

Un Sistema Multiagente de Recuperación de Objetos de Aprendizaje con Atributos de Contexto

Ana B. Gil, Francisco García

Universidad de Salamanca, Dpto. Informática y Automática
Facultad de Ciencias, Plaza de la Merced s/n, 37008, Salamanca, Spain
abg@usal.es fgarcia@usal.es

Resumen La posibilidad que ofrecen los estándares educativos de gestionar la información, facilitando su interoperabilidad y reutilización en componentes para diversas plataformas, y la existencia de lenguajes de modelado educativo que permiten además estructurarla en base a metadatos, abren nuevas e importantes posibilidades para los sistemas de educación en línea. El paradigma de Objetos de Aprendizaje centra esta nueva brecha en la gestión e intercambio de materiales docentes debido a su descripción semántica. El artículo presenta las posibilidades que se generan en base a ciertos estándares para la reutilización automática basadas en su contenido semántico. Propone además una arquitectura de recuperación de Objetos de Aprendizaje que permita la búsqueda eficiente, la selección y composición de éstos en un contexto.

1. Introducción

El crecimiento de la web la ha convertido en el mayor repositorio de información existente. Su gran dimensión unida a la heterogeneidad de la información que contiene genera serias limitaciones a la hora de realizar búsquedas y recuperar información de interés. El desarrollo de la web semántica aparece como un avance prometedor en la transición de la web actual hacia la generación de contenidos bien descritos y estructurados. Es evidente que organizar toda la información por dominios especializados estableciendo ontologías de consenso es hoy en día una utopía. Sin embargo son muchos los esfuerzos que en ese sentido se están realizando, con el desarrollo de lenguajes, herramientas, estándares e infraestructuras que proporcionen la posibilidad de generación de herramientas software que gestionen de manera eficaz y automatizada grandes volúmenes de información.

Uno de los ejes de aplicación directa de la web semántica se está desarrollando sobre los recursos educativos en la web, abriendo una importante posibilidad de mejora para los sistemas de educación en línea sobre el paradigma de los llamados objetos de aprendizaje (Learning Objects). Se trabaja en estándares educativos para gestionar la información, lenguajes de modelado educativo que describen los recursos con metadatos facilitando su gestión, interoperabilidad y

reutilización en componentes para diversas plataformas así como que permiten además estructurar los contenidos de manera que se a su sentido pedagógico.

El objetivo de este artículo es presentar ciertas singularidades del esfuerzo por la adaptación de las tecnologías de web semántica al campo de la educación en línea. Así como generar la importancia de los sistemas de recomendación en base a información semántica en la gestión ordenada de recursos educativos para los sistemas de educación en línea.

El artículo comienza con esta introducción y se organiza como sigue. La sección 2 muestra una perspectiva de los sistemas de recomendación que proviene fundamentalmente del ámbito del comercio electrónico y su traslado natural al ámbito de la enseñanza virtual. La sección 3 describe qué son los objetos de aprendizaje (OAs) así como sus aspectos semánticos y estándares de funcionamiento de los procesos que implica su gestión. La sección 4 presenta una arquitectura de recomendación de OAs como sistema intermediario de recolección y filtrado personalizado. Finalmente, la Sección 5 cierra el artículo, presentando conclusiones e ideas sobre el trabajo.

2. Una Perspectiva de los Sistemas de Recomendación

Previo a la aparición de Arquitecturas Orientadas a Servicios (Service Oriented Architecture, SOA) [1], toda la información residía en páginas estáticas en Internet. En la actualidad el desarrollo de la web semántica genera la necesidad de proveer el acceso no solo a documentos estáticos, mediante técnicas de recuperación de información, sino a los servicios web que proporcionan nuevas maneras de ofrecer información. La tecnología para la recuperación de servicios (Service retrieval technology) incrementa su importancia para proporcionar herramientas que permitan a las personas y al software encontrar fácilmente aquellos servicios que atiendan de manera personalizada los requisitos de búsqueda.

De cualquier modo las tecnologías aplicadas, resulten servicios Web o documentos que catalogar, nos obligan a hacer un estudio de las técnicas utilizadas en recuperación de información web para su recomendación. Existe un amplio número de sistemas de recomendación a lo largo y ancho de Internet. Un estudio detallado en el ámbito del comercio electrónico, donde mayoritariamente han sido desarrollados, puede consultarse en ([2], [3], [4]). Atendiendo al origen (usuario o ítem) donde fija el algoritmo de recomendación su eje se pueden clasificar en dos grandes categorías, los basados en filtrado colaborativo social y en filtrado por contenido.

2.1. Sistemas Basados en Filtrado Colaborativo Social

Se incluyen en esta clasificación, denominada también simplemente como recomendación colaborativa a aquellos sistemas que construyen la recomendación como una agregación estadística/probabilística de las preferencias de otros usuarios. Dichos sistemas realizan un modelado de usuario generando patrones de preferencias y generando grupos de usuarios como similares preferencias. Las

técnicas más utilizadas son el análisis estadístico de extracción de datos o la minería de datos y las técnicas de descubrimiento de conocimiento en grandes bases de datos (knowledge discovery in databases (KDD) techniques) haciendo uso de la monitorización del comportamiento del usuario sobre el sistema, sus puntuaciones sobre los servicios, su historial, etc. Las similitudes entre usuarios son computadas utilizando la correlación usuario a usuario. Esta técnica identifica un conjunto de “vecinos próximos” generando patrones de comportamiento social. Algunos de estos sistemas pueden encontrarse en Ringo [5] o GroupLens [6].

Estos sistemas adolecen fundamentalmente de dispersión, debido a que para generar nubes de comportamiento entre usuarios se necesita un gran volumen de usuarios y productos muestreados. La generación de recomendaciones apropiadas lleva asociada una importante masa crítica de datos. Esto resulta un problema para generar recomendación de calidad a un usuario del que no se tiene historial de ítems. Atendiendo al mecanismo de realimentación del usuario en el sistema los algoritmos empleados en sistemas de recomendación colaborativa se clasifican a su vez en algoritmos basados en memoria (Memory-based algorithms) y algoritmos basados en modelos (Model-based algorithms) [7]. Mientras los basados en memoria utilizan las puntuaciones que otros usuarios han dado a un objeto para predecir la puntuación del usuario actual los basados en modelos hacen uso de la información acumulada del usuario actual para construir un perfil o modelo, a partir del cual se realizan las predicciones.

2.2. Sistemas Basados en Filtrado por Contenido

La sugerencia del ítem se realiza atendiendo a la similitud con otros ítems que el usuario ha adquirido en el pasado. Así cuando se encuentra que el consumo de un producto A, habitualmente lleva al consumo de un producto B, en el momento que un usuario pide el producto A directamente se le recomienda el B. Esta técnica lleva asociada una serie de problemas, uno de ellos es la de recomendar un ítem nuevo en el sistema, debido a que no existe correlación con otros ítems consumidos. Para ello es necesario establecer las características detalladas de los ítems en su contexto lo que permite entonces calcular niveles de proximidad entre ellos.

Este tipo de sistemas utiliza tecnologías basadas en aprendizaje automático y mediante un clasificador establecen el grado de interés de un usuario en un artículo en base a los historiales que han consumido ese mismo ítem. Dichos clasificadores pueden ser implementados utilizando técnicas de la inteligencia artificial, tales como redes neuronales, bayesianas, árboles de decisión y en la actualidad sistemas de razonamiento basados en características semánticas que describen las particularidades de los ítems en su contexto. Los sistemas basados en contenido utilizan entonces los conocidos como algoritmos “ítem a ítem” generados mediante la asociación de reglas de correlación entre ellos. Algunos ejemplos se pueden encontrar en [8], [9], donde una serie de documentos web son clasificados mediante árboles de decisión atendiendo a su contenido sobre

Categorías	Elementos
1. General	Identificador, Título, Entrada de catálogo, Lengua, Descripción, Descriptor, Cobertura, Estructura, Nivel de agregación
2. Ciclo de vida	Versión, Estatus, Otros colaboradores
3. Meta-metainformación	Identificador, Entrada de catálogo, Otros colaboradores, Esquema de metadatos, Lengua
4. Técnica	Formato, Tamaño, Ubicación, Requisitos, Comentarios sobre la instalación, Otros requisitos para plataformas, Duración
5. Uso educativo	Tipo de interactividad, Tipo de recurso de aprendizaje, Nivel de interactividad, Densidad semántica, Usuario principal, Contexto [Nivel educativo], Edad, Dificultad, Tiempo previsto de aprendizaje, Descripción, Lengua
6. Derechos	Coste, Copyright y otras restricciones, Descripción
7. Relación [con otros recursos]	Tipo [naturaleza de la relación con el recurso principal], Recurso [recurso principal al que se refiere esta relación]
8. Observaciones	Persona, Fecha, Descripción
9. Clasificación	Finalidad, Nivel táxon (taxonómico), Descripción, Descriptor

Cuadro 1. Especificación LOM

una escala binaria, o el típico sistema de recomendación para el segundo o tercer artículo de Amazon.

El problema fundamental de estos sistemas es la superespecialización, debido a que el usuario suele ser dirigido hacia el mismo tipo de ítems que ha consumido en el pasado.

3. Los Objetos de Aprendizaje en el Contexto Educativo

El incremento actual de recursos para la educación en línea exige la gestión eficiente de los servicios y sus elementos implicados. Los mayores esfuerzos en el ámbito del e-learning se llevan a cabo para permitir la interoperabilidad de los recursos generados. En este contexto surgen los objetos de aprendizaje (OAs) con la idea de reutilizar los recursos educativos y reducir sus costes de producción. Los OAs son piezas individuales autocontenidas y reutilizables con fines instruccionales.

3.1. La Semántica de los Objetos de Aprendizaje

Los OAs deben estar etiquetados con metadatos de modo que puedan ser identificados, localizados y utilizados para propósitos educacionales en ambientes basados en web. El estándar IEEE LOM [10], especifica el esquema conceptual

que define las estructuras de los datos para las instancias de metadatos de los OAs. El esquema básico de LOM se compone de 9 categorías y 47 elementos que muestra la tabla 1.

Como consecuencia, un OA pudiera etiquetarse del modo siguiente:

```
<lom xmlns="http://ltsc.ieee.org/xsd/LOMv1p0">
  <general>
    <title>
      <string>unidad de contenido</string>
    </title>
    <language>es</language>
  </general>
  <technical>
    <format>text/html</format>
    <location type="URI">
      http://recurso.usal.es/detalle.html
    </location>
  </technical>
</lom>
```

3.2. Repositorios de Objetos de Aprendizaje

Existen repositorios de OAs (Learning Objects Repositories, LOR) tales como MERLOT (<http://www.merlot.org>), CAREO (<http://careo.ualgary.ca>), CLOE (<http://cloe.on.ca>). Estos sistemas software son catálogos digitales que bien almacenan los recursos educativos y sus metadatos o solamente estos últimos. Generalmente proporcionan algún tipo de interfaz de búsqueda que permite la recuperación de los mismos. Cualquier interacción para recuperación de OAs puede ser llevada a cabo de manera manual o ser automatizada a través de distintos sistemas software. Los principales estándares de interfaces de repositorios digitales para OAs son:

CORDRA (Content Object Repository Discovery and Registration/Resolution Architecture) [11]. Desarrolla un modelo abierto para el diseño y la implementación de sistemas software que recuperen, compartan y reutilicen contenido educativo a través de la implementación de sistemas federados en repositorios de recursos educativos. Las actividades de CORDRA están siendo coordinadas por la iniciativa Advanced Distributed Learning Initiative (ADLI) la Corporation for National Research Initiatives (CNRI) y el Learning Systems Architecture Lab (LSAL).

IMS-DRI (IMS Digital Repositories Specification) [12]. Los Repósitos IMS Digitales v1.0, tienen por objetivo proporcionar recomendaciones para la interoperabilidad de las funciones de repositorio más comunes. Estas recomendaciones deberían ser implementadas a través de servicios web para permitirles presentar un interfaz común. Esta especificación define repositorios digitales como cualquier colección de recursos educativos que son accesibles vía

una red sin el conocimiento previo de la estructura de su. Los repositorios según este estándar pueden sostener los recursos reales o los metadatos que los describen. El recurso y sus metadatos no tienen que ser almacenados en el mismo repositorio. Esta especificación es manejada para utilizar esquemas ya definidos en otra parte (p.ej., IMS Metadatos y el Empaquetado de Contenido), en lugar de introducir cualquier nuevo esquema de datos.

OKI OSID (Open-Knowledge Initiative Open Service Interface Definition) [13]. La especificación O.K.I. permite la interoperabilidad sostenible y la integración mediante la definición de estándares sustentada en Arquitectura Orientada a Servicios (SOA). OSIDs es simplemente el software de contrato y por lo tanto compatible con la mayor parte de otras tecnologías y especificaciones, tales como OAP, WSDL. Actualmente esta tecnología está siendo utilizada en algunos desarrollos de colecciones digitales en instituciones académicas, tales como Campus Project[14] en las universidades catalanas, la sección de educación de Apple, con su Apple Learning Interchange (ALI) [15].

Sin embargo la heterogeneidad del etiquetado de los recursos educativos así como las labores de búsqueda y selección de los OAs exigen, al igual que en el ámbito del comercio electrónico el desarrollo de aplicaciones que permitan extraer y catalogar de manera automática integrando toda esta información con la perspectiva de que en esta ocasión estamos en un ámbito educativo. Aparece una nueva economía en la gestión, intercambio y creación de los Objetos de Aprendizaje. Aparece entonces la necesidad de un sistema de intermediación que permita recolectar y almacenar todos los datos referentes a los OAs en un formato que permita hacer inferencia sobre ellos. El requisito fundamental para efectuar dicha inferencia es generar una base de conocimiento basado en ontologías de dominio de modo que toda la información pueda ser procesada en base a definiciones interpretables. A continuación proponemos una arquitectura para el sistema.

4. Sistema Intermediario de Recuperación de OAs

El intercambio de datos entre repositorios de OAs y el usuario implicado en la búsqueda, así como la posterior publicación automática de los resultados de dicha búsqueda, están basados en un formato de almacenamiento y publicación expresado en términos de metadatos en el modelo LOM, revisados en apartados anteriores. En el contexto educativo el alumno juega un papel fundamental en su aprendizaje [16], de modo que resulta necesario modelar y almacenar su modelo de aprendizaje y los atributos del contexto a través de metadatos en las especificaciones de los OAs. Para el modelado del aprendizaje existen especificaciones que posibilitan ciertas descripciones en base a estructuras de metadatos en XML, tales como el IMS LIP (Learner Information Package) [17] y las especificaciones PAPI (Public and Private Information for Learners) [18].

Basado en este entorno tecnológico presentamos la arquitectura de un sistema intermediario de búsqueda y catalogación de OAs incluyendo aspectos de perso-

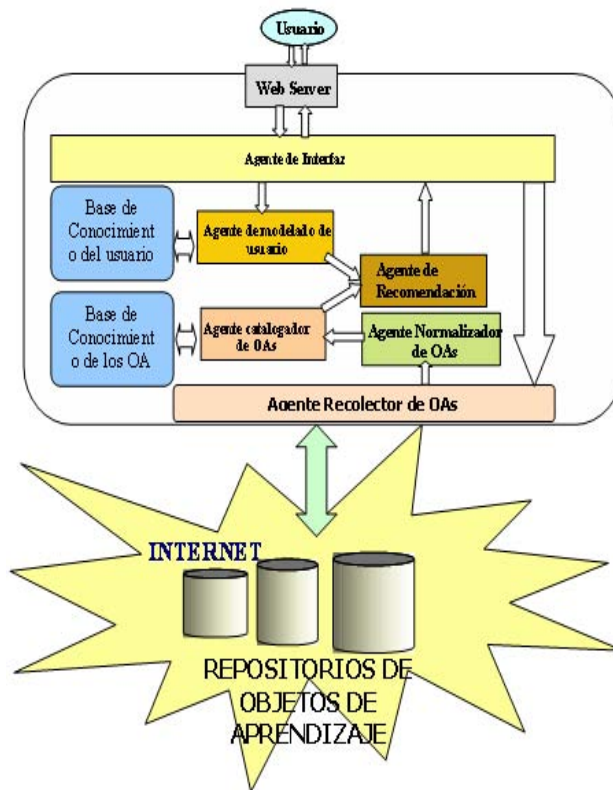


Figura 1. Arquitectura conceptual Multiagente de recuperación OAs

nalización. La arquitectura se define en base a un sistema multiagente donde los agentes software se coordinan para identificar y recuperar los OAs relevantes a la petición de información suministrada por el usuario.

En una primera aproximación, se han identificado los agentes necesarios, un total de seis, y sus relaciones, para dar soporte a la propuesta de arquitectura del sistema. Estos agentes visualizados en la Figura 1, se presentan en una arquitectura de gruesa granularidad y alto nivel de abstracción, debido a que se está definiendo la capa arquitectónica del sistema en clave de agentes para los principales componentes presentados.

Resulta necesario ajustar el contexto de idoneidad de los OAs en un ámbito de competencias que cuadre con el del usuario que realiza la petición por lo tanto el agente de modelado del usuario y el catalogador de OAs dan soporte al agente de recomendación que es el que establece la relevancia de un OA para un usuario concreto.

Los aspectos del dominio experimental para los que se va a filtrar la relevancia de los OAs, son presentados por una ontología. El ámbito puede ser intercam-

biable, tan sólo realizando un cambio en la ontología de dominio, pero para las pruebas de validación se realizarán en el área de las Bases de Datos en el ámbito de las Ciencias de la Computación.

La generación de dicha recomendación aparecerá mediante reglas de inferencia sobre varias fuentes. Por un lado la proveniente del OA, en su estructura LOM, directamente del repositorio. Por otro lado la información sobre el propio usuario que realiza las búsquedas de los OAs. Y finalmente la enriquecida con información de dominio, que resulta ser el modelo usado para operaciones de filtrado y recuperación conocido como Modelo Vectorial ([19], [20]), que recoge la relación de cada OA de una colección de N , y lo vincula con el conjunto de las n características particulares del dominio (referenciadas en la ontología de dominio). Este mecanismo calcula la entrada en el triplete Término-OA-Frecuencia, lo que permite calcular un vector, que determina en qué grado el OA satisface cada una de las características buscadas a través de unos pesos que validan el acercamiento de un OA hacia un ámbito del contexto buscado. De este modo el OA puede considerarse en primera aproximación como un vector que expresa su relación con cada una de las características de los términos de la ontología que en esta fase inicial de filtrado constituye una especie de vocabulario del contexto.

Esta información será recopilada por el sistema intermediario y almacenada en la base de conocimiento, por un lado de los OA y por otro del usuario. Ambas bases de conocimiento serán utilizadas en el proceso de recomendación por el agente de recomendación que cruzará los OAs recuperados de los repositorios y catalogados por el agente catalogador con la información almacenada del usuario permitiendo realizar un ranking de los OAs personalizado mediante la inferencia sobre los metadatos del contexto del usuario, más detalles sobre el algoritmo de recomendación aparece en [19].

El proceso de recomendación comienza para el usuario con la petición explícita de OAs mediante una interfaz de búsqueda basada el dominio. Estas palabras clave permitirán realizar la búsqueda del OA en una serie de repositorios disponibles en Internet. Todos los OAs recuperados a través del agente recolector pasarán a ser normalizados por el agente de normalización, según una ontología que define la estructura de los OAs recolectados. Es en este proceso de normalización donde se le añadirá el vector de relevancia tal como se ha indicado anteriormente. Una vez que los OAs han sido normalizados incluyendo además información de características de relevancia serán baremados de acuerdo a los metadatos del contexto del OA. El agente catalogador generará cuáles de entre ellos son relevantes o no para el usuario en el dominio seleccionado, para lo cual se realiza una comparación entre dicho OA a través de su modelo vectorial. Esta información a través de reglas de inferencia sobre el contexto del usuario añade características de personalización en la recuperación y ordenación de los OAs mostrados al usuario peticionario.

Hasta el momento se han identificado los agentes en el desarrollo de la arquitectura y su puesta inicial en funcionamiento con una ontología en el ámbito de las Bases de Datos, se encuentra en desarrollo.

5. Conclusiones

Este artículo presenta una introducción a la inclusión de semántica en los Objetos de Aprendizaje y su importancia en el contexto del e-learning. Las iniciativas actuales establecen la estructura sintáctica de los metadatos asociados a los OA, sin embargo falta la integración de cualquier OA en su contexto real, en su área de conocimiento. Para ello surge la necesidad de realizar herramientas que recolecten y filtren éstos catalogando su relevancia según su área de conocimiento y así poder integrar contextualmente los conceptos que un usuario quiere aprender en su área concreta. Las iniciativas actuales para etiquetado de OA proporcionan mecanismos de búsqueda y recuperación en base a las tecnologías derivadas de la web semántica, tales como servicios web de recuperación de OA, mecanismos para generar inferencia a través de los metadatos obtenidos, etc. Sin embargo es necesario dotar a las herramientas de inferencia con la semántica del contexto, de modo que sea posible razonar adecuadamente sobre los recursos. Establecer cómo las características de base de una recomendación han de estar basadas en parámetros relativos al dominio de especialización sobre la que trabaja el sistema representando estos elementos y su relación mediante ontologías.

Se presenta una arquitectura sobre la que se está trabajando en la actualidad basada en un sistema multiagente para la recuperación de OAs desde los Repositorios distribuidos en Internet y su catalogación, de modo que satisfaga las necesidades del usuario y se esboza el punto particular que ocupa actualmente nuestro trabajo, el diseño e implementación de un sistema de recomendación de OAs que se pueda integrar en dicha arquitectura.

Referencias

1. Web Service Architecture. W3C Working Draft 11 February 2004. <http://www.w3.org/TR/ws-arch/>
2. Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Analysis of Recommendation Algorithms for E-Commerce. ACM Conference on Electronic Commerce, pp.158-167, 2000
3. Shafer, J.; Konstan, J. A.; Riedl, J. E-Commerce Recommendation Applications. Data Mining and Knowledge Discovery, 5 (1/2), pp. 115-153, 2001
4. Montaner, M., López, B., de la Rosa, J. Ll., A Taxonomy of Recommender Agents on the Internet, Artificial Intelligence Review, Kluwer Academic Publishers. Volume 19, Issue 4, pp. 285-330. June, 2003
5. Shardanand, U.; Maes, P. Social Information Filtering: Algorithm for Automating Word of Mouth. Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, pp. 210-217, 1995
6. Konstant, J.; Miller, B.; Maltz, D.; Herlocker, J.; Gordon, L.; Riedl, J. GroupLens: applying collaborative filtering to Usenet news. Communications of the ACM, 40(3): 77-87, 1997
7. Breese, J. S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. En Proceedings of the Fourteenth Conference in Artificial Intelligence, p. 43-52, Julio 1998

8. Mooney, R. J.; Roy, L., Content-Based Book Recommending Using Learning for Text Categorization, Proceedings of the V ACM Conference on Digital Libraries, San Antonio, USA, pp.195-204, 2000
9. Pazzani M.I J, Billsus D. Content-Based Recommendation Systems. The Adaptive Web. LNCS Volume 4321/2007 pp. 325-341. Springer, Heidelberg (2007)
10. IEEE Draft Standard for Learning Object Metadata. IEEE 1484.12.1, 2002.
10. CORDRA (Content Object Repository Discovery and Registration/Resolution Architecture). <http://141.225.40.64/lisal/expertise/projects/cordra/>
11. IMS-DRI (IMS Digital Repositories Specification) <http://www.msglobal.org/digitalrepositories/>
12. Open knowledge Initiative - Open Service Interface Definitions (OKO-OSIDs) <http://www.okiproject.org/>
13. CAMPUS project <http://www.campusproject.org>
14. Apple Learning Interchange (ALI) <http://edcommunity.apple.com/ali/>
15. Frank, M., Reich, N., Humphreys, K. Respecting the Human Needs of Students in the Development of e-learning. Computers & Education 40 (2003) 57-70
16. IMS Learner Information Package Specification <http://www.msglobal.org/profiles/>
17. PAPI Learner <http://edutool.com/papi/>
18. L. Gravano, H. Garcia-Molina, and A. Tomasic. Gloss: Text-source discovery over the internet. ACM Transactions on Database Systems, 24(2):229-264, June 1999.
20. D. Grossman, O. Frieder, D. Holmes, and D. Roberts. Integrating structured data and text: A relational approach. Journal of the American Society for Information Science, 48(2), February 1997.
19. Gil, A.B., García F.J. Course Recommendation in e-learning Based in Swarm Intelligence. En las actas del IFIP TC 13 Seminar: Trends in HCI. Salamanca, 26 de Marzo de 2007. pp. 141-160, 2007.