

Ferberización de una base de datos bibliográfica

Norberto Manzanos

IDIHCS-CONICET/UNLP

Resumen: Se presenta un proyecto de software que transforma registros bibliográficos de una base de datos estándar en entidades de los distintos grupos del modelo conceptual FRBR, obras, expresiones, manifestaciones, ítems, personas, instituciones y temas, así como muchas de las relaciones entre ellas. El software convierte dinámicamente los registros originales, que corresponden a las manifestaciones e ítems de FRBR, en un árbol que parte de la obra para luego ir desplegando las restantes entidades y relaciones. En la implementación presente se ha trabajado con la base de datos de la Biblioteca Central de la Facultad de Humanidades y Ciencias de la Educación de la UNLP, pero por la modularidad con que está planteado, el software puede implementarse partiendo de cualquier otra estructura. Para poder realizar este proceso, el software construye listas de autoridades para obras, expresiones, manifestaciones, personas, instituciones y temas. En la etapa actual los datos utilizados para construir las entidades y relaciones surgen exclusivamente de la información contenida en la base de datos original. Esto implica que muchos datos no podrán ser obtenidos pues no estaban contemplados en el diseño original, por lo que se plantea una segunda etapa en la cual se recolecte esa información en bases de datos remotas. Se muestra, además de las ventajas que tiene para el usuario una visualización "ferberizada", las implicancias que tiene una estructura tal para los estudios bibliométricos.

Introducción

El modelo FRBR (*Functional Requirements for Bibliographic Records*) fue propuesto por la International Federation of Library Associations (IFLA) en el año 1998 y desde entonces ha producido un gran acumulación de artículos, infinidad de debates y algunos pocos proyectos de software. Tal vez este modelo comience a tener un impacto mayor recién ahora, con la aparición de las nuevas normas de catalogación, que –al contrario de lo que se suponía en un momento– no son una reedición de las normas angloamericanas (AACR) sino que desde su nombre (RDA, *Resources Description and Access*) se presentan como una nueva concepción en la descripción bibliográfica. Y esta nueva concepción debe mucho a FRBR: RDA debe ser visto desde FRBR para poder comprenderse.

La poca cantidad de programas que contemplen las entidades y relaciones del modelo FRBR no deja de sorprender, teniendo en cuenta que para la acelerada tecnología informática, una década es un tiempo enorme. Acaso la respuesta esté en que FRBR no es un modelo completo, no es una propuesta de implementación ni es tampoco una nueva norma a la cual ajustarse disciplinadamente. Es una propuesta conceptual que los software y las normas deben considerar en forma crítica para elaborar esquemas concretos que a la vez implementen y mejoren las ideas expresadas en el modelo. En nuestro país, el impacto de FRBR ha sido casi nulo. Cabría analizar las razones de esta indiferencia, pero no es el propósito de esta ponencia, que propone un desarrollo de software concreto adaptable a la situación particular de la descripción bibliográfica en nuestro medio.

En primer lugar explicaremos brevemente lo que consideramos las ideas principales de FRBR y luego mostraremos un software que convierte dinámicamente los registros de una base de datos tradicional, es decir, una base de *manifestaciones* en un grafo de entidades FRBR. Expondremos luego muy brevemente algunas de las diferencias y ampliaciones que venimos planteando con nuestro modelo de objetos OPUS (1), y finalmente mencionaremos como desafío para otros estudios las implicancias del uso de un modelo conceptual robusto, o de la falta del mismo en las investigaciones bibliométricas.

FRBR

Si bien es posible abordar la cuestión de las entidades de FRBR desde muchos puntos de vista, para los fines de esta ponencia escogeremos algunas definiciones centradas en su cualidad informacional. Al distinguir *obra* de *documento*, Elaine Svenonius (2) nos recuerda que un documento es la mínima entidad del universo bibliográfico, un objeto físico claramente identificable, pero que no necesariamente tiene una existencia independiente. Los documentos pueden agruparse en conjuntos que son ellos mismos otras entidades bibliográficas. Pueden agruparse según posean esencialmente la misma información, y este conjunto señala a la *obra*. Al menos intuitivamente, es claro que todos las versiones, traducciones, impresiones, etc de, digamos, El Quijote, cuentan la misma historia.

Algunas versiones tienen diferencias importantes, no en su información esencial, pero sí en sus textos; si bien dos traducciones de El Quijote cuentan la misma historia, los textos, es decir, las cadenas de signos, son totalmente distintas. Siempre que hay cambios de este tipo estamos hablando de otra entidad FRBR, la *expresión*. Otros términos propuestos para esta entidad son *texto*, no en su sentido restringido sino en su sentido semiótico y *versión*. Svenonius define *texto* como el conjunto de documentos que comparten las mismas cadenas de caracteres y los mismos símbolos.

En tanto los documentos que tiene un mismo texto, pero distintas características físicas, en el caso de los libros, por ser distintas ediciones, constituyen las *manifestaciones*. Otros autores han utilizado los términos *edición* y *publicación* para esta entidad.

Las manifestaciones comparten información asociada a las características físicas.

Aquellos libros que son exactamente iguales, por provenir de una matriz común constituyen los *ítems*. Para esta entidad a veces se prefiere el término *copia* o también el término *ejemplar*.

No podemos entrar ahora en cuestiones terminológicas u ontológicas. Utilizaremos las definiciones de Svenonius porque nos dan la posibilidad de establecer un parámetro común –la información que cada entidad porta – aunque queda totalmente afuera de esta ponencia definir qué entendemos exactamente por información y cuales serían los criterios para medirla, si es que esta medición puede ser llevada a cabo. Partimos de la base de que esta idea intuitiva de información coincide con la idea que tienen de la misma los diversos usuarios de las colecciones bibliográficas.

Ferberización básica

El objetivo de este proyecto de software es, entonces, la conversión dinámica de los registros de una base de datos bibliográfica tradicional en entidades FRBR. Sin ninguna excepción, estas bases de datos han tenido como objetivo principal registrar la entidad *manifestación* y su relación con la entidad *ítem*. En cambio, son pocas las señales que se dan sobre las otras entidades. La *obra* está implícita en la responsabilidad principal que se consigna, el autor, y eventualmente en la presencia del título original. En algunos casos el título uniforme puede estar dando información sobre la obra. La presencia de distintas expresiones, en el caso de material librístico, está señalada por la mención de la

edición original, la presencia de un traductor entre los colaboradores o la mención del tipo de edición (si es aumentada, corregida, revisada, etc.). Con estas pocas señales es posible construir, aunque en forma incompleta y sujeta a errores, un grafo FRBR en el cual se expresen las relaciones de uno a muchos entre la obra y sus expresiones y entre la expresión y sus manifestaciones. La relación entre la manifestación y sus ítems, como se dijo, es generalmente provista por las bases de datos tradicionales.

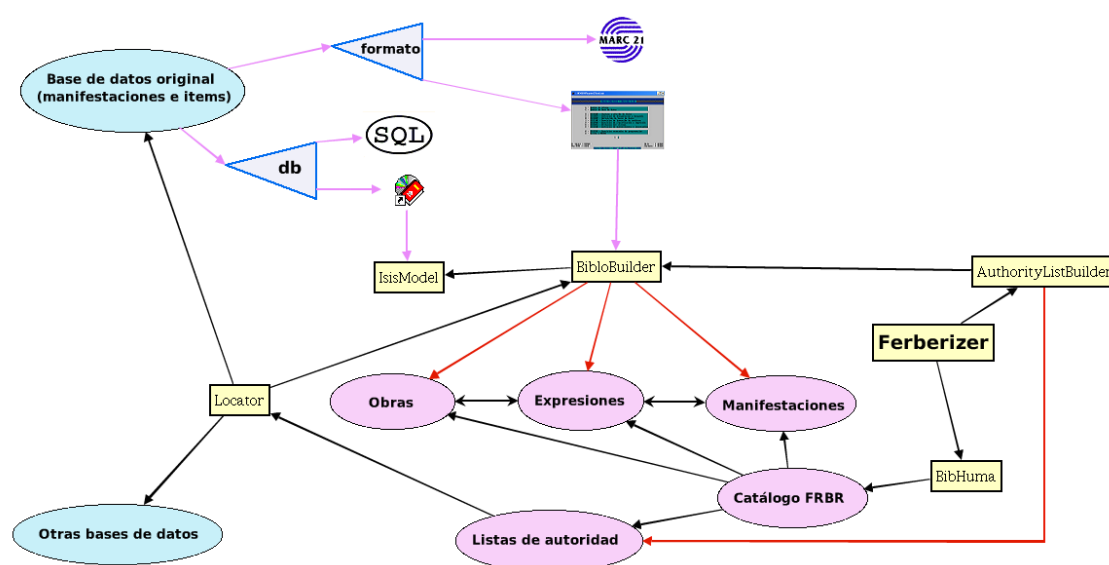
La primera etapa de este proyecto, que es la que se presenta ahora, consistió en esta conversión de los registros originales, basándose únicamente en la información que ellos portan, lo cual depende obviamente de la calidad de los registros. De nada sirve confiar en el campo *edición original* o en el campo *título uniforme* si apenas fueron usados en la carga de datos. Aun en el caso de que todos estos campos hayan sido aprovechados, persiste el eterno problema de las diferencias de criterio, de tipeo y los errores del operador. Por el momento, la comparación se basó en la igualdad de cadenas básicas, es decir, la identidad total una vez eliminadas las diferencias de mayúsculas y minúsculas, las de puntuación y los blancos sobrantes. No obstante, el diseño del software admite anexarle distintos algoritmos de comparación de cadenas para, en el futuro, incrementar la fiabilidad del proceso.

En una segunda etapa se prevé recurrir a información externa que posibilite ir rellenando los huecos. El diseño dinámico del software permite que si en alguna otra base de datos se encuentra información más confiable o completa, esta pueda reemplazar a la existente.

El software ferberizador

Para esta primera versión del software se utilizó la base de datos de la Biblioteca Central de la Facultad de Humanidades de la UNLP. No obstante, el software permite que el proceso se pueda llevar a cabo con cualquier otra base de datos, anexándole los módulos correspondientes. Para lograr esta independencia que permite una gran diversidad de situaciones, se procedió –como corresponde al programar dentro del paradigma de objetos – a separar las distintas responsabilidades detectadas y a definir distintos objetos para cumplirlas. Frente a la situación inicial de partir de una base de

datos preexistente, lo primero que se necesita es un objeto que represente la estructura de cada tipo de base de datos: si es una base de datos relacional, se utilizará un objeto que comprenda las bases de este tipo, si es una base de datos ISIS se utilizará un objeto que represente la estructura general de las bases ISIS (*IsisModel*). Como la base de datos que se utilizó es una base de datos ISIS, se utilizó este último objeto. En el diagrama se detallan, de manera esquemática, los objetos utilizados para el caso concreto, pero se muestran a modo ilustrativo otras posibilidades.



El siguiente paso es modelar el formato bibliográfico. Se procedió del mismo modo, definiendo objetos que representan a un formato bibliográfico que saben construir objetos del modelo OPUS (obras, expresiones y manifestaciones) según la especificación de cada formato. En el caso de la base de datos de la facultad de humanidades, el formato es el correspondiente al software BIBLO (3), por lo que se programó un constructor para este formato (*BibloBuilder*). Si se desea adaptar el ferberizador a otro formato, solo hay que programar el constructor correspondiente. Pero no solo es necesario saber construir las entidades FRBR sino también poder establecer cuándo dos entidades construidas a partir de los datos originales son en realidad una misma entidad. Para esto se implementó un objeto constructor de listas de autoridad (*AuthorityListBuilder*), que colabora con el *BibloBuilder* para crear listas de autoridad para obras, expresiones, manifestaciones, nombres personales, nombres

institucionales y temas.

Pero dado que, como dijimos antes, la información obtenida puede no ser fiable o no estar completa, lo que verdaderamente guardan estas listas es el modo de acceder y reconstruir estos objetos, mediante un objeto llamado *Locator*. Cuando llega el momento de mostrar una determinada entidad, este objeto se encarga de acceder a la ubicación en donde están los datos y, mediante un constructor, reconstruir la entidad FRBR. De esta manera, en la segunda etapa del proyecto se podrá implementar la recolección de información en otras bases de datos y una vez encontrados, modificar el *Locator* para que apunte al nuevo repositorio. Dado que la responsabilidad de ubicar la información y construirla están divididas entre distintos objetos, son posibles distintas combinaciones: los datos pueden estar en distinto formato, pero en una base del mismo tipo; pueden estar en el mismo formato, pero en una base de distinto tipo, o ambos – formato y tipo de base – pueden ser distintos. El objeto tope es el ferberizador (*Ferberizer*), cuya responsabilidad es –mediante un *Authority List Builder* con sus correspondientes constructores de objetos OPUS para un determinado formato bibliográfico y un determinado modelo de base de datos – construir un catálogo ferberizado para un cliente determinado, en este caso, la biblioteca de Humanidades.

El dinamismo de este esquema permite, además, que el espacio de memoria ocupado por estos objetos sea mucho menor que el de una base de datos, por lo que por el momento no utiliza ningún esquema de persistencia más que la memoria del ambiente de objetos. Como corresponde a un buen diseño de objetos, el modelo no está condicionado por la plataforma (base de datos, sistema operativo, etc.) que utiliza. Si en algún momento el uso de memoria no permite una buena performance, se podrá anexar algún esquema de persistencia, que puede ser tanto un esquema de objetos transparente, como un esquema tradicional (una base de datos) que mapee al modelo principal.

El software presentado fue programado con el lenguaje Smalltalk, en su dialecto de fuente abierta Squeak.

Ejemplos

Veamos ahora algunos ejemplos. Para esta presentación se utilizó un subconjunto de la base de datos original correspondiente al material existente en la biblioteca sobre filosofía, con el objetivo de acotar el universo representado para poder observarlo con mayor detalle. De los 76 000 registros de la base original, se ferberizaron 4400. Dado que el modelo de base de datos original es el tipo de archivos de CDS/ISIS, el primer paso de la aplicación es una consulta general a bases de datos ISIS. El acceso a este tipo de archivos es provisto por el framework Aton, desarrollado anteriormente en Caicyt para otras aplicaciones. El aspecto visual es provisto por el framework SmallFaces, desarrollado también en Caicyt por un equipo formado por Juan Burella, Hernán Morales y Norberto Manzanos. Ambos frameworks están a disposición de otros programadores smalltalk en el repositorio de software abierto squeaksource (4). Esta primera capa del software es también reutilizable y podría ser parte de un software de consulta tradicional.

Una vez realizada la consulta, se informa de los resultados y ya se puede observar aquí la división entre lo que surge de la base de datos original y lo que arrojó el proceso de ferberización. En el ejemplo que se muestra en la figura, se puede apreciar que la consulta por "Kahler\$" (es decir, todo lo que comience por la cadena de caracteres "Kahler") da 2 resultados en el catálogo original, con los cuales el ferberizador construyó 2 obras.

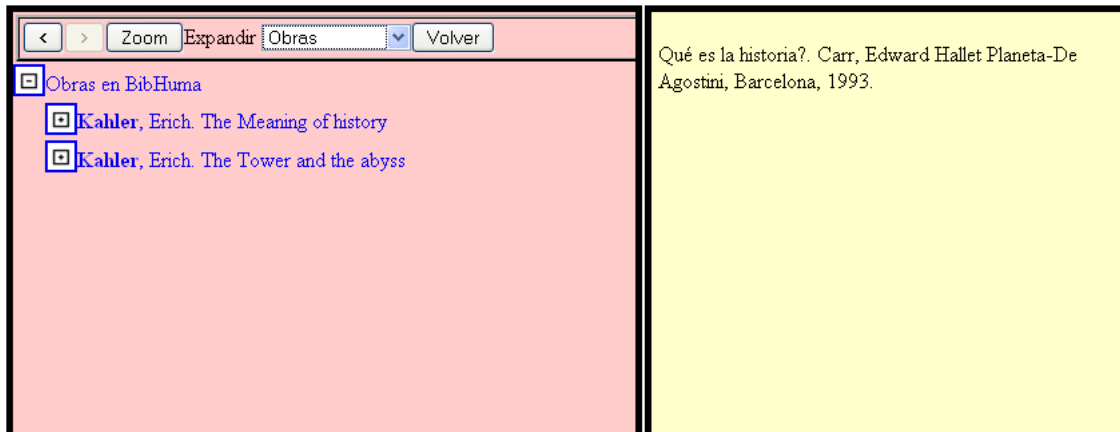
The screenshot shows a search interface with three tabs: "Busqueda simple", "Busqueda avanzada", and "Indice". The "Busqueda simple" tab is selected. Below the tabs is a search input field containing "Kahler\$" and a dropdown menu set to "Todos". There are two buttons: "Buscar" and "Mostrar". Below the search area, the results are displayed as "Catálogo: 2" and "Works: 2".

Si se opta por contar las expresiones y manifestaciones se mostrará, además, que la ferberización construyó 4 expresiones y 5 manifestaciones. Pero debe notarse que no todas estas entidades corresponden a documentos realmente existentes en la colección, sino a otros en los cuales estos están basados. El número que realmente cuenta para medir la colección concreta es el que figura a la derecha de esas cifras, que en este caso es 3.

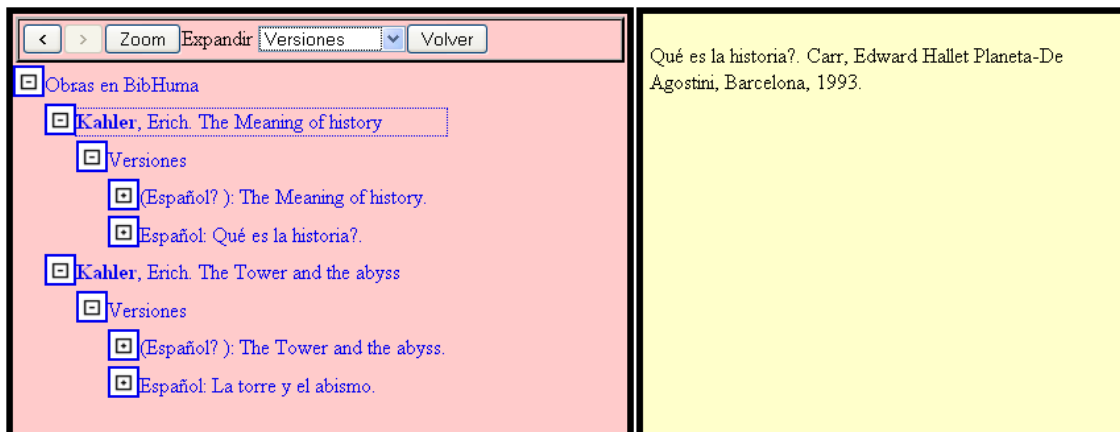
The screenshot shows a search interface with three tabs: 'Busqueda simple', 'Busqueda avanzada', and 'Indice'. The 'Busqueda simple' tab is active. A search box contains the text 'Kahler\$'. To the right of the search box is the word 'en' and a dropdown menu set to 'Todos'. Below the search box are two buttons: 'Buscar' and 'Mostrar'.

Catálogo: 2
Works: 2
Expressions: 4 / 3
Manifestations: 5 / 3
Items: 4

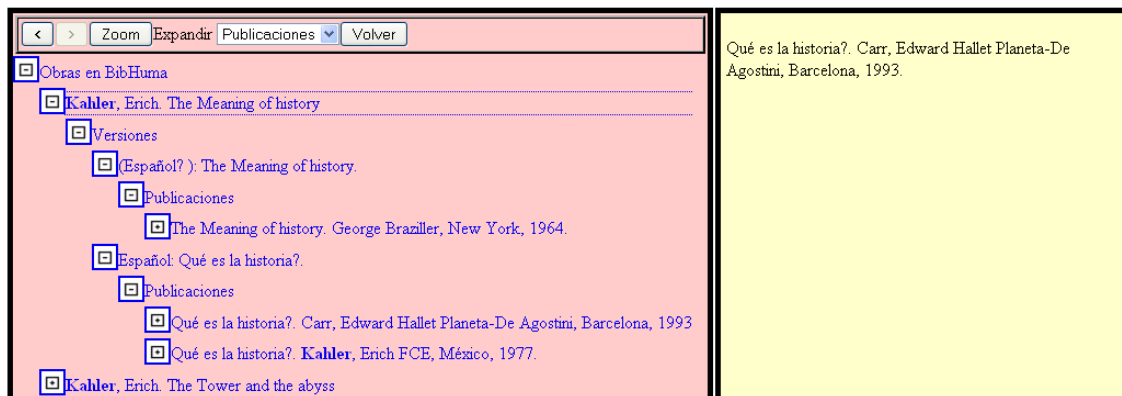
Esta divergencia se aclara cuando se pide al programa que muestre los resultados: es entonces cuando se puede apreciar más el producto de la ferberización. Lo que vemos como las dos primeras ramas de un árbol son las dos obras a las que corresponden las manifestaciones que posee la biblioteca. Como en este caso se trata de traducciones, lo que vemos son los títulos originales. En el panel de la derecha se puede ver el registro correspondiente de la base original, este sí con el título en castellano.



Se puede recorrer el árbol abriendo los distintos nodos o directamente pedir que muestre todas las expresiones, que en este caso son llamadas "versiones" por considerar que es un término más afín al usuario final.



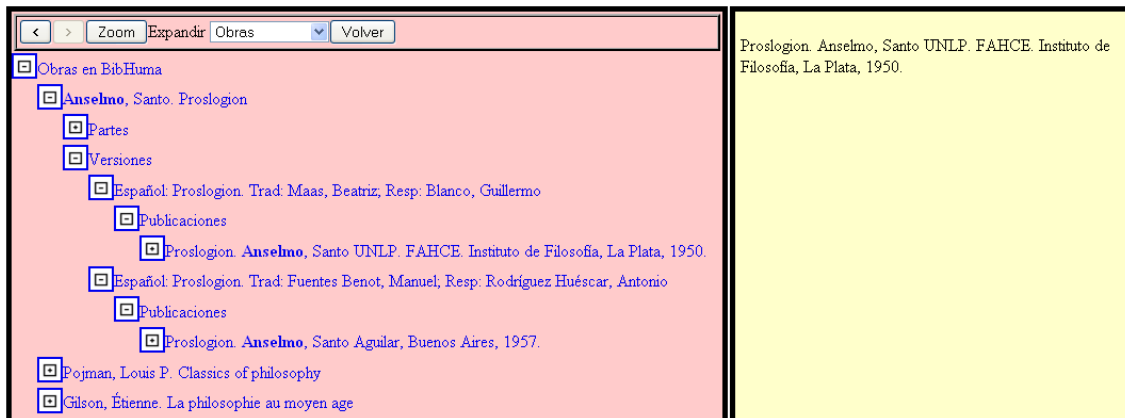
Si solicitamos que nos muestre las publicaciones (es decir, las manifestaciones) podremos apreciar dónde está la diferencia de números del conteo anterior: se trata de dos ediciones de la misma obra, de distinta editorial y con distinta fecha, de la misma obra.



Ahora hacemos la búsqueda "Jaeger\$". En este caso, la manifestación que vemos en el panel de la derecha tiene el mismo título que la obra. Esto es debido a que la biblioteca posee una versión en inglés, presumiblemente la expresión original de la obra, pero también posee una traducción.



La búsqueda de "Anselmo\$" nos muestra un caso de una obra con dos traducciones distintas.



Listas de autoridad y puntos de acceso

Como ya se mencionó brevemente, para construir este árbol FRBR el programa confecciona listas de autoridad para cada una de las entidades principales de FRBR.

Estas listas son las que permiten establecer la identidad de las entidades y de ese modo establecer las relaciones entre ellas.

Por el momento el ferberizador no brinda la posibilidad de modificarlas, por lo que mostramos solo el interior del objeto BibHuma –que representa a la Biblioteca de la Facultad de Humanidades de la UNLP– mostrando una parte de una de estas listas, el catálogo de Obras. Como continuidad de este proyecto, uno de los próximos objetivos que cumplir es agregar un módulo para la edición de estas listas que interactúe con el catálogo de la biblioteca, de manera de mejorar no solo el proceso de ferberización y el catálogo FRBR resultante, sino también el catálogo original.



El modelo OPUS

Si bien es bastante evidente la utilidad que tiene tanto para el usuario como para el investigador la discriminación que permite el modelo FRBR en los ejemplos vistos, se ha observado que el modelo es especialmente apropiado para las humanidades y no tanto para las “ciencias duras”. Pero el modelo OPUS no es una copia fiel de FRBR sino que está basado en FRBR e incorpora varios aspectos sobre los cuales IFLA no profundizó. No es posible detenernos aquí en todos los detalles, algunos de los cuales han sido presentados en otros eventos. Pero no podemos dejar de mencionar una de sus características fundamentales, la que resuelve una de las críticas que ha tenido el modelo con respecto a la falta de una subclasificación de las entidades principales. (5) Los objetos de OPUS nunca son una *obra* o una *expresión*, dado que estas son clases abstractas, sino subclases específicas de estas. En los ejemplos que hemos visto, las obras eran en realidad una subclase de *obra* denominada *Obra literaria*, las expresiones eran o bien un objeto *Edición* o bien una subclase de este, el objeto *Traducción*. De este modo, los atributos de un determinado tipo documental están expresados mediante objetos específicos. Las obras literarias y sus expresiones derivadas –como por ejemplo, las traducciones - tienen entre sus atributos el poseer un determinado lenguaje. No será el caso de una *Obra musical*, que tendrá otros atributos. Esto permite, entre otras cosas, modificar el modo de mostrarse que tiene cada uno de estos objetos. Por ejemplo, la etiqueta “Versiones” podría reemplazarse, solo para el caso de obras musicales, por “Interpretaciones”, mientras que la etiqueta “Publicaciones” podría reemplazarse por “Grabaciones”. También el modo de construir los puntos de acceso para las listas de

autoridad depende de cada subclase. Por ejemplo, en el caso de las obras musicales, el número de opus y la tonalidad de la obra son datos específicos de este tipo de obras utilizados para construir puntos de acceso. Otras características de OPUS, como el detallado esquema de relaciones entre las entidades, no han podido aún ser incorporados al feberizador. Si se tiene en cuenta que una de esas relaciones es la relación de *cita*, se comprenderá por qué el modelo también puede ser de mucha utilidad para las ciencias duras.

Aplicaciones bibliométricas

Si bien no es nuestro propósito utilizar la información provista por el feberizador para realizar estadísticas bibliométricas, y la aplicación aún no está en un estado que permita fiarse totalmente de los resultados que se obtienen, es evidente que una base de datos feberizada da un panorama muy distinto que el de las bases de datos tradicionales que solo registran manifestaciones.

Si partimos de la definición de obra de Svenonius dada inicialmente y consideramos que la información es el objeto de interés de los usuarios, es evidente que cualquier medición que se haga sobre la base de una colección de manifestaciones no nos arroja un resultado válido para evaluar la colección desde el punto de vista de la información que provee. Es plausible suponer que ningún usuario estará interesado en obtener dos manifestaciones de una misma expresión, ya que, a menos que le interesen las eventuales pequeñas diferencias textuales entre distintas ediciones, encontrará en ambas exactamente lo mismo. No es muy común que un usuario obtenga dos expresiones distintas de una misma obra, aunque en muchos ámbitos puede suceder que algunos usuarios deseen comparar traducciones o verificar cuáles son los agregados de una edición aumentada. De todas formas no es posible reconocer y discriminar estas situaciones partiendo de una colección de manifestaciones. De este modo, no solo queda oculto el valor informativo de la colección sino también los distintos usos que se hacen de ella.

Del mismo modo, si se realiza un recorte temático, no se obtendrán números fiables. A modo de ejemplo ilustrativo mostramos los resultados obtenidos por una serie de búsquedas. Se indica primero el resultado en la consulta a la base de datos original, es

decir, las manifestaciones, y a la derecha la cantidad de obras reconocidas por el ferberizador.

Filosofía Cultural 100/90
Filosofía de la Educación 108/98
Filosofía Práctica 109/85
Filosofía Literaria 97/89
Foucault, Michel 115/75

Se puede observar que la relación entre obras y manifestaciones no es constante, y que en algunos casos, los resultados son divergentes. Hay más registros originales sobre “Filosofía Práctica” que sobre “Filosofía de la Educación”, pero es al revés si se consideran las obras. La búsqueda por “Foucault, Michel” da más manifestaciones que todas las otras consultas, pero menos obras que en todos los casos. Es evidente que las estadísticas de una colección bibliográfica basada en las manifestaciones nos pueden estar hablando de muchas cosas, como de la relación entre la biblioteca y el mundo editorial, pero no nos dicen mucho sobre el valor informativo, es decir, el valor de uso que tienen para los usuarios los documentos de la colección. Debe tenerse en cuenta, además, que el modelo FRBR es más potente mientras más grande es la colección. En la medida en que los catálogos se intercomunican y surjan catálogos colectivos más amplios estas divergencias pueden llegar a una situación paradójal: mientras más se incrementa la cantidad de datos y por más que estos estén prolijamente ordenados, menor será la cantidad de información.

Conclusión

Si bien hemos presentado un programa realizado sobre una base de datos determinada, se trata de un programa adaptable a cualquier otra situación, si se realiza el trabajo de desarrollo informático necesario para cada situación. Para utilizar el ferberizador en una base de datos que utilice el sistema BIBLO apenas es necesario agregar un objeto que represente a la institución. Si se trata de una base de datos ISIS, se trata de programar un objeto *Builder* al cual simplemente se le debe especificar su estructura de campos y subcampos y cómo construir los objetos OPUS a partir de esta estructura. Un trabajo no mucho mayor implicaría convertir datos de archivos MARC. En el caso de bases de datos relacionales, habría que anexarle alguno de los muchos paquetes que hay

disponibles en Smalltalk para tal fin y programar un *Builder* para el formato bibliográfico particular.

El ferberizador estará disponible a principios de 2011 en la web, probablemente con la incorporación de otras bases de datos, para que la comunidad bibliotecaria pueda experimentar el uso de FRBR y realizar sugerencias y aportes para mejorar el programa.

Bibliografía

- Presentaciones del modelo Opus
- Carsen, Tatiana M., Garcia, Hugo, Mabragaña, Cecilia y Manzanos, Norberto (2003) *Modelo de descripción documental basado en el paradigma de objetos*. In José Antonio, Frías Montoya and Travieso Rodríguez, Crispulo, Eds. Proceedings VI Congreso del Capítulo Español de ISKO, Salamanca (Spain). <http://eprints.rclis.org/archive/00008585/>
- Manzanos, N. *El impacto de FRBR en Argentina: Implementación de un modelo de objetos basados en FRBR, CRM y FRBRoo en CAICyT-CONICET*. En 3er. Encuentro Internacional de Catalogadores: Tendencias actuales en la organización de la información, Buenos Aires (Argentina), October 29 2007. ISBN: 978-987-9350-45-4. <http://eprints.rclis.org/12534/>.
- Flores, A; Manzanos, N. *OPUS: Sistema de registro bibliográfico basado en FRBR*. I Encuentro Nacional de Catalogadores. Biblioteca Nacional, Buenos Aires 2008. <http://www.bn.gov.ar/encuentro/ponencias08/ponencia-Flores-Manzanos.pdf>.
- Svenonius, Elaine. (2001). *The Intellectual Foundation of Information Organization*. The MIT Press, Cambridge.
- Balerdi, G., Fushimi M, Manzanos, N, Zaritzky E. *BIBLO: sistema integral para bibliotecas y centros de información*. En: II Jornadas nacionales y I Latinoamericanas y del Caribe sobre Micro ISIS Actas. Buenos Aires : Comisión Nacional de Energía Atómica (CNEA), 1992. p. 13-22.
- <http://www.squeaksoure.com>

I Jornada de Intercambio y Reflexión acerca de la Investigación en Bibliotecología,
La Plata, 6-7 de diciembre de 2010. La Plata: Facultad de Humanidades y Ciencias de la Educación de la
Universidad Nacional de La Plata.

- Le Boeuf, Patrick. Brave new FRBR world. En: First IFLA Meeting of Experts on an International Cataloguing Code, July 28 - 30, 2003, Frankfurt am Main, Germany