

NextLib: un sistema de software basado en ontologías para la consulta automática de OPACs desde sedes web especializadas

[[Versió catalana](#)]



LLEDÓ FELIP VIDAL

Becaria de documentación en el Servicio d'Investigació i Estudis Sociolingüístics (SIES)
Conselleria de Educación de la Generalitat Valenciana
felip.ile@gva.es

ENRIQUE ORDUÑA-MALEA 

Estudiante de postgrado y colaborador
Facultad de Informática de la Universidad Politécnica de Valencia
enorma@posgrado.upv.es

Opciones

    [<meta />](#)
[Imprimir](#) [Recomanar](#) [Citació](#) [Estadístiques](#) [Metadades](#)

Resumen [[Resum](#)] [[Abstract](#)]

Las extensiones para navegadores web (*Add-ons*) están ampliando las funcionalidades de éstos. Entre ellas cabe destacar la reciente aparición de LibX, herramienta que permite la interacción entre el OPAC de una biblioteca y la Web. Se estudian las limitaciones de esta herramienta en cuanto a detección automática de datos y automatización de tareas para proponer un sistema (NextLib) que, mediante la combinación de diferentes aplicaciones, permita superarlas. Este sistema estaría formado por LibX, una base de datos semántica, un software de detección automática de datos (Data Detector) y un programa que aprende por ejemplos (*learning by example*, LBE). La integración de estas herramientas permitiría una mayor interacción entre el usuario y una sede web, y entre ésta y un OPAC.

1 Introducción

"La Web Semántica es un conjunto de iniciativas, tecnológicas en su mayor parte, destinadas a crear una futura World Wide Web en la cual los ordenadores puedan procesar la información, esto es, representarla, encontrarla, gestionarla, como si los ordenadores poseyeran inteligencia"(Codina, 2003).

Como se puede interpretar de esta definición, la web semántica no sería más que técnicas de inteligencia artificial aplicadas a gestionar la información contenida en las páginas web de forma que el usuario, al navegar por Internet, pudiera recuperar esta información de manera más eficiente que como lo puede realizar hoy día, puesto que los programas que usaría para dicha recuperación serían capaces de razonar y aprender por sí mismos.

Esto significa que los sistemas de recuperación de información (motores de búsqueda, directorios, etc.) serían capaces de responder a preguntas del usuario pese a que, tal vez, la respuesta no se encontrara explícitamente en la base de datos. Podrían, gracias a una representación y manipulación eficiente de la información, inferir la respuesta.

Por estas razones y por ser los buscadores esos programas que permiten al usuario realizar preguntas y recuperar las páginas web donde se puede encontrar la información que responda a éstas, la mayoría de iniciativas hacia la web semántica han estado orientadas a mejorarlos, considerando a éstos como la pieza clave en la Web.

Sin embargo, no sólo los motores de búsqueda deberían ser tenidos en cuenta. Los navegadores (programas que permiten visualizar una página web) han ido evolucionando hasta comenzar a proporcionar cada vez más servicios y funcionalidades (Monistrol; Codina, 2007). Estas nuevas mejoras (inclusión de barras de tareas, herramientas específicas, etc.) están ayudando a mejorar la interactividad del usuario cuando navega por Internet.¹

En el campo de la información y documentación, se destaca la reciente aparición de las barras de herramientas LibX, las cuales permiten realizar búsquedas de términos seleccionados en una página web en el OPAC de una biblioteca, produciendo una relación usuario-información-navegador-OPAC.

LibX hace posible la expansión de los servicios que una biblioteca puede ofrecer a los usuarios en Internet. Es una forma de acercar el OPAC al usuario, navegue por donde navegue éste.

Pese a sus innumerables ventajas, se constatan ciertas limitaciones. En este sentido se pretende mostrar cómo aplicando ciertas técnicas de inteligencia artificial se podrían superar. Así pues, los objetivos de este trabajo son los siguientes:

- Analizar el funcionamiento de LibX para mostrar ciertas limitaciones en su uso.
- Mostrar ciertas técnicas de inteligencia artificial que podrían mejorar las prestaciones de LibX, analizando a su vez las limitaciones que éstas poseen.
- Proponer una integración de todas las aplicaciones vistas de forma que unas complementen a las otras formando un sistema, que hemos denominado NextLib.

En primer lugar se estudia la barra de herramientas LibX, su aparición, servicios que proporciona, bibliotecas que lo implementan, funcionamiento y restricciones de uso.

Después se presentan las aplicaciones de detección automática de datos como posible vía para solucionar algunas de las limitaciones de LibX. Se estudia su incorporación en navegadores a través de Smart Tags y AutoLink, analizando éste último. Se indican a su vez las limitaciones que estos programas tienen en cuanto a semántica, realización de acciones y ciertos problemas éticos asociados a su uso.

Posteriormente, se estudia el uso de las bases de datos semánticas como herramienta que, a su vez, puede solucionar las limitaciones de los Data Detector. Se señala la conveniencia de desarrollar este tipo de bases de datos para su aplicación en dominios concretos, mediante la construcción de ontologías.

Finalmente, tras mostrar dos aplicaciones pioneras (Miro y Creo), se propone, partiendo de la base de estas investigaciones, un sistema (NextLib) basado en el uso combinado de LibX, Data Detector, bases de datos semánticas y programas que aprenden con ejemplos (*learning by example*, LBE) como método para mejorar LibX y conseguir una navegación más orientada al objetivo del usuario.

2 Material y métodos

Para el análisis de LibX se procedió a la descarga e instalación en el navegador Firefox de la versión que proporciona el consorcio Madroño, en concreto la correspondiente a la Biblioteca de la Universidad Complutense de Madrid, elegida de forma aleatoria entre el resto de las ofertadas, siendo todas ellas igualmente válidas para este estudio).

Tras su instalación se comprobó su funcionamiento mediante la selección de una palabra en la propia página web de inicio del navegador <<http://www.google.es/firefox>> y su consulta en el catálogo de la Biblioteca de la Universidad Complutense de Madrid.

Para el análisis del Data Detector AutoLink, se procedió a la descarga de la barra de tareas de Google (versión inglesa) en el navegador Internet Explorer. Su funcionamiento fue analizado a través de la detección del ISBN tras la consulta, al catálogo de la Universidad Politécnica de Valencia y de la Biblioteca Nacional de España, de una misma obra. La elección de dichos catálogos vino motivada por el hecho de que cada uno de ellos visualizaba el ISBN de forma diferente.

Para el estudio de la base de datos semántica *ConceptNet*, se descargó el kit completo desde la web del proyecto <<http://web.media.mit.edu/~hugo/conceptnet/#download>>.

Finalmente, el estudio directo del Data Detector Miro y del LBE Creo no se pudo realizar, pues los autores nos comunicaron que el software no estaba disponible. Su estudio se basa, por lo tanto, en los trabajos publicados.

Todas estas tareas fueron realizadas durante el mes de mayo de 2007.

3 Firefox Extension For Libraries (LibX)

LibX es una herramienta desarrollada en 2005 por la Virginia Tech University Library y el VT Department of Computer Science. Funciona como una extensión del navegador Firefox que permite un acceso integrado a los recursos de una biblioteca desde el navegador, a modo de barra de tareas (Bailey; Godmar, 2006).

Una de las principales motivaciones por las que nació este proyecto fueron los resultados de un estudio realizado por la OCLC en 2005, titulado *College students': perceptions of libraries and information resources* (De Rosa, 2005), donde se indicaba que aproximadamente el 89% de estudiantes comenzaba un proceso de búsqueda (sobre un tema concreto) mediante un motor de búsqueda, mientras que

únicamente un 2% lo comenzaba a partir del OPAC de una biblioteca.

Mientras que los motores de búsqueda proporcionan indudables ventajas como son la velocidad, facilidad de uso y bajos costes para el usuario, las bibliotecas proporcionan a sus búsquedas una mayor precisión y credibilidad en sus resultados.

Así pues, la idea fue intentar ofrecer una herramienta que fuera capaz de integrar todas las ventajas: la búsqueda en un catálogo de una biblioteca desde un navegador a través de una barra de tareas.

Los servicios que proporciona son básicamente los siguientes:

- Acceso al catálogo de la biblioteca mediante una barra de tareas integrada en el navegador.
- Búsquedas directas de términos seleccionados en Google Scholar.
- Posibilidad de crear ecuaciones de búsqueda directamente en la barra de tareas.
- Posibilidad de soportar múltiples OPACs de forma simultánea.
- Búsquedas mediante el servicio XISBN (instalación adicional).
- Diversas posibilidades de selección de términos de búsqueda.

En la actualidad aproximadamente 60 bibliotecas (la mayoría de EE.UU.) tanto académicas como públicas ofrecen servicios de LibX y cerca de 90 están en período de prueba.²

En España, la mayor iniciativa al respecto la ha tomado el Consorcio Madroño, quien ha desarrollado una adaptación de la barra de tareas LibX para las bibliotecas que forman parte del mismo.

Se ofrece la posibilidad de descargar una barra de tareas común a todas las bibliotecas del consorcio o bien una barra de tareas para cada biblioteca. Con el objetivo de probar el servicio, se procedió a la descarga de la barra LibX correspondiente a la Biblioteca de la Universidad Complutense de Madrid.



Figura 1. Selección de texto (fuente: elaboración propia)

En la parte superior de la figura 1 se encuentra la barra LibX, instalada e insertada correctamente en el navegador.

El usuario debe elegir qué palabra o frase desea mediante su selección directa con el ratón. En este caso se ha seleccionado la palabra "navegación" que aparece en la página de inicio de Firefox. La selección se puede realizar desde cualquier página web que esté consultando el usuario.

La elección se puede realizar de dos formas diferentes:

- Mediante un clic con el botón derecho del ratón.
- Seleccionando y arrastrando la palabra hacia las distintas zonas de búsqueda de la barra ("Scholar", "Search catálogo completo", "Cuadro de búsqueda").

En el caso del "Cuadro de búsqueda", se tiene que escoger el rol del texto elegido. Por ejemplo, especificaremos si deseamos que "navegación" sea considerado como materia, como descriptor, etc. En este caso, la función es más obvia, aunque si hubiésemos escogido un nombre de persona no lo sería tanto.

Tras realizar la selección, pulsando "intro", el navegador abre automáticamente el catálogo de la Biblioteca de la Universidad Complutense en una nueva ventana (aunque se puede configurar de forma que abra una nueva pestaña, o se abra en la misma ventana), realiza la búsqueda automáticamente y muestra el resultado.

Además, podemos abrir cuantos cuadros de búsqueda deseemos, de forma que la ecuación de búsqueda sea más completa; aunque de momento no permite operadores booleanos (por defecto realiza un "AND").



Figura 2. Cuadro de búsqueda LibX extendido (fuente: elaboración propia)

Para el ejemplo anterior, obtenemos la pantalla de resultados que se muestra en la Figura 2, donde el sistema ha recuperado un resultado mediante la búsqueda a través de la materia "navegación". En la parte superior se muestra el aspecto que tienen los cuadros de búsqueda, preparados para que el usuario introduzca nuevos términos.

Si el término elegido se arrastra al icono "Scholar", situado en la parte superior derecha de la pantalla, el sistema realiza una búsqueda automática del término en Google Scholar.

A continuación se muestra la pantalla de resultados que obtendríamos para "navegación":



Figura 3. Búsqueda automática en Google Scholar (fuente: elaboración propia)

Como se ha podido observar, el uso de las barras de tareas LibX permite combinar la facilidad y rapidez de la navegación en la Web, con la precisión y confianza que proporcionan las búsquedas realizadas en el OPAC de una biblioteca. La posibilidad de seleccionar una palabra o frase y poder realizar búsquedas en catálogos o en otros lugares abre una vía para ofrecer una mayor interactividad al usuario.

Sin embargo, pese a las indudables ventajas que ofrece, existen una serie de limitaciones:

- El usuario debe elegir una palabra o frase concreta y decidir su rol. Esto provoca que ciertas acciones resulten complejas, fundamentalmente aquéllas que precisen de la selección de un elevado número de términos.
- Las funciones son limitadas y no ampliables sin programación.
- Las acciones están exclusivamente relacionadas con recursos del OPAC.
- El sistema no recuerda las acciones del usuario.

Muchas de las limitaciones de LibX se podrían resolver mediante el uso de un tipo de programas llamados Data Detector. Pero éstos, a su vez, presentan una serie de problemas relacionados con la semántica y las acciones de usuarios.

4 Data Detector

Un Data Detector es un software capaz de reconocer estructuras de datos y permitir la realización de operaciones múltiples con ellos. Su función es relacionar los tipos de datos detectados con acciones particulares para optimizar tiempo en la realización de acciones rutinarias para el usuario y ganar en la usabilidad de aplicaciones.

El desarrollo de este tipo de aplicaciones empezó en los años 90 centrándose en la cantidad de clases diferentes de datos que podían identificar y en la flexibilidad a la hora de configurar acciones coherentes con los tipos de datos seleccionados. Desde entonces hasta la actualidad los Data Detector han ido evolucionando, desde su primer empleo en el escritorio y documentos del usuario, hasta su reciente salto a la Web.

Aunque la tecnología que hace posible su funcionamiento está ampliamente superada en la actualidad, su implementación en los navegadores abre un mundo de posibilidades por descubrir y convierte a los Data Detector en la base para la construcción de interfaces web orientadas a objetivos de usuario.

Los Data Detector llegaron a los navegadores de la mano de grandes firmas en el sector de la informática e Internet como son Microsoft con Smart Tags y Google con AutoLink. La enorme aceptación de los productos comerciales de estas empresas por los usuarios levantó las sospechas y recelos de los desarrolladores de sitios web que veían en el uso de estas aplicaciones un aprovechamiento indebido de los contenidos web (Hugues; Carr, 2002).

Smart Tags, lanzadas en 2001 como parte del Windows XP, permitían a Microsoft insertar sus propios enlaces en cualquier página web que fuera visualizada a través del navegador Internet Explorer. Esos enlaces aparecían como un subrayado discontinuo de color púrpura, para diferenciarlos de los enlaces originales. Cuando el cursor se colocaba sobre una palabra marcada como Smart Tag, aparecía una lista de enlaces relacionados con esa palabra. Ante las enormes críticas recibidas, Microsoft puso la tecnología Smart Tag disponible para los usuarios que la desearan, pero no como parte integrante de su sistema operativo, al tiempo que desarrolló una etiqueta "meta" que permitía a los desarrolladores web deshabilitar la función Smart Tags en sus páginas.

AutoLink, lanzado en 2005, es capaz de escanear una página web y detectar 3 tipos diferentes de datos: direcciones postales, identificadores de vehículos (VIN), e identificadores de libros (ISBN).

Aunque en principio sólo funcionaba para EE.UU., poco a poco sus funcionalidades van aumentando, aunque no existe aún mucha información disponible acerca de cómo funciona realmente.

A pesar de que sólo es capaz de detectar tres tipos de datos, AutoLink tiene un interés especial para el documentalista al ser capaz de detectar ISBN y asociarles una acción.

Un ejemplo de uso en la detección de ISBN se muestra seguidamente:



Figura 4. Ejemplo de AutoLink (fuente: elaboración propia)

En esta figura observamos como, para cierta búsqueda realizada en el catálogo de la Universidad Politécnica de Valencia <<http://www.upv.es/bib>> de la obra *Archivese*, en la parte superior se activa la función de AutoLink mostrando el mensaje: "Show Book Info". Tras pulsar, detecta los ISBN escritos en la pantalla del navegador.

En este caso se ilumina el correspondiente ISBN en la ficha catalográfica. Colocando el cursor encima del ISBN se observa que éste se ha transformado en un hipervínculo que apunta a una URL, en este caso al catálogo compartido *WorldCat*.

Pulsando el nuevo link se abre el catálogo *WorldCat*, donde el sistema ha realizado automáticamente una búsqueda en su catálogo con ese ISBN, mostrando el resultado.

Pese a todo, este sistema presenta una serie de inconvenientes de cara a conseguir una verdadera interactividad con los usuarios. Estos inconvenientes son:

- Búsquedas predeterminadas: el usuario no puede añadir nuevos enlaces.
- Algoritmo de detección de ISBN no muy eficiente.

Veamos con ejemplos estos dos problemas:

Ejemplares	Ubicación	Préstamo
7-31/10B	Biblioteca Central	Sala Préstamo 7 días
7-31/897B	Biblioteca Central	Sala Prestado hasta 07/05/07
7-31/1764B	Biblioteca Central	Sala Prestado hasta 07/11/06
7-31/2028B	Biblioteca Central	Sala Préstamo 7 días

Figura 5. La detección del ISBN de AutoLink (fuente: elaboración propia)

Para el mismo ejemplo anterior, se muestran las posibilidades que ofrece AutoLink. Como se comentó, está predeterminado para *WorldCat*. El usuario no puede en ningún momento modificar estas opciones, ni añadir nuevos enlaces de interés, por lo que la interactividad es limitada.

Respecto al algoritmo de detección, se ha realizado una búsqueda del mismo libro en el catálogo de la Biblioteca Nacional de España <<http://www.bne.es>>. En la Figura 6 podemos observar como el ISBN no se muestra de la misma forma que antes (los 10 números seguidos) sino que intercala guiones para separar las cuatro áreas del código ISBN. AutoLink no detecta el ISBN, mostrando el mensaje "No Items Found".

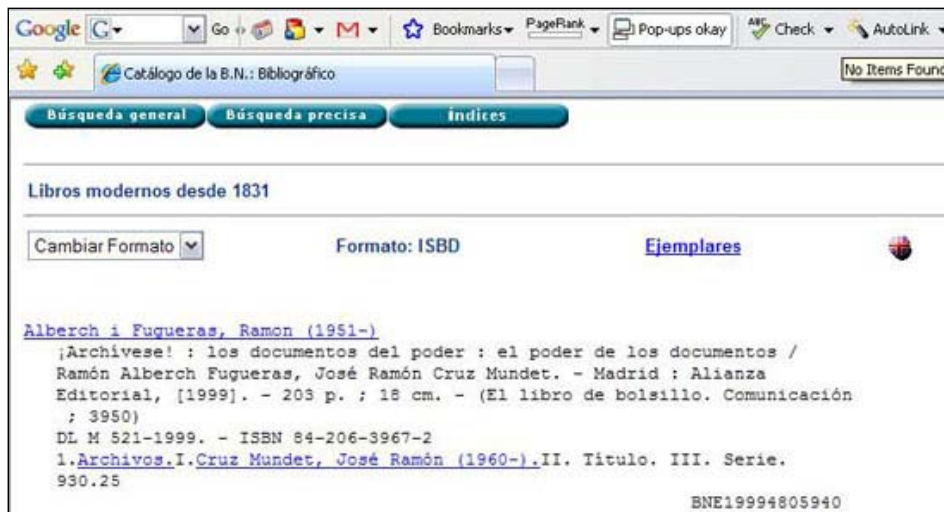


Figura 6. Catálogo de la Biblioteca Nacional de España (fuente: elaboración propia)

4.1 Limitaciones y problemas

Los Data Detector presentan una serie de problemas relacionados directamente con las dos principales tareas que desarrollan: la identificación de estructuras de datos y la asociación de acciones a cada una de éstas. Ambas tareas están programadas desde el principio, por lo que son sistemas cerrados. Sin embargo, la dificultad de programar a priori todos los tipos de datos que van a querer usarse y todas las acciones que pretenden realizarse con ellos puede superarse (en dominios acotados) con el uso de técnicas de inteligencia artificial.

Las aplicaciones LBE son sistemas de aprendizaje que permiten grabar secuencias de acciones asociadas al objetivo del usuario y que se pueden reproducir cuando el objetivo se repite. Permitirían generalizar acciones de usuario. Aunque el principal problema que presentan es, justamente, la generalización; ésta se va perfeccionando por la constante evolución de estos sistemas y su uso asociado a bases de conocimiento semántico.

La detección de los datos tiene como obstáculo la ambigüedad del lenguaje. El uso de bases de datos con conocimiento semántico y la concreción a un campo más específico de conocimiento expandiría el número de datos capaces de ser reconocidos y eliminaría la ambigüedad, respectivamente.

5 Bases de datos semánticas

Las bases de datos con contenido semántico intentan representar las relaciones que los humanos establecen entre las palabras, procurando no sólo realizar ontologías y controlar el vocabulario desde un punto de vista léxico sino desde el punto de vista conceptual.

De modo natural, el ser humano establece relaciones entre los conceptos que van más allá de la forma sintáctica o gramatical de una frase, relaciones basadas en su experiencia y en aspectos espaciales, físicos, temporales y psicológicos del día a día. En inteligencia artificial este modo de concebir el mundo ha recibido el nombre de *commonsense knowledge* (el conocimiento del sentido común) (Liu; Singh, 2004a) y es el que pretenden representar estas bases de conocimiento.

Desde los inicios de la inteligencia artificial hace 50 años se está intentando representar la red conceptual del ser humano (Pitrat, 1998). El lenguaje es el principal inconveniente que se ha encontrado a la hora de intentar que un ordenador imite el comportamiento y la manera de pensar humanos.

Se pueden destacar tres proyectos de bases semánticas que abordan el estudio e intento de representar la red conceptual del ser humano: *Cyc*, *WordNet* y *ConceptNet* (Liu; Singh, 2004b). Aunque el fin es el mismo, la aproximación a la representación conceptual parte de supuestos distintos: *Cyc* pretende representar el *commonsense* en un marco formal y lógico. *WordNet* pone su énfasis en el léxico y la clasificación del mismo, mientras que *ConceptNet* intenta abarcar el lenguaje natural y representar su capacidad de realizar inferencias a través de las relaciones establecidas entre los términos.

5.1 Limitaciones

Aunque las bases de conocimiento semántico suponen grandes avances hacia el ordenador "inteligente", existe un obstáculo evidente que plantea una serie de limitaciones a las potenciales

funcionalidades de éstas, esto es, la imposibilidad de abarcar todo el conocimiento. ¿Cuánto tiempo es necesario para "capturar" toda la realidad?, ¿existen los medios técnicos para poder representar, aplicar y usar óptimamente los millones de conceptos y relaciones del *commonsense knowledge*?

En este sentido, Lluís Codina (2007) es bastante explícito:

"Insistir en que las ontologías permitirán razonar a los ordenadores y realizar inferencias sobre los contenidos, no de un dominio del conocimiento bien delimitado y para unas tareas específicas (como ya sucede), sino de la web en su conjunto y en todos los ámbitos del conocimiento y de la realidad, es, hoy por hoy, no querer ver precisamente la simple realidad. No existe ninguna evidencia empírica ni base conocida alguna para que esto pueda funcionar, ni a corto ni a medio plazo".

Por tanto se considera que para un uso adecuado y funcional de las bases semánticas se debería restringir la aplicación de éstas a un dominio concreto, donde los conceptos y relaciones que entraran en juego fueran limitados y, así, ser más fácilmente representables.

6 NexT-Lib (Navigator's Extension Tool for Libraries)

Vista la interactividad ofrecida entre la web y el OPAC de una biblioteca gracias a LibX, se considera que su integración con las diferentes aplicaciones vistas anteriormente supondría una mejora considerable.

6.1 Estudio previo

La primera gran experiencia en combinar un Data Detector con técnicas de inteligencia artificial para su aplicación en la detección de datos en la web fue el desarrollo, en 2006, por Alexander Faaborg y Henry Lieberman (MIT), de Creo (un LBE) y Miro (un Data Detector) para usarse combinados con *ConceptNet* (Faaborg; Lieberman, 2006).

Gracias al uso de la base semántica, Miro es capaz de expandir los tipos de datos que detecta a un número infinito, en concreto a todo el conocimiento almacenado en *ConceptNet*. Miro realiza un mapeo de los términos y frases que aparecen en la página web que está visualizando el usuario y realiza generalizaciones gracias a la base de datos semántica. Estas generalizaciones las compara con las generalizaciones hechas previamente con Creo, que describen tareas de usuario. De esta forma es capaz de asociar la generalización de un tipo de dato con una acción, que en ningún caso es predeterminada, sino que el ordenador, gracias a Creo, es capaz de ir creando y aprendiendo en función de las tareas que realiza un usuario.

Los resultados obtenidos por estos investigadores hacen presagiar que esta línea de investigación puede proporcionar unos resultados muy interesantes y servir como base para conseguir una interacción entre usuario y web basada en los objetivos de éste.

6.2 El sistema NextLib

Partiendo del análisis previo y teniendo en cuenta las ventajas e inconvenientes de las herramientas que se desean usar, se propone una integración de las mismas, NextLib: un sistema formado por LibX, un Data Detector, un LBE y una base de datos semántica, que permitiría aumentar las prestaciones de LibX (u otra aplicación similar).

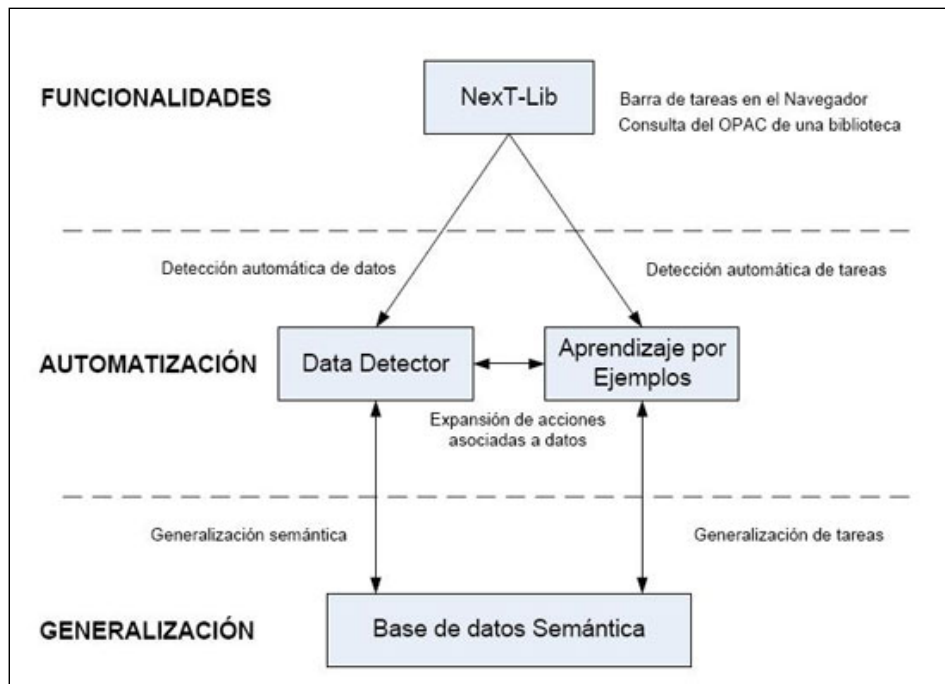


Figura 7. Modelo NextLib (fuente: elaboración propia)

6.3 Componentes

El Data Detector permitiría seleccionar, en una determinada página web, todos los conceptos que correspondiesen o se relacionaran con un tipo de dato (autor, editor, título, materia o ISBN). Por tanto, no sería necesario que el usuario seleccionara la palabra y le indicara su rol (aunque podría hacerlo). Asimismo, se detectarían sólo aquellos conceptos que integraran la base de conocimiento semántico, resultando esta consideración fundamental para comprender el sistema.

El LBE (no entra en los objetivos del trabajo el análisis completo de este tipo de aplicaciones) favorecería la adaptación del sistema al gusto del usuario permitiendo relacionar determinadas acciones a los tipos de datos, grabándolas y ampliando la lista de acciones básicas a aquellas que el usuario considerara de interés, por ejemplo añadir enlaces de interés a páginas relacionadas con un determinado concepto.

El uso de la base de conocimiento semántico queda justificado por la aplicación de los Data Detector y los LBE. Ambos programas son dependientes de las bases de conocimiento semántico, pues se detectan los conceptos que pertenecen o se relacionan con una clase determinada (autor, título,...) y las acciones se asocian a esas mismas clases.

6.4 Funcionamiento

NextLib podría realizar búsquedas en el OPAC de una o varias bibliotecas a partir de palabras seleccionadas (como LibX) o detectadas en una página de una determinada sede web y ejecutar nuevas acciones y nuevas búsquedas con conceptos relacionados.

La idea es aplicar NextLib a un sitio web concreto con contenido especializado en un ámbito determinado del conocimiento, permitiendo de esta forma la construcción de una base de datos semántica (pretender abarcar un conocimiento más general no sería viable), y relacionar automáticamente los términos detectados con el OPAC de una biblioteca, centro de documentación, librería, etc., con unos fondos acordes a dicha temática.

Una vez el usuario navegara dentro de este *website*, podría poner en funcionamiento NextLib y detectar automáticamente los términos de interés (únicamente los referidos al ámbito de conocimiento en cuestión, no todos los contenidos en la sede) de una determinada página web. Gracias a la base de datos semántica se podrían realizar las generalizaciones apropiadas ofreciendo unos términos generales y unas acciones asociadas, una de las cuales podría ser la consulta en un OPAC determinado (tal como ya realiza LibX).

Suponiendo, por ejemplo, la navegación por la sede web de un museo y el desarrollo correspondiente de una ontología sobre Arte -parte de la cual podría ser la mostrada en la figura 8-, un usuario que navegara por dicha sede, podría seleccionar (o detectar) la palabra "Expresionismo Abstracto" -como ya realiza LibX. El Data Detector entraría en acción consultando la base semántica que, por inferencia, llegaría a Lichtenstein (y otros pintores), que es el nodo central, ofreciendo posibles acciones -

mejorando por ello LibX-, como por ejemplo:

- Búsqueda en la Wikipedia.
- Búsqueda en el OPAC, con sus posibles variantes:
 - Por el propio término, si existe como nodo central (materia, título) en la base semántica.
 - Por los términos inferidos, esto es, Lichtenstein y el resto de artistas así relacionados en la base semántica.
- Acciones programadas por el propio usuario.

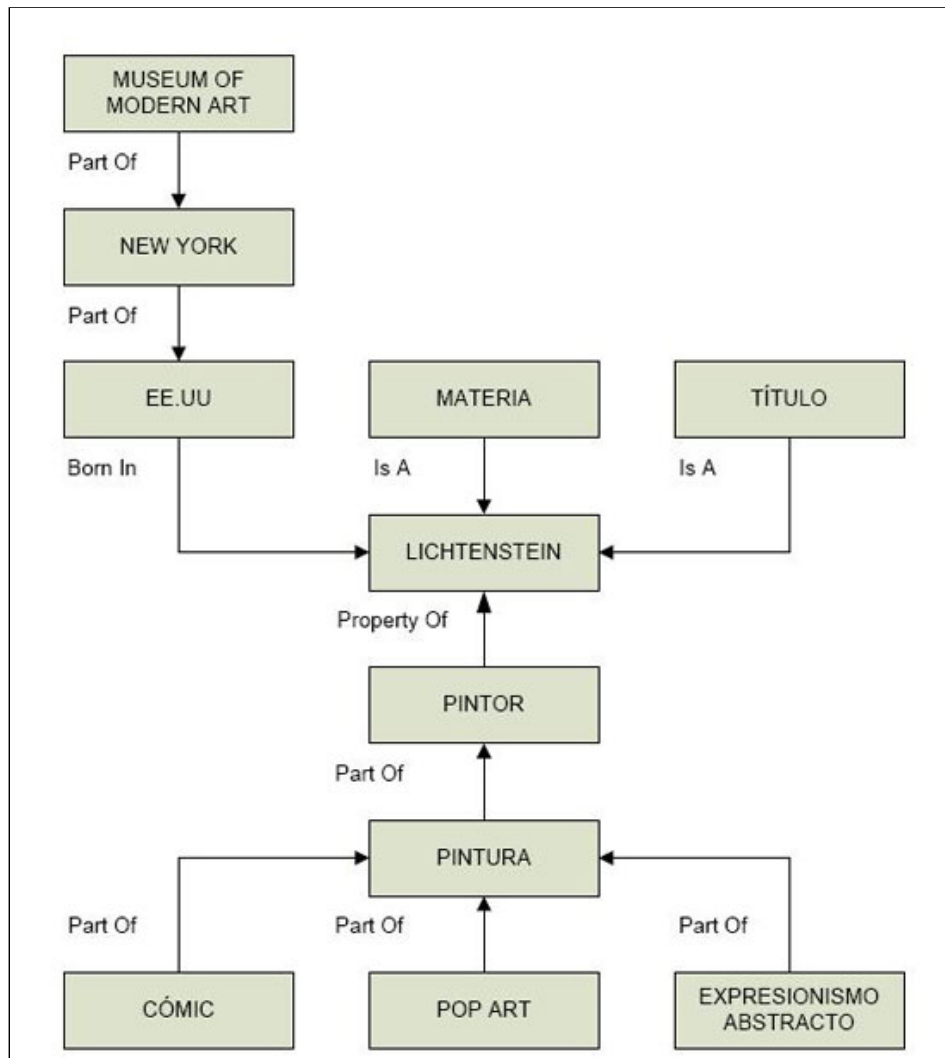


Figura 8. Relaciones semánticas sobre arte moderno (fuente: elaboración propia)

Escapa a los objetivos de este trabajo el diseño completo de las relaciones que compondrían esta red semántica, la figura precedente es tan sólo un ejemplo, a modo orientativo, usando las relaciones de *ConceptNet*.

Debemos establecer una clara diferencia entre la base de datos bibliográfica, accesible desde el OPAC, que recoge todos los datos relacionados con una colección, y la base de conocimiento semántico, que establece las relaciones conceptuales entre los términos explícitos e implícitos que aparecen en una determinada sede web, con contenido acotado a una rama del conocimiento concreto.

Sus propósitos y funciones son diferentes: la base de datos semántica infiere datos y, a través del Data Detector y el LBE, ofrece opciones de búsqueda, mientras que la base de datos bibliográfica, consultada desde el OPAC, recupera registros bibliográficos en función de dichas opciones.

6.5 Ventajas

Las principales ventajas de usar un sistema de estas características serían las siguientes:

- La ampliación de los servicios de una biblioteca (o librería) a los usuarios de una sede web especializada.

- Una mayor interacción del usuario, a través del navegador, con la información especializada contenida en una página web, favoreciendo el acceso tanto a otras (gracias a acciones de enlace), como a material disponible en determinadas bibliotecas (gracias a acciones de búsqueda en el OPAC).
- Mayor facilidad para encontrar información dentro de la propia sede web. Los desarrolladores de la misma podrían preconfigurar NextLib con enlaces internos entre conceptos importantes.
- Posibilidad de cooperación entre sedes con una misma temática, al compartir una misma ontología.

6.6 Limitaciones

Pese a las ventajas que supondría este sistema, su implementación debería tener en cuenta una serie de limitaciones. Destacan, entre otras:

- La detección del ISBN: los algoritmos de detección deberían tener en cuenta las distintas formas en que aparecen representados los ISBN.
- La convergencia OPAC-NextLib: Las prestaciones de NextLib deberían estar en consonancia con las que ofreciera el OPAC de la biblioteca con el fin de evitar ecuaciones de búsquedas en el navegador imposibles de ejecutar en el OPAC.
- Sería necesario un control del vocabulario en la redacción de contenidos en las páginas del *website*, de acuerdo a la ontología creada.
- Así como LibX extiende su uso a cualquier página web por la que se navegue, NextLib debería circunscribirse a un sitio web concreto, pues su uso estaría limitado a la ontología con la que trabajase.
- Sólo funcionaría en determinadas sedes con contenido muy específico donde una ontología pudiera ser elaborada, como el ejemplo mostrado de un museo, portales temáticos, científicos, etc.

7 Conclusiones

Las herramientas como LibX, elaboradas con el fin de facilitar la interacción entre la Web, el usuario y el OPAC de la biblioteca, son un primer acercamiento a las interfaces de usuario orientadas al objetivo, aplicadas al mundo de la documentación. Pese a sus ventajas, presenta limitaciones en la detección automática de datos y la automatización de tareas que podrían resolverse con la aplicación de Data Detector y programas LBE.

La implementación de Data Detector en la web (AutoLink y Smart Tag) ha supuesto un reto en la detección de nuevas tipologías de datos. El análisis realizado de estas herramientas ha mostrado su poca precisión a la hora de detectar conceptos –e incluso códigos (ISBN)– derivada de la ambigüedad del lenguaje. Por ello, las bases de datos semánticas juegan un papel indiscutible para favorecer, ampliar y adecuar las nuevas necesidades de detección de los usuarios, pero acotadas a campos concretos del saber.

NextLib es un sistema de aplicaciones, constituido por la combinación de proyectos y programas informáticos que, aisladamente, resuelven pequeñas dificultades y que, juntos, conforman un sistema potencialmente más interactivo, eficiente y orientado al usuario, que los existentes hasta ahora.

Como trabajo futuro de investigación, sería necesario un estudio directo del Data Detector Miro y del LBE Creo (y/o aplicaciones similares) para poder contrastar nuestro planteamiento, así como un análisis de la interacción entre LibX y el resto de las aplicaciones del sistema NextLib.

Asimismo, estudios sobre la dificultad de uso de un sistema de estas características por parte de los usuarios, tomando como base el realizado por A. Faaborg, y H. Lieberman con Creo y Miro, serían de interés.

Los resultados expuestos a lo largo del trabajo evidencian las posibilidades que la investigación en esta línea ofrece tanto a las bibliotecas como a los desarrolladores de sedes web.

Bibliografía

Bailey, A.; Godmar, B. (2006). "LibX: a Firefox extension for enhanced library access". *Library hi tech*, vol. 24, no. 2, p. 290–304.

Cantera, J. M.; Hierro, J. J.; Romo, P. A. (2006). "Web semántica: tecnologías y arquitectura". *Comunicaciones de telefónica I+D*, n.º 39. <http://www.tid.es/documentos/revista_comunicaciones_i%2Bd/numero39.pdf>. [Consulta: 09/05/07].

Codina, L. (2003). "Internet invisible y web semántica: ¿el futuro de los sistemas de información en línea?". *Tradumática*, núm. 2. <<http://www.fti.uab.es/tradumatica/revista/num2/articles/06/06central.htm>>.

[Consulta: 09/05/07].

Codina, L. (2007). "La web semántica y el vaporware del siglo xxi". *Anuario thinkEPI 2008* [en prensa].

Dans, E. *El blog de Enrique Dans*. <<http://www.enriquedans.com>>. [Consulta: 09/05/07].

De Rosa, C. et al. (2005). "Perceptions of libraries and information resources: a report to the OCLC membership". Dublin (Ohio): OCLC. <http://www.oclc.org/reports/pdfs/Percept_all.pdf>. [Accessed: 09/05/07].

Dey, A.; Abowd, G.; Wood, A. (1998). "CyberDesk: a framework for providing self-integrating context-aware services". *Proceedings of the International Conference on Intelligent User Interfaces (IUI 98)*. <<ftp://ftp.cc.gatech.edu/pub/gvu/tr/1997/97-10.pdf>>. [Accessed: 09/05/07].

Dix, A.; Beale, R.; Wood, A. (2000). "Architectures to make simple visualisations using simple systems". *Proceedings of advanced visual interfaces (AVI 00)*.

Faaborg, A. (2005). "A goal-oriented user interface for personalized semantic search" [masters thesis]. Massachusetts Institute of Technology: Cambridge, USA. <<http://agents.media.mit.edu/projects/semanticsearch>>. [Accessed: 09/05/07].

Faaborg, A.; Lieberman, H. (2006). "A goal-oriented web browser". *Proceedings of the SIGCHI conference on human factors in computing systems*. Montréal, Québec, Canadá, p. 751–770. <http://alumni.media.mit.edu/~faaborg/files/thesis/draft/complete/CHI06_goalOrientedWebBrowser.pdf>. [Accessed: 09/05/07].

García Marco, F. J. (2007). "El nacimiento y el despegue de la investigación moderna sobre ontologías". *Anuario thinkEPI 2008* [en prensa].

Hughes, G.; Carr, L. (2002). "Microsoft smart tags: support, ignore or condemn them?". *Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia*. College Park: Maryland, USA. <http://eprints.ecs.soton.ac.uk/7351/01/gvh_smart_tags_ht02.pdf>. [Accessed: 09/05/07].

La flecha: tu diario de ciencia y tecnología. <<http://www.laflecha.net>>. [Consulta: 09/05/07].

LibX: a Firefox extension for libraries. <<http://www.libx.org>>. [Accessed: 09/05/07].

Liu, H.; Singh, P. (2004a). "ConceptNet: a practical commonsense reasoning tool-kit". *BT technology journal*, vol. 22, no 4. <<http://web.media.mit.edu/~hugo/publications/papers/BTTJ-ConceptNet.pdf>>. [Accessed: 09/05/07].

Liu, H.; Singh, P. (2004b). "Commonsense reasoning in and over natural language". *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES'2004)*. Wellington, New Zealand. September 22–24. Lecture notes in artificial intelligence, Springer. <<http://web.media.mit.edu/~hugo/publications/papers/KES2004-csr-nl.pdf>>. [Accessed: 09/05/07].

Madroño. <<http://www.consorciomadrono.net>>. [Consulta: 09/05/07].

Monistrol, M.; Codina, L. (2007). "Los navegadores de la web 2.0: Firefox, Opera y Explorer". *El profesional de la información*, vol. 16, n.º 3, p. 261–267.

Nardi, B., Miller, J.; Wright, D. (1998). "Collaborative, programmable intelligent agents". *Communications of the ACM*, vol. 41, no. 3. <<http://portal.acm.org/citation.cfm?id=272331&coll=portal&dl=ACM>>. [Accessed: 09/05/07].

Pitrat, J. (1998). "El nacimiento de la inteligencia artificial". *Mundo científico*, vol. 8, n.º 53, p. 1196–1209.

Pandit, M.; Kalbag, S. (1997). "The selection recognition agent: instant access to relevant information and operations". *Proceedings of the International Conference on Intelligent User Interfaces (IUI 97)*. <<http://portal.acm.org/citation.cfm?id=238285&coll=portal&dl=ACM>>. [Accessed: 09/05/07].

Ros, M.; Marco, M. E. *El documentalista Enredado*. <<http://www.documentalistaenredado.net>>. [Consulta: 09/05/07].

The Concept Net Project [V.2.1]: a very-large semantic network of common sense knowledge. <<http://web.media.mit.edu/~hugo/conceptnet>>. [Consulta: 09/05/07].

Fecha de recepción: 05/09/2007. Fecha de aceptación: 10/12/2007.

Notas

¹ En la siguiente URL se puede consultar el listado de Add-ons disponibles actualmente para Firefox: <<https://addons.mozilla.org/es-ES/firefox>>. [Consulta: 12/10/07].

² A fecha de 07/08/07.

Facultat de Biblioteconomia i Documentació
Universitat de Barcelona
Barcelona, juny de 2008
<http://www.ub.edu/biblio> •  [Comentaris](#)

[Recomanar](#) • [Citació](#) • [Estadístiques](#) • [Metadades](#)
Els textos publicats a *BiD* estan subjectes a una llicència de [Creative Commons](#)
[Política de privadesa](#)
[UB](#) • [Facultat](#) • [BiD](#)

