentación científica*, v26, n1.

Castillo Vidal, J. 2003. Catálogos colectivos, redes de bibliotecas y catálogos virtuales. *El profesional de la información*, v12, n2.

Celestino Angulo, S. 2007. Los centros de recursos para el aprendizaje y la investigación: de la biblioteca al CRAI. Jornadas Nacionales *Información y Comunicación en Ciencias de la Salud*. Zaragoza. Disponible em: http://www.iacs.aragon.es/econocimiento/documentos/biblioteca/jornadasbiblios_alud/mesa-redonda-crais/sonsoles-celestino.pdf [Consult. 8 Jun.2013].

Cerbo, M. A. 2011. Is there a future for library catalogers? *Cataloguing & Classification Quarterly*, 49: 323-327.

Chapman, A. 1994. Up to Standard? a study of the quality of records in a shared cataloguing database. *Journal of Librarianship and Information Science*, 26 (4).

Clarke, S. G., Zeng, M. L. 2012. From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modeling. *ISQ Information Standards Quarterly*, v24 (1): 21-26.

- Clavel-Merrin, G. 2004. MACS (Multilingual access to subjects): A virtual authority file across languages. *Cataloguing & Classification Quarterly*, v39, 1-2: 323-330.
- Cleverdon, C. W. 1962. *ASLIB Cranfield research project – report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*. Disponível em: <http://dspace.lib.cranfield.ac.uk/handle/1826/836> [Consult. 20 Jun. 2013].
- CLIP-Compatibilização de Linguagens de Indexação em Português. *Projecto Clip: fontes para o controlo de terminologia*. Lisboa, 1995.
- Collantes, L. Y. 1995. Degree of agreement in naming objects and concepts in information retrieval. *Journal of the American Society for Information Science*, 46 (2): 116-132.
- Coyle, K. 2007. The Library Catalog: some possible futures. *The Journal of Academic Librarianship*, v33, n3: 414-416.
- Cristán, A. 2003. Subject gateways and SACO. *International Conference on Authority Control: Definition and international experiences*. Florence, Italy. Disponível em: <http://www.sba.unifi.it/ac/en/program.htm> [Consult. 20 Dez. 2012]
- Dal'Evedove, P. R., Fujita, M. S. L. 2008. O estudo da cognição profissional pelo protocolo verbal de catalogadores de assunto em contexto de biblioteca universitária: uma abordagem sociocognitiva pela análise de domínio. *Revista de Iniciação Científica da FFC*, 8: 249-262.
- Danskin, A. 2013. Linked and open data: RDA and bibliographic control. *JLIS.it Italian Journal of Library and Information Science*, v4, n1. Disponível em: <http://leo.cilea.it/index.php/jlis/article/view/5463> [Consult. 29 Jul. 2013].

- Dean, R. J. 2003. FAST: Development of simplified headings for metadata. *International Conference on Authority Control: Definition and international experiences*. Florence, Italy. Disponível em: <http://www.sba.unifi.it/ac/en/program.htm> [Consult. 30 Dez. 2012]
- Delandés Palomares, M. 2012. *Gestión de bibliotecas y centros de recursos para el aprendizaje y la docencia en el nuevo entorno docente*. Universidad Politécnica de Valencia.
- Dicionário Priberam da Língua Portuguesa. [Em linha]. Disponível em: <http://www.priberam.pt/dlpo/>. [Consult. 13 Jun. 2013].
- Drabenstott, K. M. 1996a. Authority control in digital libraries. In Authority control in the 21st century: an invitational conference: *proceedings of the OCLC Symposium*.
- Drabenstott, K. M. 1996b. Enhancing a new design for subject access to online catalogs. *Library Hi Tech*, 14 (1): 87-109.
- Drabenstott, K. M., Simcox, S., Fenton, E. G. 1999. End-user understanding of subject headings in library catalogs. *Library Resources & Technical Services*, 43 (3): 140-160.
- Extremeño, A. 1999. Calidad de la indización e incidencia de errores en la base de datos ECOSOC. *Revista Española de Documentación Científica.*, 22 (2): 157-173.
- Extremeño, A., Moscoso, P. 1998. El control de la calidad en bases de datos de Ciencias Sociales. *Boletín de la ANABAD*, n1: 231-253.
- Fidel, R. 1986. The possible effect of abstracting guidelines on retrieval performance of free-text searching. *Information Processing and Management*, v22, n4: 309-316.

- Fidel, R. 1991a. Searcher's selection of search keys. I. The selection routine. *Journal of the American Society for Information Science*, 42 (7): 490-500.
- Fidel, R. 1991b. Searchers' selection of search keys: II. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science*, 42 (7): 501-514.
- Fidel, R. 1992. Who needs controlled vocabulary. *Special Libraries*, 83(1): 1-9.
- Foskett, A. C. 1997. *The subject approach to information*. London: Library Association.
- Fujita, M. S. L. 2003. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, v1, n1. Disponível em: <http://eprints.rclis.org/6266/> [Consult. 12 Abr. 2012].
- Fujita, M. S. L., Gil- Leiva, I. 2009. As linguagens de indexação em Bibliotecas Nacionais e sistemas de informação na América Latina. XVI Seminário Nacional de Bibliotecas Universitárias e II Seminário Internacional de Bibliotecas Digitais. Rio de Janeiro, Brasil. Disponível em: <http://eprints.rclis.org/15137/> [Consulta 12 Ago. 2012].
- Functional Requirements for Authority Data: a conceptual model*. Working Group on Functional Requirements and Numbering of Authority Records. IFLA, 2009. Disponível em: <http://www.ifla.org/publications/ifla-series-on-bibliographic-control-34> [Consult. 16 Ago. 2012].
- Functional Requirements for Bibliographic Records: final report*. IFLA Study Group on the functional requirements for bibliographic records. München: Saur, 2008. Disponível em: http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf [Consult. 16 Ago. 2012].

- Functional Requirements for Subject Authority Data (FRSAD): a conceptual model.* Working Group for Subject Authority Records. IFLA, 2009. Disponível em: <http://nkos.slis.kent.edu/FRSAR/report090623.pdf>. [Consult. 20 Ago. 2012].
- Galvão, R. M. 2003. UNIMARC Autoridades: novos desenvolvimentos. Jornadas Porbase (9^{as}). *Novos contextos para a cooperação*. Lisboa. Disponível em: <http://purl.pt/331/1/programa.html> [Consult. 12 Ago. 2012].
- Galvão, R. M., Machado, A. 2004. Avaliação de qualidade: em busca da excelência nas bibliotecas. Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas (8^o). *Nas encruzilhadas da Informação e da Cultura: (Re)Inventar a Profissão*. Estoril. Disponível em: <http://www.apbad.pt/Edicoes/EdicoesCongresso8.htm> [Consult. 12 Ago. 2012].
- Galvão, R. M., Cordeiro, M. I. 2010. Novos princípios, modelos e normas para o futuro dos serviços de informação bibliográfica. Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas (10^o), Guimarães. *Políticas de Informação na Sociedade em Rede: actas*. Lisboa: B.A.D. Disponível em: <http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/197> [Consult. 10 Ago. 2013].
- García Marco, F. J., Agustín Lacruz, C., Caro-Castro, C. [et al.] 2007. Proyectos internacionales de reforma y ampliación de las normas sobre tesauros para su adaptación a los nuevos contextos de integración e interoperabilidad en el entorno digital. In *La interdisciplinareidad y la transdisciplinaridad en la organización del conocimiento científico: actas del VIII Congreso ISKO-España*. (Eds.) Blanca Rodríguez Bravo, M^a Luisa Alvite Díez. León: Universidad de León: 289-298.
- Gil-Leiva, I. 1997. *La automatización de la indización, propuesta teórico-metodológica: aplicación al área de biblioteconomía y documentación*. Universidad de Murcia.

- Gil-Leiva, I. 2001. Consistencia en la asignación de materias en Bibliotecas Públicas del Estado. *Boletín de la Asociación Andaluza de Bibliotecarios*, n63: 69-86.
- Gil-Leiva, I. 2002. Consistencia en la indización de documentos entre indizadores noveles. *Anales de documentación*, n5: 99-111.
- Gil-Leiva, I. 2008. Consistência na indexação em bibliotecas universitárias brasileiras. *TransInformação*, 20 (3): 233-253.
- Gómez Hernández, J. A. 1995. *La función de la biblioteca en la Educación Superior. Estudio aplicado a la Biblioteca Universitaria de Murcia*. Universidad de Murcia.
- Gorman, M. 2003. Authority control in the context of bibliographic control in the electronic environment. *International Conference on Authority Control: Definition and international experiences*. Florence, Italy. Disponible em: <http://www.sba.unifi.it/ac/en/program.htm> [Consult. 30 Dez. 2012].
- Hancock-Beaulieu, M. 1990. Evaluating the impact of an online library catalogue on subject searching behaviour at the catalogue and at the shelves. *Journal of Documentation*, 46(4): 318-338.
- Hatsek, I. N. 2012. *Resource Description and Access (RDA): evolução no conceito de catalogação*. Florianópolis: Universidade Federal de Santa Catarina.
- Heiner-Freiling, M. 2000. Survey on subject heading languages used in national libraries and bibliographies. *Cataloging & Classification Quarterly*, 29 (1/2): 189-198.
- Henzler, R. G. 1978. Free or controlled vocabularies. *International Classification*, v5, n1: 21-26.

- Hjørland, B. 1997. *Information seeking and subject representation: an activity-theoretical approach to information science*. London: Greenwood Press.
- Hjørland, B. 2007. Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41 (1): 367-405.
- Hjørland, B. 2008. *Lifeboat for Knowledge Organization* [Em linha]. Disponível em: http://www.iva.dk/bh/Lifeboat_KO/home.htm [Consult. 25 Fev. 2012].
- Hjørland, B. 2012. Is classification necessary after Google? *Journal of Documentation*, v68, n3: 299-317.
- Hooper, R. S. 1965. *Indexer consistency test-origin, measurements, results and utilization*. Bethesda: IBM Corporation.
- Hunter, R. N. 1991. Successes and failures of patrons searching the online catalog at a large academic library: a transaction log analysis. *RQ*, v30, n3: 395-402. Disponível em: <http://www.jstor.org/stable/i25828801> [Consult. 2 Fev. 2013].
- IFLA 2009. *Guidelines for Multilingual Thesauri*. Working Group on Guidelines for Multilingual Thesauri. IFLA Classification and Indexing Section. Disponível em: <http://www.ifla.org/publications/ifla-professional-reports-115> [Consult. 17 Jul. 2012].
- IFLA 2011. *Guidelines for Subject Access in National Bibliographies: draft 2011 for worldwide review*. Working Group on Guidelines for Subject Access by National Bibliographic Agencies. IFLA Classification and Indexing Section.
- Ingwersen, P. 1982. Search procedures in the library: analysed from the cognitive point of view. *Journal of Documentation*, v38, n3: 165-191.

- Ingwersen, P. 2002. *Information Retrieval Interaction* [Em linha]. Disponível em: [http://pure.iva.dk/en/persons/peter-ingwersen\(06732ae6-b7ec-4427-8049-6b087f1e5f6d\).html](http://pure.iva.dk/en/persons/peter-ingwersen(06732ae6-b7ec-4427-8049-6b087f1e5f6d).html) [Consult. 05 Ago. 2013]
- ISO25964/1:2011 – Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization.
- ISO2788:1986 - Documentation, guidelines for the establishment and development of monolingual thesauri. Geneva: International Organization for Standardization.
- Keen, E. M., Digger, J. A. 1972. *Report of an information science index languages test*. College of Librarianship, Department of Information Retrieval Studies.
- Knapp, S. D., Cohen, L. B., Juedes, D. R. 1998. A natural language thesaurus for the humanities: the need for a database search aid. *Library Quarterly*, v68, n4: 406-30.
- Lancaster, F. W. 1968. *Evaluation of the MEDLARS demand search service*. U.S. Department of Health, Education, and Welfare.
- Lancaster, F. W. 1991. *Indexing and abstracting in theory and practice*. University of Illinois, Graduate School of Library and Information Science.
- Lara, M. L. G. 2013. Documentary languages and knowledge organization systems in the context of the semantic web. *Transinformação*, v.25 (2): 145-150.
Disponível em: http://www.scielo.br/scielo.php?script=sci_issuetoc&pid=0103-378620130002&lng=es&nrm=iso
- Larson, R. R. 1991. The decline of subject searching: long term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science*, 42(3): 197-215.

- Library Linked Data Incubator Group Final Report*. W3C Incubator Group Report 2011. [Em linha]. Disponível em: <http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025/> [Consult.23 Out. 2012].
- Lopes, C. 2006. *Qualidade de serviço em bibliotecas universitárias: desenvolvimento e validação de um instrumento de avaliação*. Universidade de Salamanca.
- Lopes, I. L. 2002. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. *Ciência da Informação*, v31(n1): 45-52.
- Lopes, M. M., Galvão R. M. 2012. Da ficha bibliográfica aos dados relacionados: uma revolução em curso. [Em linha]. *Integração, Acesso e Valor Social: actas do 11º Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*. Lisboa. Disponível em: <http://www.bad.pt/publicacoes/index.php/congressosbad/index> [Consult. 16 Mar. 2013].
- Lopes, M. M., Freire, N., Fontes, D. [et al.] 2007. UNIMARC: Estudo da sua utilização na PORBASE. *Bibliotecas e Arquivos- Informação para a Cidadania, o Desenvolvimento e a Inovação: actas do 9º Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*. Açores. Disponível em: <http://www.apbad.pt/Edicoes/EdicoesCongresso9.htm> [Consult. 10 Out. 2012].
- Markey, K. 1984. *Subject searching in library catalogs: before and after the introduction of online catalogs*. Dublin, Ohio: OCLC.
- Markey, K. 2007a. Twenty-five years of end-user searching, part 1: research findings. *Journal of the American Society for Information Science and Technology*, 58(8): 1071-1081.

- Markey, K. 2007b. Twenty-five years of end-user searching, part 2: future research directions. *Journal of the American Society for Information Science and Technology*, 58(8): 1123-1130.
- Martín Gavilán, C. 2009. *Temas de Biblioteconomía: Lenguajes documentales. Principales tipos de clasificación. Encabezamientos de materia, descriptores y tesauros*. Disponible em: <http://eprints.rclis.org/14817/> [Consult. 15 Feb. 2012].
- Martins, C. 2009. A prática de indexação em bibliotecas: estudo de caso do Instituto Politécnico de Portalegre: tornar possível a indexação. (não publicado), Universidade de Salamanca.
- Martínez, D. 2004. El Centro de Recursos para el Aprendizaje CRAI. El nuevo modelo de biblioteca universitaria. *Educación y biblioteca*, v16, 144: 98-108. Disponible em: <http://www.ucm.es/BUCM/biblioteca/doc6202.pdf> [Consult. 8 Jun. 2013].
- Martínez Méndez, F. J., Rodríguez Muñoz, J. V. 2004. Reflexiones sobre la evaluación de los sistemas de recuperación de información: necesidad, utilidad y viabilidad. *Anales de Documentación*, n7: 153-170.
- Martínez, A. M., Mangiaterra, N. E., Fortunato, R. H. [et al.] 1997. Control de autoridades en catálogos en línea. *Investigación Bibliotecológica*, v11, n23: 81-101.
- Melo, L. B. 1994. Avaliação de sistemas de recuperação de informação: breve retrospectiva dos principais projectos. *Cadernos BAD*, v2: 29-46.
- Méndez Rodríguez, E. M., Moreira González, J. A. 1999. Lenguaje natural e indización automatizada. *Ciencias de la Información*, 30(3): 11-24.
- Miler, U., Teitelbaum, R. 2002. Pre-coordination and post-coordination: past and future. *Knowledge organization*, v29, n2: 87-93.

- Montalvo, M. M. 2011. LCSH, FAST y DELICIOUS: vocabularios normalizados y nuevas formas de catalogación temática. *Anales de Documentación*, v14, n1. Disponible em: <http://revistas.um.es/analesdoc/issue/view/9791> [Consult. 21-03-2013].
- Moreira, M. P. 2005. *Ambiente para geração e manutenção semi-automática de tesouros*. Belo Horizonte: Escola de Ciência da Informação da UFMG.
- Moreiro, J. A., Sánchez-Cuadrado, J., Morato, J., Moreno, V. 2009. Desarrollo de una aplicación ontológica para evaluar el mercado de trabajo español en Biblioteconomía y Documentación. *Revista Española de Documentación Científica*, v32, n1: 81-98.
- Moreno, F. P. 2006. *Requisitos Funcionais para Registros Bibliográficos – FRBR: um estudo no catálogo da Rede Bibliodata*. Universidade de Brasília: FACE - Departamento de Ciência da Informação e Documentação.
- Moreno, F. P., Brascher, M. 2007. Requisitos Funcionais para Registros Bibliográficos – FRBR: um estudo no catálogo da Rede Bibliodata. VIII ENANCIB - *Encontro Nacional de Pesquisa em Ciência da Informação*. Salvador, Bahía.
- Moreno, F. P. 2009. O modelo conceitual FRBR: discussões recentes e um olhar sobre as tarefas do usuário. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*. [Em linha]. v14, n27: 47-68. Disponible em: <https://periodicos.ufsc.br/index.php/eb> [Consult. 14-08-2012].
- Moscoso, P., Extremeño, A. 1999. Análisis del lenguaje controlado en tres bases de datos internacionales. *Información, Cultura y Sociedad*, n2: 45-64.

- Muddamalle, M. R. 1998. Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics. *Journal of the American Society for Information Science*, 49(10): 881-887.
- NP 3715. 1989. Documentação – *Método para análise de documentos, determinação do seu conteúdo e selecção de termos de indexação*. Lisboa: IPQ.
- NP 4036. 1992. Documentação - *Tesauros monolíngues: directivas para a sua construção e desenvolvimento*. Lisboa: IPQ.
- NP 4285-3. 2000. Documentação e informação. *Vocabulário parte 3: Aquisição, identificação e análise de documentos e de dados*. Lisboa IPQ.
- NP 4285-4. 2000. Documentação e informação. *Vocabulário. parte 4: Linguagens documentais*. Lisboa. IPQ.
- Olensky, M. 2010. Semantic interoperability in Europeana. An examination of CIDOC CRM in digital cultural heritage documentation. *Bulletin of IEEE Technical Committee on Digital Libraries*. [Em linha], v2, 2. Disponível em: <http://www.ieee-tcdl.org/Bulletin/v6n2/Olensky/olensky.html> [Consult. 14 Ago. 2012].
- Olson, H. A., Boll, J. J. 2001. *Subject analysis in online catalogs*. Englewood, CO.: Libraries Unlimited.
- Pacheco, Emília L.M. 2003. *A biblioteca híbrida: o acesso ao conteúdo das publicações periódicas científicas portuguesas nas bibliotecas universitárias*. Lisboa: ISCTE: Departamento de Ciências e Tecnologias da Informação.

- Poo, D. C. C., Khoo, C. 1997. Subject searching in online catalog systems. In *Encyclopedia of Library and Information Science*. A. Kent (Ed.) New York: Marcel Dekker, v60 (23): 324-341.
- Regras portuguesas de catalogação*. Lisboa: Biblioteca Nacional de Portugal, 2000.
- Requisitos funcionais dos registos bibliográficos: relatório final*. IFLA. Lisboa: Biblioteca Nacional de Portugal, 2008.
- Requisitos funcionais para dados de autoridade: um modelo conceptual*. IFLA. Lisboa: Biblioteca Nacional de Portugal, 2011.
- Ribeiro, F. 1996. Subject indexing authority control in archives: the need for subject indexing in archives and for an indexing policy using controlled language. *Journal of the Society of Archivists*, 17(1): 27-54.
- Ribeiro, F. 2005. Novos caminhos para a avaliação da informação. Seminário *Avaliação de documentos a serviço da administração*. Recife. Disponível em: <http://ler.letras.up.pt/site/default.aspx?qry=id07id145&sum=sim> [Consult. 8 Ago. 2013].
- Ríos Hilario, A. B. 2003. *La estructura conceptual del registro bibliográfico: análisis de la funcionalidad de las "Reglas de Catalogación" españolas y del formato Ibermarc bibliográfico*. Salamanca: Ediciones Universidad de Salamanca.
- Ríos Hilario, A. B. 2006. Diseño de un modelo de descripción para los recursos audiovisuales basado en el modelo FRBR. *VIII Jornadas de Gestión de la Información*. Madrid: SEDIC: 127-136.
- Rodríguez Yunta, L. 1998. Evaluación e indicadores de calidad en bases de datos. *Revista Española de Documentación Científica*, 21(1): 9-23.

- Rodríguez Yunta, L., Alonso-Gamboa, J. O. 2011. Normalización del vocabulario de indización em bases de datos multidisciplinares de ciencias sociales y humanas: análisis comparativo de la experiencia de ISOC (España) y CLASE (México) desde la década de 1970 hasta 2010. In *actas del X Congreso ISKO-España*. Ed. lit. M^a Carmen Pérez Pais, M^a G. Bonome. Ferrol: Universidade da Coruña: 209-226.
- Rothman, J. 1983. Is indexing obsolete? Keyword indexing and free-text searching. In *Indexing specialized formats and subjects*. London: Scarecrow Press.
- Rowley, J. 1994. The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. *Journal of Information Science*, 20 (2): 108-119.
- Salton, G. 1967. *The SMART Project - Status Report and Plans*. Disponible em: http://www.sigir.org/museum/pdfs/ISR-12/pdfs/p1-section_i.pdf [Consult. 15 Set. 2012].
- Salton, G. 1986. On the use of knowledge-based processing in automatic text retrieval. In *ASIS'86: Proceedings of the 49th Annual Meeting of the American Society for Information Science*. Chicago: American Society for Information Science.
- Salton, G. 1996. A new horizon for information science. *Journal of the American Society for Information Science*, 47(4): 333.
- Salton, G., Buckley, C., Fox, E. A. 1983. Automatic query formulations in information retrieval. *Journal of the American Society for Information Science*, 34 (4): 262-280.
- Salton, G., McGill M. J. 1983. *Introduction to modern information retrieval*. New York: McGraw-Hill.

- Sánchez-Cuadrado, S., Colmenero-Ruiz, M. J., Moreiro, J. A. 2012. Tesauros: Estándares y Recomendaciones. *El profesional de la información*, v21, n3: 229-235.
- Santos, M. 2006. *Indexação: terminologia e controlo de autoridades (manual)*. Lisboa: Biblioteca Nacional de Portugal.
- Santos, M., Almeida, A. C. 2003. Ficheiro de autoridade e qualidade da indexação. Jornadas Porbase (9^{as}), *Novos contextos para a cooperação*. Lisboa.
- Santos, M. L. F. N. d. 2007. *Organização do conhecimento e representação de assuntos: caminhos para uma efectiva recuperação da informação em ambientes de rede*. Lisboa: Biblioteca Nacional de Portugal.
- Savoy, J. 2004. Bibliographic database access using free-text and controlled vocabulary: an evaluation. *Information Processing and Management*, 41: 873-890.
- Sequeda, J. 2012. *Introduction to: SKOS* [Em linha]. Disponível em: http://semanticweb.com/introduction-to-skos_b33086 [Consult. 21 Jun. 2013].
- Serrano, A. 2003. Avaliação de serviços de informação do ensino superior. *Páginas a&b*, 12: 7-65.
- Shields, G. 2005. What are the main differences between human indexing and automatic indexing? [Em linha]. In *LI-842 Automatic Indexing Assignment*. Disponível em: <http://www.docstoc.com/docs/95492730/What-are-the-main-differences-between-human-indexing-and> [Consult. 30 Ago. 2012].
- Sievert, M. C., Andrews, M. J. 1991. Indexing consistency in information science abstracts. *Journal of the American Society for Information Science*, 42(1): 1-6.

- Simão, J. V., Santos, S. M., Costa, A. A. 2003. *Ensino Superior: uma visão para a próxima década*. Lisboa: Gradiva.
- Simão, J. V., Costa, A. A. 2000. *O ensino politécnico em Portugal*. Lisboa: Conselho Coordenador dos Institutos Superiores Politécnicos.
- Simões, M. d. G. 2008. *Da abstracção à complexidade formal: relações conceptuais num tesouro*. Coimbra: Almedina.
- SIPORbase – Sistema de Indexação em Português: manual, 1992. Grupo de Trabalho de Indexação; Comissão Coordenadora do Projecto Clip. Lisboa: Biblioteca Nacional de Portugal.
- SIPORbase - Sistema Indexação em português: manual, 1998. Área de Classificação e Indexação. Lisboa: Biblioteca Nacional de Portugal.
- Soergel, D. 1985. *Organizing information. Principles of data base and retrieval systems*. Orlando, FI: Academic Press.
- Soler Monreal, M. C. 2009. *Evaluación de vocabularios controlados en la indización de documentos mediante índices de consistencia entre indizadores*. Universidad Politecnica de Valencia.
- Strehl, L. 1998. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. *Ciência da Informação*, v27, n3: 329-335.
- Svenonius, E. 1986. Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, 37(5), 331-340.
- Svenonius, E. 2000. LCSH: Semantics, syntax and specificity. *Cataloging & Classification Quarterly*, 29(1-2): 17-30.

Svenonius, E. 2003. Design of controlled vocabularies. In *Encyclopedia of Library and Information Science*. New York: Marcel Dekker: 823-838.

Taladriz Mas, M. 2006. La biblioteca como centro de recursos: nuevos modelos para realidades cambiantes. In Jornada profesional *Bibliotecas y Educación: una relación a debate*. (org.) Biblioteca Nacional de España, Fundación Germán Sánchez Ruipérez, Sociedad Española de Documentación e Información Científica (SEDIC). Madrid. Disponible em: http://www.sedic.es/Margarita_Taladriz.pdf [Consult. 10 Jun. 2013].

Taladriz Mas, M. 2007. CRAIS: Nuevos espacios bibliotecarios para el aprendizaje y la investigación. *Jornadas Nacionales de Información y Documentación en Ciencias de la Salud (12^{as})*. Zaragoza.

Tartaglia, S. 2003. Authority control and subject indexing languages. *International Conference on Authority Control: Definition and international experiences*. Florence, Italy. Disponible em: <http://www.sba.unifi.it/ac/en/program.htm> [Consult. 30-12-2012].

Tillet, B. B. 2003. Authority control: state of the art and new perspectives. *International Conference on Authority Control: Definition and international experiences*. Florence, Italy. Disponible em: <http://www.sba.unifi.it/ac/en/program.htm> [Consult. 30-12-2012].

Tonta, Y. 1991. A Study of indexing consistency between Library of Congress and British Library Catalogers. *Library Resources & Technical Services*, v35, n2: 177-185. Disponible em: <http://eprints.rclis.org/9464/> [Consult. 5 Ago. 2013].

UNIMARC formato bibliográfico: versão abreviada da edição de 2008. IFLA. Lisboa, Biblioteca Nacional de Portugal, 2009.

- UNIMARC: formato bibliográfico - manual*. IFLA. Lisboa: Biblioteca Nacional de Portugal, 2008.
- UNIMARC: formato autoridade - manuals*. IFLA. Lisboa: Biblioteca Nacional de Portugal, 2011.
- Van Slype, G. 1991. *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide.
- Vargas-Quesada, B., Moya Anegón, F., Olvera Lobo, M. D. 2002. Enfoques en torno al modelo cognitivo para la recuperación de información: análisis crítico. *Ciência da Informação*, v31, n2: 107-119.
- Vasconcelos, D. 2004. Biblioteca do Conhecimento Online: um novo impulso para a sociedade da informação. *Cadernos BAD* (1): 117-123.
- Vatant, B. 2010. Porting library vocabularies to the Semantic Web, and back: A win-win round trip. *World Library and Information Congress: 76th IFLA General Conference and Assembly*. Gothenburg, Sweden.
- Vicente, P., Reis, E., Ferrão, F. 1996. *Sondagens: a amostragem como factor decisivo de qualidade*. Lisboa: Edições Sílabo.
- Villén-Rueda, L. 2006. Indización y recuperación por materias en los opacs de las bibliotecas españolas: ¿dos décadas de evaluación? *El profesional de la información*, v15, n2: 87-98.
- Weber, J. 2003. LEAF. Linking and Exploring Authority Files (www.leaf-eu.org). *International Conference on Authority Control: Definition and international experiences*. Florence, Italy. Disponível em: <http://www.sba.unifi.it/ac/en/program.htm> [Consult. 30-12-2012].

Wellisch, H. H. 1994. Book and periodical indexing. *Journal of the American Society for Information Science*, v45, n8: 620-627.

White, H. D., Griffith, B. C. 1987. Quality of indexing in online databases. *Information Processing and Management*, 23: 211-224.

Wolverton Jr., R. E. 2006. Becoming an authority on authority control: an annotated bibliography of resources. *Library Resources & Technical Services*, 50 (1): 31-41

Yee, M., Layne, S. S. 1998. *Improving online public access catalogs*. Chicago: American Library Association.

Zapparoli, Z. M., Camlong, A. 2002. *Do léxico ao discurso pela informática*. São Paulo: Editora da Universidade de São Paulo.

Zavalina, O. L. 2012. Subject Access: Conceptual Models, Functional Requirements, and Empirical Data. *Journal of Library Metadata*, v12 (2-3). Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/19386389.2012.699829> [Consult. 22 Jul. 2013].

Nota: A bibliografia apresentada segue as regras 'Harvard Referencing Style'. Disponível em: <http://www.library.auckland.ac.nz/instruct/ref/harvard.htm>. [Consult. 11 Set. 2013].

9. Anexos

Anexo I. Tabelas abreviadas SIPORbase

As tabelas abreviadas que a seguir se apresentam pretendem sintetizar pontos básicos para orientação da tarefa de indexação. Remetem nalguns casos para a secção correspondente do Manual SIPORbase, que deverá ser consultado para verificação de questões mais aprofundadas. Estas tabelas são constituídas essencialmente por exemplos, que poderão facilitar a pesquisa e localização de termos por parte dos profissionais, nos índices de assuntos do catálogo da BNP, construído de acordo com as regras do sistema.

Tab. SIP1 – Política de indexação

Análise do documento

Identificação do número de assuntos

Identificação do conceito que expressa cada assunto

Seleção dos conceitos

Representação (princípios de indexação)

Princípio terminológico

Princípio da entrada única

Princípio da entrada direta

Princípio da entrada específica

Princípio de pré-coordenação

Tab. SIP2 – Cabeçalhos

Formas do cabeçalho não subdividido

- Formado por substantivo
- Constituído por expressão adjetiva
- Constituído por expressão prepositiva
- Constituído por expressão conjuntiva
- Cabeçalho com qualificador
- Termos compostos

Formas do cabeçalho subdividido

- Subdivisões de assunto
- Subdivisões geográficas
- Subdivisões cronológicas
- Subdivisões de forma
- Formas de pré-coordenação

Tab. SIP3 - Terminologia

- Termos sinónimos
- Termos homónimos
- Singular
- Plural
- Qualificadores parentéticos
- Terminologia de NOMES GEOGRÁFICOS

Terminologia de OUTROS NOMES PRÓPRIOS

Nomes de edifícios e outras estruturas fixas

Nomes de famílias e dinastias

Personagens, figuras lendárias e mitológicas

Acontecimentos

ASSUNTOS ESPECIAIS

Literatura

Cultura

Civilização

História

Filosofia

Entidades geográficas

Obras sobre artes visuais

Obras religiosas e sobre religião

DOCUMENTOS ESPECIAIS

Obras de referência

Obras para crianças e jovens

Manuscritos e Livro antigo

Manuais de ensino

Documentos visuais gráficos

Publicações em série

Tab.SIP1 – POLÍTICA DE INDEXAÇÃO

A) Análise do documento

Analisar o documento para definição do seu conteúdo, identificação dos conceitos que expressem cada assunto e seleção dos que forem considerados pertinentes para representação em linguagem documental.

B) Identificação do número de assuntos

- Identificar o número de assuntos com base na sumarização do conteúdo da obra, efetuada em conformidade com as diretrizes da NP 3715 (ISO 5963);
- Se forem identificados mais de quatro assuntos, proceder à análise dos mesmos com vista à sintetização do conteúdo da obra. Para isso, verificar se existe entre os assuntos sobreposição semântica.
- Se existir essa sobreposição, decidir sobre a possibilidade de agrupar os assuntos específicos em assuntos mais gerais, desde que a síntese resultante não produza assuntos em número superior a quatro, e desde que não haja prejuízo para a cobertura total do conteúdo da obra.

C) Identificação do(s) conceito(s) que expressa(m) cada assunto

Assunto simples = expresso por um só conceito

Assunto complexo = expresso por mais do que um conceito

Num assunto complexo categorizar os conceitos em:

- Assunto básico – conceito que designa o próprio assunto, subordinante dos conceitos que designam aspetos ou facetas (conceito sem o qual o assunto não pode ser definido; conceito que designa o elemento que sofre influência)
- Aspetos ou facetas – conceitos subordinados do assunto básico (conceitos que são fatores ocasionais relativamente ao assunto básico)

D) Seleção dos conceitos

1. Selecionar os conceitos relevantes para a descrição de cada assunto de acordo com os seguintes critérios de pertinência:

Pertinentes: os que constituem limitações intencionais ou evidentes do autor na sua abordagem ao assunto, e desde que tenham, na obra em causa, expressão significativa relativamente ao assunto básico;

Não pertinentes: aos que não constituem parâmetros de limitação evidentes, surjam como informações pontuais, com funções ilustrativas ou acessórias relativamente ao assunto básico.

E) Representação

2. Representar em linguagem documental os conceitos seleccionados pela análise como pertinentes para a descrição do assunto, observando os seguintes princípios gerais de indexação:

Princípio terminológico – o(s) conceito(s) que representa(m) um assunto deve(m) ser expresso(s) por termo(s) conforme(s) às instruções de terminologia (Cf.4)

Princípio da entrada única – cada assunto deve ser representado por um só cabeçalho coextensivo com o conteúdo temático a que se aplica.

Princípio da entrada directa – a entrada de assunto deve ser constituída por um cabeçalho em que o assunto é dado directamente pelo termo que o representa, não recorrendo à interposição de termo(s) mais genérico(s);

Princípio da entrada específica – a entrada de assunto deve ser constituída por um cabeçalho específico, isto é, cujo sentido não seja nem mais lato nem mais restrito que o conteúdo temático a que se aplica. (ver Tab. 2 - Cabeçalhos);

Princípio de pré-coordenação – um assunto complexo, isto é, não suscetível de ser representado apenas por um termo de indexação, deve ser representado através da combinação dos termos correspondentes aos vários conceitos.

Ver ‘Termos compostos’ e ‘Formas de pré-coordenação’ na tab.SIP2 CABEÇALHOS, seguinte.

Tab.SIP2 – CABEÇALHOS¹

Observar o **Princípio do Cabeçalho específico**: Cada assunto deve ser representado por um cabeçalho cujo sentido não seja nem mais lato nem mais restrito que esse assunto.

- ✓ Pode implicar o **princípio da pré-coordenação**: Um assunto não suscetível de ser representado apenas por um termo de indexação, deve ser representado através da combinação dos termos correspondentes aos vários conceitos. *Ver adiante* ‘Cabeçalho subdividido’.

A) Cabeçalho não subdividido

A forma mais simples do cabeçalho não subdividido é o formado por um substantivo (o assunto é expresso por um conceito, representado por um só substantivo ou locução nominal).

Além da forma simples, os cabeçalhos não subdivididos podem ainda ter as seguintes formas:

- Constituído por **expressão adjetiva** (um conceito não passível de ser representado pelo substantivo sem o modificador – substantivo modificado por adjetivo).
- Constituído por **expressão prepositiva** (um conceito não passível de ser representado por um só substantivo, ou assunto expresso por uma relação de fase entre conceitos distintos):
 - Relações de Exposição - Cabeçalhos constituídos por expressões contendo a **preposição ‘em’** (assunto constituído por dois conceitos em relação de fase que expresse a abordagem ou representação dum tema na literatura, na arte, na mitologia, nos meios de comunicação social, ou que designe todas as mulheres ou homens associados a uma disciplina ou atividade profissional, ou não profissional). (ex.: Mulheres na medicina).
 - Relações de Direção (audiência) - Cabeçalhos constituídos por expressões contendo a **preposição ‘para’** (assunto constituído por dois conceitos em relação de fase, em que o assunto é abordado de forma orientada para um tipo de audiência com a qual normalmente não está associado, excetuando os casos de audiência infantil e juvenil). (ex.: Informática para bibliotecários).
- Constituído por **expressão conjuntiva** (um assunto expresso por uma relação de fase entre conceitos distintos - substantivos ligados por conjunção):

¹ Cabeçalho é o elemento da linguagem documental que representa um assunto; pode ser constituído por um só termo de indexação, se se tratar de um assunto simples (expresso por um só conceito), ou por dois ou mais termos de indexação pré-coordenados, se se tratar de um assunto complexo (expresso por mais do que um conceito). Tem como função, no registo bibliográfico, como elemento de descrição e acesso ao conteúdo do documento catalogado; no catálogo, como elemento de ordenação e acesso por assunto a um conjunto de registos bibliográficos.

Relações de Influência recíproca - Cabeçalhos formados por expressões contendo a **conjunção ‘e’** (quando existe uma relação de influência a partir das perspectivas de ambos os conceitos); estas relações de fase constituem em si assuntos consagrados e não uma relação de conceitos ocasionalmente desenvolvida (ex.: Educação e crime; Igreja e Estado).

Relações de Comparação - Cabeçalhos constituídos por expressões contendo a **conjunção ‘como’** (assunto constituído por conceitos que designem duas classes de pessoas em relação de fase de comparação. Por exemplo, classes de pessoas envolvidas em duas profissões ou atividades (médicos como políticos / professores como escritores); Classes de pessoas profissionais ou não, e classes não relacionadas com trabalho (bibliotecários como automobilistas); Classes normalmente consideradas como não tendo uma profissão e classes profissionais ou de atividade (crianças como colecionadores /deficientes como atletas).

Cabeçalho com qualificador

- Ao qual é acrescentado, entre parênteses curvos, expressão pra a resolver situações de homonímia ou ambiguidade.

3. Termos compostos

- Um termo composto é uma locução nominal (que constitui uma só unidade semântica), na forma gramatical de expressão adjetiva ou prepositiva, que designa um único conceito obtido por interseção de outros. O primeiro e o segundo elementos do termo composto expresso em linguagem natural, designam-se respetivamente o foco e o modificador.
- Para verificação dos termos na linguagem natural que devem ser decompostos na linguagem documental, seguir as orientações seguintes:

✓ **Manter na forma da linguagem natural se:**

- o modificador aplicado ao foco determina uma subclasse do mesmo (ex. aves de rapina)
- se o resultado da decomposição levar a perda de significado ou ambiguidade
- se o termo composto é de uso corrente generalizado na linguagem natural e se prevê que seja pesquisado nessa forma

✓ **Decompor os termos compostos se:**

- o foco constitui uma propriedade, qualidade, parte ou material, e o modificador representa o todo ao qual pertence o elemento foco (ex. aparelho digestivo do coelho)
- o foco representa uma ação transitiva – o modificador constitui o elemento que sofre a ação (ex. concursos de fotografia)

B) Cabeçalho Subdividido

- Um assunto constituído por dois ou mais conceitos pertencentes a duas categorias semânticas é representado pelo conceito que define o próprio assunto (assunto básico), e pelos conceitos que constituem aspetos ou facetas desse assunto.
- Um cabeçalho subdividido resulta da pré-coordenação de um cabeçalho, nesta situação designado como cabeçalho principal, com uma ou mais subdivisões.
- O cabeçalho principal representa o(s) conceito(s) eleito(s) pela análise para designação do assunto básico e as subdivisões os elementos que se seguem ao cabeçalho principal ou a outra subdivisão, para representar o(s) conceito(s) definido(s) como aspeto(s) ou faceta(s) do assunto básico.
- Ordem da apresentação das subdivisões (quando ocorrerem os vários tipos):
cabeçalho principal -- subdivisão de assunto -- subdivisão geográfica -- subdivisão cronológica -- subdivisão de forma

Formas do Cabeçalho subdividido

- Subdivisões de assunto: representam os aspetos de limitação temática do assunto básico
- Subdivisões geográficas: representam os aspetos de espaço a que se limita o assunto básico; Designam a entidade ou área limitadora e são expressas na mesma forma que for estabelecida para uso como cabeçalho de nome
- Subdivisões cronológicas: representam os aspetos de tempo a que se limita o assunto básico; Usam formas numéricas ou combinações de termos e datas, prevista em instrução própria
- Subdivisões de forma: representam a forma física duma obra e/ou forma bibliográfica segundo a qual se organiza o seu conteúdo

Formas de pré-coordenação:

- Representar em cabeçalho principal: os conceitos considerados como designativos do assunto básico, e os conceitos que designam o elemento que sofre a influência, quando se trata de uma relação de fase de influência não recíproca entre conceitos.

- Representar em subdivisões: os conceitos considerados como aspetos ou facetas do assunto básico e os conceitos que designam o agente que exerce influência quando se trata de uma relação de fase não recíproca entre conceitos (ex. Produção agrícola – Efeitos do clima).

Tab.SIP3 – TERMINOLOGIA

A) Terminologia (forma gramatical)

Termos sinónimos²: usar o termo...

- Que for mais familiar ao utilizador [ex. Catálogos de autores UP Catálogos onomásticos]
- Que for menos polissémico [ex. Disciplinas UP Cadeiras (Ensino)]
- Que for mais generalizado na língua portuguesa; [ex. Casamento UP Matrimónio]
- Na forma de substantivo simples [ex. Ética UP Filosofia moral]

Termos homónimos³:

- Aplicar qualificador parentético que dê um significado unívoco e que constitui parte integrante do termo a que se aplica [ex. Alfaiates (insetos)]

Singular: usar para...

- Noções abstratas como fenómenos, propriedades, disciplinas, etc. [ex. Elasticidade, Filosofia]
- Entidades concretas não contáveis (quanto?) [ex. Cobre, Dinheiro, Papel]
- Partes do corpo (quando exista apenas uma) [ex. Boca, Cabeça, Nariz]

Plural: usar para...

- Noções abstratas que designem uma classe [ex. Ciências médicas, Reações químicas]
- Entidades concretas contáveis (quantos?) [ex. Laranjas, Mesas]

Nota: Quando as formas do singular e plural do mesmo termo designam noções diferentes, podem ser ambas usadas [ex. Teatro (atividade); Teatros (instituições)]

Qualificadores parentéticos: Palavra ou expressão, colocada entre parênteses curvos,

² Sinónimos: termos diferentes com o mesmo significado

³ Homónimos: termos com a mesma grafia, com mais do que um significado (polissémicos)

acrescentada a um termo de indexação com a função de esclarecer ou especificar o sentido.

Formular qualificadores nas situações em que se verifique necessidade de:

- distinguir homónimos; esclarecer termo obscuro; limitar o sentido de um termo de modo a torna-lo mais específico.

Limitações ao uso de qualificadores – não acrescentar qualificador a um termo usado na linguagem documental no seu sentido mais corrente, mesmo que na linguagem natural tenha outros significados.

B) Terminologia de NOMES GEOGRÁFICOS (*seção 48 do Manual SIPORbase*)

Nomes geográficos jurisdicionais

São nomes próprios de entidades que constituem áreas de competência administrativa (países, estados, províncias, distritos, concelhos, cidades, vilas, ruas, etc.)

Exemplos:

Distrito da Guarda (Portugal)

Paris (França)

Avenida de Roma (Lisboa, Portugal)

Nomes de áreas associados com cidades

Exemplos:

Área metropolitana de Lisboa (Portugal)

Área suburbana de Coimbra (Portugal)

Região de Aveiro (Portugal)

Nota: usar em cabeçalho principal, quando constitui assunto básico, em subdivisão geográfica, quando constitui aspeto geográfico, ou como qualificador geográfico; usar em qualificador apenas nomes de entidades geográficas jurisdicionais de nível superior, em relação à entidade que qualificam.

Nomes geográficos não jurisdicionais

São nomes próprios de entidades naturais (desertos, grutas, montanhas, lagos, rios, mares, reservas naturais, etc.); Incluir sempre que possível o termo que designa o tipo de entidade

Exemplos:

Bosque de Bolonha (Paris, França)

Baía de Cascais (Portugal)

Serra de São Mamede (Distrito de Portalegre, Portugal)

Nomes de entidades jurisdicionais passadas

(cidades, vilas que deixaram de existir, impérios, etc.)

Exemplos:

Conímbriga (Portugal: loc. ant.)

Império Carolíngio, 800-843

Nomes de rios e áreas associadas com rios

(estuários, vales, bacias hidrográficas, etc.)

Exemplos:

Rio Mondego (Portugal)

Rio Guadiana (Espanha e Portugal)

Rio Amazonas

Nota: não qualificar rios situados em mais do que dois países.

Nomes de ilhas e arquipélagos

(se a ilha ou arquipélago pertencer a mais do que dois países, não incluir qualificador)

Exemplos:

Ilha de São Miguel (Portugal)

Ilha de Chipre

Nomes de regiões geográficas

(área que se define pela zona de influência da entidade básica que lhe dá o nome.)

Exemplos:

Região mediterrânica

Região da Serra da Estrela (Portugal)

C) Terminologia de OUTROS NOMES PRÓPRIOS (*seção 49 do Manual SIPORbase*)

Nomes de edifícios e outras estruturas fixas

(edifícios, monumentos, pontes, barragens, estradas, jardins, ruínas)

Exemplos:

Arco do Triunfo (Paris, França)

Palácio de Versalhes (França)

Barragem de Belver (Concelho de Gavião, Portugal);

Templo de Diana (Évora, Portugal)

Nota: nas entidades portuguesas qualificar pela jurisdição de nível inferior a país.

Nomes de família e dinastias

Exemplos:

Bragança, Casa de

Palmela, Duques de

Melo, Família

Dinastia de Avis, 1385-1580 (Portugal)

Personagens, figuras lendárias e mitológicas

(nomes de personagens de obras literárias ou artísticas.) Deve-se acrescentar ao nome o qualificador.

Exemplos:

Gato das Botas (Personagens)

Galo de Barcelos (Figuras lendárias)

Diana (Figuras mitológicas)

Afonso Henriques, Rei de Portugal, 1111-1185 (Personagens)

Nota: se uma pessoa real é personagem de uma obra de ficção com o seu nome, estabelece-se o cabeçalho de acordo com as regras para cabeçalhos de nome, acrescentando-se o qualificador.

Acontecimentos

(catástrofes, eventos, festividades, comemorações, etc.)

Exemplos:

Terramoto de 1755 (Lisboa, Portugal)

Jogos Olímpicos, 20, Munique, 1972

Marchas Populares de Lisboa (Portugal)

25 de Abril, 1974 (Portugal)

D) ASSUNTOS ESPECIAIS

Literatura / Cultura / Civilização / História / Filosofia / Entidades geográficas/ Obras sobre artes visuais/ Obras religiosas e sobre religião.

Literatura

- Obras sobre literatura

Literatura portuguesa (*autores de nacionalidade portuguesa*)

Exemplos:

Literatura portuguesa -- Alentejo (Portugal)

Literatura portuguesa -- Idade Média

Literatura de viagens portuguesa

Literatura -- Europa

Literatura de língua portuguesa (*autores que escrevem em português*)

Literatura infantil

Luís, Agustina Bessa, 1922- -- A Sibila--Críticas literárias

- Obras literárias

Há obras literárias sem assunto, que não revelem características temáticas pertinentes para a sua recuperação.

- Obras literárias com assunto: (*ex.*) Péron, Eva, 1919-1952--[Peças de Teatro]

- Obras em géneros não especificamente literários
(biografias, autobiografias, diários, memórias, cartas, ensaios, crónicas, discursos, narrativas de viagens, banda desenhada, entrevistas, etc.)

Exemplos:

Sena, Jorge de, 1919-1978--Correspondência

Portugal—Política--1850-1956--[Cartas]

Escritores--[Biografias]

Portugal—História--[Banda desenhada]

Pomar, Júlio, 1926- -- [Entrevistas]

Cultura – (Ter em conta a seguinte classificação)

(às obras que tratem da discussão teórica do conceito de Cultura, atribuir o cabeçalho Cultura)

- Cultura individual dos povos: (ex.) Cultura portuguesa

- Cultura limitada a um período lato da história

Exemplos:

Cultura portuguesa – Séc. 19-20

Cultura portuguesa – Idade Média

- Formas de cultura

Exemplos:

Cultura científica

Cultura popular

Cultura popular portuguesa – Algarve (Portugal)

- Identidade cultural de classes de pessoas

Exemplos:

Estudantes americanos – Identidade cultural

Minorias étnicas – Identidade cultural

Civilização

(às obras que tratam do conjunto dos elementos económicos, jurídicos, técnicos, culturais, morais e religiosos, suscetíveis de serem aplicados à totalidade da espécie humana, atribuir o cabeçalho *Civilização*; usar o plural *Civilizações* no caso de mais de quatro civilizações desaparecidas, seguido de subdivisão cronológica)

- Civilização na Idade Moderna e/ou Contemporânea

Exemplos:

Civilização--Séc. 16-18

Civilização--Séc. 19-20

- Civilização [modificador adjectivo que designe povo ou raça]

Exemplos:

Civilização Grega
Civilização egípcia
China--História (*e não*) Civilização chinesa
Civilização árabe (*e não*) Civilização islâmica

História

- Obras sobre a história como ciência

Exemplos:

História económica
História social
História natural

- História geral da humanidade (ex.) História mundial--Séc. 20

- História individual dos povos

Exemplos:

Portugal--História
Algarve (Portugal)--História
Brasil--História—Séc. 16-18
Judeus--História

- Obras sobre períodos latos da história

Exemplos:

História mundial--Séc. 16-18
Paleolítico
Portugal--Pré-história
Portugal--História--Idade Média

- Assuntos abordados numa perspetiva histórico-descritiva

Exemplos:

Agricultura--História
Portugal--Economia--História
Arte--História

Filosofia

(às obras que tratam da filosofia em geral atribuir o cabeçalho Filosofia subdividido ou não cronologicamente)

- Obras sobre filosofia em geral

Exemplos:

Filosofia alemã
Filosofia cristã
Filosofia ocidental

- Divisões da filosofia

Exemplos:

Metafísica
Filosofia do conhecimento
Ética

- Ramos da filosofia

Exemplos:

Filosofia do espírito
Filosofia da natureza
Filosofia da linguagem

- Obras sobre doutrinas e correntes filosóficas

Exemplos:

Marxismo
Existencialismo
Filosofia transcendental

- Obras sobre conceitos filosóficos ou outros conceitos abordados no âmbito da filosofia

Exemplos:

Dialética
Razão

- Obras sobre filósofos

Exemplos:

Sócrates--Filosofia
Aristóteles--Lógica
Sartre, Jean-Paul--Existencialismo
Dialéctica--Ponto de vista de Hegel
Hegel--[Biografias]

Entidades geográficas

- Entidades geográficas jurisdicionais como assunto básico

Exemplos:

Portugal--1990-1999

Minho (Portugal)

Europa--Séc. 20

- Assuntos modificados pelo fator de nacionalidade

Exemplos:

Arte francesa

Emigrantes portugueses

Literatura italiana

Cultura portuguesa

- Entidades geográficas como subdivisão geográfica

Exemplos:

História--Portugal

Agricultura--Portugal

Grupos marginais--Lisboa (Portugal)

- Entidades fisiogeográficas

Exemplos:

Serra da Estrela (Portugal)--Flora

Rio Tejo (Espanha e Portugal)--Fauna

Obras sobre artes visuais – *(ter em conta a seguinte classificação)*

- Conjuntos de produção artística, de acordo com as várias formas de arte (Pintura; Escultura; Gravura; Artes Gráficas)

Exemplos:

Lisboa (Portugal)--Fotografias aéreas

Catarina de Bragança, Rainha de Inglaterra, 1662-1685—Gravuras

Obras religiosas e sobre religião

...

E) DOCUMENTOS ESPECIAIS *(seção 9 do SIPORbase)*

Obras de referência / Obras para crianças e jovens / Manuais de ensino

Obras de referência

Dicionários / Glossários / Enciclopédias / Bibliografias / Catálogos / Inventários / Repertórios / Anuários / Almanques / Guias / Terminologias / Índices

Exemplos:

Língua inglesa--[Dicionários]

Arte--[Enciclopédias]

Ciência--[Enciclopédias juvenis]
Literatura--[Bibliografias]
Pintura--Exposições--[Catálogos]
Património mundial--[Inventários]
Distrito da Guarda (Portugal)--[Anuários]
Vestígios arqueológicos--Portugal--[Guias]

Obras para crianças e jovens

- Obras sobre publicações infantis e juvenis

Exemplos:

Enciclopédias juvenis
Física--Publicações juvenis
Vieira, Alice, 1943- --Romances infantis
Andersen, Hans Christian, 1805-1875--Literatura infantil

- Publicações infantis ou juvenis

(com valor literário: a mesma política que está estabelecida para as obras literárias em geral)

(sem valor literário: com assunto – atribuir o cabeçalho adequado à representação do assunto seguido da subdivisão de forma –[Publicações juvenis] ou –[Publicações infantis]; as obras sem assunto pertinente para recuperação, não são indexadas)

Manuscritos e livro antigo

(devem ser analisados com vista a determinar se possuem assunto pertinente para efeitos de recuperação; A determinação de assunto exige ao indexador o conhecimento apurado das coleções em que se inserem e o recurso a obras de referência a fim de situar os documentos no seu contexto e de distinguir o que é essencial do que é acessório no seu conteúdo).

Manuais de ensino

(obra que contém uma matéria sistematizada de acordo com um programa curricular e que se destina a acompanhar o ensino dessa matéria).

Exemplos:

Matemática--[Manuais de ensino básico]
Língua portuguesa--Gramática--[Exercícios]
Língua francesa--[Testes de avaliação]
Direito administrativo--[Manuais de ensino superior]
Língua inglesa--[Manuais de ensino]

Documentos visuais gráficos

(representações visuais a duas dimensões, geralmente sobre papel e acessíveis a olho nu: gravuras, reproduções, fotografias, postais ilustrados, cartazes, etc.; Os DVG que não revelem assuntos pertinentes para recuperação, não são indexados; Aos que revelem assuntos pertinentes devem ser atribuídos cabeçalhos de assunto conforme as normas gerais).

Publicações em série

(jornais, revistas, boletins, relatórios e planos de atividades de instituições, séries monográficas, etc.)

Exemplos:

Arte--[Periódicos]

Penamacor. Câmara Municipal--Periódicos]

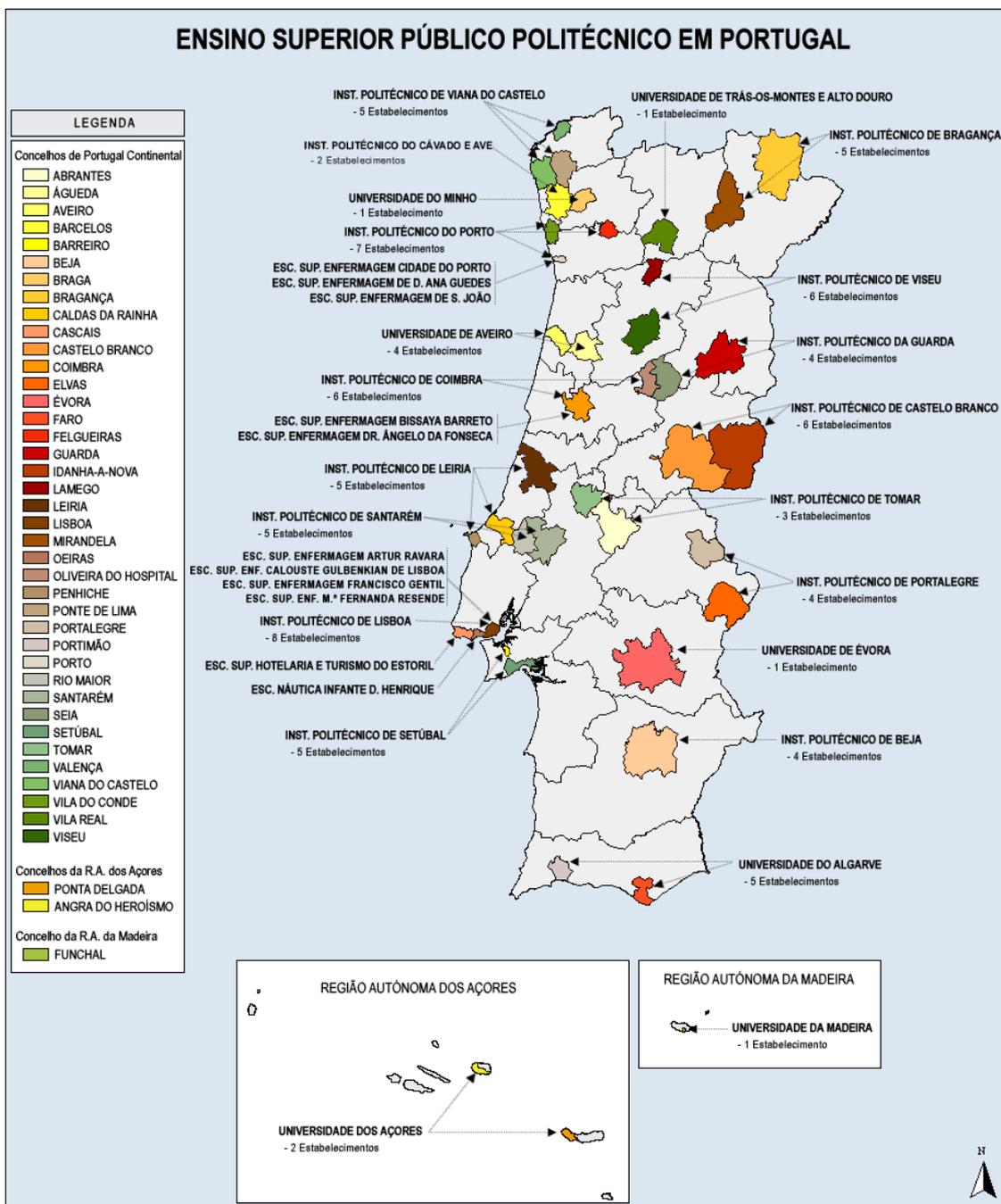
Jovens--[Publicações periódicas juvenis]

Melgaço (Portugal)--[Periódicos]

Educação--[Séries monográficas]

Anexo II. Rede do Ensino Superior Politécnico em Portugal¹

DGES Direcção Geral do Ensino Superior
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR



¹ Mapa da Rede de Ensino Superior Público Politécnico – representação geográfica, <http://www.dges.mctes.pt/DGES/pt/Instituicoes/>

Anexo III. Apresentação do estudo e pedido de resposta a inquérito

Ex.mo(a) Senhor(a)

Sou bibliotecária e estou atualmente a desenvolver uma tese de doutoramento na Universidade de Salamanca, sobre sistemas de indexação em bibliotecas.

No ano de 2009 foi apresentado um inquérito no âmbito de um estudo exploratório, cujas respostas deram ao tema uma contribuição valiosa. Sendo agora este o tema central da tese, é muito importante a colaboração de todas as instituições para o desenvolvimento da investigação.

O inquérito pretende analisar o sistema de indexação utilizado nas bibliotecas do ensino superior politécnico em Portugal, com o objectivo de fundamentar a necessidade de apoio profissional sistematizado nestas tarefas, em relação às quais se sentem tantas dificuldades.

O questionário vai estar disponível cerca de um mês e destina-se a ser respondido pelo responsável da(s) biblioteca(s). As respostas serão confidenciais, e só a conjugação de resultados será tornada pública para fins académicos.

Assim, solicito a amabilidade de resposta ao inquérito através do *link* indicado, e agradeço antecipadamente a atenção dispensada.

Com os melhores cumprimentos,
Maria Catarina Barradas Martins

QUESTIONÁRIO

P. S. Caso tenha alguma dificuldade em responder agradeço a informação através do *e-mail* catarinamartins03@gmail.com.

Anexo IV. Questionário *online*

Sistema de indexação

Este inquérito tem como objectivo conhecer o sistema de indexação nas bibliotecas do ensino politécnico em Portugal, e identificar dificuldades inerentes a esta tarefa.

*Obrigatório

Instituição * (Nome do Instituto)

Unidade Orgânica (Escola)

Designação da Unidade de Informação * (Biblioteca, Centro Documental, etc.)

1 - QUANTAS PESSOAS TRABALHAM NA BIBLIOTECA EM HORÁRIO COMPLETO? * (Total de funcionários administrativos e técnicos)

1.1 - QUANTOS TÉCNICOS PROFISSIONAIS DE BD * (Biblioteca e Documentação)

1.2 - QUANTOS TÉCNICOS SUPERIORES DE BD * (Biblioteca e Documentação)

2 - QUANTOS PROFISSIONAIS DESENVOLVEM TAREFAS DE INDEXAÇÃO? *

2.1 - OS PROFISSIONAIS QUE DESENVOLVEM TAREFAS DE INDEXAÇÃO TÊM FORMAÇÃO NESSA ÁREA? *

SIM ALGUNS NÃO

2.2 - SE A RESPOSTA FOI "sim" OU "alguns" QUAL A FORMAÇÃO MAIS QUALIFICADA? * Formação da Associação BAD / Formação da Biblioteca Nacional / Licenciatura (área Ciências Documentais) / Especialização em Ciências Documentais / Mestrado (área Ciências Documentais) / Doutorado (área Ciências Documentais) / Outra formação

2.2.1- SE A RESPOSTA FOI "outra formação" QUAL?

2.3 - O(S) PROFISSIONAL(AIS) COM FORMAÇÃO NESSA ÁREA DESENVOLVE(M) SÓ ESSA TAREFA? * (Dedicação total ao processo de indexação)
SIM ALGUNS NÃO

3 - NO PROCESSO DE INDEXAÇÃO É USADO ALGUM INSTRUMENTO NORMATIVO? * (Normas / Regras)

SIM NÃO

3.1 - SE A RESPOSTA FOI "sim", QUAL OU QUAIS?

4 - O PROCESSO DE INDEXAÇÃO É FEITO COM AUXÍLIO DE UM LÉXICO DE ASSUNTOS OU DE UM THESAURUS? *

SIM NÃO

4.1 - SE A RESPOSTA FOI "sim" QUAL OU QUAIS?

5 - NO PROCESSO DE INDEXAÇÃO É FEITO O CONTROLO DE AUTORIDADES DE ASSUNTOS? * (Existência de ficheiro de autoridades no software)

SIM NÃO

6 - QUAL O SOFTWARE DE GESTÃO BIBLIOGRÁFICA UTILIZADO? * Porbase
Bibliobase Docbase Horizon Aleph outro

6.1 - SE A RESPOSTA FOI "outro" QUAL?

7 - QUAIS AS MAIORES DIFICULDADES QUE IDENTIFICA NAS TAREFAS DE
INDEXAÇÃO? *

Tecnologia Google Docs

Anexo V. Terminologia dos cabeçalhos de Nomes Comuns.¹

Ocorrência de termos em tesauro(s) e índices BNP (por bases bibliográficas)

Base B1

Termos	Tesauro	Índices BNP
Administração do ensino	Eurovoc	(BN) Ensino -- Gestão
Ambiente físico	Eurovoc	-
Animação	-	(BN) <i>não ocorre isolado</i>
Animação sociocultural	Eurovoc	BN
Anuários	-	(BN) [forma]
Artes e técnicas de expressão	-	-
Atlas	Eurovoc	(BN) Atlas mundial
Avaliação de projecto	Eurovoc	(BN) Avaliação de projectos
Banda desenhada	-	BN
Belas-artes	Eurovoc	(BN) Artes
Bibliografias	-	(BN) [Bibliografias]
Biografia	Eurovoc	(BN) [Biografias]
Carta das Nações Unidas	Eurovoc	(BN) ONU
Ciências da natureza	-	BN
Ciências da terra	Eurovoc	(BN) Ciências da terra e da vida
Ciências da vida	Eurovoc	BN
Ciências físicas	Eurovoc	(BN) Ciências físico-químicas
Ciências sociais	Eurovoc	BN
Comunicação	Eurovoc	BN
Comunicação educativa	-	-
Condição feminina	Eurovoc	(BN) Mulheres – Condições sociais
Desenvolvimento curricular	-	(BN) Currículos -- Desenvolvimento
Diabetes	Eurovoc	BN
Dicionário	-	(BN) [Dicionários]
Dicionário de línguas	-	-
Direitos da criança	Eurovoc	(BN) Crianças -- Direitos
Direitos do indivíduo	Eurovoc	(BN) Direitos do homem
Documentação cinzenta	Eurovoc	-
Doença crónica	Eurovoc	(BN) Doenças --
Educação	Eurovoc	BN
Educação de adultos	Eurovoc	(BN) Adultos -- Educação
Educação permanente	Eurovoc	BN
Educação pré-escolar	Eurovoc	BN
Emprego	Eurovoc	BN
Enciclopédia	Eurovoc	(BN) [Enciclopédias]
Enciclopédias gerais e especializadas	-	-
Ensino	Eurovoc	BN

¹ A terminologia aqui registada à data da pesquisa, pode-se encontrar alterada posteriormente alterada.

Ensino de línguas	Eurovoc	BN
Ensino na área das ciências	-	(BN) Ciências – Ensino/Aprendizagem
Ensino primário	Eurovoc	(BN) Ensino básico primeiro ciclo
Estatística	Eurovoc	BN
Estatísticas	-	(BN) [Dados estatísticos]
Exclusão social	Eurovoc	BN
Experiência pedagógica	Eurovoc	(BN) Experiências pedagógicas
Expressão dramática e teatro	-	-
Ficção	-	-
Filosofia	Eurovoc	BN
Física e química	-	(BN) Física -- Ensino
Floresta	Eurovoc	(BN) Florestas
Formação de professores	Eurovoc	(BN) Professores – Formação profissional
Gestão	Eurovoc	BN
Gestão de empresas	Eurovoc	(BN) Empresas -- Gestão
Gestão e administração escolares	-	-
Grupo sociocultural	Eurovoc	-
Guia turístico	-	(BN) Guias turísticos
História	Eurovoc	BN
História e filosofia da educação	-	-
História medieval	-	(BN) Idade Média
História moderna	Eurovoc	-
História universal	Eurovoc	BN
Histórias gerais e tratados	-	-
Identidade cultural	Eurovoc	BN
Imprensa	Eurovoc	BN
Indústria cinematográfica	Eurovoc	(BN) Cinema
Informática	Eurovoc	BN
Inovação	Eurovoc	BN
Inovação pedagógica	Eurovoc	BN
Investigação	Eurovoc	BN
Investigação em educação	-	BN
Investigação pedagógica	Eurovoc	(BN) Pedagogia --
Jornalismo	Eurovoc	BN
Legislação	Eurovoc	BN
Língua estrangeira	-	-
Língua materna	Eurovoc	BN
Linguística	Eurovoc	BN
Literatura	Eurovoc	BN
Literatura de autores estrangeiros	-	-
Literatura de autores portugueses	-	(BN) Literatura portuguesa
Literatura infantil e juvenil	-	-
Manifestação cultural	Eurovoc	(BN) Manifestações culturais
Manuais escolares	-	(BN) Manuais de ensino
Marca comercial	Eurovoc	(BN) Marcas comerciais

Marketing	Eurovoc	BN
Matemática	Eurovoc	BN
Meteorologia	Eurovoc	BN
Método de aprendizagem	Eurovoc	(BN) Aprendizagem --
Obras de informação	-	-
Organização não governamental	Eurovoc	(BN) ONG (BN) Organizações não governamentais
Pais	-	BN
Participação social	Eurovoc	BN
Património	Eurovoc	BN
Património natural	-	BN
Petróleo	Eurovoc	BN
Pluralismo cultural	Eurovoc	(BN) Multiculturalismo
Pobreza	Eurovoc	BN
Política da educação	Eurovoc	(BN) Política de educação
Política de emprego	-	BN
Política e sociologia da educação	-	-
População	-	BN
Profissional da comunicação	Eurovoc	(BN) Profissionais da informação
Psicologia	Eurovoc	BN
Psicologia da educação	-	BN
Psicologia e educação	-	-
Publicidade	Eurovoc	BN
Recursos naturais	Eurovoc	BN
Regime político	Eurovoc	(BN) Regimes políticos
Relação escola-vida profissional	Eurovoc	(BN) Relação escolas-
Relações internacionais	Eurovoc	BN
Relações públicas	Eurovoc	BN
Saúde pública	Eurovoc	BN
Semiótica	-	BN
Serviço europeu de luta anti-fraude	Eurovoc	-
Serviço social	Eurovoc	BN
Sociologia	Eurovoc	BN
Técnicas de ensino	-	(BN) Ensino/Aprendizagem
Tecnologia	Eurovoc	BN
Teoria e história da literatura	-	(BN) Literatura -- História
Tese de doutoramento	-	(BN) [Teses]
Tese de mestrado	-	(BN) [Teses]
Teses	-	(BN) [Teses]
Teses de doutoramento	-	(BN) [Teses]
Turismo	Eurovoc	BN
Vida profissional	Eurovoc	-

Terminologia dos cabeçalhos – (ocorrência em tesouro e índices BNP)

Base B2

Termos	Tesouro	Índices BNP
Abdome	Decs	-
Aborto	Decs	BN
Acesso aos serviços de saúde	Decs	(BN) Serviços de saúde --
Acidentes de trânsito	Decs	(BN) Acidentes rodoviários
Administração de medicamentos	-	(BN) Medicamentos -- Consumo
Administração financeira de hospitais	-	(BN) Hospitais -- Administração
Adolescente	Decs	(BN) Adolescentes
Agentes antibacterianos	Decs	-
Alcoolismo	Decs	BN
Alfabetização	-	BN
Alimentação	-	BN
Alongamento	-	-
Alta do paciente	Decs	(BN) Doentes --
Alunos com necessidades educativas especiais	-	BN
Amplitude de movimento articular	Decs	-
Análise de conteúdo	-	BN
Análise estatística	Decs	BN
Anatomia	Decs	BN
Anatomia & histologia	-	-
Anatomia humana	-	(BN) Corpo humano -- Anatomia
Animação sociocultural	-	BN
Anticoncepção	Decs	(BN) Contracepção
Anticoncepcionais	Decs	(BN) Contracepção
Aparelho locomotor	-	BN
Aplainamento radicular	Decs	-
Aptidão física	Decs	BN
Asma	Decs	BN
Aspectos bioéticos	-	-
Assédio sexual	Decs	-
Assistência à saúde	Decs	(BN) Assistência médica
Assistência domiciliar	Decs	(BN) Assistência domiciliária
Assistência odontológica	Decs	(BN) Odontologia --
Atenção primária à saúde	Decs	(BN) Cuidados de saúde primários
Atitudes perante a morte	-	(BN) Morte -- Atitudes ...
Atlas	Decs	(BN) Atlas mundial / [Atlas]
Atlas geográfico	-	(BN) Atlas mundial
Avaliação da capacidade de trabalho	Decs	(BN) -- Avaliação
Avaliação de currículos	-	(BN) Currículos -- Avaliação
Badajoz	-	(BN)

Bibliografia	-	(BN) [Bibliografias]
Bioética	Decs	BN
Boca	Decs	BN
Braquiterapia	Decs	BN
Cálculos renais	Decs	(BN) Doenças dos rins
Cárie dentária	Decs	BN
Casas de saúde	Decs	(BN) Casa de Saúde (<i>nome</i>)
Catatonía	Decs	-
Censos	Decs	BN
Cicatrização de feridas	-	(BN) Feridas – Cuidados de enfermagem
Ciclo menstrual	Decs	BN
Ciências de enfermagem		(BN) Enfermagem
Ciências sociais	Decs	BN
CIPE	-	-
Classificação Decimal Universal	-	BN
Classificação internacional	-	(BN) Classificação internacional de cuidados primários
Cobertura de serviços de saúde	Decs	(BN) Serviços de saúde -- Gestão
Cocaína	Decs	BN
Cocaína crack	Decs	-
Códigos de ética	Decs	(BN) Ética
Coeficiente de mortalidade	Decs	(BN) Taxa de mortalidade
Colo do útero	Decs	-
Competência profissional	Decs	BN
Complicações	-	-
Complicações e acidentes	Decs	-
Complicações na gravidez	Decs	(BN) Gravidez --
Comunicação interpessoal	-	BN
Conceito de si	-	BN
Condução de veículo	Decs	(BN) Automóveis
Conduta de saúde	Decs	-
Congressos	-	BN
Consumo	-	BN
Consumo de bebidas alcoólicas	Decs	(BN) Bebidas alcoólicas -- Consumo
Contraceção	-	BN
Controlo da dor	-	-
Cuidados continuados	-	(BN) Cuidados continuados de saúde
Cuidados de enfermagem	Decs	BN
Cuidados paliativos	Decs	BN
Custos de cuidados de saúde	Decs	(BN) Cuidados de saúde --
Dente	Decs	(BN) Dentes
Deontologia	-	BN
Depósitos dentários	-	-
Depressão pós-parto	Decs	(BN) Depressão pós-natal
Desenvolvimento de programas	-	-
Desenvolvimento fetal	Decs	(BN) Fetos -- Crescimento

Desenvolvimento infantil	Decs	(BN) Crianças -- Desenvolvimento
Desenvolvimento ósseo	Decs	-
Diabete mellitus tipo 1	Decs	(BN) Diabetes mellitus
Diabetes mellitus	Decs	BN
Diagnóstico da situação	-	(BN) Diagnóstico médico
Diagnóstico e tratamento	-	(BN) Diagnóstico médico
Diagnóstico pré-natal	Decs	BN
Diarreia infantil	Decs	(BN) Crianças -- Doenças
Dicionário	-	(BN) [Dicionários]
Dicionários de línguas	-	-
Dieta	Decs	BN
Direito	-	BN
Direito biomédico	-	-
Direitos reprodutivos	Decs	-
Dissertações académicas	Decs	BN
Distocia	Decs	-
Distúrbios alimentares	-	BN
Distúrbios cardiovasculares	-	(BN) Cardiologia – (BN) Doenças cardiovasculares
Distúrbios endócrinos	-	(BN) Endocrinologia
Distúrbios gastrointestinais	-	-
Distúrbios musculoesqueléticos	-	-
Distúrbios neurológicos	-	(BN) Neurologia --
Distúrbios psiquiátricos	-	(BN) Doenças mentais --
Distúrbios respiratórios	-	(BN) Aparelho respiratório --
Doença crónica	Decs	(BN) Doenças --
Doença de Alzheimer	Decs	BN
Doença pulmonar obstrutiva crónica	Decs	BN
Doenças da boca	Decs	-
Doenças dos genitais femininos	Decs	-
Doenças maxilomandibulares	Decs	
Doenças periodontais	Decs	-
Doenças renais	-	(BN) Doenças dos rins
Doenças respiratórias	Decs	BN
Dor pélvica	Decs	-
Drogas	-	BN
Drogas ilícitas	Decs	(BN) Drogas
Educação	Decs	BN
Educação cívica	-	(BN) Educação para a cidadania
Educação de adultos	-	(BN) Adultos -- Educação
Educação de pacientes como assunto	-	-
Educação em enfermagem	Decs	(BN) Enfermagem – Ensino/Aprendizagem
Educação em saúde	Decs	-
Educação em saúde oral	-	(BN) Saúde oral
Educação permanente	-	BN

Educação sanitária	-	-
Educação superior	Decs	-
Electrofisiologia cardíaca	-	(BN) Doenças cardíacas
Enfermagem	Decs	BN
Enfermagem de reabilitação	-	(BN) Enfermagem de recuperação
Enfermagem obstétrica	Decs	BN
Engenharia humana	Decs	(BN) Ergonomia
Ensino superior	-	BN
Envelhecimento	Decs	BN
Epistemologia	-	BN
Equidade	Decs	-
Equilíbrio ácido-básico	-	-
Equilíbrio hidroelectrolítico	-	-
Ergonomia	-	BN
Erros médicos	Decs	(BN) Erros em medicina
Escola de enfermagem	-	-
Escola superior de enfermagem de Artur Ravara	-	-
Escolaridade obrigatória	-	(BN) Educação escolar obrigatória
Escultura	-	BN
Espanha	Decs	BN
Estatística	Decs	(BN)
Estatística & dados numéricos	-	-
Estatísticas de saúde	-	(BN) -- [Dados estatísticos]
Estatísticas demográficas	-	(BN) Demografia --
Estratégias da aprendizagem	-	(BN) Aprendizagem -- Estratégias
Ética	Decs	BN
Etiologia	-	-
Exame	-	-
Exercício	Decs	(BN) Actividades físicas
Exercício físico	-	(BN) Actividades físicas
Exercícios	-	(BN) Actividade físicas
Factores de risco	-	-
Família	Decs	(BN) Famílias
Fármacos	-	(BN) Medicamentos
Fase conceptual	-	-
Fase empírica	-	-
Fase metodológica	-	-
FEDER	-	-
Fibrose quística	-	BN
Financiamento da educação	-	BN
Fisiologia	Decs	BN
Fisiopatologia	-	BN
Fisioterapia	Decs	BN
Fluoretos tópicos	Decs	-
Formação de professores	-	(BN) Professores – Formação profissional
Formação profissional	-	BN

FSE	-	(BN) Fundo Social Europeu
Funcionário público	-	(BN) Funcionários públicos
Fundo de coesão	-	-
Fundos estruturais	-	(BN) Comunidade Europeia. CEE – Fundos estruturais
Garantia da qualidade dos cuidados de saúde	-	(BN) Cuidados de saúde -- Qualidade
Genética	Decs	BN
Genoma humano	Decs	-
Geriatría	Decs	BN
Gestão de empresas	-	(BN) Empresas -- Gestão
Gestão de stocks	-	BN
Ginecologia	Decs	BN
Glomerulonefrite	Decs	-
Gramática	-	(BN) [Gramáticas]
Gravidez	Decs	BN
Grupos de risco	Decs	-
Hábitos alimentares	Decs	BN
Hematologia pediátrica	-	(BN) Pediatria --
Higiene oral	-	BN
Higienistas dentários	Decs	-
Hipertensão renovascular	Decs	-
Histerectomia	Decs	-
Histologia	Decs	-
História	Decs	BN
História da enfermagem	Decs	(BN) Enfermagem -- História
Hospitais	Decs	BN
Hospitais psiquiátricos	Decs	BN
Hospitalização	Decs	BN
Humanização	-	BN
Idoso	Decs	(BN) Idosos
Impostos	Decs	BN
Infecções bacterianas	Decs	BN
Infecções urinárias	Decs	-
Infertilidade	Decs	BN
Inquéritos	-	BN
Instrumentação	-	BN
Insuficiência renal aguda	-	(BN) Insuficiência renal
Insuficiência renal crónica	Decs	BN
Internamento hospitalar	-	BN
Intubação gastrointestinal	Decs	-
Investigação	-	BN
Investigação científica	-	(BN) Investigação
Investigação em ciências de enfermagem	-	(BN) In vestigação em ciências da saúde
Investigação em educação	-	BN
Investigação qualitativa	-	BN
Investigação-acção	-	-
Legislação	Decs	BN
Legislação & jurisprudência	-	-
Liderança	Decs	BN

Língua inglesa	-	BN
Lisboa e Vale do tejo	-	(BN) Lisboa (Portugal)
Listas de espera	Decs	-
Literatura de autores portugueses	-	(BN) Literatura portuguesa
Logística	Decs	(BN) Logística empresarial
Manipulações musculoesqueléticas	Decs	-
Manuais	Decs	(BN) [Manuais]
Marca-passo artificial	Decs	-
Marketing	Decs	BN
Materiais dentários	Decs	-
Medicamentos	-	BN
Medição da dor	Decs	-
Medicina de emergência	Decs	BN
Medicina neonatal	-	(BN) Crianças – Cuidados de saúde
Medicina oral	-	-
Meia-idade	Decs	-
Menopausa	Decs	BN
Métodos	Decs	BN
Métodos pedagógicos	-	BN
Microbiologia	Decs	BN
Mortalidade	Decs	BN
Morte	Decs	BN
Morte súbita	Decs	-
Mulheres	Decs	BN
Nas infecções	-	(BN) Infecções
Nefrologia	Decs	BN
Nefropatias	Decs	BN
Neonatologia	Decs	BN
Novas tecnologias	-	(BN) Tecnologias da informação e comunicação
Número de leitos em hospital	Decs	-
Nutrição	-	BN
Nutrição da criança	Decs	(BN) Crianças -- Alimentação
Nutrição do adolescente	Decs	(BN) Adolescentes -- Alimentação
Nutrição do lactente	Decs	(BN) Nutrição
Nutrição em saúde pública	Decs	(BN) Saúde pública --
Nutrição enteral	Decs	(BN) Nutrição entérica
Nutrição pré-natal	Decs	(BN) Nutrição
Obesidade	Decs	BN
Obstetrícia	Decs	BN
Oclusão dentária	Decs	BN
Odontologia	Decs	BN
Odontologia preventiva	Decs	(BN) Odontologia
Odontopatias	Decs	-
Oligoelementos	Decs	-
Oncologia	Decs	BN
Organização & administração	-	-

Organização e administração	Decs	-
Paciente	-	(BN) Doentes
Parto	Decs	BN
Patologia	Decs	BN
Pedagogia	-	BN
Pediatria	Decs	BN
Pessoas mentalmente doentes	Decs	(BN) Doentes mentais
Piperacilina	Decs	-
Placa dentária	Decs	-
Placenta	Decs	-
Planeamento estratégico	-	BN
Planeamento familiar	-	BN
Política comunitária	-	(BN) União Europeia -- Política
Política de saúde	Decs	BN
Política educativa	-	(BN) Política de educação
Política governamental	-	BN
Portalegre	-	(BN) Portalegre (Portugal)
Portugal	Decs	BN
Pré-operatória	-	-
Prestação de cuidados de saúde	-	-
Prevenção	Decs	(BN) -- Prevenção
Prevenção & controle	-	-
Prevenção de acidentes	Decs	(BN) Acidentes -- Prevenção
Previsão	-	(BN) <i>não ocorre isolado</i>
Profilaxia dentária	Decs	(BN) Terapêuticas
Profissionais de saúde	-	BN
Profissional de saúde	-	(BN) Profissionais de saúde
Programas de educação	-	(BN) Programas de ensino
Prolapso uterino	Decs	-
Promoção da saúde	Decs	(BN) Educação para a saúde
Psicologia	Decs	BN
Puerpério	-	BN
Qualidade do ensino	-	(BN) Ensino -- Qualidade
Quimioterapia	Decs	BN
Radioterapia	Decs	BN
Reabilitação	Decs	BN
Recém-nascido	Decs	(BN) Recém- nascidos
Recenseamento	-	(BN) Recenseamento da população (BN) Recenseamento eleitoral
Recursos humanos	Decs	BN
Redução da gravidez multifetal	-	-
Referência e consulta	Decs	-
Região Alentejo	-	(BN) Alentejo (Portugal)
Região Algarve	-	(BN) Algarve (Portugal)
Região Centro	-	(BN) Portugal (Centro)
Região Norte	-	(BN) Portugal (Norte)
Regime jurídico	-	BN
Regulamento CE	-	(BN) Comunidade Europeia. CEE
Relação com o doente	-	(BN) Relação enfermeiros-

		doentes
Relação profissionais de saúde- pessoas de luto	-	BN
Relações enfermeiro-paciente	Decs	(BN) Relação enfermeiros- doentes
Relações interpessoais	Decs	BN
Reprodução medicamente assistida	-	BN
Resolução de problemas	Decs	BN
Risco	Decs	(BN) <i>não ocorre isolado</i>
Saúde	Decs	BN
Saúde da mulher	Decs	(BN) Mulheres -- Saúde
Saúde do trabalhador	Decs	(BN) Trabalhadores --
Saúde escolar	Decs	BN
Saúde oral	-	BN
Saúde pública	Decs	BN
Saúde reprodutiva	Decs	(BN) Reprodução humana
Serviço hospitalar de enfermagem	Decs	(BN) Hospitais – Serviços de enfermagem
Serviços de assistência domiciliar	-	-
Serviços de saúde	Decs	BN
Serviços de saúde mental	Decs	BN
Sexualidade	Decs	BN
Sida	-	BN
Síndrome de locked-in	-	-
Sistema de saúde	-	(BN) Sistemas de saúde
Sistemas de saúde	Decs	BN
Sucesso escolar	-	BN
Tabaco	Decs	BN
Tabagismo	Decs	BN
Tabelas	Decs	BN Tabelas ... [Tabelas]
Teatro	-	BN
Técnicas de exercício e de movimento	Decs	-
Técnicas de fisioterapia	-	(BN) Fisioterapia
Terminologia	Decs	(BN) <i>não ocorre isolado</i>
Teses	-	BN [Teses]
Toxicodependência	-	BN
Trabalho de parto	Decs	(BN) Parto
Transplante renal	-	(BN) Rins -- Transplante
Tratamento	-	(BN) Terapêuticas
Traumatismos da medula espinal	Decs	(BN) Traumatismos físicos
Úlcera da perna	Decs	(BN) Úlceras de pressão
Úlcera por pressão	Decs	(BN) Úlceras de pressão
Úlceras de pressão	-	BN
Universidades	Decs	BN
Uso terapêutico	-	(BN) Terapêuticas
Utilização da injeção intracitoplasmática de	-	-

espermatozóides no ovócito		
Viagens	-	BN

Terminologia dos cabeçalhos – (ocorrência em tesouro e índices BNP)

Base B3

Termos	Tesouro	Índices BNP
Administração local	Eurovoc	BN
Administração pública	Eurovoc	BN
Advogados	-	BN
Alentejo	Eurovoc	(BN) Alentejo (Portugal)
Algarve	Eurovoc	(BN) Algarve (Portugal)
Ambiente	-	BN
Análise de dados	-	-
Análise matemática	-	(BN) vt Análise numérica
Análise numérica	-	-
Anuário	Eurovoc	(BN) Anuários
Aparelho científico	Eurovoc	-
Aparelho de medição	Eurovoc	-
Aparelhos indicadores	-	-
Armazenagem de energia	Eurovoc	(BN) Energia --
Arquitectura	Eurovoc	BN
Arquitectura de interiores	-	BN
Artes do espectáculo	Eurovoc	(BN) Espectáculos (BN) Artes dramáticas
Artes visuais	Eurovoc	BN
Auditoria da qualidade	-	(BN) Qualidade -- Gestão
Avaliação	Eurovoc	BN
Banda desenhada	-	BN
Base de dados	Eurovoc	-
Bioengenharia	-	-
Bioética	Eurovoc	BN
Biofísica	-	BN
Biologia humana	-	BN
Biomaterial	Eurovoc	-
Biomecânica	-	BN
Biometria	Eurovoc	BN
Bioquímica	Eurovoc	BN
Biotecnologia	Eurovoc	BN
CAE Classificação das actividades económicas	-	(BN) Actividades económicas --
Cães	-	BN
Cálculo financeiro	-	BN
Catálogo em linha	-	BN

Cerâmica técnica	Eurovoc	(BN) Cerâmica
Ciências biológicas	-	(BN) Biologia
Ciências da educação	-	BN
Ciências médicas	Eurovoc	BN
Cinema de animação	-	(BN) Cinema --
Circuitos eléctricos	-	BN
Classificação Decimal Universal	-	BN
Classificação económica	-	-
Cliente	-	(BN) Clientes
Clientela	Eurovoc	(BN) Clientes
Climatização	Eurovoc	-
Código do trabalho	Eurovoc	(BN) Direito do trabalho
Competitividade	Eurovoc	BN
Comunicação de marketing	-	(BN) Marketing --
Comunicação visual	-	BN
Condição feminina	Eurovoc	(BN) Mulheres – Condições sociais
Congresso	-	(BN) Congressos
Consolidação de contas	-	BN
Consumidor	Eurovoc	(BN) Consumidores -- Educação
Contabilidade	Eurovoc	BN
Contrato de trabalho	Eurovoc	(BN) Contratos individuais de trabalho
Controlo	-	BN
Controlo de gestão	Eurovoc	-
Cor	-	(BN) Cores
Correspondência	Eurovoc	BN
Crítica literária	-	BN
Cultura	Eurovoc	BN
Cultura organizacional	Eurovoc	BN
Cultura popular	Eurovoc	(BN) Cultura popular portuguesa
Desenho (Metodologia do design)	-	(BN) Desenho
Desenvolvimento pessoal	Eurovoc	(BN) Desenvolvimento individual
Desenvolvimento psicomotor	-	BN
Desenvolvimento regional	Eurovoc	BN
Desenvolvimento rural	Eurovoc	BN
Desenvolvimento social	Eurovoc	BN
Design	-	BN
Design de embalagens	-	-
Design de engenharia	-	-
Design de interiores	-	-
Design de publicidade	-	-
Design gráfico	-	BN
Design industrial	-	BN
Designers	-	BN
Desktop publishing	-	-
Diagnóstico médico	Eurovoc	BN
Dicionário	-	(BN) [Dicionários]
Didáctica	-	BN
Difusão da cultura	Eurovoc	(BN) Cultura -- Difusão

Direcção de empresa	Eurovoc	(BN) Empresas -- Administração
Direito bancário	Eurovoc	(BN) Bancos (Instituições) --
Direito civil	Eurovoc	BN
Direito do trabalho	Eurovoc	BN
Direito fiscal	Eurovoc	BN
Doença crónica	Eurovoc	-
Eco-design	-	-
Ecologia	Eurovoc	BN
Economia	Eurovoc	BN
Economia de energia	Eurovoc	(BN) Energia eléctrica -- Consumo
Economia internacional	Eurovoc	(BN) Economia mundial
Economia regional	Eurovoc	BN
Educação	Eurovoc	BN
Educação ambiental	Eurovoc	BN
Educação artística	Eurovoc	(BN) Formação artística
Energia	-	BN
Energia não poluente	Eurovoc	(BN) Energias renováveis
Energia renovável	Eurovoc	(BN) Energias renováveis
Enfermagem	-	BN
Engenharia civil	Eurovoc	BN
Engenharia eléctrica	-	(BN) Engenharia electrónica
Ensino	Eurovoc	BN
Ensino na área das ciências	-	(BN) Ciências – Ensino/Aprendizagem
Ensino oficial	Eurovoc	BN
Ensino primário	Eurovoc	(BN) Ensino básico primeiro ciclo
Ensino superior	Eurovoc	BN
Espanha	Eurovoc	BN
Estática	-	BN
Estatística	Eurovoc	BN
Estatística descritiva	-	BN
Estatísticas	-	(BN) – [Dados estatísticos]
Estatísticas das pescas	Eurovoc	(BN) – [Dados estatísticos]
Estrutura das organizações	-	(BN) Organizações -- Estrutura
Ética	Eurovoc	BN
Ética nos negócios	-	(BN) Negócios
Exercícios	-	BN
Férias	Eurovoc	BN
Filosofia	Eurovoc	BN
Finanças locais	Eurovoc	(BN) Concelhos -- Finanças
Fiscalidade	Eurovoc	BN
Física	-	BN
Formação	-	BN
Formação de professores	Eurovoc	(BN) Professores – Formação profissional
Fotografia	-	BN
Fundo de coesão	Eurovoc	-
Fundos estruturais e de coesão	-	(BN) Comunidade Europeia. CEE – Fundos estruturais
Gares marítimas	-	BN
Geologia	Eurovoc	BN

Gestão	Eurovoc	BN
Gestão administrativa	-	-
Gestão da comunicação	-	(BN) Informação -- Gestão
Gestão da imagem	-	-
Gestão das organizações	-	(BN) Organizações -- Gestão
Gestão de empresas	-	(BN) Empresas -- Gestão
Gestão de resíduos	Eurovoc	(BN) Resíduos -- Gestão
Gestão de stocks	-	BN
Gestão do material	Eurovoc	-
Gestão financeira	Eurovoc	(BN) Finaças -- Gestão
Gramática	-	(BN) -- Gramática – [Manuais]
Guia	-	(BN) Guias ...
Hidrogénio	Eurovoc	-
História	Eurovoc	BN
História de Portugal	-	(BN) Portugal -- História
Homossexualidade feminina	-	(BN) Homossexualidade
Homossexualidade masculina	-	(BN) Homossexualidade
Hotelaria	-	(BN) Indústria hoteleira
Iluminação	Eurovoc	BN
Ilustração	-	BN
Imagem corporativa	-	-
Imagens de marca	-	(BN) Imagens de publicidade
Imigração	Eurovoc	BN
Impacto da informática	Eurovoc	(BN) Informática --
Imposto sobre as sociedades	Eurovoc	-
Imposto sobre o rendi. das pess. sing.	Eurovoc	(BN) Imposto sobre o rendimento
Imposto sobre os rendimentos	Eurovoc	(BN) Imposto sobre o rendimento
Indústria automóvel	Eurovoc	BN
Indústria cinematográfica	Eurovoc	(BN) Cinema
Indústria da construção civil	Eurovoc	(BN) Construção civil
Indústria portuguesa	-	(BN) Indústria -- Portugal
Informática aplicada	Eurovoc	-
Informática médica	Eurovoc	-
Instrumentos de medição	-	-
Instrumentos para a medição da temperatura	-	-
Internet	Eurovoc	BN
INTERREG	-	-
Investigação	Eurovoc	BN
Investigação operacional	-	BN
IRS	-	-
Leader	-	(BN) LEADER (Programas da União Europeia)
Língua portuguesa	-	BN
Linguagem de programação	Eurovoc	(BN) Linguagens de programação
Linguística	Eurovoc	BN
Lisboa	-	(BN) Lisboa (Portugal)
Literatura	Eurovoc	BN
Literatura portuguesa	Eurovoc	BN
Logística	-	(BN) <i>não ocorre isolado</i>

Manómetros	-	-
Manuais	-	(BN) [Manuais]
Máquina eléctrica	Eurovoc	(BN) Máquinas eléctricas
Marca comercial	Eurovoc	(BN) Marcas comerciais
Marketing	Eurovoc	BN
Marketing ecológico	-	-
Materiais	-	(BN) <i>não ocorre isolado</i>
Mecânica aplicada	-	BN
Merchandising	Eurovoc	(BN) Merchandising (Marketing)
Método de aprendizagem	Eurovoc	(BN) Aprendizagem --
Método de estudo	-	(BN) Estudo --
Método de investigação	Eurovoc	(BN) Investigação -- Metodologia
Metodologia do design	-	(BN) Design --
Microsoft powerpoint	-	(BN) PowerPoint (Programas de computador)
Monforte	-	(BN) Monforte (Portugal)
Movimento de capitais	Eurovoc	-
Normalização	Eurovoc	BN
Optimização energética	-	-
Organismos de normalização	-	-
Organização de eventos	-	(BN) Eventos -- Organização
Painéis murais	-	(BN) Graes marítimas – Painéis murais
Poder de compra	Eurovoc	BN
Política da educação	Eurovoc	(BN) Política de educação
Política de emprego	-	BN
Política de produção	Eurovoc	(BN) Produção --
Portugal	Eurovoc	BN
Prática da redacção	Eurovoc	(BN) Redacção -- Aprendizagem
Processamento de dados	Eurovoc	(BN) Processamento electrónico de dados
Processo de Bolonha	-	(BN) Ensino superior --
Produção audiovisual	Eurovoc	BN
Professor	Eurovoc	(BN) Professores
Programa comunitário	Eurovoc	(BN) Programas da União Europeia
Protecção do ambiente	Eurovoc	(BN) Ambiente -- Protecção
Psicologia	Eurovoc	BN
Publicidade	Eurovoc	BN
Qualidade	-	BN
Qualidade do ambiente	Eurovoc	(BN) Ambiente -- Qualidade
Química	Eurovoc	BN
Racismo	Eurovoc	BN
Recursos energéticos	Eurovoc	(BN) Energia --
Recursos hídricos	Eurovoc	BN
Rede energética	Eurovoc	(BN) Redes Energéticas Nacionais
Rede informática	Eurovoc	(BN) Redes de computadores
Registadores de temperatura	-	-
Relações públicas	Eurovoc	BN
Relatório	Eurovoc	(BN) [Relatórios]
Responsabilidade social das empresas	Eurovoc	(BN) Empresas – Responsabilidade social

Satisfação do cliente	-	(BN) Clientes -- Satisfação
Séc. 21	-	-
Serviço de interesse geral	Eurovoc	-
Sistema de ensino	Eurovoc	-
Sistemas de sinalização	-	-
Sociedade da informação	Eurovoc	BN
Sociologia	Eurovoc	BN
Sociologia do design	-	-
Software	Eurovoc	BN
Tabelas	-	(BN) [Tabelas]
Tauromaquia	-	BN
Técnica de gestão	Eurovoc	-
Técnico de contas	-	(BN) Contabilidade
Tecnologia da informação	Eurovoc	(BN) Tecnologias da informação e comunicação
Tecnologia química	-	BN
Tempos livres	Eurovoc	BN
Teoria do design	-	BN
Teorias organizacionais	-	-
Termómetros	-	-
Tese de mestrado	-	(BN) [Teses]
Tipografia	-	BN
Trabalho social	Eurovoc	(BN) Serviço social
Tratado de Lisboa	Eurovoc	-
Tratamento da água	Eurovoc	(BN) Água --
Turismo	Eurovoc	BN
União Europeia	Eurovoc	BN
Urbanismo	Eurovoc	BN
Utilizador informático	Eurovoc	(BN) Informática --
Visual Basic.net 2005	-	(BN) Visual Basic (Linguagens de programação) – [Manuais]

Terminologia dos cabeçalhos – (ocorrência em tesauro e índices BNP)

Base B4

Termos	Tesauro	Índices BNP
Abóbora	Agro	-
Actividades equestres	-	-
Acupunctura equina	-	-
Agricultura de precisão	-	(BN) Agricultura de regadio
Agronomia	Agro	BN
Alcobaça	-	(BN) Alcobaça (Portugal)
Alentejo	-	(BN) Alentejo (Portugal)
Algarve	-	(BN) Algarve (Portugal)
Alimentação animal	-	(Bn) Animais -- Alimentação

Anestesia	Agrov	(BN) <i>animais</i>
Anestesia veterinária	-	(BN) Anestesia
Animais de companhia	-	(BN) Animais domésticos
Animais exóticos	-	-
Aparelho genital feminino	-	-
Aranhão vermelho	-	-
Atlas	-	(BN) Atlas mundial
Bem-estar animal	Agrov	(BN) Bem-estar
Biodiesel	Agrov	-
Biotecnologia vegetal	Agrov	(BN) Biotecnologia
Boas práticas	Agrov	-
Bovinicultura	-	-
Bovino alentejano	-	(BN) Bovinos
Cão	Agrov	BN
Caprinos	-	BN
Caracterização genética do cavalo	-	-
Carne	Agrov	BN
Carne fresca	-	-
Carraça	-	-
Catálogo	-	(BN) [Catálogos]
Cavalo	Agrov	(BN) Cavalos
Cavalo de desporto	-	-
Cavalo lusitano	-	(BN) Cavalos de raça lusitano
Ciência da carne	-	-
Cirurgia	-	BN
Cirurgia ortopédica	-	-
Cirurgia veterinária	-	(BN) Medicina veterinária
Citricultura	-	BN
Coelho	-	-
Colza	Agrov	-
Comportamento animal	Agrov	BN
Concentrado de tomate	-	(BN) Tomate --
Construção de espaços verdes	-	(BN) Espaços verdes --
Cooperativismo	-	BN
Criação de empresa	-	(BN) Empresas – Criação
Cuidados de enfermagem	-	(BN) Medicina veterinária
Cuidados intensivos	-	(BN) <i>animais</i>
Cuidados veterinários	-	-
Culturas arvenses	-	-
Cunicultura	-	-
Diagnóstico	Agrov	(BN) -- Diagnóstico / Diagnóstico ...
Diagnóstico e tratamento	-	-
Doenças das plantas	-	-
Dor	Agrov	-
Economia	Agrov	BN
Economia portuguesa	-	(BN) Portugal -- Economia
Égua	Agrov	BN
Embalagem	Agrov	BN
Enfermagem veterinária	-	-
EPRAL	-	-

Equinicultura	-	-
Equipamento agrícola	-	(BN) Equipamentos agrícolas
Equipamentos de rega	-	(BN) Irrigação
Equitação	-	BN
Erosão do solo	-	(BN) Erosão
Espaços verdes	-	BN
Espécies pecuárias	-	(BN) Pecuária --
Espécies piscícolas	-	(BN) Peixes
Estado crítico	-	-
Esterilização	-	-
Évora	-	(BN) Évora (Portugal)
Exercícios de economia	-	(BN) Economia – [Exercícios]
Farmacologia	Agrov	BN
Faro	-	(BN) Faro (Portugal)
Fauna	Agrov	BN
Febre catarral	-	-
Febre da carraça	-	-
Fisiologia veterinária	-	-
Fisioterapia veterinária	-	-
Formação	Agrov	BN
Forragens	-	-
Fotografia aérea	Agrov	BN
França	Agrov	BN
Fundação Alentejo	-	-
Gato	Agrov	(BN) Gatos
Germinação	Agrov	-
Gestação	Agrov	(BN) Animais --
Gestão	Agrov	BN
Gestão agrícola	-	-
Ginja	-	(BN) Ginja
GPS	-	BN
Horticultura	Agrov	BN
Horticultura biológica	-	BN
Horticultura protegida	-	-
Incidência	-	-
Indústria corticeira	-	(BN) Indústria da cortiça
Inspecção sanitária	-	(BN) Inspecção de carnes / Inspecção do pescado
Jardins	-	BN
Leishmaniose canina	-	(BN) Leishmaniose (BN) Animais -- Doenças
Manual de horticultura	-	(BN) Horticultura [Manuais]
Marcadores biológicos de tumores	-	(BN) Cancros
Marketing agrícola	-	-
Marketing agrícola	-	-
Mecanização agrícola	-	(BN) Mecanização
Medicina de urgência	-	(BN) Medicina veterinária
Medicina desportiva	-	(BN) <i>animais</i>
Medicina veterinária	Agrov	BN
Milho	Agrov	BN

Motores	-	BN
Neonatais	-	(BN) <i>animais</i>
Neonatologia	-	(BN) <i>animais</i>
Nutrição animal	Agrov	(BN) Animais -- Alimentação
Obesidade	Agrov	(BN) Animais --
Oftalmologia veterinária	-	(BN) Medicina veterinária
Olea europaea	Agrov	-
Olival	-	-
Oliveira	-	(BN) Oliveiras
Olivicultura	Agrov	BN
Osteoartrite	Agrov	(BN) <i>animais</i>
Ovino	Agrov	-
Ovinos	-	BN
Parasitologia	Agrov	BN
Parque Natural	-	(BN) Parques naturais
Pastagens	-	BN
Património rural	-	BN
Peixe	Agrov	(BN) Peixe – Receitas culinárias
Piscicultura	Agrov	BN
Plantas aromáticas	-	BN
Plantas medicinais	-	BN
Plantas ornamentais	-	BN
Política de desenvolvimento rural	-	(BN) Desenvolvimento rural
Porco alentejano	-	(BN) Suínos de raça Alentejana
Portugal	Agrov	BN
Praga das plantas	-	(BN) Praga
Prestação de cuidados pediátricos	-	-
Preventic	-	-
Primeiros socorros	Agrov	(BN) <i>animais</i>
Produção de carne	Agrov	-
Produção de forragem	-	(BN) Pastagens
Projecto de instalação	-	(BN) Projectos
Prunus cerasus	Agrov	-
Qualidade ambiental	-	(BN) Ambiente -- Qualidade
Qualificação profissional	-	(BN) Qualificações profissionais
Reacção	-	-
Reciclagem	Agrov	BN
Recursos hídricos	-	BN
Relatório de estágio	-	(BN) [Relatórios]
Reprodução animal	-	(BN) Animais – Reprodução
Rhizobium	Agrov	-
Saneamento	-	(BN) Saneamento básico
Saúde animal	-	-
Serra da Estrela	-	(BN) Parque Natural da Serra da Estrela (Distrito da Guarda, Portugal)
SIG	-	(BN) Sistemas de informação geográfica
Simposium veterinário	-	-

Sistema de rega	-	BN
Sistemas de rega	-	-
Sociologia rural	Agrov	BN
Soja	Agrov	-
Sorgo forrageiro	-	-
Taludes	-	BN
Tencologias da informação	-	(BN) Tecnologias da informação e comunicação
Tomateiro	-	(BN) Tomateiros
Transusão de sangue	-	-
Tratado de Lisboa	-	-
Tratamento da dor	-	(BN) Terapêuticas
Urgência veterinária	-	(BN) Medicina veterinária
Variabilidade espacial	-	-
Veterinária	-	(BN) Medicina veterinária
Viabilidade económica	Agrov	-
Vitelos	-	-
Viticultura	Agrov	BN

Terminologia dos cabeçalhos – (ocorrência em tesouro e índices BNP)

Base B5

Termos	Tesouro	Índices BNP
[Bibliografias]	-	BN
25 de Abril, 1974 (Portugal)	-	BN
31 de Janeiro, 1891 (Porto, Portugal)	-	BN
31 de Janeiro, 1891 (Porto, Portugal) nos jornais	-	(BN) 31 de Janeiro, 1891 (Porto, Portugal)
Aceitação	-	-
Administração	-	BN
Administração militar	-	BN
Administração portuguesa – 1320-1433	-	-
África	Eurovoc	BN
Alfabetização	Eurovoc	BN
Aliança Luso-Britânica	-	(BN) Aliança Luso-Britânica Colonialismo
Alunos do ensino básico	-	BN
Alunos do ensino secundário	-	BN
Anarquismo	Eurovoc	BN
Angola	Eurovoc	BN
Arquivo Casa de Bragança	-	BN
Arquivo Nacional da Torre do Tombo	-	BN
Asma	-	BN
Avaliação	Eurovoc	BN
Brasil	Eurovoc	BN

Burguesia	-	BN
Cabo Verde	Eurovoc	BN
Cantigas de amigo	-	BN
Caracterização sócio profissional	-	(BN) Caracterização socioprofissional
Carta Constitucional, 1826 (Portugal)	-	-
Cartografia portuguesa	-	BN
Causalidade	-	BN
Centros de formação	-	(BN) <i>não ocorre isolado</i>
China	Eurovoc	BN
Ciência	-	BN
Ciências da educação	-	BN
Civilização	Eurovoc	BN
Colonização portuguesa	-	BN
Comunismo	Eurovoc	BN
Concelho de Coimbra (Portugal)	-	BN
Condições económicas	Eurovoc	BN
Congressos	-	BN
Conselhos de guerra	-	-
Consequências	-	-
Constituição, 1787 (Estados Unidos)	-	(BN) Constituições – Estados Unidos da América
Contribuições da educação escolar	-	(BN) Educação escolar --
Costumes	-	BN
Crianças	-	BN
Crónicas	-	BN
Cultura portuguesa	-	BN
Cursos	-	BN
Descobrimiento do Brasil, 1500	-	BN
Descobrimientos portugueses	-	(BN) Descobrimientos portugueses – Séc. 15-16
Desempenho profissional	-	BN
Desenvolvimento	-	BN
Desenvolvimento individual	-	BN
Desenvolvimento social	Eurovoc	BN
Desportos	-	BN
Dicionários	-	BN
Direito	-	BN
Disciplina de história	-	BN
Educação	Eurovoc	BN
Educação cívica	Eurovoc	(BN) Educação para a cidadania
Educação das crianças	-	BN
Educação intercultural	-	BN
Educação pré-escolar	Eurovoc	BN
Educador	-	(BN) Educadores
Eleitorado	Eurovoc	BN
Emprego	-	BN
Empresas	-	BN
Ensino	Eurovoc	BN
Ensino básico	-	BN

Ensino médio	-	-
Ensino superior	Eurovoc	BN
Ensino superior politécnico	-	BN
Epistemologia	-	BN
Escolas	-	BN
Espanha	Eurovoc	BN
Estado Novo, 1933-1974 (Portugal)	-	BN
Estados Unidos	Eurovoc	(BN) Estados Unidos da América
Estrutura	-	BN
Estruturas morfológicas	-	BN
Ética política	-	BN
Europa	Eurovoc	BN
Evangelização	-	BN
Evangelização portuguesa	-	(BN) Evangelização – Participação dos missionários portugueses
Expansão portuguesa	-	BN
Expressão escrita	-	BN
Famílias	-	BN
Fascismo	Eurovoc	BN
Filosofia	Eurovoc	BN
Filosofia do conhecimento	-	BN
Filosofia política	Eurovoc	BN
Filósofos	-	BN
Formação profissional	Eurovoc	BN
Geologia	Eurovoc	BN
Guerra Civil de Espanha, 1936-1939	-	BN
Guerra colonial portuguesa, 1961-1974	-	BN
Guerra da Restauração, 1640-1668	-	(BN) Guerra da Restauração, 1641-1668 (Portugal)
Guiné-Bissau	Eurovoc	BN
Hemofilia	-	BN
Hipersensibilidade	-	BN
História	Eurovoc	BN
História Universal	Eurovoc	BN
Humanismo	-	BN
Idade média	-	BN
Ideias políticas	-	(BN) Ideologias políticas
Ilhas (habitação) – Porto (Portugal)	-	BN
Implantação da República, 1910 (Portugal)	-	BN
Imprensa operária	-	BN
Índia	Eurovoc	BN
Índia portuguesa (Jur. pass.)	-	(BN) Índia Portuguesa (Índia: jur. pass.)
Indivíduo	-	(BN) Pessoas
Influências da cultura inglesa	-	BN
Instituições militares	-	BN
Jesuítas (Religiosos)	-	BN
Jornalismo	-	BN
Liceus	-	BN

Língua portuguesa	-	BN
Linguística portuguesa	-	BN
Literatura portuguesa	-	BN
Lunda (Angola)	-	-
Lutas Liberais, 1828-1834 (Portugal)	-	BN
Macau (China)	-	BN
Macau (Portugal)	-	BN
Maçonaria	-	BN
Manuais de ensino superior	-	BN
Marketing	Eurovoc	BN
Marxismo	Eurovoc	BN
Missões	-	BN
Movimentos de libertação nacional	-	-
Movimentos revolucionários	-	BN
Mudanças sociais	-	BN
Mulheres	-	BN
Operações navais	-	-
Origens	-	BN
Países de língua portuguesa	-	BN
Pará (Brasil)	-	BN
Pareceres	-	BN
Parque Natural da Serra da Estrela (Portugal)	-	(BN) Parque Natural da Serra da Estrela (Distrito da Guarda, Portugal)
Patrocínio	Eurovoc	BN
Pedagogia	-	BN
Pernambuco (Brasil)	-	BN
Política	-	BN
Política colonial	-	BN
Política de educação	-	BN
Política de formação profissional	-	BN
Pontos de vista	-	-
Portalegre	-	(BN) Portalegre (Portugal)
Porto (Portugal)	-	BN
Portugal	-	BN
Portugueses	-	BN
PREC, 1974-1976 (Portugal)	-	(BN) PREC
Professores	-	BN
Professores portugueses	-	BN
Psicopedagogia	-	(BN) Psicologia da educação
Reformas do ensino	-	(BN) Ensino -- Reformas
Reino Unido	Eurovoc	BN
Relações família-escola	-	(BN) Relações famílias-escolas
Relações internacionais	Eurovoc	BN
Relatórios	-	BN
Religiões	-	BN
Renascimento	-	BN
Restauração, 1640 (Portugal)	-	BN
Revolução de Outubro, 1917 (URSS: Jur. pass.)	-	BN

Revolução espanhola, 1936-1939	-	(BN) Guerra Civil de Espanha, 1936-1939
Revolução Francesa, 1789-1799	-	BN
Séc. 15-16 ²	-	-
Séc. 15-20	-	-
Séc. 16-18	-	-
Séc. 18	-	-
Séc. 21	-	-
Serviço social	Eurovoc	BN
Sistema constitucional	-	(BN) Constituições
Socialismo	Eurovoc	BN
Sociedade	Eurovoc	BN
Sociedade portuguesa	-	(BN) Portugal -- Sociedade
Sociedade portuguesa na poesia portuguesa	-	-
Sociologia	Eurovoc	BN
Teorias	-	BN
Terapia domiciliária	-	BN
Tese de mestrado	-	(BN) [Teses]
Timor Loro Sae	-	BN
Tradições	-	(BN) <i>não ocorre isolado</i>
Universidades	-	BN
Verbo deixar	-	-
Verbos modais	-	BN
Vida quotidiana	-	BN
Visitas pastorais	-	BN

² A designação de ‘séc....’ em números arábicos só aparece como ^z – modificador cronológico; em assunto principal aparece com numeração romana.

Anexo VI. Terminologia dos cabeçalhos de Nomes Próprios

Ocorrência de Nomes Próprios nos cabeçalhos do bloco 6xx Unimarc

UNI	Ocorrência na base		Índices BNP
600	Agiuar-Branco, Fernando	X	Branco, Fernando Aguiar, 1923-
600	Agiuar-Branco, Fernando, 1923-	X	Branco, Fernando Aguiar, 1923-
600	Almeida, António de	X	Almeida, António de, 1900-1984
600	Barata Feyo	X	Feio, Salvador Carvão de Eça Barata, 1902-1990
600	Breyner, Tomás de Melo, Conde de Mafra, 1866-1933	X	Breyner, Tomás de Melo, 1866-1933
600	Costa, Helena, 1913-	X	Costa, Helena Moreira de Sá e, 1913- 2006
600	Dali, Salvador	X	Dali, Salvador, 1904-1989
600	Dias, Henrique, 1600-1662	X	Dias, Henrique, ?-1661
600	Einstein, Alfred, 1879-1955	X	Einstein, Albert, 1879-1955
600	Fernando Pessoa	X	Pessoa, Fernando, 1888-1935
600	Freitas, José Joaquim de, 1840-1896	X	Freitas, Rodrigues de, 1840-1896
600	Godard, Jean Luc	X	Godard, Jean-Luc, 1930-
600	Jane Austen	X	Austen, Jane, 1775-1817
600	João VI, Rei de Portugal	X	João VI, Rei de Portugal, 1767-1826
600	José Régio	X	Régio, José, pseud.
600	Le Corbusier, 1887-1965	X	Le Corbusier, pseud.
600	Lévinas, Emmanuel	X	Lévinas, Emmanuel, 1906-1995
600	Lou Reed	X	Reed, Lou
600	Natividade, J. Vieira	X	Natividade, J. Vieira, 1899-1968
600	Neto, João Cabral de Melo	X	Neto, João Cabral de Melo, 1920-1999
600	Niemeyer, Oscar	X	Niemeyer, Oscar, 1907-
600	Pessoa, Joaquim Seabra	X	Pessoa, Joaquim Seabra, 1850-1893
600	Sá, Moreira de, 1853-1924	X	Sá, Bernardo Valentim Moreira de, 1853-1924
600	Sena da Silva	X	Silva, Sena da, 1926-2001
600	Silveira, Gonçalo da, 1523- 1561	X	Silveira, Gonçalo da, 1521-1561
600	Soares, Mário – [Biografias]	X	Soares, Mário, 1924-

600	Suggia, Guilhermina, 1885-1950	X	Suggia, Guilhermina, 1888-1950
600	Xavier, Alfredo Augusto Caldas, 1853-1896	X	Xavier, Alfredo Augusto Caldas, 1852-1896
600	Batista, Alexandre		
600	Berkeley, George, 1685-1753		
600	Caeiro, Francisco da Gama, 1928-1994		
600	Camões, Luís de, 1524?-1580		
600	Carneiro, Mário de Sá, 1890-1916		
600	Chagas, Pinheiro, 1842-1895		
600	Coimbra, Leonardo, 1883-1936		
600	Coimbra, Leonardo, 1883-1936		
600	Cutileiro, João, 1937-		
600	Dias, Bartolomeu, ?-1500		
600	Fonseca, Manuel Pinto da, Frei, 1681-1773		
600	Gaudí, Antoni, 1852-1926		
600	Gomes, António Ferreira, 1906-1989		
600	Laranjo, José Frederico, 1846-1910		
600	Leonardo da Vinci, 1452-1519		
600	Martins, Oliveira, 1845-1894		
600	Medina, Henrique, 1901-1988		
600	Medina, Henrique, 1901-1988		
600	Meireles, Cecília, 1901-1964		
600	Negreiros, Almada, 1893-1970		
600	Nemésio, Vitorino, 1901-1978		
600	Neto, Agostinho, 1922-1979		
600	Nóbrega, Manuel da, 1517-1570		
600	Pereira, Miguel Baptista, 1929-		
600	Pessoa, Fernando, 1888-1935		
600	Pompeu, Orlando, 1956-		
600	Portugal, Manuel de, pseud.		
600	Queirós, Eça de, 1845-1900		
600	Queirós, Pedro Fernandes de, 1575-1614		
600	Quental, Antero de, 1842-1891		
600	Régio, José, pseud.		
600	Sartre, Jean-Paul, 1905-1980		
600	Sérgio, António, 1883-1969		
600	Souto, A. Meireles do, 1900-1972		

600	Woodhouse, Roberto Guilherme, 1828-1876	
601	Confagri	X Confederação Nacional das Cooperativas Agrícolas de Portugal
601	Escola Superior Agrária de Elvas	X Instituto Politécnico de Portalegre. Escola Superior Agrária de Elvas
601	Escola Superior de Educação	X Instituto Politécnico de Portalegre. Escola Superior de Educação
601	Escola Superior de Educação de Portalegre	X Instituto Politécnico de Portalegre. Escola Superior de Educação
601	Escola Superior de Saúde	X Instituto Politécnico de Portalegre. Escola Superior de Saúde
601	Escola Superior de Tecnologia e Gestão	X Instituto Politécnico de Portalegre. Escola Superior de Tecnologia e Gestão
601	Escola Superior de Tecnologia e Gestão de Portalegre	X Instituto Politécnico de Portalegre. Escola Superior de Tecnologia e Gestão
601	Instituto Superior Técnico	X Universidade Técnica de Lisboa. Instituto Superior Técnico
601	Portugal. Presidente da República, 2006- (Aníbal Cavaco Silva)	X Portugal. Presidente da República, 2006- (Cavaco Silva)
601	Évora. Diocese	
601	Fundação Calouste Gulbenkian	
601	Fundação Luso-Americana	
601	Instituto Politécnico de Portalegre	
601	Portugal. Administração Regional de Saúde do Norte	
601	Portugal. Conselho de reflexão sobre a saúde	
601	Portugal. Instituto do Emprego e Formação Profissional. Centro de Emprego de Évora	
601	União Europeia. Comissão	
601	Universidade de Lisboa. Faculdade de Medicina – Museu de Medicina	
601	Universidade Federal de Santa Catarina	
607	Alcobaça	X Alcobaça (Portugal)
607	Alentejo	X Alentejo (Portugal)
607	Algarve	X Algarve (Portugal)
607	Alqueva	X Barragem do Alqueva (Conselho de Portel, Portugal)
607	Alvão	X Alvão (Concelho de Mondim de Basto, Portugal)
607	Badajoz	X Província de Badajoz (Espanha)

607	Beja	X	Beja (Portugal)
607	Castelo Branco	X	Castelo Branco (Portugal)
607	Costa Alentejana	X	Alentejo (Portugal) -- Litoral
607	Elvas	X	Elvas (Portugal)
607	Ermelo	X	Ermelo (Concelho de Mondim de Basto, Portugal)
607	Esposende	X	Esposende (Portugal)
607	Évora	X	Évora (Portugal)
607	Extremadura	X	Extremadura (Espanha)
607	Faro	X	Faro (Portugal)
607	Herdade da Ajuda Nova	X	Herdade da Ajuda Nova (Concelho de Vendas Novas, Portugal)
607	Leiria	X	Leiria (Portugal)
607	Lisboa	X	Lisboa (Portugal)
607	Litoral Alentejano	X	Alentejo (Portugal) -- Litoral
607	Manteigas	X	Manteigas (Portugal)
607	Nisa	X	Nisa (Portugal)
607	Portalegre	X	Portalegre (Portugal)
607	Região Alentejo (Portugal)	X	Alentejo (Portugal)
607	Ria Formosa	X	Ria Formosa (Portugal)
607	Ribeira da Póvoa	X	Ribeira da Póvoa (Coimbra, Portugal)
607	Sado	X	Sado (Concelho de Setúbal, Portugal)
607	Serra da Estrela	X	Serra da Estrela (Portugal)
607	Vila Real	X	Vila Real (Portugal)
607	África		
607	Angola		
607	Brasil		
607	Cabo Verde		
607	Calecut (Índia)		
607	Espanha		
607	Portugal		
607	Portugal -- Porto		
607	Portugal –1870-1910		

607	Salvador (Brasil)
607	São Tomé e Príncipe