

Organizar la recuperación de la información para organizar el conocimiento: el caso práctico de UBUCAT, catálogo de la Biblioteca de la Universidad de Burgos.

Fernando Martín Rodríguez

Director de la Biblioteca

Universidad de Burgos

Biblioteca Universitaria

Pza. Infanta Dña Elena, s/n

C.P. E-09001 Burgos

C.e.: bubdir@ubu.es

Mónica Casado Candelas

Jefa de Sección de Automatización y Proceso Técnico

Universidad de Burgos

Biblioteca Universitaria

Pza. Infanta Dña Elena, s/n

C.P. E-09001 Burgos

C.e.: bubaut@ubu.es

RESUMEN:

La Biblioteca de la Universidad de Burgos inició en Junio de 2001 los trabajos de implementación de un nuevo Sistema Integrado de Gestión de la Biblioteca (Innopac Millennium) que sustituiría a Libertas, utilizado desde 1996. Para ello fue necesario migrar los registros bibliográficos del formato LCMARC a MARC 21, lo que obligó a realizar una reorganización de los datos codificados con el fin de adaptarlos al nuevo formato. Simultáneamente, con la migración de los datos se definieron los índices que posibilitaran a los usuarios recuperar la información por medio del uso de las búsquedas diseñadas para tal fin, lo que en conjunto supuso una nueva indización de la base de datos. La metodología de trabajo consistió en establecer una correspondencia entre etiquetas de uno y otro formato, adaptando aquella información que no estaba contemplada en el formato MARC21 en unos casos y codificando en nuevas etiquetas, campos y subcampos datos que estaban mal codificados en Libertas. Una vez mapeadas las nuevas etiquetas se realizó una carga de prueba en la que se pudo comprobar la tasa de éxito de la migración. Posteriormente se definieron el número de subcatálogos o scopes en los que se dividió el conjunto de la base de datos, los tipos de búsquedas que se ofrecerían al usuario para recuperar la información contenida en el catálogo, y las etiquetas, campos y subcampos recuperables desde cada una de las búsquedas diseñadas. La implementación del sistema Innopac Millennium ha dado como resultado inmediato un nuevo producto: UBUCAT, el catálogo de la Biblioteca de la Universidad de Burgos. Una vez abierto el nuevo catálogo se ha establecido un sistema de control y seguimiento del uso del mismo, basado en la técnica del análisis de las transacciones que los usuarios realizan sobre el OPAC. Para ello se dispone de una aplicación que ofrece información sobre el comportamiento de los usuarios a la hora de utilizar las distintas posibilidades que se ofrecen para la recuperación de los registros. Con este sistema se pretende obtener información que permita evaluar la migración e indización llevada a cabo, así como mejorar el sistema de referencias de autoridades del catálogo y ampliar la lista de encabezamientos de materia utilizada en la biblioteca.

PALABRAS CLAVE: Diseño de OPACS ; MARC21 ; INNOPAC MILLENNIUM ; Análisis transaccional ; Indización ; Recuperación de la información

Organising the recovery of information for the organisation of knowledge: the practical case of the UBUCAT, the catalogue of the University of Burgos Library.

Fernando Martín Rodríguez

Library Manager

Universidad de Burgos
Biblioteca Universitaria
Pza. Infanta Dña Elena, s/n
C.P. E-09001 Burgos
C.e.: bubdir@ubu.es

Mónica Casado Candelas

Head of the Automation and Technical Process Section

Universidad de Burgos
Biblioteca Universitaria
Pza. Infanta Dña Elena, s/n
C.P. E-09001 Burgos
C.e.: bubaut@ubu.es

ABSTRACT:

In June 2001 the University of Burgos began work on implementing a new Integrated Library Management System (Innopac Millennium) to replace Libertas, in use since 1996. To do this it was necessary to migrate the bibliographic records from the format LCMARC to MARC 21, which meant reorganising the coded data to adapt them to the new format. Simultaneously with the migration of data, indexes were defined which enable users to recover information using the searches designed for the purpose, which meant providing the data base with a new indexing system. The work methodology consisted of establishing a correspondence between labels in the two formats, adapting that information which was not contemplated in the MARC 21 format in some cases and coding in new labels, fields or sub-fields data which was badly coded in Libertas. Once the new labels had been mapped, a trial run was performed to check the degree of success of the migration of data. Later, definitions were given for the number of subcatalogues or scopes into which the whole of the data base was divided, the types of search to be offered to the user to recover information contained in the catalogue, and the labels, fields and subfields recoverable from each of the searches designed. Implementation of the Innopac Millennium system has had as an immediate result a new product: UBUCAT, the University of Burgos Library catalogue. Once the new catalogue had been opened, a system of control and monitoring was established for its use, based on the technique of analysis of the transactions users make on the OPAC. The system contains an application which offers information as to users' behaviour when using the different possibilities available for the recovery of records. The purpose of this system is to obtain information which enables one to evaluate the migration and indexing done, and to improve the system of references of the authorities of the catalogue and extend the list of subject headings used in the library.

KEY WORDS: Design of OPACS ; MARC21 ; INNOPAC MILLENNIUM ; Transactional analysis ; Indexing ; Information retrieval

1. INTRODUCCIÓN

La Biblioteca Universitaria de Burgos comenzó la automatización de sus catálogos en Mayo de 1996, cuando adquiere el Sistema Integrado de Gestión Bibliotecaria Libertas, de la empresa SLS. Dicho sistema se mantiene en funcionamiento durante 5 años, tiempo en el que se consigue automatizar el 100% de los fondos bibliográficos ubicados en las bibliotecas de las Facultades y Escuelas Universitarias, además de procesar alrededor del 30% de los fondos bibliográficos almacenados en los seminarios de los Departamentos y las Áreas. La envergadura de la base de datos bibliográfica alcanza los 110.000 ejemplares de monografías y 2.600 registros de publicaciones periódicas, consultables a través del catálogo Libertas en modo telnet. Simultáneamente se automatiza el préstamo en los cinco puntos de servicio con los que cuenta la Biblioteca Universitaria y se gestionan las compras de material bibliográfico a través del módulo de adquisiciones. Se puede afirmar que este primer proyecto de automatización cumple con creces sus objetivos, puesto que se consigue automatizar la mayor parte de los procesos que se realizan en la biblioteca, quedando excluidos de Libertas tan solo el préstamo interbibliotecario, los ficheros cardex de publicaciones periódicas y la consulta de bases de datos bibliográficas por limitaciones del sistema. En 1998 la empresa informa a la Universidad de la paralización del desarrollo de Libertas, y oferta otro producto más avanzado tecnológicamente: Innopac Millennium. Para tomar conciencia del volumen de negocio que supone el mercado de los sistemas automatizados, el informe realizado por Bilal, Barry y Penniman (1) indica que durante ese año se vendieron alrededor de 1.000 sistemas de automatización de bibliotecas por un valor estimado de 475 millones de dólares, correspondiendo 70 millones de dólares a las ventas realizadas por Innovative Interfaces en concepto de su producto Innopac Millennium. Después de las valoraciones que lleva a cabo personal de la Universidad de Burgos se decide su adquisición, siendo esta decisión tomada de forma conjunta por las 4 universidades públicas de Castilla y León y por la Universidad Pontificia de Salamanca, con el propósito de facilitar acuerdos de cooperación teniendo como base el mismo sistema de gestión de bibliotecas.

2. ALCANCE DEL PROYECTO

Mhill (2) estima que la duración de un proyecto de migración de sistema puede prolongarse alrededor de 12 meses. En nuestro caso el proyecto UBUCAT comienza en mayo de 2001, y finaliza su primera fase en septiembre de 2002, fecha en la que se abre el OPAC al público. Un mes después de su adquisición se mantiene con responsables técnicos de la empresa una reunión en la que se fijarán los grupos de trabajo y el calendario, con el objetivo final de implementar los módulos correspondientes a Catalogación, Circulación, Adquisiciones, Publicaciones Seriadas y Web Opac. Se estima que las labores de migración del sistema y la implementación del conjunto de los módulos adquiridos se prolonguen desde en una segunda fase hasta enero de 2003

Las etapas 1ª (conversión de los datos), 3ª (definición de los SCOPES), 4ª (diseño de la base de datos) y 7ª (Diseño del Web Opac) (*fig. 1*) tienen una relación directa con el éxito de los usuarios en sus búsquedas y forman parte, junto con el motor de búsqueda de la base de datos de lo que podríamos denominar el sistema de recuperación de la información del catálogo, por lo que el desarrollo de la comunicación se centrará en estos 4 fases del cambio al sistema.

2.1 Conversión de los datos

Consiste en la migración de los datos codificados en Libertas MARC al nuevo sistema, que codifica los datos en un nuevo formato: MARC21. Para la conversión de los datos y su migración al nuevo

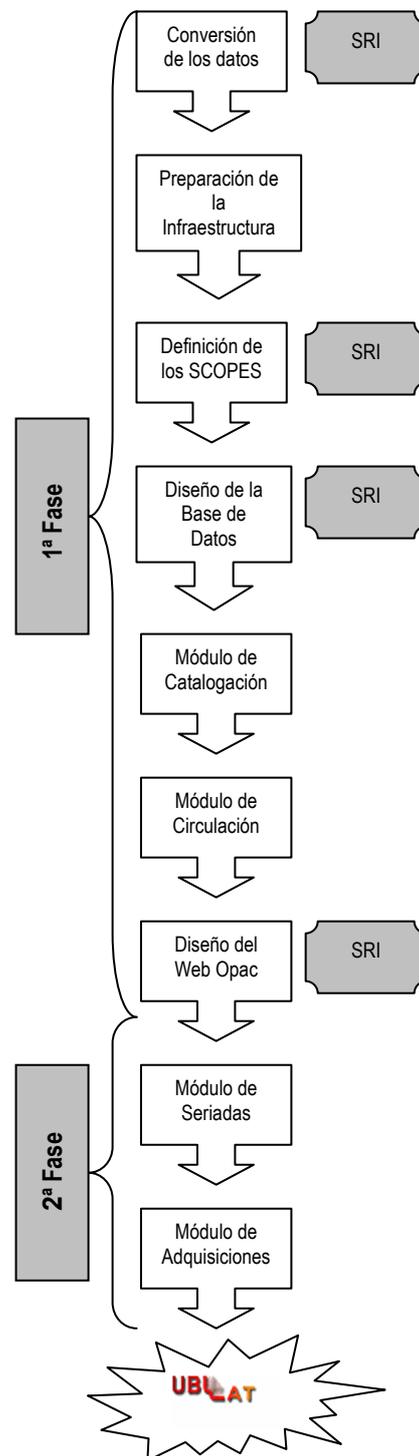


Fig.1: Fases del Proyecto de Implementación. Están marcadas las fases de diseño del Sistema de Recuperación de la Información (SRI)

sistema se tuvieron en cuenta, siguiendo el orden en el que se enuncian, los criterios siguientes:

1. No perder información codificada en Libertas MARC al pasarla a MARC 21
2. No invalidar etiquetas del formato MARC 21 por recoger información codificada en Libertas MARC
3. Ser fieles al uso asignado a cada uno de las etiquetas, campos y subcampos del formato MARC 21

Para ello se realizó un análisis de cada una de las etiquetas de Libertas MARC y se trazó una correspondencia con las etiquetas recogidas en el formato MARC 21¹. Una vez analizado uno y otro formato se cumplimentaron unas hojas de trabajo en las que se indicó en qué subcampo Y etiqueta del formato MARC 21 debía recogerse la información recogida en las etiquetas del formato empleado por Libertas. Se realizó un fichero de prueba de conversión sobre 10.000 registros con el fin de detectar posibles errores. Hay que tener en cuenta que un error en la codificación de la información puede repercutir en el éxito de los usuarios en la recuperación de la información. Los errores de conversión fueron del tipo siguiente:

Pérdida de información: P. ej.: En las etiqueta 248 \$a ; \$g ; \$e se detectó que no pasaba la información al nuevo formato codificada en el subcampo repetible \$e

Errores de puntuación: P. ej.: En la etiqueta 440 \$a ; \$v ; \$l, al pasar la información a la nueva etiqueta del formato MARC 21 440 \$a ; \$v ; \$p se detectó que la información codificada en el \$p debía de ir precedida por un punto: 440.XX\$aLibro de bolsillo;\$v373.\$pSección Historia.

Se realizó una segunda carga de prueba con los errores depurados y finalmente se aceptó la conversión global de toda la base de datos.

2.2 Definición de los SCOPES

El SIGB permite segmentar el catálogo en diferentes partes denominadas SCOPES. Mediante el establecimiento de estos subcatálogos se facilita al usuario poder emplear

¹ Para establecer las correspondencias entre etiquetas se emplearon los documentos de trabajo de la Biblioteca de la Universidad Complutense nº 93/5, 98/8 y 2001/2 en los que se detallan y explica el uso de las etiquetas y subcampos de los formatos Libertas MARC y MARC 21. Ver bibliografía (3), (4) y (5).

todas las herramientas del catálogo, en lo que a tipos de búsqueda y presentación de registros se refiere sobre un conjunto de registros más o menos amplio que comparten características comunes, atendiendo a criterios de tipo:

Ubicación física de los documentos: Se establecen tantos SCOPES como número de bibliotecas que participan de una única instalación de Innopac Millennium y que pertenecen a un mismo sistema o red de biblioteca. Al buscar sobre un SCOPE, el sistema trabaja únicamente sobre los registros ubicados en la Biblioteca seleccionada a la que se le ha asignado el SCOPE.

Tipología formal de los documentos: En este caso el SCOPE se establece en función de la tipología documental compartida por todos los registros que forman el catálogo, y que viene descrita en los indicadores correspondientes de las etiquetas 006 y 007 del formato.

La biblioteca de la Universidad de Burgos está concebida como un sistema de bibliotecas único para toda la comunidad universitaria, y en la actualidad la instalación adquirida por la UBU sólo presta servicio para las 6 bibliotecas que componen el sistema, por lo que carecía de sentido utilizar SCOPES en función de la ubicación física de los documentos. Por el contrario se consideró implementar esta herramienta para segmentar el catálogo atendiendo al segundo criterio, configurando así dos SCOPES:

Catálogo General: Las búsquedas sobre este SCOPE actúan sobre el conjunto de los 110.000 registros que actualmente contiene el catálogo, con independencia de su ubicación o de su tipología formal.

Catálogo de Revistas: Agrupa a los más de 2.600 registros correspondientes a la colección de publicaciones periódicas reunidas por la Biblioteca Universitaria. Las razones que nos llevó a establecer el SCOPE de revistas son las siguientes:

- Las revistas son consideradas por el personal docente e investigador de la Universidad como una publicación esencial para poder desarrollar su labor docente e investigadora, de tal forma que adquieren de esta forma una especial relevancia sobre el resto de las publicaciones.

- Se detectó que al recuperar un título de revista desde el SCOPE de Catálogo General cuyo título abreviado (etiqueta 210) título clave (etiqueta 222) y título propiamente dicho (etiqueta 245) fueran coincidentes el sistema devolvía como resultado de la búsqueda tres entradas distintas que llevaban al mismo registro, y se consideró una inconveniencia para los usuarios.

En un futuro está previsto añadir un SCOPE más que aglutine al conjunto de registros que contienen una etiqueta 856 (acceso a recurso electrónico) independientemente de la naturaleza de las publicaciones (revista electrónica, libro electrónico, página web, documento de vídeo, fotografía...) con el fin de crear una auténtica biblioteca electrónica.

2.3 Diseño de la base de datos

Quizás sea esta fase la más comprometida con el diseño del sistema de recuperación de la información, ya que consiste en crear cada una de las reglas de indexación que harán posible que la información codificada en formato MARC21 sea recuperable por uno u otro índice establecido.

2.3.1 Índices en Innopac Millennium

Hay tres tipos de índices en Millennium: número de registro, frase y búsqueda avanzada

Índice por número de registro: Está definido por el sistema y contiene el identificador único, es decir, el “número ID” generado por el sistema de cada registro, para cada uno de los siete tipos de registro que emplea Innopac.

Índices de frase: Toman una cadena específica de texto de un campo de longitud variable y lo indexan como una unidad. Las búsquedas se realizan por comparación entre los

ETIQUETA DE ÍNDICE	TIPO DE REGISTRO
Autor	Bibliográfico, autoridad
Código de barras	Ejemplar, usuario
Signatura	Bibliográfico, ejemplar, fondos
Materias	Bibliográfico, autoridad
ID Familia	Usuario
Nº Documento del gobierno	Bibliográfico
Nº Estandar	Bibliográfico
Título clave	Bibliográfico
Nombre	Usuario
Nº Control	Bibliográfico
Profesor	Reserva por curso
Curso	Reserva por curso
Título	Bibliográfico, autoridad
ARN	Autoridad

Fig. 2: Índices y su correspondiente tipo de registro sobre el que actúan.

términos introducidos por el usuario y las entradas del índice de frase, comenzando por la izquierda. Automáticamente el sistema realiza un truncado por la derecha. Ej.: Si la cadena de caracteres “*native american animal stories*” se indexa en un índice de frase, se puede recuperar buscando por “*native american anim*”, pero el sistema no recuperará el registro si la búsqueda planteada es “*animal stories native*”.

Hay que señalar que el sistema cuenta con un solo conjunto de índices para la totalidad de los siete tipos de registros. Algunos índices pueden ser relevantes para un solo tipo de registro (índice de profesores para registros de cursos) y alguno pueden serlo para más de uno (índice de códigos de barras para registros de ejemplar y de usuario).

Índice de búsqueda avanzada: Toma una cadena específica de texto de un campo de longitud variable e indexa cada palabra por separado. En una búsqueda introducida por el usuario en el sistema cada palabra es comparada por separado con las entradas del índice de búsqueda avanzada. Este índice contiene sólo datos de registros bibliográficos.

Este índice está dividido en cuatro segmentos, correspondientes cada uno de ellos a autor, título, materia y notas. Los tres primeros se derivan de sus respectivos índices de frases, mientras que el cuarto, se utiliza para indexar otras etiquetas y subcampos MARC no incluidos en los anteriores. Los campos numéricos (ISBN, ISSN, signatura...) no se indexan en este índice.

2.3.2 Reglas de Indexación

Teniendo en cuenta la estructura de índices de Millennium, el paso siguiente fue el de establecer las correspondientes reglas de indexación de los registros. Una regla de indexación determina en qué índice va a poder ser recuperada la información de un registro codificada mediante el formato MARC.

A la hora de diseñar las reglas de indexación se tomó como punto de partida las tablas de índices diseñados por el personal del Servicio de Archivos y Bibliotecas de la Universidad de Salamanca. Se analizó la estructura de los índices y se realizaron las modificaciones oportunas estableciendo los siguientes índices con sus correspondientes reglas de indexación:

Índice de autor: Recupera por nombre de autor personal, nombre de autor corporativo, nombre de congreso, nombre de persona, entidad o congreso como serie y nombre de persona, entidad o congreso como materia.

Índice de título: Recupera el título en los campos de: título, título uniforme título abreviado, título clave, variante de título, título anterior o variante de título, colección autorizada o no, secundaria de colección autorizada o no.

Índice de materia: Recupera por todos los campos 6XX del formato MARC 21 así como por cada uno de los subencabezamientos de materia.

Índice de colección: Recupera por las etiquetas de Nombre de persona, entidad o congreso como colección, título uniforme como entrada secundaria de serie de colección, tanto las que están codificadas en las correspondientes etiquetas 4XX como en las etiquetas 8XX.

Índice de número normalizado: Recupera por ISBN, ISSN, NIPO y otros identificadores normalizados (ISRN, ISMN...),

Índice de palabra clave: Recupera por palabra clave de título, colección, materia y nota.

Índice de signatura: Recupera por la información codificada en las etiquetas 080, 081 y 083, correspondientes a la notación CDU y signatura local.

Se realizó una carga de prueba y el equipo encargado de evaluarla realizó búsquedas por cada uno de los índices definidos empleando como elementos de búsqueda la información codificada en cada una de las etiquetas determinantes de reglas de indexación. Se detectaron errores en la recuperación de información codificada en subcampos repetibles (p. ej en la etiqueta 245.XX, subcampo Id de título distintivo), así como errores en el funcionamiento del propio motor de búsqueda del sistema. Una vez que se consideró que las reglas de indexación funcionaban correctamente se reinició la base de datos de prueba y se realizó la carga de datos completa.

2.4 Diseño del Web Opac

Esta fase supone el diseño del interface de búsqueda que se sitúa entre el usuario y la base de datos. Se optó por denominar de manera singular al catálogo de la Universidad, pasando a llamarse UBUCAT, con su correspondiente logotipo. Desde la pantalla de presentación del catálogo se da la opción al usuario de seleccionar el SCOPE sobre el que se desea efectuar la búsqueda, y la pantalla inmediatamente siguiente ya permite ejecutar las estrategias de búsqueda. Finalmente se han hecho visibles para los usuarios los índices correspondientes a autor, título, colección, materias, números oficiales y palabras clave. Se ha omitido la búsqueda por signatura, que a pesar de su importancia, se ha estimado eliminarla para la visualización por parte de los usuarios ya que la Biblioteca tiene como proyecto a corto plazo cambiar el sistema de clasificación (pasar de un sistema "propietario" a la CDU).

También se ha diseñado una búsqueda híbrida por autor/título, cuyo funcionamiento está basado en la combinación de las reglas de indexación que forman los índices de autor y de título. Con el propósito de simplificar la complejidad del uso del catálogo se ha puesto a disposición de los usuarios ayudas en línea para cada uno de los tipos de búsqueda así como una guía general de uso.

3. EVALUACIÓN DE LA INDEXACIÓN

El trabajo de organización de los índices no finaliza con la puesta en marcha del nuevo sistema, ya que se ha diseñado un sistema de evaluación de la indexación que permite conocer al equipo de trabajo de proceso técnico el uso que los usuarios realizan del catálogo y comprobar si, por ejemplo, utilizan el índice adecuado en función de los elementos que cuentan para localizar los registros. Para ello se emplea una herramienta del propio programa Innopac que permite visualizar los términos que los usuarios emplean para sus búsquedas y los índices seleccionados en cada caso. En realidad se trata de emplear la técnica del análisis transaccional para evaluar el OPAC, que En sentido estricto, ha sido definido como "el estudio de las interacciones registradas electrónicamente entre los sistemas de recuperación de información en línea y las personas que buscan información contenida en estos sistemas" (6) . Indica las distintas características del comportamiento así como las dificultades encontradas y, como ha señalado Wall , puede "ayudarnos a dar una respuesta a lo mejor de la interacción humano-humano en nuestros sistemas y podría incluso ayudarnos a superar algunos de los aspectos negativos de la interacción humana"(7).

Trabajos como los de Ferry, Fallon y Scout (8) son un buen ejemplo de cómo esta técnica para tomar decisiones de gestión que implican cambios en las modificaciones de los sistemas automatizados. Desde el uno de enero de 2003 se recogen los datos siguientes:

- Numero total de búsquedas realizadas con el sistema
- Porcentaje de uso de cada uno de los índices
- Tasa de éxito en cada uno de los índices
- Análisis de las búsquedas por materia con resultado 0
- Análisis de los hits de mayor popularidad utilizados como términos de búsqueda en cada uno de los índices

Peters (9) señala que mediante el análisis de las transacciones es posible determinar el número y tipos de SCOPE del catálogo, y en nuestro caso la observación de las búsquedas por título, serie y materia nos permitió comprobar que alrededor de un 10% de búsquedas con resultado 0 tenían como propósito localizar proyectos fin de carrera y tesis leídas en la Universidad de Burgos. Este dato nos animó a ofrecer en el menú principal de búsqueda dos búsquedas predefinidas (búsqueda de proyectos de fin de carrera y búsqueda de tesis doctorales), que muestran a los usuarios el conjunto de ambos tipos de documentos con los que cuenta la biblioteca, permitiendo las herramientas de búsqueda del sistema filtrar por alguna materia concreta o por palabra clave que interese al usuario. El análisis del uso de estas búsquedas nos permitirá tomar la decisión sobre la conveniencia o no de implementar un SCOPE que aglutine las consultas sobre tesis y proyectos fin de carrera.

4. AGRADECIMIENTOS

UBUCAT se abrió el 26 de septiembre de 2002, con el módulo de catalogación, OPAC y circulación en marcha. Está previsto que la apertura de los módulos de publicaciones seriadas y adquisiciones se realice a finales del mes de enero de 2003. Los autores de la comunicación queremos agradecer el trabajo realizado a los miembros de los equipos de automatización y proceso técnico, diseño de Web Opac, circulación, publicaciones seriadas, adquisiciones y Servicio de Informática y Comunicaciones que han participado activamente en la consecución del proyecto. De igual forma deseamos agradecer la ayuda prestada por los equipos de automatización de la Universidad Complutense de Madrid y Universidad de Salamanca.

Bibliografía:

- (1) Bilal, Dania ; Barry, Jeff ; Penniman, David. A balancing act. *Library Journal*, 124, 6, apr. 1999, 45-54.
- (2) Myhill, Martin. Time for change. A personal insight into library systems' implementation : experiences at Exeter University Library. *Program*, 31, 1, jan. 2000, 89-101.
- (3) LIBERTAS MARC : monografías : basado en el Libertas MARC Manual, v.3, Julio 1992. Madrid, Universidad Complutense, 1998. Documentos de Trabajo U.C.M. Biblioteca, 98/8
- (4) LIBERTAS MARC : publicaciones seriadas : basado en el Libertas MARC Manual,v.3, Julio 1992. Madrid, Universidad Complutense, 1993. Documentos de Trabajo U.C.M. Biblioteca, 93/5
- (5) MANUAL de US MARC para la Biblioteca de la Universidad Complutense : versión 2.0. Madrid, Universidad Complutense, 2001. Documentos de Trabajo U.C.M. Biblioteca, 2001/2.
- (6) Peters, Thomas A. The history and development of transaction log analysis. *Library Hi Tech*, 11(2), 1993, 41-66.
- (7) Wall, C. Edward. Editorial: A tool for improving machine-human interaction. *Library Hi Tech*, 11(2), 1993, 5, 36.
- (8) Scott, Janne ; Trimble, Jeffrey A. ; Fallon, Fleming L. '@*!*@ this computer and the horse it rode in on: patron frustration and failure at the OPAC. Proceedings of the ACRL 7th National Conference, Pittsburg, PA, march 29 through april, 1, 1995. Chicago, ACRL, 1995.
- (9) Peters, Thomas A. Using transaction log analysis for library Management information. *Library Administration and Management*, 10, 1, win. 1996, 21-25.