

## BIBCONT 2.1: UN PROGRAMA PARA EL ANÁLISIS DE CONJUNTOS DE REFERENCIAS BIBLIOGRÁFICAS

Pedro M. Valero y J. Gabriel Molina  
Universitat de València

### RESUMEN

BIBCONT 2.1 es una nueva versión del programa BIBCONT el cual es capaz de hacer recuentos de las referencias generadas por el traductor Psyclittotab. Ello permite realizar análisis cuantitativos de las salidas generadas por las bases de datos de referencias bibliográficas. Esta nueva versión es una base de datos de tipo relacional, la cual permite seleccionar, buscar y ordenar las referencias bibliográficas además de hacer los recuentos, además de otras posibilidades. Otras funciones que permanecen por añadir son análisis estadísticos sencillos y conexiones con el paquete estadístico SPSS.

### INTRODUCCIÓN.

Bibcont 2.1 es una nueva versión experimental del programa Bibcont, el cual forma parte de un conjunto de herramientas informáticas para el análisis de conjuntos de referencias bibliográficas que han ido siendo desarrolladas por el equipo de investigación de Métodos informáticos aplicados a la Psicología de la Facultad de Psicología de la Universitat de València (Valero et al, 1992)(Valero et al., 1992a). En esta comunicación esbozaremos en primer lugar la trayectoria seguida a la hora de confeccionar estas herramientas, en segundo lugar el estado actual del proyecto y por último algunos de los problemas que deseamos solucionar antes de poder ofrecer una herramienta terminada definitivamente.

### Experiencias previas: Bibcont y traductores.

Durante las últimas fechas (Valero and Molina, 1993a, Valero, et al., 1992a, Valero, et al., 1992b) hemos llevado a cabo una tarea que se ha ceñido a dos áreas principalmente:

1) En primer lugar, desarrollo de programas *traductores* encargados de convertir la información obtenida a través de una petición a una base de datos bibliográfica a un formato basado en tabuladores, un estandar dentro del sistema operativo Macintosh. Esta traducción proporciona los datos que permiten llevar a cabo análisis posteriores. En concreto los traductores disponibles son: Psyclittotab (Valero, et al., 1992a), Ericotab, Medlinetab y SCCLitotab (artículos en preparación) los cuales permiten la traducción desde respectivamente las bases de datos Psyclit, Eric, Medline y SSCI.

2) En segundo lugar, se ha desarrollado un programa capaz de hacer recuentos de esos campos tanto individualmente como de sus cruces. El resultado de estas dos herramientas combinadas permite averiguar la siguiente información a partir de un conjunto de referencias bibliográficas: frecuencia de autores, revistas, palabras claves, descriptores, años y el cruce entre estos campos como por ejemplo, frecuencia de palabras clave por cada año o frecuencia de aparición de autores en determinadas revistas. Otros programas desarrollados posteriormente permiten transformar ciertos campos en inputs adecuados para el cálculo de índices de similitud que sean la base

de análisis de la estructura de interrelaciones de un conjunto de autores o palabras clave (Valero y Molina, 1993b).

Todos estos programas, no obstante, presentan una serie de limitaciones que hemos pretendido superar. Estas son: dependencia de otros programas tanto para la introducción de datos como para visualizar los resultados finales, incapacidad de seleccionar la información a analizar y, por último, problemas de falta de RAM cuando los conjuntos de referencias superaban el millar.

Además, y por encima de todo, a estos programas les faltaba la capacidad de funcionar como herramientas interactivas a través de las cuales los usuarios pudieran llevar a cabo de modo completo sus análisis cuantitativos de la información. Para lograr ello era necesario plantear el problema en términos más complejos y sofisticados.

Esta transformación, de momento, ha estado centrada en el programa encargado de hacer los recuentos de la información (Bibcont) aunque, no obstante, esperamos que el desarrollo progresivo de este programa lleve a incluir los traductores de las bases de datos dentro de aquél, integrando por tanto los distintos módulos en uno solo.

### El Bibcont actual.

Bibcont 2.1 es una versión ligeramente modificada de Bibcont 2.0 el cual fue presentado anteriormente (Valero y Molina, 1993a). En el gráfico 1 es posible ver la estructura de ficheros y campos que lo componen. A continuación sigue una descripción de este esquema:

- Ficheros: En **negrita** se encuentra el nombre de los archivos que componen la base de datos. A la izquierda del gráfico 1 se encuentra **REFERENCIAS**, el archivo principal y en el que se encuentra la mayor parte de la información. A la derecha del gráfico se hallan representados **AUTORES, DESCRIPTORES, PALABRAS CLAVE y CITAS** los cuales almacenan la información "múltiple", es decir, aquella compuesta por varios elementos por referencia (p.e. los autores del artículo, los descriptores o las citas bibliográficas). La conexión entre estos archivos se lleva a cabo por medio de una clave (denominada simplemente número de referencia).

- Campos que componen los archivos.

Los nombres de los campos se encuentran también en la figura 1. Añadiremos ahora una breve descripción de su contenido, su tipo y sus características, pero antes introduciremos una breve discusión acerca de las distintas posibilidades barajadas en ellos.

Los campos de Bibcont pueden tratarse de campos indexados o no. Cuando el campo está indexado, la base de datos mantiene un registro interno de cada una de las palabras que se encuentran en ese campo para todos los registros. De este modo, ciertos procesos tales como las búsquedas no disminuyen su velocidad cuando la masa de referencias aumenta. No obstante, otras tareas tales como importar o añadir información sí disminuyen su velocidad. Además, el espacio de almacenamiento en disco crece también, alcanzando los archivos un tamaño apreciable.

Por otra parte, un buen número de las operaciones de recuento que Bibcont lleva a cabo tienen como condición que los campos correspondientes se encuentren indexados, así que los inconvenientes que presenta esta característica resultan difícil de soslayar. A cambio, estas tareas necesitan una cantidad de RAM muy inferior a la que necesitaba la versión anterior de Bibcont, por lo que llevar a cabo análisis de grandes masas de información no implica una problemática especial añadida a la del tiempo de procesamiento proporcionalmente mayor.

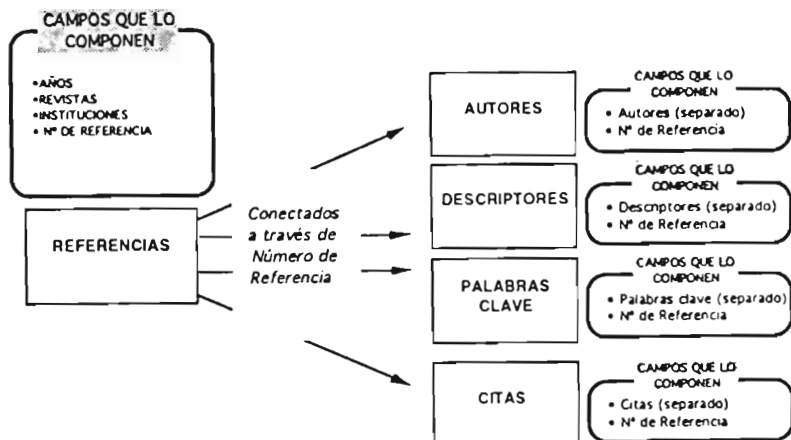
Veremos ahora una descripción de cada uno de los campos y de sus características:

- Fichero REFERENCIAS

- **Años:** Se trata de un campo de tipo numérico e indexado.
- **Revistas:** Se trata de un campo indexado.
- **Instituciones:** Se trata de un campo indexado.
- **Citas (Unidos):** Actualmente todavía por configurar.

-**Número de referencias:** Se trata de un campo indexado que permite a los otros archivos de la base de datos relacionar su información respectiva.

Figura 1. Esquema de archivos en Bibcont 2.1



- Otros ficheros

Los otros ficheros separados simplemente incluyen aquella información de los campos compuestos de varios elementos por referencias. Así, **Autores (separado)** el campo fundamental del fichero AUTORES incluye a los autores separando cada uno de ellos en un único registro. El campo Nº de referencia permite la conexión de esa información con el archivo general.

#### Diferencias entre Bibcont 2.0 y Bibcont 2.1

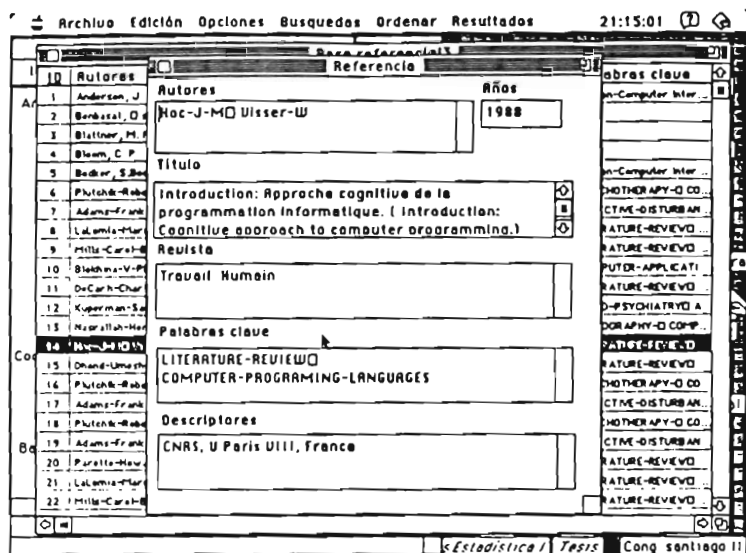
La diferencia fundamental entre estas dos versiones radica en que la información correspondiente a archivos con valores múltiples por campo se encontraba duplicada al encontrarse en el archivo central y en los secundarios. Ello suponía un tamaño mucho mayor de los archivos. Actualmente, al no haber información duplicada, el espacio en disco necesario para mantener la base de datos se ha reducido considerablemente. Asimismo, se han llevado a cabo otras mejoras más leves en orden a mejorar la presentación y el funcionamiento del programa.

## Funcionalidad de Bibcont.

### Introducción de información.

Bibcont 2.1 permite la introducción de datos de modo manual (desaconsejada) o bien su importación desde ficheros de texto (aconsejada). La primera de las opciones tiene el inconveniente de introducir errores de teclado que trastoran los resultados a realizar además de ser, naturalmente, más laboriosa y difícil. En un próximo futuro esperamos que Bibcont 2.1 añada una serie de opciones que faciliten la entrada de datos, tales como elaboración de índices, control de valores en blanco, etc. No obstante, la mejor forma de introducir los datos es importándolos desde ficheros de texto, los cuales a su vez tengan su origen en bases de datos de información bibliográfica. Por medio de los traductores que hemos desarrollado, las peticiones en esas bases pueden transformarse al formato basado en tabuladores ampliamente extendido en el mundo Macintosh el cual es importado fácilmente a Bibcont. Una pantalla que muestra la ventana de introducción de datos aparece en la figura 2.

Figura 2: Pantalla de Bibcont 2.1



### Recuentos de información en campos.

Uno de las capacidades más importantes en Bibcont 2.1 es poder hacer recuentos del número de veces que los elementos en un campo aparecen a lo largo de todos los registros. Por ejemplo, Bibcont indica el número de veces que cada autor aparece en un conjunto de referencias. Esta posibilidad es el elemento principal de funcionalidad de este programa y lo que añade respecto a bases de datos de otro tipo.

Asimismo, Bibcont 2.1 es capaz de hacer recuentos de cruces entre campos, detectando por ejemplo en cuantas referencias apareció un determinado autor en un año dado o cuantas veces se ha utilizado un descriptor dado en relación con él.

Es interesante comentar que Bibcont 2.1 es en ciertos aspectos inferior en este apartado a Bibcont 1.0. Así, Bibcont 1.0 era más rápido y permitía hacer recuentos de campos más complicados (hasta tres cruces) que Bibcont 2.1, debido a que su estructura de ficheros, y en general las operaciones llevadas a cabo internamente eran más sencillas.

En pocas palabras, el algoritmo utilizado por Bibcont 2.1, al estar basado en el indexamiento de cada campo, permite reducir el problema a pequeños subproblemas que son solucionados individualmente. Ello hace que Bibcont 2.1 presente tres ventajas que justifican este aumento en el tiempo empleado en sus cálculos: 1) Cuando el número de palabras a contar aumenta enormemente, ello no supone una disminución de la velocidad tal y como le ocurría a Bibcont 1.0, 2) No presenta problemas de RAM aunque el número de referencias sea muy grande y 3) Permite la selección de información, de modo que no se emplea tiempo en contar aquello que no interesa al usuario. En la figura 3 es posible ver el un ejemplo de recuento de frecuencias en Bibcont 2.1.

Figura 3: Un ejemplo de recuento llevado a cabo con Bibcont.

The screenshot shows the Bibcont 2.1 interface with a search results window. The window title is 'Base referenci...' and 'Frecuencias'. The table has columns: 'ID', 'Autores', 'Titulo', 'Años', 'Firma', 'F', and 'Clase'. The data rows list various authors and their works with corresponding years and frequency counts. A callout box with a pointer to the 'Firma' column contains the text: 'BIBCONT HACE RECuentOS DE LOS ELEMENTOS EN LOS CAMPOS QUE SE LE INDIQUE.' Below the main window, another window shows a similar table with a different set of data.

**Búsquedas.**

Bibcont 2.1 permite hacer búsquedas sobre la información introducida. Estas búsquedas tienen ante todo la finalidad de seleccionar información para luego llevar a cabo análisis y recuentos sobre ella. Por ejemplo, dada una masa de referencias, es posible solicitar aquellas que firma un determinado autor. Una vez hecho es posible pedir el recuento del número de firmas que otros autores comparten con él, obteniendo de este modo un cuadro general de la posición de ese autor. Es posible llevar a cabo similares selecciones y recuentos con el resto de los campos de información. En la figura 4 es posible ver un ejemplo de búsquedas.

Figura 7: Haciendo búsquedas en Bibcont 2.1

Archivos Edición Opciones **Busquedas** Ordenar Resultados 21:17:48 (7)

Base referencial				Frecuencias	
ID	Autores	Titulo	Años	Elemento	F
1	Anderson, J. R. & Jeffrey...	Navice LISP Errors Und...	1985	Adams-Frank	3
2	Benbasat, D. & Sammartino...	The Influence of Color an...	1986	Aggarwal-J-K	1
3	Blalock,			Anderson, J.R.	3
4	Bloom,				
5	Badker,				
6	Plutchik,				
7	Adams-				
8	Laloma-				
9	Mills-C				
10	Blekhin,				
11	DeCarli,				
12	Kupern,				
13	Nasrati,				
14	Moo-				
15	Dhend-				
16	Plutchik,				
17	Adams-				
18	Plutchik,				
19	Adams-				
20	Paretti-Howard-P & Van	Use of microcomputers	1990	Perceptual and Motor	LITERATURE-REVIEW
21	Laloma-Mary-JO Side-	Measurements of compu	1990	International Journal o	LITERATURE-REVIEW
22	Mills-Carel-BD Velden-	Reading text from comp.	1987	ACM Computing Surveys	LITERATURE-REVIEW

**Búsquedas**

Campos: Autores: Bloom, C. P.  
 revista: Human-Computer I...  
 palabras:  
 desc:  
 Number:

Titulo: Procedures for Obtai...

Buscar en campo:  
 Significativa  
 No significativa

Clave  Rango  
 1990

7 Registros hallados

Hallar Añadir Interseccionar Omitir OK

Estadística / Tesis / Conq santiago

#### Otras posibilidades pendientes de incorporar.

Partiendo de la estructura anterior, Bibcont 2.1 irá incorporando otras características que vayan añadiendo más posibilidades de análisis de la información. Algunos proyectos en marcha son:

- Conexión con SPSS (SPSS Inc., 1991).
- Generación de tablas para los campos cruzados: Actualmente, la información en Bibcont 2.1 para los cruces entre campos aparece en columnas. Dado que esta presentación es peor que la basada en tablas rectangulares procuraremos añadirla en un futuro próximo.
- Cálculo de índices bibliométricos (Carpintero, H. y Peiró, J. M., 1981): Estos índices (tales como el número de autores medio por referencia o la frecuencia media de aparición de un descriptor) serán añadidos en breve plazo.
- Campos de usuario: Se añadirá próximamente campos vacíos, de modo que estos puedan añadir palabras información a las bases de datos, permitiendo por ejemplo categorizar las referencias o indicar los grupos de cofirmación.

#### REFERENCIAS BIBLIOGRÁFICAS

- Carpintero, H. y Peiró, J. (1981). *Psicología contemporánea. Teoría y métodos cuantitativos para el estudio de su literatura científica*. Valencia: Alfaplus.
- Microsoft Corporation (1988). *Microsoft QuickBasic for Apple® Macintosh® Systems*.
- SPSS Inc. (1991). *SPSS Categories*. Chicago, IL.

- Valero, P. M. & Molina, J. G. (1993a). Bibcont 1.1: Programa informático orientado al análisis de conjuntos de referencias bibliográficas. In pp. 115). Comunicación presentada al III Simposium de Metodología de las C.C. Sociales y del Comportamiento.:
- Valero, P. M. & Molina, J. G. (1993b). Un análisis cuantitativo de las publicaciones aparecidas en la revista "Behavior Research Methods, Instruments and Computers" llevado a cabo mediante el programa Bibcont 1.1. In pp. 116). Comunicación presentada al III Simposium de Metodología de las C.C. Sociales y del Comportamiento.:
- Valero, P. M., Molina, J. G. & Sanmartín, J. (1992a). Bibcont: un Programa Informático para el Análisis de conjuntos de Referencias Bibliográficas. *Revista de Historia de la Psicología*, 13(4), 19-26.
- Valero, P. M., Molina, J. G. & Sanmartín, J. (1992b). Un Grupo de Herramientas Informáticas para el Análisis de conjuntos de Referencias Bibliográficas. *Revista de Historia de la Psicología*, 13(1), 93-103.