
PLAN DE INVESTIGACIÓN

PROGRAMA DE DOCTORADO EN

FORMACIÓN EN LA SOCIEDAD DEL CONOCIMIENTO

UNIVERSIDAD DE SALAMANCA

Título:

**Análisis y utilización del Big Data en bibliotecas.
Caso de uso.**

Autor: Abdón Martín-Zarco Gallego

Director: Ana Belén Ríos Hilario

2/06/2015

INTRODUCCIÓN Y JUSTIFICACIÓN DEL TEMA OBJETO DE ESTUDIO (MÁXIMO 50 LÍNEAS):
INTRODUCTION AND JUSTIFICATION OF THE TOPIC OF STUDY (50 LINE MAXIMUM):

El concepto de nuevas tecnologías se ha quedado pequeño a la hora de analizar la sociedad actual, así como sus demandas y necesidades. La sobrecarga de información digital en la que estamos inmersos (y expuestos) está encabezada por el océano de Internet y los mares que representan las redes sociales, bases de datos, comercio electrónico, medios de comunicación digitales, aplicaciones móviles, sensores, estaciones meteorológicas, smart-phones, satélites, GPS o los más de “55 millones de imágenes y 340 millones de tuits” que subimos a la Nube y a Internet (Funk-e Studios, 2013, 12 nov.), que permiten obtener enormes cantidades de datos en tiempo real, difíciles de procesar pero que, combinados de la manera más adecuada, correcta y favorable, posibilitan establecer predicciones sobre conductas, hábitos y rutinas de los consumidores, usuarios y potenciales clientes, que añaden valor al producto o servicio de determinada entidad, asociación o empresa, suponiendo un beneficio a éstas. Se trata del Big Data.

El Big Data es un “conjunto de datos tan extenso y complejo que resulta difícil de procesar utilizando los sistemas tradicionales de procesamiento” (Ríos Hilario, Ana Belén y Díaz Rodríguez, Jesús, 2014, p. 32). Están producidos por las nuevas tecnologías que, impulsadas por Internet, permiten conectar a las personas y facilitar a organizaciones y empresas la recolección de datos útiles para sus investigaciones. El Big Data se caracteriza, según Russom (2011) por las tres Vs: volumen (cantidad de datos), velocidad (con la que se producen) y variedad (procedencia de los datos). A esto, se le puede sumar una cuarta V, que sería la del valor que puede extraerse de los datos y la combinación de éstos (Martínez-Martínez, Silvia y Lara-Navarra, Pablo, 2014).

En el año 2002, se almacena más información en formato digital que en soportes analógicos (Big Bang Data, 2015), hecho que ha impulsado el Big Data, siendo a partir de 2012 cuando los distintos sectores de actividad se sienten atraídos por las capacidades de este nuevo descubrimiento capaz de lograr beneficios y ventajas en investigaciones académicas, negocios, empresas, organizaciones... donde la toma de decisiones se basa en un enfoque predictivo y previsor a partir de datos estructurados y sin estructurar, obtenidos en tiempo real, que combinados, resultan una clara ventaja competitiva dentro de esa organización, empresa o institución (Schmarzo, Bill, 2014).

El Big Data ya se está utilizando en determinados sectores, tales como medicina, gobierno comercio, finanzas, telecomunicaciones, energía, transporte u hostelería, entre otros, llegando a mejorar la calidad de vida de las personas, aunque a la vez, suponen un reto para la privacidad de éstas.

El Big Data se organiza como una biblioteca de datos, por ello que estas instituciones, se constituyan como un depósito de Big Data, donde los bibliotecarios son los administradores que proporcionan acceso a la información, instruyen a los investigadores y son los principales defensores de la privacidad de los datos (Reinhalter, Lauren y Wittmann, Rachel J., 2014). Ana Santos, actual directora de la Biblioteca Nacional de España, es consciente de ello, ya que en la BNE se almacena toda la información generada desde el dominio .es, lo que puede generar nuevo conocimiento (García Calero, Jesús, 2015).

En este punto, los bibliotecarios juegan un papel muy importante en la adaptación de estas nuevas fuentes de conocimiento, así como en el almacenamiento, la preservación y el acceso a la información, participando de forma activa en la elaboración de estándares, prácticas y esquemas que guíen a otras bibliotecas (Johnston, Leslie, 2012), al fin y al cabo, los bibliotecarios tienen experiencia en normalización y estándares: Reglas de Catalogación, Formato MARC, Dublin Core... la justificación de este trabajo se origina en las bibliotecas como depósitos del conocimiento humano.

Pero no solo los bibliotecarios se constituyen como profesionales de la información, sino que el Big Data conforma un fenómeno heterogéneo que exige de la colaboración de otros perfiles profesionales: estadistas, informáticos, programadores... Con el Big Data, estos perfiles evolucionan, surgiendo otros nuevos como el bibliotecario de datos o los científicos de datos, que son aquellos con destreza en matemáticas, estadística e ingeniería informática (Tascón, Mario, 2013), abriéndoseles un nuevo campo de trabajo y explotación, del que, a día de hoy, aún no son del todo conscientes.

Según Mayer-Schönberger, Viktor y Cukier, Kenneth (2013, p. 232), “los datos masivos están en el punto de remodelar nuestro modo de vivir, trabajar y pensar”, y dado que el conocimiento se almacena en las bibliotecas, los bibliotecarios son las personas más idóneas para descubrir y mostrar esas nuevas formas de vivir, de trabajar y de pensar.

HIPÓTESIS DE TRABAJO Y PRINCIPALES OBJETIVOS A ALCANZAR (MÁXIMO 50 LÍNEAS):
WORKING HYPOTHESIS AND PRINCIPAL OBJECTIVES SOUGHT (50 LINE MAXIMUM):

Una revisión bibliográfica sobre la temática escogida, ha permitido definir las siguientes hipótesis:

1.- El Big Data se presenta como un paso más en la escala evolutiva de la tecnología, empleándolos en empresas e instituciones de manera que éstas obtienen grandes beneficios y un carácter diferenciador ante la competencia.

2.- Las bibliotecas deben asumir el Big Data para obtener una mejora en la calidad de sus servicios y procesos, lo que les permita competir en el actual mercado de información.

Para llevar a cabo las hipótesis, se precisará definir indicadores y variables que ayuden a demostrar si el Big Data resulta exitoso o no en una institución, organización o empresa.

El objetivo general que se persigue mediante esta investigación, trata de demostrar la utilidad del Big Data en bibliotecas.

A partir de este objetivo general, se pueden definir una serie de objetivos específicos:

- Definir el Big Data dentro de la Sociedad del Conocimiento.
- Explicar la utilidad del Big Data a lo largo de la historia hasta la actualidad.
- Determinar la situación actual del Big Data en distintos entornos: histórico, económico, político, tecnológico y social.
- Señalar los diferentes medios y tecnologías empleadas en el entorno bibliotecario, así como las técnicas y métodos de recogida de datos en bibliotecas.
- Identificar el Big Data en la biblioteca y su potencial uso.
- Describir y analizar datos de uso, procesos, usuarios, tecnologías etc. obtenidos en una muestra, en una biblioteca física.
- Clarificar datos relativos a densidad de población, hábitos de lectura, intereses de la población, aficiones etc. obtenidos de fuentes externas, susceptibles de ser combinados con el conjunto de datos obtenidos en la muestra de la biblioteca física.
- Construir e implementar un sistema de aplicación de Big Data en una biblioteca piloto.
- Analizar los resultados obtenidos mediante la aplicación del Big Data en la biblioteca piloto y estudiar la conveniencia de incorporar el Big Data en las bibliotecas.
- Enumerar una serie de recomendaciones y predicciones para el futuro empleo del Big Data en bibliotecas.

METODOLOGÍA A UTILIZAR (APORTAR CONFORMIDAD/INFORMES/PROTOCOLOS GARANTIZANDO BIOÉTICA/BIOSEGURIDAD SI EL TIPO DE EXPERIMENTACIÓN LO REQUIERE) (MÁXIMO 50 LÍNEAS):
METHODOLOGY TO BE USED (PROVIDE CONSENT FORMS/REPORTS/PROTOCOLS GUARANTEEING BIOETHICS/BIOSECURITY IF REQUIRED BY THE TYPE OF EXPERIMENTATION) (50 LINE MAXIMUM):

La metodología que se llevará a cabo a lo largo de la investigación, se basará en múltiples técnicas:

- Revisión bibliográfica de la literatura relevante, tanto nacionales como internacionales, en la temática seleccionada, el Big Data. Con ello, se facilitará un acercamiento al estado de la cuestión del Big Data, así como una definición y alcance de éstos.
Esto, también nos permitirá precisar una serie de indicadores a utilizar en el caso de uso en el entorno bibliotecario.
- Implementación en una biblioteca a determinar (biblioteca piloto), de un sistema de recogida de datos durante un periodo de tiempo limitado. Para la recogida de esta información, se utilizarán medios tradicionales y/o digitales, ya sean encuestas, chats, apps etc., y/o recogiendo directamente en los perfiles de Twitter o Facebook de la propia biblioteca mediante programas informáticos de recogida de información de estas fuentes, ya sean crawlers u otras herramientas como Topsy o Tweetdeck.
- El análisis de la información conseguida, arrojará información sobre intereses de los usuarios en materia de lectura, gustos o aficiones de los usuarios de la biblioteca, así como información útil al bibliotecario sobre adquisiciones futuras, funcionamiento de los servicios de la biblioteca o rendimiento de la misma.
La combinación de esta información con datos obtenidos de fuentes públicas (INE, REBIUN, Observatorio de la Lectura y el Libro, DILVE...) permitirá elaborar un servicio o sistema que aporte un beneficio en el rendimiento, función o servicios de la biblioteca en cuestión, incorporando un valor añadido que antes no existía.
- Implementación en la biblioteca del nuevo sistema de mejora del rendimiento, función o servicios basado en Big Data.
- Análisis de los resultados conseguidos al implementar la tecnología del Big Data en la biblioteca.
- Establecimiento de las conclusiones, así como una serie de recomendaciones que permitan el máximo rendimiento de la tecnología Big Data en la biblioteca.

MEDIOS Y RECURSOS MATERIALES DISPONIBLES (MÁXIMO 50 LÍNEAS):
MATERIAL MEANS AND RESOURCES AVAILABLE (50 LINE MAXIMUM):

La presente investigación se surtirá de la bibliografía y recursos disponibles en centros de documentación y bibliotecas (tanto físicos como digitales), necesarios para complementar la tesis sobre Big Data que se produzcan a lo largo del tiempo hasta la finalización de la misma.

Para la realización del caso de uso, se buscará la colaboración de organismos institucionales preferentemente públicos o, privados y se utilizarán la información, datos y recursos que éstos pongan a disposición del doctorando, siempre sujetos a un presupuesto o a la posible financiación que pueda obtenerse en un futuro.

Señalar que dicho proyecto tiene la intencionalidad de presentarse a convocatorias de proyectos de investigación, con el fin de conseguir apoyo tanto económico como recursos humanos.

En determinadas circunstancias, puede hacerse necesario consultar con un experto informático, un matemático, un estadista o un ingeniero de sistemas para poder continuar avanzando en la elaboración de la tesis.

Así pues, los recursos correrán a cuenta de la autonomía del doctorando, así como de la información y recursos que la directora pueda proveer, siempre contando con los materiales que la Universidad de Salamanca pone a disposición de sus estudiantes.

PLANIFICACIÓN TEMPORAL AJUSTADA A TRES AÑOS / CINCO AÑOS (Tiempo parcial) (MÁXIMO 50 LÍNEAS):

TIMING SCHEDULE OVER THREE YEARS / FIVE YEARS (Part time)(50 LINE MAXIMUM):

	Curso 2014/2015		Curso 2015/2016		Curso 2016/2017		Curso 2017/2018		Curso 2019/2020	
	Sept.-Ene.	Feb.-Jun.	Sept.-Ene.	Feb.-Jun.	Sept.-Ene.	Feb.-Jun.	Sept.-Ene.	Feb.-Jun.	Sept.-Ene.	Feb.-Jun.
Revisión bibliográfica										
Asistencia a seminarios/cursos/congresos										
Publicaciones en revistas/congresos										
Redacción de la tesis										
Definición del tema										
Preparación Plan de Investigación										
Elaboración índice provisional										
Definición de variables e indicadores a analizar										
Recogida de datos en biblioteca piloto										
Análisis datos biblioteca piloto										
Recogida de datos en fuentes externas										
Análisis datos de fuentes externas										
Aplicación Big Data en biblioteca Piloto										
Análisis de resultados obtenidos Big Data										
Elaboración conclusiones										
Defensa de la tesis										

REFERENCIAS BIBLIOGRÁFICAS (MÁXIMO 50 LÍNEAS):
BIBLIOGRAPHICAL REFERENCES (50 LINE MAXIMUM):

- Big Bang Data (2015). [Exposición]. Madrid: Fundación Telefónica
- Funk-e Studios (2013, 12 nov.). What is Big Data and how does it work?. [Archivo de vídeo]. Consulta: 12/03/2015. Disponible en: <https://goo.gl/cxDrGT>
- García Calero, Jesús. (2015, 13 abr.). Ana Santos: "El Big Data español está aquí, en la Biblioteca Nacional". ABC.es, *cultura*. Consulta: 22/04/2015. Disponible en: <http://goo.gl/SXrldj>
- Johnston, Leslie (2012). Digital collection as Big Data. Consulta: 12/05/2015. Disponible en: <http://goo.gl/4bc26k>
- Martínez-Martínez, Silvia y Lara-Navarra, Pablo. (2014). El big data transforma la interpretación de los medios sociales. *El profesional de la información*. Vol. 23, nº 6, noviembre-diciembre, 575-581
- Mayer-Schönberger, Viktor y Cukier, Kenneth. (2013). *Data: La revolución de los datos masivos*. Madrid: Turner
- Reinhalter, Lauren y Wittmann, Rachel J. (2014). The library: Big Data's Boomtown. *Serials librarian: From the printed page to the digital age*. Vol. 67, nº 4, 363-372
- Ríos Hilario, Ana Belén y Díaz Rodríguez, Jesús. (2014). Big Data Big Match. *Archivamos*. Nº 94, pp. 29-36.
- RUSSOM, Phillip (2011). Defining Big Data Via the Three Vs. *Big data analytics. TDWI Best practices report*. 4th quarter. Consulta: 25/01/2015. Disponible en: <http://goo.gl/QPuVMR>
- Schmarzo, Bill. (2014). *Big Data, el poder de los datos*. Madrid: Anaya Multimedia
- Tascón, Mario. (2013). Introducción Big Data: pasado, presente y futuro. *Revista TELOS: Cuadernos de Comunicación e Innovación*. Vol. 23, nº 6, junio-septiembre, 47-50