

DBME_3: ADQUISICIÓN DE DATOS, COMPOSICIÓN Y BASE DE DATOS NEBRIJA-VALDÉS

JORDI PORTA ZAMORANO
Universidad Autónoma de Madrid¹

Resumen

En este artículo se presentan la metodología y las herramientas empleadas para la adquisición de datos bibliográficos, la composición digital del *Diccionario bibliográfico de metalexigrafía del español* (DBME_3) y la creación del catálogo bibliográfico de metalexigrafía del español o base de datos Nebrija-Valdés.

Palabras clave: Composición digital, bases de datos bibliográficas.

DBME_3: DATA ACQUISITION, COMPOSITION AND NEBRIJA-VALDÉS DATABASE

Abstract

This article describes the methodology and tools employed in the acquisition of bibliographic data, the digital composition of the *Diccionario bibliográfico de metalexigrafía del español* (DBME_3), and the creation of the Spanish metalexigraphic catalogue or Nebrija-Valdés bibliographic database.

Keywords: Digital composition, bibliographic databases.

¹ Departamento de Ingeniería Informática. Escuela Politécnica Superior. Este trabajo complementa los artículos publicados en este mismo volumen a cargo de I. Ahumada Lara, «Metalexigrafía del español: clasificación orgánica y tipología de los diccionarios en el *Diccionario Bibliográfico de la Metalexigrafía del Español* (DBME)», págs. 5-24, y de E.J. Jacinto García, «El *Diccionario Bibliográfica de Metalexigrafía del Español* como obra de consulta: estructura, fuentes y funciones», págs. 147-169.

1. INTRODUCCIÓN

Este artículo se describe el trabajo realizado para la composición digital del tercer volumen del *Diccionario bibliográfico de metalexigrafía del español* (DBME_3) y para la creación del catálogo bibliográfico de metalexigrafía del español o base de datos Nebrija-Valdés. Además, se describen la metodología de trabajo colaborativa y las herramientas que fueron empleadas para la adquisición de las referencias bibliográficas del DBME_3.

2. METODOLOGÍA DE TRABAJO

El DBME_3 es el resultado del trabajo de varios colaboradores. Desde el primer momento fueron objetivos evitar el intercambio de información ineficaz y complejo mediante el uso de medios como el correo electrónico y evitar la proliferación de copias y versiones. Para ello se crearon un esquema y un entorno de trabajo distribuido. En el esquema de trabajo se establecieron dos roles: compilador y editor. Los compiladores se encargaron de hacer acopio del material bibliográfico, de clasificarlo orgánicamente y de introducir adecuadamente toda la información en registros con unos campos determinados usando una misma herramienta y con unas convenciones acordadas y documentadas en un manual. El editor, por su parte, se encargó de revisar periódicamente el material introducido por cada colaborador, de comprobar la coherencia y la consistencia de los datos, y de generar informes con los problemas detectados. Finalmente, la información de los distintos colaboradores fue fusionada para comprobar su coherencia global.

Un ejemplo de los problemas principales observados en la adquisición de datos es el de la variación en los nombres de los autores, ya que algunas veces la firma de un autor varía de una publicación a otra. En unos casos se añaden dos apellidos, en otros uno, a veces estos apellidos pueden ir coordinados, otras veces no. Incluso el nombre de pila puede variar por estar escrito en otra lengua vernácula o por ser este compuesto, en especial los nombres que contienen *María*, habitualmente abreviado por *M.^a*, o por contener partículas de enlace entre el nombre y el primer apellido. Toda esta variación en los nombres es sometida a un proceso de armonización de manera que se asegure la su unicidad. Esta unicidad de nombres permite que en una base de datos bibliográfica el número nombres distintos coincida con el número de autores distintos. También ha permitido que en la edición en papel del DBME_3 la bibliografía esté correctamente alfabetizada. Las fechas o el lugar de publicación, los números de página, las revistas, y muchos otros campos de las referencias bibliográficas fueron sometidos a un proceso similar, detectándose no pocos errores.

Tanto los compiladores como el editor compartieron una carpeta de trabajo especial en sus ordenadores cuyos contenidos estuvieron permanente sincronizados a través de un servicio de hospedaje de ficheros en la nube. En este caso fue el servicio proporcionado por Dropbox (<<https://www.dropbox.com>>) el que se usó para implementar el esquema de trabajo del *DBME_3*. Además, para evitar solapamientos e información redundante o inconsistente, fue decisión de la dirección del proyecto que los compiladores trabajaran en particiones disjuntas y especializadas de la bibliografía: artículos en revistas, homenajes, congresos, u otros.

3. ADQUISICIÓN DE REFERENCIAS BIBLIOGRÁFICAS

La introducción de los datos bibliográficos fue llevada a cabo por los compiladores mediante el programa informático JabRef para el manejo de bibliografías. En la fig. 1 puede verse el aspecto de este programa, que es configurable al gusto del usuario y en el que pueden distinguirse dos zonas bien diferenciadas: una, situada en la parte de arriba, que permite ver la lista de referencias con su información compactada y que permite la ordenación por los distintos campos bibliográficos, y una segunda zona en la que puede editarse toda la información de una referencia bibliográfica. El formato en el que JabRef representa la información bibliográfica es el mismo que el definido por el sistema BibTeX.

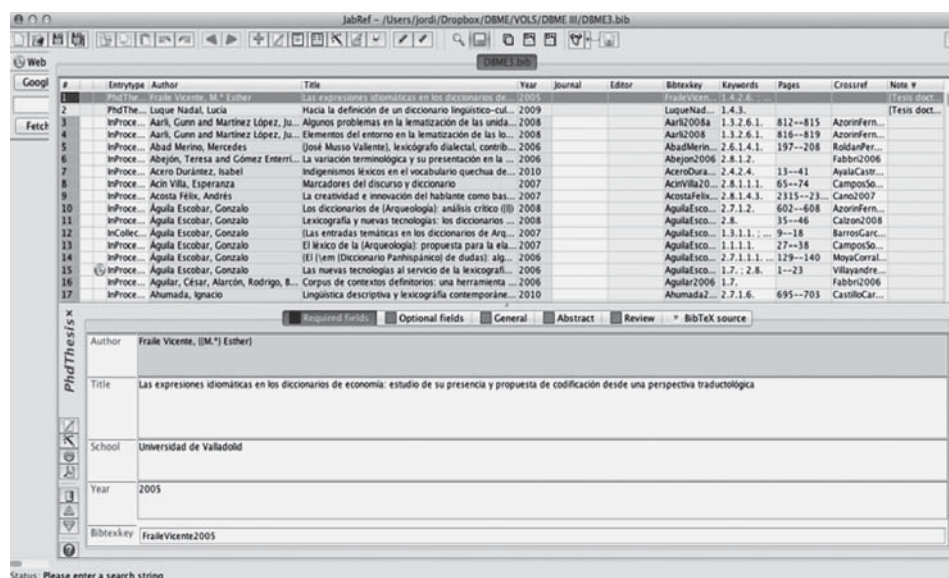


Figura 1. Editor JabRef.

El sistema BibTeX (<<http://www.bibtex.org>>) es en realidad un conjunto de herramientas informáticas para el manejo de bibliografías que incluye un sistema de marcación que se ha convertido ya en un estándar. Uno de los principios de diseño de BibTeX consiste en la separación de la información bibliográfica de la presentación o aspecto de esta información. En BibTeX los datos bibliográficos o campos correspondientes a los distintos tipos de publicaciones se indican mediante una serie de marcas estándar, como *author* para el nombre de los autores, *title* para el título del trabajo o *journal* para el título de la revista científica. BibTeX reconoce una tipología de más de una docena de registros bibliográficos distintos que permite representar la información de artículos en revistas o publicaciones periódicas, libros, capítulos de libros, artículos en compilaciones, actas de congresos, artículos en congresos, tesis, manuales, y otros. Para cada uno de estos registros hay una serie de campos que son requeridos y otros que son opcionales. Por ejemplo, la información requerida para un libro es el autor, editor, título, editorial y año. En la fig. 2 puede verse un ejemplo con una entrada bibliográfica generada con JabRef en formato BibTeX correspondiente a un artículo publicado en una revista en el que se incluye, además un resumen, la información sobre el compilador responsable y la fecha de alta, localizables ambas al final del registro en la fig. 2.

4. COMPOSICIÓN DEL DBME _ 3

Los originales en formato Word (<<https://products.office.com/es/word>>) de las partes *Bibliografía orgánica* y *Bibliografía alfabética* de la versión en papel del DBME _ 3 se han obtenido de manera automática. A partir de una bibliografía BibTeX, se genera un documento LaTeX con las citas y referencias que se compone con la propia bibliografía y el estilo bibliográfico del DBME.

LaTeX (<<http://www.latex-project.org>>) es un sistema de composición para textos académicos y científicos cuya calidad tipográfica resultante es comparable con la de las mejores editoriales científicas. LaTeX es un sistema de composición lógica, por oposición a los sistemas de composición visual, o sistemas WYSIWYG², en los que el autor escribe al mismo tiempo que ve aproximadamente el aspecto que tendrá el documento una vez impreso. En LaTeX hay una disociación entre la escritura del documento y la visualización del mismo documento compuesto diseñada para permitir al autor concentrarse en el contenido sin las distracciones ocasionadas al intentar conseguir la apariencia adecuada.

² WYSIWYG es el acrónimo para *What You See Is What You Get* («lo que ves es lo que obtienes»).

```

@Article{EdoMarza2009,
  Title = {{The Generation of Active Entries in a Specialised, Bilingual, Corpus-
based Dictionary of the Ceramics Industry: What to Include, Why and How}},
  Author = {Edo Marzá, Nuria},
  Journal = {Ibérica},
  Year = {2009},
  Pages = {43--70},
  Volume = {18},
  Keywords = {2.8.1.2.},
  Abstract = {La elaboración de entradas de diccionario útiles es siempre una labor
ardua por lo complicado de decidir qué incluir y cómo incluirlo. Así pues, esta
investigación tiene como principal objetivo mostrar cómo se han generado “entra-
das activas” en el caso concreto de la elaboración de un diccionario especializado
bilingüe basado en corpus en el campo específico de la cerámica industrial. De
este modo, el presente artículo ilustra cómo se han diseñado las entradas finales
y cómo se han tomado decisiones teniendo a los usuarios potenciales del diccio-
nario especialistas y traductores en el sector cerámico industrial como referencia.
Este artículo reflexiona además sobre cómo las entradas activas complementan
creaciones terminológicas previas con la inclusión de información adicional per-
tinentes y sobre los intrincados procesos de toma de decisiones necesarios para la
elaboración de este tipo de entradas. En la primera parte del artículo, se exponen
las consideraciones teóricas adoptadas mientras que la segunda parte muestra
las entradas teóricas propiamente dichas y la manera en que se han completado
los distintos campos; es decir, cómo las diferentes informaciones sobre contextos
de uso, implicaciones pragmáticas, clasificación semántica y definiciones, entre
muchos otros, han sido incluidos en las entradas con el fin de satisfacer las ne-
cesidades del usuario.},
  Owner = {lpnunez},
  Timestamp = {2013.05.19}
}

```

Figura 2. Entrada bibliográfica del *DBME_3* en BibTeX.

Por un lado, para la parte de la *Bibliografía orgánica*, se generan automáticamente las secciones correspondientes a cada uno de los valores de la clasificación orgánica. En cada sección se listan y agrupan alfabéticamente las citas bajo el mismo punto de la clasificación orgánica. En algunos casos se añade información sobre el país. Por otro lado se genera la lista alfabética de referencias bibliográficas citadas. El resultado es como el siguiente:

- 1.1. *El diccionario: concepto*
 Anaya Revuelta, 2007; Ávila, 2008a.
 Guerrero Ramos, 2006a; Gutiérrez Cuadrado, 2007.
 Ibarretxe Antuñano, 2010.

López Facal, 2010.

Moral, 2006.

En la sección *Bibliografía alfabética* se genera un listado alfabético por autor en el que se numeran las referencias bibliográficas, como en el siguiente ejemplo, en el que se manejan también referencias cruzadas:

5648

LAHUERTA GALÁN, Javier (2006): «El diccionario para la Primaria: herramienta para la descodificación», en Bernal/DeCesaris (ed.), 2006, pp. 125-137.

Tanto las citas como las referencias, siguen el formato establecido en el *DBME* en cuanto al orden de la información y tipografía. Para conseguirlo fue necesaria la definición de un estilo propio, ya que ninguno de los disponibles para BibTeX coincide con el requerido para el *DBME*. El sistema BibTeX incluye también herramientas para la programación de nuevos estilos bibliográficos para citas y referencias.

5. BASE DE DATOS DE CONSULTA A LA BIBLIOGRAFÍA A TRAVÉS DE INTERNET: NEBRIJA-VALDÉS

La versión electrónica del *DBME_3* tiene forma de base de datos consultable desde internet, la Nebrija-Valdés. Esta base de datos ha sido implementada con Refbase, otro software de libre distribución para repositorios de referencias bibliográficas y que es gestionable y consultable a través de Internet. Refbase importa y exporta referencias bibliográficas a distintos formatos estándar comunes como BibTeX, Endnote o RefWorks, entre otros. En nuestro caso, las referencias del *DBME_3* son directamente importadas en BibTeX.

El sistema de consulta utiliza un motor de búsqueda que permite combinar distintos criterios de búsqueda sobre cualquiera de los campos de los registros bibliográficos incluido en campo de palabras clave que contiene la clasificación orgánica de la referencia bibliográfica o el campo de resumen, que es consultable por cualquier combinación de palabras. El interfaz actual de Nebrija-Valdés se ha configurado y localizado al español y tiene el aspecto que se muestra en la fig. 3.

En el futuro próximo se incorporarán a Nebrija-Valdés los volúmenes previos del *DBME* así como avances de un futuro volumen.

 **Base de datos Nebrija-Valdés. Catálogo bibliográfico de la metalexigrafía del español**
 Página principal | Mostrar todo | Búsqueda simple | Búsqueda avanzada
 101-105 de 745 registros encontrados para su consulta (RSS | historia) palabras clave 1 Login

Search & Display Options

Seleccionar todos | Deseleccionar todo [11-20] << 21 22 23 24 25 26 27 28 29 30 >> [31-40] Vista lista | Citaciones | Detalles

Autor	Título	Año	Publicación	Palabras clave	Volumen	Páginas	Enlaces
<input type="checkbox"/> Castañer Martín, M.ª Rosa; Languéns, Vicente (eds)	De moneda nunca usada. Estudios dedicados a José M.ª Enguita Utrilla	2010		2.1.1.			
<input type="checkbox"/> Quilis Merín, Mercedes	De Nebrija a Franciosini (1497-1620): presencia y consolidación de las locuciones adverbiales en la lexicografía histórica del español	2008	Actas del VII Congreso Internacional de Historia de la Lengua Española. Mérida, Yucatán, 4-8 septiembre de 2008	2.3.1.1.; 2.4.2.2.; 2.1.1.	I	1017-1034	
<input type="checkbox"/> Borrás, Laura; Turner, Sergi	De nuevo sobre la marca de figurado en los diccionarios	2006	Palabra por palabra. Estudios ofrecidos a Paz Battaner	1.2.3.4.5.; 2.8.1.1.1.; 2.1.1.		33-45	
<input type="checkbox"/> Santos López, Javier Luis	Definición del léxico político en los DDMM: algunas dificultades	2008	Límites. Lexicografía y lexicología de las lenguas de especialidad	2.8.1.1.; 2.1.1.		343-358	
<input type="checkbox"/> Soler i Bou, Joan	Definición lexicográfica, vocabulario definidor y metalinguaje	2007	Reflexiones sobre el diccionario	1.3.2.2.; 2.1.1.		315-334	

Seleccionar todos | Deseleccionar todo [11-20] << 21 22 23 24 25 26 27 28 29 30 >> [31-40] Vista lista | Citaciones | Detalles

Cite & Export Options

Página principal | Búsqueda COI | Buscar en la Biblioteca | Mostrar registro | Extraer citas Ayuda

Figura 3. Interfaz de consulta de la Base de datos Nebrija-Valdés.