

Digitalización de obras protegidas: software para la detección de obras fuera del circuito comercial

Digitization of protected works: Software for the detection of out of commerce works

Manuel Blázquez-Ochando; Luis-Fernando Ramos-Simón

Cómo citar este artículo:

Blázquez-Ochando, Manuel; Ramos-Simón, Luis-Fernando (2019). "Digitalización de obras protegidas: software para la detección de obras fuera del circuito comercial". *El profesional de la información*, v. 28, n. 6, e280624.

<https://doi.org/10.3145/epi.2019.nov.24>

Artículo recibido el 27-02-2019
Aceptación definitiva: 27-06-2019



Manuel Blázquez-Ochando ✉

<https://orcid.org/0000-0002-4108-7531>

Universidad Complutense de Madrid
Facultad de Ciencias de la
Documentación
Santísima Trinidad, 37
28010 Madrid, España
manublaz@ucm.es



Luis-Fernando Ramos-Simón

<https://orcid.org/0000-0003-2267-8405>

Universidad Complutense de Madrid
Facultad de Ciencias de la
Documentación
Santísima Trinidad, 37
28010 Madrid, España
lframoss@ucm.es

Resumen

La digitalización de obras protegidas en el contexto de la preservación cultural está cobrando una mayor importancia, según la última propuesta de directiva de la *Comisión Europea* sobre derechos de autor en el mercado único digital, pendiente de aprobación definitiva. Ello puede representar una oportunidad para las bibliotecas europeas, que pueden crear colecciones digitales con aquellas obras que manifiestamente están fuera de los canales comerciales. Para ello se requiere un conjunto de programas informáticos capaces de extraer información de los catálogos, y que faciliten una primera aproximación de los titulares de las obras y datos de edición. En esta investigación se aborda la metodología para el desarrollo de un software capaz de cruzar la información de los catálogos bibliográficos de las bibliotecas, con los catálogos comerciales, a fin de determinar la presencia o ausencia de sus libros. En su desarrollo se explican las dificultades y soluciones empleadas para su construcción, derivadas de la heterogeneidad de los catálogos consultados. Finalmente se concluye que resulta factible y de gran utilidad, la creación de este tipo de aplicaciones informáticas, ya que puede obtenerse una media de aciertos superior al 90% en la distinción de obras fuera de comercio. Sin embargo, aún se encuentran problemas a la hora de diferenciar las ediciones o incluso de interpretar adecuadamente los falsos positivos, derivados de aspectos tales como los algoritmos de sugerencia automática de obras.

Palabras clave

Obras fuera del circuito comercial; Comparador bibliográfico; Programas informáticos; Catálogos bibliográficos; *Web scraping*; *Parser*; Obras descatalogadas; Bibliotecas; Libro; Digitalización.

Abstract

The digitization of protected works in the context of cultural preservation is becoming more important, according to the latest proposal of the European Commission directive on copyright in the digital single market, pending final approval. This may represent an opportunity for European libraries, which can create digital collections with those works that are manifestly outside commercial channels. This requires a set of computer programs capable of extracting information from catalogs, and provide a first detection of the owners of the works. This research approach the methodology for the development of a software capable of crossing information from library bibliographic catalogs, with commercial catalogs, in order to determine the presence or absence of their books. In its development, the difficulties and solutions used for its construction are explained, derived from the heterogeneity of the consulted catalogs. Finally, it is concluded that the creation of this type of computer applications is feasible and very useful, since an average of more than 90%

correct answers can be obtained in the distinction of non-commercial works. However, there are still problems when trying to differentiate the editions or even interpreting the false positives, derived from aspects such as the algorithms of automatic suggestion of works.

Keywords

Out-of-commerce works; Bibliographic catalogues; Bibliographic comparator; Web scraping; Parser; Out-of-print-works; Computer programmes; Libraries; Book; Digitization.

1. Introducción

En el marco de la digitalización de patrimonio cultural, la *Comisión Europea* viene desarrollando sus trabajos de iniciativa legislativa para favorecer la preservación, difusión, reutilización de la documentación y materiales bibliográficos dentro de sus fronteras. El proyecto de *Biblioteca Digital Europea* depende en gran medida de los avances legales para salvaguardar la producción libraria en cada país miembro, y ello encuentra obstáculos que deben ser atendidos (**Janssens; Tryggvadóttir**, 2014, pp. 12-16).

Además de los aspectos legales recientes (contenidos generados por los usuarios, véase **Díaz-Noci; Tous-Roviroso**, 2012) o más tradicionales (consentimiento de los titulares, plazos de protección, respeto a los derechos de las obras en coautoría, delimitación de los derechos entre autores, traductores y editores, etc.), propios de una época de transición (**Iglesias-Rebollo**, 2012), hay una serie de barreras bibliográficas y catalográficas que dificultan en gran medida esos procesos y sin los cuales no se puede efectuar una digitalización sin las debidas garantías. Se hallan dos tipos de obras que presentan dificultades en este apartado:

- obras huérfanas: si ninguno de los titulares de los derechos sobre dicha obra o fonograma está identificado o si, de estarlo uno o más de ellos, ninguno está localizado a pesar de haber efectuado una búsqueda diligente. Están reguladas desde 2012 (*Parlamento Europeo*, 2012);
- obras fuera de circuito comercial: también están protegidas, pero se encuentran descatalogadas y/o agotadas, aunque sus titulares son conocidos.

En muchos casos las obras huérfanas son también obras que están fuera de los circuitos comerciales.

En términos generales, en España, entran en dominio público las obras de los autores fallecidos hace 80 años (antes de 1939). A pesar de que el artículo 26 del texto vigente de la LPI establece que los derechos de explotación de la obra durarán toda la vida del autor y setenta años después de su muerte o declaración de fallecimiento, hay que tener en cuenta la disposición transitoria cuarta (autores fallecidos antes del 7 de diciembre de 1987) de la LPI, según la cual, los derechos de explotación de las obras creadas por autores fallecidos antes de esa fecha tendrán la duración prevista en la Ley de 10 de enero de 1879 sobre Propiedad Intelectual que era de 80 años. Es decir, las obras de los autores fallecidos antes de 1987 entrarán en dominio público a los 80 años del óbito. Asimismo, hay que tener en cuenta la disposición transitoria cuarta (autores fallecidos antes del 7 de diciembre de 1987). Los derechos de explotación de las obras creadas por autores fallecidos antes del 7 de diciembre de 1987 tendrán la duración prevista en la *Ley de 10 de enero de 1879 sobre Propiedad Intelectual*.

Las obras descatalogadas son obras fuera de distribución comercial por decisión del editor. En general, se suelen mezclar con las obras agotadas o que ya no se distribuyen, es una categoría difícil de delimitar. En España, la práctica profesional, la costumbre, identifica a las obras descatalogadas con las agotadas: cuando se ha terminado de vender la edición o quedan pocos ejemplares, el editor retira la obra del catálogo. A partir de entonces sólo es posible encontrarla en las librerías de lance o de viejo. Estos conceptos que identifican a las obras huérfanas, descatalogadas, agotadas y/o fuera de distribución se suelen contraponer a los títulos en el mercado o títulos vivos. La costumbre así parece confirmarlo, siempre referido al ámbito editorial tradicional, ya que no se han abordado planes para la digitalización de este tipo de obras. Por contra, en Francia, con vistas a su digitalización, se denominan obras indisponibles, definición en la que se incluye todo libro publicado en Francia

“antes del 1 de enero de 2001 que ya no sea objeto de comercialización por un editor y que actualmente no esté publicado en formato impreso o digital”,

según el artículo 134-1 del *Código de Propiedad Intelectual*, creado por la *Loi n° 2012-287 du 1er mars 2012 relative à l'exploitation numérique des livres indisponibles du XXe siècle*.

En Alemania se consideran obras fuera de circuito comercial a las obras impresas publicadas antes de 1960. Esta denominación es la que se ha elegido en la nueva directiva europea para abordar la digitalización del patrimonio cultural. Se considera que una creación está fuera del circuito comercial cuando la totalidad de la obra u otra prestación, en todas sus traducciones, versiones y manifestaciones, no está a disposición del público a través de los canales comerciales habituales y no pueda esperarse razonablemente que lo esté. La definición se aplica a todo tipo de obras, sean libros, revistas especializadas, audiovisuales, artes visuales o sonido.

El interés común por este tipo de obras es que puedan ser digitalizadas y accesibles en el contexto de las instituciones

documentales públicas (*Comisión Europea*, 2016). Para su identificación, ambos tipos de obras exigen previamente confirmar que no se encuentran disponibles en los canales comerciales habituales y, por tanto, se necesita conseguir los datos básicos sobre su autoría y divulgación (**Curto-Polo**, 2017, pp. 731-741).

El interés por las obras fuera de los canales comerciales, proviene de la *Directiva del Parlamento Europeo* (2016) sobre los derechos de autor y el mercado único digital

El asunto tiene un alcance más general. Hace unos años José Antonio Millán se preguntaba ¿Qué está pasando con los libros? En la respuesta mostraba su extrañeza por las esperanzas frustradas en el impacto de la mediación digital del libro. Nos interesa subrayar su percepción de que la reubicación digital del libro podría

“liquidar un circuito que venía teniendo un funcionamiento aceptable, pero consensuadamente mejorable, sustituyéndolo por oligopolios que suministran a consumidores encadenados” (**Millán**, 2015).

Enseguida este autor cita como ejemplo de la convulsión de esa revolución digital el área de los libros descatalogados y huérfanos. Alude en concreto que, junto al macro proyecto de digitalización masiva de *Google*, otras instituciones, como las bibliotecas universitarias, se han lanzado también a digitalizar obras de sus fondos.

Por otro lado, en el contexto de la Unión Europea, desde hace años la *Comisión* lleva a cabo iniciativas para favorecer la preservación, difusión y reutilización de materiales bibliográficos con el objetivo –claramente expresado en la iniciativa *Bibliotecas Digitales 2010* (*Comisión Europea*, 2005; 2006), concretado en la creación de *Europeana* (2008) y en otras iniciativas posteriores (**Tallova**, 2014)– de hacer posible la digitalización del patrimonio cultural europeo y su puesta online.

Tanto las grandes iniciativas privadas, como las institucionales, sean locales o promovidas por la *Comisión Europea*, se tropiezan con barreras estrictamente legales (consentimiento de los titulares, plazos de protección, respeto a los derechos de las obras en coautoría, delimitación de derechos entre autores, traductores y editores, etc.) que muchas veces ocultan el problema bibliográfico de fondo, tal como apunta **Bensamoun** (2014) con las obras fuera de comercio en el caso de Francia. Sin duda, las principales dificultades se centran en las llamadas obras descatalogadas (ahora redefinidas como obras fuera del circuito comercial para dar cabida a otros formatos no librarios) y en las obras huérfanas, aunque muchas veces ambos tipos de obras comparten las mismas características. Para una visión de las características de ambas, véanse los trabajos recientes de **Evangelio-Llorca** (2018) y **Ramos-Simón y Blázquez-Ochando** (2018a; 2018b), entre otros.

Desde la perspectiva bibliográfica y catalográfica los problemas en este tipo de obras radican en que por su naturaleza son obras con cierta antigüedad desde su publicación y que en muchos casos carecen de suficiente soporte catalográfico o bien han tenido una difusión muy reducida, por lo que no disponen de una información bibliográfica contrastada, incluso dentro de las instituciones librarias existentes cuando se publicaron. No debemos pasar por alto que estas dificultades también las padecen obras que ya han entrado en dominio público, aunque la posible divulgación digital de esas obras no encuentre problema legal alguno.

Hay que tener en cuenta que este tipo de obras, por diversas circunstancias, muchas veces fueron publicadas con escasa información bibliográfica (autores y editores aficionados, ediciones no venales, imprentas sin infraestructura, ausencia de boletines bibliográficos o depósitos oficiales, falta de presencia en los catálogos de las bibliotecas, etc.). Todo ello dificulta en gran medida la identificación y localización, tanto de los autores como de otros titulares de derechos (ilustradores, fotógrafos, traductores...). Estas circunstancias son determinantes para resolver de un modo pacífico la integración de esas obras en el patrimonio digitalizado y conseguir que el libro digitalizado tenga una nueva vida, incorporando plenamente su origen como obra impresa.

Como decimos, todas estas obras (descatalogadas, huérfanas y en dominio público) afrontan los mismos tipos de carencias bibliográficas, las cuales pueden tener graves efectos sobre el patrimonio documental una vez digitalizado. En concreto, nos referimos a:

- Carencia de información para una correcta identificación de las obras en los catálogos bibliográficos y bibliotecarios (fechas, ediciones, autores, coautores, traductores, reimpresiones, editoriales...)
- Dificultades para la construcción de catálogos de las obras una vez digitalizados: extracciones erróneas, vacías, duplicadas...
- Carencia de datos en un gran número de obras y dificultades para identificar en soporte digital la obra original, situación que se puede dar con facilidad en las ediciones bajo demanda (*print on demand*, POD). Así la misma obra utilizada por dos bibliotecas puede parecer que son obras o ediciones diferentes.
- Para hacer frente a estos problemas, más pronto que tarde se hará necesario consolidar catálogos funcionales. Es decir, ha de haber unas pocas instituciones que validen la autenticidad de las ediciones digitales. Las más llamadas a ocupar este papel son las instituciones bibliotecarias que dispongan de ediciones impresas y digitales que además dispongan de una organización eficaz. Este problema es poco discutible: ¿Quién se atrevería a validar el rigor de un trabajo académico realizado sobre la edición de las obras completas de Azorín de 1919 en el que se ha empleado una edición *print on demand* o electrónica, fechada en India en 2018?

Este contexto que hemos descrito hace necesaria la puesta en marcha de un software que facilite la extracción, comparación y combinación entre distintos catálogos bibliográficos y bibliotecarios.

Esta realidad, tan bien descrita por Millán, como hemos glosado en la introducción, ha evolucionado y es el punto de partida de nuestra investigación, cuyo objeto es la elaboración de un programa informático, que ayude a definir qué obras pueden estar fuera de los circuitos comerciales, mediante la ejecución de consultas cruzadas en los catálogos bibliográficos. El programa deberá proporcionar listas de libros que potencialmente pueden estar fuera de los circuitos comerciales. Esto es, una relación de libros ausentes en los catálogos comerciales, identificados en los catálogos funcionales (propios de las universidades, la base de datos del ISBN, *Rebiun* y la *Biblioteca Nacional de España*). De esta forma, se facilita la identificación temprana de las obras susceptibles de digitalización y puesta online, de acuerdo con la política y normativa comunitaria en este contexto.

La detección de estos libros permitiría su aprovechamiento en proyectos de digitalización europeos y la preservación de la herencia cultural común

2. Metodología

Se ha creado un programa en lenguaje PHP, para la búsqueda automática de obras descatalogadas o fuera de los cauces comerciales. Estas funciones de búsqueda no se encuentran sistematizadas en la bibliografía consultada, y fueron desarrolladas expresamente, planteando la siguiente premisa: Todo libro que no figure en los principales catálogos y portales de comercialización bibliográfica, puede considerarse fuera del circuito comercial. También tendrán la misma consideración, aquellos libros que presenten el estado “descatalogado”, o bien no figuren en las fuentes bibliográficas oficiales, como el catálogo del ISBN del *Ministerio de Educación, Cultura y Deporte*. En virtud de estas consideraciones, el programa ha sido diseñado con las siguientes funciones:

- Importación de los registros bibliográficos que fueren objeto de análisis.
- Anexión de datos estadísticos vinculados, tales como el número de préstamos, ejemplares, consultas y reservas para cada libro.
- Corrección de duplicaciones, errores, registros vacíos y marcación de obras traducidas.
- Búsqueda y detección automática de presencia o ausencia en los catálogos comerciales (dedicados a la venta de libros) y funcionales (dedicados a labores consultivas).
- Generación de informes y estadísticas sobre la presencia de los registros bibliográficos en los catálogos, libros descatalogados, para reutilización y digitalización.

En segundo lugar, se ha aplicado un método de prototipado (**Pfleeger; Atlee**, 2010, pp. 48-57), para confeccionar la rutina de extracción de datos y su comparación en un catálogo bibliográfico de referencia. Esto conlleva la detección de la presencia o ausencia del libro, para determinar con posteridad su situación. Finalmente se utiliza el método de desarrollo incremental (**Pressman**, 2005, pp. 35-36), adaptando dicha técnica al caso de cada catálogo. En este programa se ha decidido trabajar con los catálogos más representativos, *Amazon*, *Todos tus libros*, *Iberlibro*, *ISBN*, *BNE*, *BNEdatos*, *Rebiun* y *Google books*.

El prototipo del programa considera los siguientes puntos:

1) Delimitación y preparación de los libros sometidos al análisis

El programa trabaja con colecciones bibliográficas universitarias, dada la accesibilidad y disponibilidad de sus catálogos online. En concreto, la entrada de datos está orientada al caso de la Biblioteca de la *Universidad Complutense de Madrid*, por motivos de filiación, aunque ésta podría adaptarse a la de cualquier biblioteca universitaria. Teniendo en cuenta que la finalidad es discernir la disponibilidad de los documentos en los catálogos comerciales (*Amazon*, *Todos tus libros*, *Iberlibro*), la selección de libros será efectuada por el investigador en torno a alguno de los siguientes criterios:

- obras más prestadas;
- obras publicadas en fechas conflictivas, anteriores o recientes a la incorporación del ISBN o el depósito legal;
- colección bibliográfica completa de la biblioteca.

En estos casos, el programa puede proporcionar información para determinar el estado de los derechos de las obras más prestadas de la biblioteca, y por tanto con mayor uso. También, la situación de las obras más antiguas, cuyos derechos de autor aún no han expirado, pero cuyo registro y normalización aún puede ser confuso. Finalmente, el porcentaje de obras descatalogadas respecto al total de la colección.

2) Creación de la base de conocimiento

El método para la extracción de datos bibliográficos puede variar levemente en función del código fuente de cada catálogo OPAC. En este caso, el catálogo de la *Biblioteca Complutense* dispone de un número de control *OCLC*, que coincide con el código de enlace permanente de las fichas catalográficas. Por ello, el programa utiliza esta singularidad, extrayendo la información bibliográfica, mediante la técnica de *web-scraping*, muy conocida y documentada por autores como **Ward** (2013) y **Pautasso** (2013) entre otros. En la tabla 1 se muestra la rutina del programa encargada de esta tarea. En primer

lugar, se realiza una petición HTTP del enlace permanente, mediante la función “file_get_contents”, obteniendo una respuesta, que es su código fuente HTML.

Tabla 1. Rutina para la obtención del código fuente HTML

```

/* Ejemplo de enlace permanente $urlUCM="https://ucm.on.worldcat.org/search?queryString=1024651745"; /**/

// Obtener el código fuente HTML
$dataHTML = file_get_contents("$urlUCM");

// Transformaciones del código HTML para facilitar la extracción de datos
$dataHTML = preg_replace("&nbsp;","",$dataHTML);

// Se guarda la información en el objeto DOM
$dataDOM = new DOMDocument();
@$dataDOM->loadHTML("$dataHTML");
$pathDATA = new DOMXPath($dataDOM);

// Título principal
$itemTitle = $pathDATA->query("//div[@class='title']/a")->item(0)->nodeValue;

// Título adicional
$itemAlternateTitle = $pathDATA->query("//dd[@class='additionalTitle']")->item(0)->nodeValue;

// Enlace
$itemLink = $pathDATA->query("//div[@class='title']/a/@href")->item(0)->nodeValue;
$itemLink = "https://ucm.on.worldcat.org$itemLink";
$itemLink = preg_replace("&?.*$/","",$itemLink);

// Autor principal
$itemAuthor1 = $pathDATA->query("//span[@class='author-string-primary']")->item(0)->nodeValue;

// Autor secundario
$itemAuthor2 = $pathDATA->query("//span[@class='author-string-secondary']")->item(0)->nodeValue;

// Formato
$itemFormat = $pathDATA->query("//dd[@class='item-type']")->item(0)->nodeValue;

// Edición
$itemEdition = $pathDATA->query("//dd[@class='edition']")->item(0)->nodeValue;

// Serie/Colección
$itemCollection = $pathDATA->query("//dd[@class='series']")->item(0)->nodeValue;

// Fecha de publicación
$itemPublicationYear = $pathDATA->query("//dd[@class='publication-year']")->item(0)->nodeValue;

// Descripción física
$itemPhysicalDesc = $pathDATA->query("//dd[@class='physical-description']")->item(0)->nodeValue;

// ISBN
$itemIsbn = $pathDATA->query("//dd[@class='isbn']")->item(0)->nodeValue;
$itemIsbn = trim($itemIsbn);
$itemIsbn = preg_replace("/.*$/","",$itemIsbn);

// Materias
for($i=0; $i<=10; $i++){
    $itemSub = $pathDATA->query("//dd[@class='subjects']/ul/li/a")->item($i)->nodeValue;
    if($itemSub==""){ break; }
    $fullSubjects .= "$itemSub;";
    unset($itemSub);
}

// Permalink
$itemPermalink = $pathDATA->query("//input[@id='permalink-field-$code']/@value")->item(0)->nodeValue;

```

A continuación se requieren transformaciones sobre el código fuente para facilitar la extracción de los datos bibliográficos. En concreto, se eliminan las entidades HTML referidas a espacios y tabulaciones presentes con el código “ ”. En segundo lugar, el código fuente HTML se carga en un objeto DOM (Zandstra, 2013), desde el cual se realizan consultas XPath para extraer los datos bibliográficos. Esto es, su título principal, título secundario o adicional, autor principal, autor secundario, formato, edición, serie o colección, fecha de publicación, descripción física, ISBN y materias. Una vez se obtienen los datos de la catalogación se almacenan en la tabla principal de la base de datos. En la tabla 2 se aprecia

la consulta SQL empleada a tal efecto. Obsérvese que, cerrando la rutina de inserción de datos, se rellena el campo de indexación “indexer”, que contiene los textos del registro, facilitando su recuperación con el método *Fulltext FTS* de MySQL (Widenius; Axmark; Arno, 2002, pp. 548-553).

Tabla 2. Muestra del código SQL de inserción de datos del programa

```
INSERT INTO bibliocat SET id='378',originFILE='input/1940.txt',code='b1439728',newcode='b1439728',url='https://ucm.on.worldcat.org/search?databaseList=&queryString=911855520,author='Einaudi, Luigi, 1874-1961',authorsec=',title='Principii di scienza della finanza. Luigi Einaudi',edition=',trad=',publication='Torino : Giulio Einaudi, 1940',physical='XXVIII,539p. ; 25cm',subject='Hacienda pública -- Italia',collection='Collezione di opere scientifiche di economia e finanza',permalink='https://ucm.on.worldcat.org/oclc/911855520',isbn=',presenceAmazon=',linkAmazon=',reliAmazon=',presenceDilve=',linkDilve=',reliDilve=',presenceTuslibros=',linkTuslibros=',reliTuslibros=',presencelberlibro=',linklberlibro=',reliiberlibro=',presenceRebiun=',linkRebiun=',presenceBne=',linkBne=',reliBne=',presenceBnedatos=',linkBnedatos=',reliBnedatos=',presencelsbn=',linklsbn=',reliISBN=',descatlsbn=',presenceGooglebooks=',linkGooglebooks=',reliGooglebooks=',indexer='einaudi luigi 1874-1961 principii di scienza della finanza. luigi einaudi torino giulio einaudi 1940 collezione di opere scientifiche di economia e finanza hacienda publica italia';
```

Otro aspecto reseñable es que, en este estadio, todavía no se pueden rellenar todos los campos del registro. En concreto se trata de los campos con prefijo “presence”, “link” y “reli”, ya que están destinados a su compleción posterior:

- los campos de tipo “presence” registran la presencia o ausencia de los libros con un valor lógico;
- los campos de tipo “reli” sirven para anotar el porcentaje de similitud con respecto al resultado obtenido en la consulta;
- los campos de tipo “link” registran el enlace de consulta para cada catálogo. De esta forma, si fuera necesario, puede comprobarse manualmente el grado de acierto del programa, accediendo a la página del catálogo en la que se encuentra el libro.

3) Revisión automática en el catálogo de prueba

Registrados los datos bibliográficos de la muestra, el programa presenta un método normalizado de consulta para comprobar la presencia o ausencia de los libros en los catálogos. Se resume en los siguientes pasos:

- Listar los libros registrados y preparar sus datos bibliográficos para la consulta: esto es, realizar cambios en los sets de caracteres, supresión de saltos de línea, reemplazo de caracteres no aceptados conforme a las especificaciones de consulta de cada catálogo.
- Diseño de la estrategia de consulta automatizada: usando el protocolo RESTful (Rodríguez, 2015) se identifican las variables de consulta pública de cada catálogo.

El principal problema reside en la identificación de estas obras, que presentan carencias bibliográficas, por una catalogación deficiente, o bien presentan un versionado confuso

Tabla 3. Variables de consulta pública de los catálogos

Catálogo / Variables	Título	Autor	Fecha de publicación	ISBN
Amazon	&field-title=	&field-keywords=		
Todos tus libros	&keyword=			
Iberlibro	&tn=	&an=	&kn=	&isbn=
ISBN	¶ms.liConceptosExt[0].texto=		¶ms.reAnnoPublica.annoDesde= ¶ms.reAnnoPublica.annoEntre=	¶ms.cisbnExt=
Google books	&q=			
Rebiun	&q=			
BNE	&searchdata1=			
BNEdatos	?s=			

En la tabla 3 se observa que las variables de consulta difieren según el catálogo y ello determina o limita el modo de búsqueda. Por ejemplo, el catálogo *Todos tus libros* hace posible la recuperación a través de la variable “&keyword=” a modo campo de búsqueda a texto completo, igual que la variable “&q=” de *Google books* y *Rebiun*. Sin embargo, *Iberlibro* consta de variables específicas para precisar mejor las búsquedas. Sin duda, ya se advierte que el éxito de los resultados depende en parte del sistema de recuperación en cada catálogo, sus campos de búsqueda, la precisión de su algoritmo, pero también de la cadena de consulta utilizada en cada caso. Las consultas por defecto en el programa son las siguientes:

- Q1: Título, Autor, Fecha de publicación.
- Q2: Título reducido, Autor reducido, Fecha de publicación.
- Q3: ISBN.

Tabla 4. Ejemplo de consultas automáticas en Amazon

```

$urlQuery1="https://www.amazon.es/s/ref=nb_sb_noss?url=search-alias=stripbooks&field-title=$varTitle1&field-keywords=$varAuthor1%20$varPublicationYear";

$urlQuery2="https://www.amazon.es/s/ref=nb_sb_noss?url=search-alias=stripbooks&field-keywords=$varTitle1%20$varPublicationYear%20$varEditorial";

$urlQuery3="https://www.amazon.es/s/ref=nb_sb_noss?url=search-alias=stripbooks&field-keywords=$varIsbn";

```

El propósito de las consultas es la localización del libro en el catálogo. Si bien el caso de *Google books* y *Amazon* resulta más sencillo, en virtud de la precisión y nivel de desarrollo de sus algoritmos, en otros casos esto puede ser una tarea compleja. Por ello, el programa analiza la extensión, número de caracteres y palabras del título y el autor, para determinar si se puede reducir la extensión de la consulta, sin que ello afecte a la significación y univocidad. Ello se consigue estableciendo límites de extensión de 250 caracteres en los campos de título y autor, o bien eliminando iniciales del nombre, manteniendo los apellidos, suprimiendo palabras vacías, o asegurando al menos 5 palabras en cada consulta.

El objeto de la investigación es el desarrollo de un programa capaz de asistir en la detección de obras fuera de los canales comerciales

c) *Web-scraping* de las consultas. Definidas las consultas, éstas se ejecutan secuencialmente usando el método de *web-scraping*, hasta agotar todas las posibilidades de búsqueda planteadas. Esto es, en caso de que la primera consulta arroje resultados negativos, se procederá a la siguiente de forma recurrente hasta obtener un resultado positivo de la presencia del libro.

Tabla 5. Método de identificación de libros en catálogos bibliográficos

```

$refQuery="$varTitle1 $varAuthor1 $varPublicationYear";

$dataHTML = file_get_contents("$urlQuery1");
$dataHTML = preg_replace("/\&nbsp;/", " ", $dataHTML);

if(preg_match_all("/( resultado| resultados|nico resultado)/im", $dataHTML) && !preg_match_all("/ 0 resultados/", $dataHTML)){

    $dataDOM = new DOMDocument();
    @$dataDOM->loadHTML("$dataHTML");
    $xpathDATA = new DOMXPath($dataDOM);

    for($i=0; $i<=10; $i++){
        $item = $xpathDATA->query("//div[@class='s-item-container']")->item($i)->nodeValue;
        similar_text($refQuery, $item, $percent);
    }
}

```

d) Identificación de resultados. La presencia o ausencia de un libro en un catálogo online puede detectarse a través de fragmentos textuales y códigos denotativos. Por ello se aplican las siguientes reglas:

- ¿En el código fuente se encuentran frases o palabras unívocas que confirman la presencia de resultados? Por ejemplo, “n resultados”, “n resultado”, “resultado único”, “único resultado”. Esto se consigue definiendo expresiones regulares que permitan filtrar el código fuente de la página de resultados de cada catálogo. Véase función “preg_match_all” de la tabla 5.
- En caso positivo, ¿el grado de similitud del texto que conforman el título, autor y fecha de publicación supera el umbral de precisión estipulado por el investigador?
- Si el coeficiente de similitud supera el porcentaje establecido, entonces el programa marca el libro consultado, como “presente” en el catálogo. Tal efecto se registra, así como las consultas realizadas, según se muestra en la tabla 6.

El programa busca automáticamente todos los libros, sometidos a revisión, en los catálogos comerciales y funcionales, para determinar si se encuentran presentes y diagnosticar su estado actual

Tabla 6. Actualización de datos correspondiente al proceso de comparación

```
UPDATE FROM bibliocat SET presenceAmazon='1', linkAmazon='https://www.amazon.es/s/ref=nb_sb_noss?url=search-alias=stripbooks&field-title=Perspectiva+para+artistas&field-keywords=Benedicts+U|https://www.amazon.es/s/ref=nb_sb_noss?url=search-alias=stripbooks&field-keywords=Perspectiva+para+artistas%20Benedicts+U|https://www.amazon.es/s/ref=nb_sb_noss?url=search-alias=stripbooks&field-keywords=', reliAmazon='49.655172413793', presenceDilve='', linkDilve='', reliDilve='', presenceTuslibros='1', linkTuslibros='https://www.todostuslibros.com/busquedas/?keyword=Perspectiva+para+artistas%20Benedicts+U|https://www.todostuslibros.com/busquedas/?keyword=Perspectiva+para+artistas%2019%20Benedicts+U|https://www.todostuslibros.com/busquedas/?keyword=', reliTuslibros='11.075949367089', presencelberlibro='1', linkiberlibro='https://www.iberlibro.com/servlet/SearchResults?sts=t&tn=Perspectiva+para+artistas&n=Benedicts+U&sts=t|https://www.iberlibro.com/servlet/SearchResults?sts=t&isbn=', relilberlibro='67.692307692308', presenceRebiun='1', linkRebiun='http://rebiun.baratz.es/rebiun/search?q=Perspectiva+para+artistas%20Benedicts+U|http://rebiun.baratz.es/rebiun/search?q=Perspectiva+para+artistas%2019%20Benedicts+U|http://rebiun.baratz.es/rebiun/search?q=', reliRebiun='8.5510688836105', presenceBne='', linkBne='', reliBne='', presenceBnedatos='', linkBnedatos='', reliBnedatos='', presencelsbn='1', linklsbn='http://www.mcu.es/webISBN/buscarLibros.do;jsessionid=9EEC1D21B2330646D0AF4924EBEC757A?params.claseRegistroFormId=33&params.liConceptosExt[0].concepQueryFormIdExt=1&params.liConceptosExt[0].texto=Benedicts%20&params.liConceptosExt[1].concepQueryFormIdExt=5&params.liConceptosExt[1].texto=Perspectiva%20artistas%20&params.cdispo=A&params.cisbnExt=&params.reAnnoPublica.annoDesde=&params.reAnnoPublica.annoEntre=&params.orderByFormId=1&action=Buscar&language=es&prev_layout=busquedaisbn&layout=busquedaisbn|http://www.mcu.es/webISBN/buscarLibros.do;jsessionid=9EEC1D21B2330646D0AF4924EBEC757A?params.claseRegistroFormId=33&params.liConceptosExt[0].concepQueryFormIdExt=1&params.liConceptosExt[0].texto=&params.liConceptosExt[1].concepQueryFormIdExt=5&params.liConceptosExt[1].texto=Perspectiva%20artistas%20&params.reAnnoPublica.annoDesde=&params.reAnnoPublica.annoEntre=&params.cdispo=A&params.cisbnExt=&params.orderByFormId=1&action=Buscar&language=es&prev_layout=busquedaisbn&layout=busquedaisbn|none', relilsbn='28.729281767956', descatsbn='0', presenceGooglebooks='1', linkGooglebooks='https://www.google.es/search?hl=es&tbo=p&tbm=bks&q=Perspectiva+para+artistas+Benedicts+U+19|none|none', reliGooglebooks='30.357142857143' WHERE id='4';
```

Funcionamiento

El programa presenta una entrada de datos en forma de archivo de texto, en el que se pueden proporcionar los números de control de los registros bibliográficos de la biblioteca en cuestión. Esta información puede obtenerse a través de la biblioteca que será sometida al estudio. También es posible editar directamente la tabla principal con los datos bibliográficos de los libros y autores en los que interesa centrar el estudio, usando para ello programas de gestión de bases de datos, tales como *PhpMyAdmin*. Esta información se reutiliza para su consulta en el catálogo OPAC de la biblioteca a fin de alcanzar su completación bibliográfica. Esto significa cumplimentar al menos los campos título, autor, autor secundario, edición, publicación, descripción física, colección y número ISBN. También se permite la introducción de datos suplementarios en tablas relacionadas como por ejemplo el número de préstamos, ejemplares disponibles, lista de préstamos por ejemplar, entre otros. Esta información favorece la ordenación de los libros en función de su interés, distinguiendo qué libros fuera del circuito comercial tienen mayor ratio de préstamos y cuáles podrán ser susceptibles de digitalización.

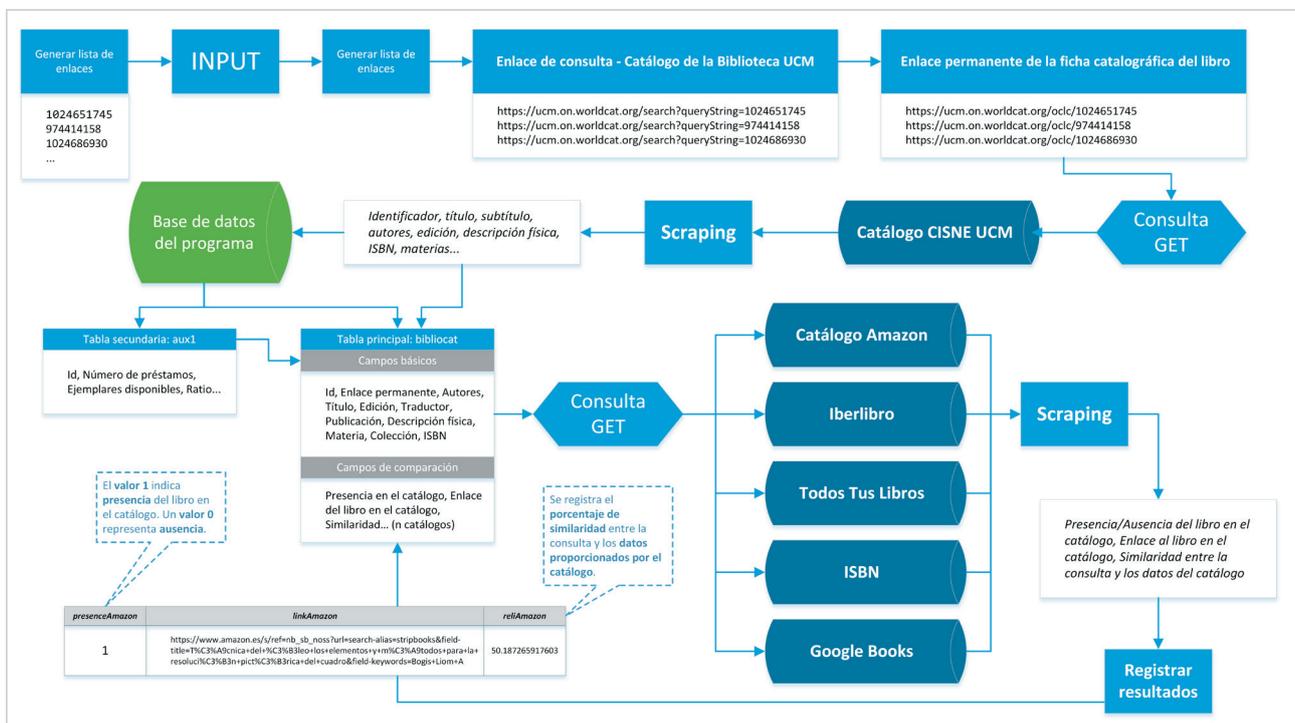


Figura 1. Funcionamiento del programa

El programa realiza un análisis de los libros que tienen traductores. Resulta relevante distinguir esta propiedad, a fin de asegurar que el libro no presenta derechos de propiedad añadidos que dificulten su posterior reutilización. Así pues, el programa descarta automáticamente estos libros, anotando en la base de datos la presencia de un traductor. Completados estos pasos, se procede a la tarea principal de comparación en otros catálogos. Esto es, detectar la presencia o ausencia de los libros previamente registrados en catálogos comerciales y funcionales. Se consideran catálogos funcionales aquellos que proporcionan una información de referencia para la comparación de los registros bibliográficos, pero que no determinan el estatus descatalogado o fuera de circuito comercial. En base a esta consideración, el programa contrasta la información en los catálogos de *Amazon*, *Todos tus libros* e *ISBN* (soportados por *Dilve*), *Iberlibro*, *Rebiun*, *BNE Datos*, *BNE* y *Google books*. La comparación se realiza en base a un algoritmo que diseña varias combinaciones de consulta, según la casuística de cada catálogo, con los datos del título, autor, fecha de publicación e ISBN del libro.

Figura 2. Interfaz del programa

Usando nuevamente el método de *web scraping*, se recuperan las páginas de resultados de los catálogos para cada consulta, hasta localizar el registro más exacto. Esto se consigue cuando el resultado es igual a 1 y el coeficiente de similitud relativo para dicho catálogo es elevado. Dicho coeficiente se obtiene al comparar el texto del primer resultado obtenido con el de la consulta, comprobando la presencia o ausencia de sus palabras clave. Si bien una coincidencia exacta debería proporcionar el 100% de similitud, también se recuperan palabras y textos de relleno o contextuales que reducen este porcentaje. Sin embargo, se llega a promediar un valor de similitud media, por el

Un libro es susceptible de reutilización digital, siempre que esté ausente en los catálogos comerciales, no sea una edición traducida o versionada y figure como descatalogado en la base de datos del ISBN, o bien no se encuentre en la misma

cual los registros recuperados realmente coinciden con los buscados. Este valor varía en cada catálogo y determina un margen de confianza aproximado sobre el cuál puede fundamentarse el resultado del programa. Un libro cuya presencia ha sido detectada, y que devuelva un porcentaje de similitud elevado para la media, supondrá un libro encontrado. Este hecho es registrado por el programa con un valor lógico 0 o 1, así como las consultas utilizadas para su diagnóstico. Si bien todos los catálogos cumplen este criterio operativo, el caso del catálogo del ISBN presenta una notable diferencia: este catálogo proporciona, además, el dato “descatalogado” que indica claramente la situación oficial del libro. De esta forma un libro presente en el catálogo de *ISBN*, que figure como descatalogado, o bien que no figure en dicha base de datos, se considera fuera del circuito comercial español.

Obtenidos los datos, el programa puede generar informes y estadísticas que incluyen la lista de libros presentes y ausentes en cada catálogo, libros fuera de los canales comerciales y presencia en los catálogos funcionales. Aparte, el programa aplica una fórmula de filtrado para generar la lista de libros descatalogados, fuera de comercio y cuyo contenido puede ser reutilizado para su digitalización en el marco de la biblioteca universitaria. Así pues, un libro es susceptible de reutilización digital siempre que esté ausente en los catálogos comerciales, no sea una edición traducida o versionada y figure como descatalogado en la base de datos del ISBN, o bien no se encuentre en la misma. Esto conduce a la fórmula para discriminar obras fuera de comercio, disponible en la tabla 7.

Tabla 7. Fórmula para discriminar las obras fuera de comercio

$$\begin{aligned} & \textit{Obra fuera de comercio} \\ & = (\textit{Libro} \notin \textit{Amazon}) \cap (\textit{Libro} \notin \textit{Iberlibro}) \cap (\textit{Libro} \notin \textit{Todos Tus Libros}) \\ & \cap ((\textit{Libro} \notin \textit{ISBN}) \cup (\textit{Libro} \in \textit{ISBN} \cap \textit{Libro} \ni \textit{ISBN Descatalogado})) \cap \textit{Libro} \nrightarrow \textit{Traductor} \end{aligned}$$

En otras palabras, el programa determina que un libro está fuera de comercio porque después de configurar diversas consultas en base a los datos bibliográficos disponibles, no se recupera en los catálogos comerciales ya mencionados. Su condición se certifica al comprobar que tampoco se encuentra disponible en el catálogo del ISBN, ya sea por omisión o por consideración de su estatus “descatalogado”, lo que indica inequívocamente que el libro ya no se comercializa.

4. Prueba de fiabilidad

Para determinar la validez del software y el método de discriminación de obras fuera de comercio, se ha elaborado una prueba de fiabilidad. En primer lugar se obtuvo la lista de los 80.000 libros más prestados de la *Biblioteca de la Universidad Complutense de Madrid* entre el año 2010 y 2017 de obras publicadas en España antes del año 2000. El sesgo de la muestra pretendía destacar el dato de que estas obras de ya cierta antigüedad en el mercado (al menos 10 años) son todavía muy demandadas entre los estudiantes universitarios e investigadores.

En el siguiente paso se tomaron muestras aleatorias de entre 80 y 100 registros de las obras más prestadas cuyas fechas de publicación declaradas eran 1947, 1967, 1987 y un último grupo de las 100 obras más prestadas en el conjunto del período (éstas son las obras más prestadas entre 2010 y 2017, publicadas en España antes del año 2000), a fin de comprobar tanto el volumen de obras demandadas, como la precisión del programa en varios periodos, previos y posteriores a la implantación del ISBN. Una vez preparadas las muestras, se ejecuta el programa, para obtener sus resultados de discriminación, presencia y ausencia en los catálogos. Finalmente se verifica manualmente uno a uno cada registro, comprobando los errores y aciertos del programa. El porcentaje de error es igual al número de errores dividido por el número total de registros de la muestra, multiplicado por 100.

Tabla 8. Resultados de la prueba de fiabilidad

	Muestra 1947	Muestra 1967	Muestra 1987	Muestra de los títulos más prestados
Número de registros	86	99	100	100
Fuera de comercio	68	69	59	58
Fuera de comercio (control manual)	76	72	66	56
Errores de diagnóstico	8	3	7	2
Porcentaje de error	9,3%	3,0%	7,0%	2,0%
Porcentaje de acierto	90,7%	97,0%	93,0%	98,0%
Libros digitalizados	54	52	32	17

Globalmente, los resultados obtenidos muestran una gran aproximación entre los datos arrojados por el programa y la prueba de fiabilidad realizada manualmente, ya que hay varios errores repetidos, que se contrarrestan entre sí. Nos referimos a los errores en los catálogos, de diversa índole: títulos de obras y nombres de autores acortados, datos incompletos de las obras, cambios en las grafías, errores en títulos y subtítulos. Es decir, una notable falta de homogeneidad de la información que contienen los catálogos para localizar una misma obra. Asimismo, al haber analizado catálogos comerciales junto a catálogos bibliotecarios, se pone de relieve la predisposición de aquellos por forzar resultados positivos en las búsquedas. Por ejemplo, ante una determinada búsqueda en *Amazon*, de la que no se hallan resultados, en lugar de devolver una consulta vacía, el catálogo presenta la respuesta: “sé el primero en opinar sobre este producto”. Si no se comprueba esa información, el programa necesariamente responderá que ha habido un resultado a la consulta, cuando en realidad se trata de un falso positivo. Esto se debe a los métodos de sugerencia automática, que emplean la información del historial de búsquedas, para devolver resultados patrocinados (**Ricci; Rokach; Shapira**, 2015). Su objetivo es mejorar la tasa de conversión, por lo que el interés comercial prevalece en todo caso. También se da la situación en la que un libro figura casi sin existencias, y al cual se le atribuye, un precio irreal. Costó tiempo entender este comportamiento, a tenor de los datos obtenidos en todas nuestras pruebas.

El software de la investigación se encuentra disponible para su aprovechamiento, revisión y reedición, en forma de código libre en:
<https://sourceforge.net/projects/ocw-test>

Por otro lado, la decisión sobre si una obra está comercializada envuelve un amplio margen de subjetividad. En la decisión nos hemos orientado por:

- la disponibilidad de ejemplares de la edición;
- la actualización de los datos: la reproducción de la portada es un indicio de la puesta al día de las obras en los catálogos comerciales;
- la presencia de las obras en varios catálogos.

Por lo tanto, podemos afirmar que los resultados son correctos, aunque haya un margen de error provocado, no tanto por el funcionamiento del programa, como por el comportamiento de los catálogos y los errores que hemos señalado.

Por último se ha aprovechado la amplia disponibilidad de los datos para muestrear el porcentaje de obras de estos años que ya están digitalizadas. Llama la atención que, si nos remontamos a 1967 y años anteriores, el porcentaje de libros digitalizados sea superior al 50%. Pero eso no quiere decir, ni mucho menos, que los libros sean accesibles online. Organizar el acceso a esas obras debe empezar a considerarse una tarea prioritaria.

5. Conclusiones

El programa proporciona una primera aproximación para resolver la problemática de la detección de obras fuera de comercio. Es capaz de realizar búsquedas pre-configuradas en los principales catálogos comerciales, para detectar su presencia y diagnosticar automáticamente su estado. En este sentido, se define que una obra fuera de comercio es aquella que no está presente en los principales catálogos comerciales y tampoco figura en el catálogo del ISBN o bien figura en calidad de libro descatalogado. Además, para que el libro pueda ser reutilizado y formar parte de los programas de digitalización auspiciados por la directiva europea (*Comisión Europea*, 2016), se añade la condición de que el libro no presente traductores, ya que en tal caso implicaría nuevos derechos de autor que no necesariamente han decaído o a los cuáles no siempre se puede aplicar la citada directiva.

En los tests de fiabilidad efectuados sobre las muestras de libros aleatorias de los años 1947, 1967 y 1987, se obtiene un acierto medio en el diagnóstico de obras fuera de comercio, superior al 90%. Sin embargo, cabe señalar que se producen falsos positivos entre los resultados, debido a los mecanismos de sugerencia automática, y a la falta de completitud en las descripciones bibliográficas. En concreto, resulta muy difícil contrastar el tipo y número de edición en los catálogos comerciales, ya que no están preparados para incluir tales parámetros de búsqueda. Esto afecta a la distinción y precisión de las obras más versionadas.

La recuperación y detección de los libros en los catálogos, dependen de los siguientes factores:

- variables y campos de consulta RESTful disponibles;
- estrategia de combinación de datos bibliográficos para confeccionar la consulta;
- limpieza de los textos de consulta y adecuación al set de caracteres;
- algoritmo de recuperación integrado en cada catálogo;
- nivel de ofuscación del código fuente o de la información presente en las páginas de resultados de cada catálogo;
- grado de completitud y calidad de la información bibliográfica.

Debido a que el programa realiza búsquedas automáticas con diversas combinaciones de consulta para detectar los libros en cada catálogo, ello podría aprovecharse en tareas de evaluación y mejora del rendimiento de los buscadores y los OPACS bibliográficos. En concreto, los catálogos comerciales analizados, presentan protocolos API con limitadas variables de consulta, dificultando la búsqueda exacta de las obras. Si bien presentan campos de búsqueda básicos como

el título y el autor, en muchos casos no se puede filtrar la edición, el idioma o bien la fecha de publicación. En consecuencia, sería deseable la incorporación de un modo de búsqueda bibliográfico que facilitara los procesos de búsqueda diligente, proporcionando una ficha descriptiva paralela a la puramente comercial, en la que se aseguraran los datos básicos de identificación bibliográfica.

Las pruebas efectuadas sobre muestras de libros más prestados, y listas aleatorias de libros publicados en 1947, 1967 y 1987, obtienen un porcentaje de acierto aproximado del 90%

A pesar de las dificultades se demuestra que es posible desarrollar programas que asistan la búsqueda diligente de obras fuera de comercio. Sin embargo, aún se requerirá la supervisión humana, a fin de confirmar los resultados que se obtienen. Todo ello viene limitado por el grado de exhaustividad de las fichas catalográficas de las colecciones que se desean revisar y de la fidelidad de la información reseñada en los catálogos comerciales. Por tanto, resulta lógico que la realización de búsquedas diligentes, aplicadas a la detección de obras huérfanas, de forma automática, puede resultar una tarea aún más complicada, dada la heterogeneidad descriptiva de los catálogos online disponibles para su consulta con los que se verifican las ediciones y la autoría de los libros.

6. Bibliografía

Bensamoun, Alexandra (2014). "The French out-of-commerce books law in the light of the European orphan works directive". *Queen Mary journal of intellectual property*, v. 4, n. 3, pp. 213-225.
<https://doi.org/10.4337/qmjip.2014.03.03>

Comisión Europea (2005). *Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: I2010: Bibliotecas Digitales: [Sec(2005) 1194] [Sec(2005) 1195]*.
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex:52005DC0465>

Comisión Europea (2006). "Recomendación de la Comisión 24 de agosto de 2006 sobre la digitalización y la accesibilidad en línea del material cultural y la conservación digital (2006/585/CE)". *Diario oficial*, n. L 236/28 de 31/08/2006, pp. 1279-1281.
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32006H0585>

Comisión Europea (2008). *Comunicación de la Comisión al Consejo, al Parlamento Europeo, al Comité Económico y Social Europeo y al Comité de las Regiones - El patrimonio cultural europeo a un clic del ratón: avances en la digitalización y el acceso en línea al material cultural y en la conservación digital en la UE [Sec(2008) 2372]*.
<https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52008DC0513>

Comisión Europea (2016). *Propuesta de Directiva del Parlamento Europeo y del Consejo sobre los derechos de autor en el mercado único digital*.
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52016PC0593>

Curto-Polo, Mercedes (2017). "El destino de los libros olvidados". *European papers*, v. 2, n. 2, pp. 731-741.
<https://doi.org/10.15166/2499-8249/156>

Díaz-Noci, Javier; Tous-Roviroa, Anna (2012). "La audiencia como autor: narrativas transmedia y propiedad intelectual del público. Algunas reflexiones jurídicas". *El profesional de la información*, v. 21, n. 5, pp. 458-467.
<https://doi.org/10.3145/epi.2012.sep.03>

Evangelio-Llorca, Raquel (2018). "Estrategias de la Unión Europea para facilitar el uso de obras fuera del circuito comercial por instituciones de patrimonio cultural en el mercado único digital". *PEI. Revista de propiedad intelectual*, n. 60, pp. 13-66.
<https://www.pei-revista.com/numeros-publicados/numero-60-septiembre-diciembre-2-018>

Iglesias-Rebollo, César (2012). "Derecho e información". *El profesional de la información*, v. 21, n. 5, pp. 449-452.
<https://doi.org/10.3145/epi.2012.sep.01>

Janssens, Marie-Christine; Tryggvadóttir, Ran (2014). "Facilitating access to orphan and out of commerce works to make Europe's cultural resources available to the broader public". *SSRN papers*, 15 December, pp. 12-16.
<https://doi.org/10.2139/ssrn.2538097>

Millán, José-Antonio (2015). "Edición y difusión del libro". *El profesional de la información*, v. 24, n. 6, pp. 699-703.
<https://doi.org/10.3145/epi.2015.nov.01>

Parlamento Europeo (2012). "Directiva 2012/28/UE del Parlamento Europeo y del Consejo, de 25 de octubre de 2012, sobre ciertos usos autorizados de las obras huérfanas (Texto pertinente a efectos del EEE)". *Diario oficial*, n. L299/5 de 27/10/2012, pp. 253-260.
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex:32012L0028>

Pautasso, Cesare (2013). "RESTful web services: Principles, patterns, emerging technologies". In: Bouguettaya, Athman; Sheng, Quan; Daniel, Florian. *Web services foundations*. New York: Springer, pp. 31-51. ISBN: 978 1 4614 7517 0
https://doi.org/10.1007/978-1-4614-7518-7_2

Pfleeger, Shari-Lawrence; Atlee, Joanne (2010). *Software engineering: theory and practice*. Pearson, pp. 48-57. ISBN: 978 0 136061694

Pressman, Roger (2005). *Software engineering: A practitioner's approach*. McGraw-Hill, pp. 35-36. ISBN: 978 0 073655789
<https://doi.org/10.1002/0471028959.sof529>

Ramos-Simón, Luis-Fernando; Blázquez-Ochando, Manuel (2018a). "Soluciones para el acceso en línea de obras con derechos de autor pertenecientes a las colecciones de las instituciones documentales y de patrimonio cultural". *Revista general de información y documentación*, v. 28, n. 2, pp. 525-549.
<https://doi.org/10.5209/RGID.62838>

Ramos-Simón, Luis-Fernando; Blázquez-Ochando, Manuel (2018b). "Patrimonio digital: Obras renacidas, una nueva vida para las obras fuera del circuito comercial". *Scire*, v. 24, n. 2, pp. 13-23.
<https://www.iberid.eu/ojs/index.php/scire/article/view/4591>

Ricci, Francesco; Rokach, Lior; Shapira, Bracha (2015). "Recommender systems: Introduction and challenges". In: Ricci, Francesco; Rokach, Lior; Shapira, Bracha. *Recommender systems handbook*. Boston: Springer, pp. 1-34. ISBN: 978 1 4899763761
https://doi.org/10.1007/978-1-4899-7637-6_1

Rodríguez, Alex (2015). "Restful web services: The basics". *IBM developerWorks*, n. 33.
<https://developer.ibm.com/articles/ws-restful>

Tallova, Lydie (2014). "Copyright aspects of disclosure of works within the Europeana Digital Library". In: *Sgem2014 Conference on political sciences, law, finance, economics and tourism*, pp. 561-568.
<https://doi.org/10.5593/sgemsocial2014/B21/S5.074>

Ward, Jacob (2013). *Instant PHP web scraping*. Packt Publishing. ISBN: 978 1 782164760

Widenius, Michael; Axmark, David; Arno, Kaj (2002). "MySQL full-text search". In: Widenius, Michael; Axmark, David; Arno, Kaj. *MySQL reference manual: documentation from the source*. O'Reilly. ISBN: 978 0 596002653
<https://learning.oreilly.com/library/view/mysql-reference-manual/0596002653/ch06s08.html>

Zandstra, Matt (2013). "A simple parser". In: Zandstra, Matt. *PHP objects, patterns, and practice*. Berkeley: Apress, pp. 467-488. ISBN: 978 1 430260318
https://doi.org/10.1007/978-1-4302-6032-5_23



Si te interesan los
INDICADORES EN CIENCIA Y TECNOLOGÍA,
y todos los temas relacionados con la medición de la ciencia, tales como:
Análisis de citas, Normalización de nombres e instituciones, Impacto de la ciencia en la sociedad, Indicadores, Sociología de la ciencia, Política científica, Comunicación de la ciencia, Revistas, Bases de datos, Índices de impacto, Políticas de open access, Análisis de la nueva economía, Mujer y ciencia, etc.
Entonces **INCYT** es tu lista. Suscríbete en:
<http://www.rediris.es/list/info/incyt.html>