

Sistema experto hipermedia para el reconocimiento, indización y recuperación de literatura gris

Emilia Currás

Universidad Autónoma de Madrid

0.1. Resumen

Sistema experto con interfaz hipertextual para la identificación, catalogación, indización y recuperación de documentos de “literatura gris”, cuya importancia y cantidad ha aumentado considerablemente en los últimos años. El sistema ha sido programado por Norbert Henrichs, profesor de la Universidad de Düsseldorf, en el entorno de desarrollo Knowledge Pro para Windows 95. Se definieron formatos de catalogación para 63 tipos de documentos distintos, de los que se presentan algunos ejemplos. También se desarrolló un módulo que facilita la determinación de las instituciones productoras de los documentos. Además de asistir en el tratamiento y recuperación de la información de “literatura gris”, puede resultar muy útil como material didáctico en la enseñanza.

Palabras clave: Literatura gris. Indización. Recuperación. Sistemas expertos. Sistemas hipermedia.

0.2. Abstract

Expert system with a hypertext interface to help in the definition, cataloging, indexing and retrieval of grey literature, whose importance is increasing every day in a considerable way. The system was developed by Norbert Henrichs, professor of the University of Düsseldorf, in ‘Knowledge Pro’ for Windows 95. 63 grey literature documents types were defined, of which several examples are provided. Also, a module to assist in the identification of the institution which produced them was developed. Besides providing a tool for cataloging and retrieving grey literature, it can also be very useful as an educational and pedagogical material.

Keywords: Grey literature. Indexing. Grey literature retrieval.. Hypermedia systems. Experts systems.

1. Preliminares

La aparición cada vez mayor de documentos no identificados por medios editoriales u oficializados va en aumento progresivamente desde hace unos cuantos años a esta parte. Este hecho empieza a preocupar a usuarios y profesionales de la información y la documentación, que se han venido ocupando principalmente de su almacenamiento y recuperación siguiendo los métodos tradicionales.

En este trabajo se trata de dar un paso más, es decir, de determinar cuáles pueden ser esos documentos de “literatura gris” o “no convencionales”, definiéndose hasta 63 de ellos. También se trata de fijar los parámetros para su indización y posterior recuperación, útiles para la construcción y uso de bases de datos. Para cada una de las clases de documentos se fijan cuáles son los datos que se deben tomar para su indización y recuperación. Como ejemplo, se citan los necesarios para los textos de conferencias o charlas (Fig. 1). Asimismo, se citan las entidades que los producen estableciendo relaciones entre éstas y los documentos de literatura gris.

Para realizar este trabajo se ha utilizado un sistema experto-hipermedia realizado en Knowledge Pro, adaptado a Windows 95, ideado por el Prof. Dr. Norbert Henrichs, Catedrático y Director del Informationswissenschaft Abteilung de la Heinrich Heine Universität en Düsseldorf. Se trata por tanto de un programa original, que está protegido en sus derechos de copia. Se ha obtenido, como producto final, un programa multicolor, organizado en ventanas, entre las cuáles resulta fácil desplazarse de un lugar a otro. (Henrich y Currás, 1998).

Se hace hincapié en que su utilidad no radica única y principalmente en el provecho que usuarios y profesionales de la información y la documentación puedan obtener de él; sino que puede resultar un instrumento didáctico muy provechoso en las Facultades y Escuelas de Biblioteconomía y Documentación.

2. Definiciones

Para mejor situarnos en el tema, empecemos por citar algunas definiciones, recopiladas entre las existentes en la bibliografía, las cuáles coinciden unas con otras casi exactamente.

1. La denominación de “literatura gris” se utiliza para aquellos documentos que ofrecen una información informal, en formato no convencional. En la mayoría de los casos no llevan una clasificación específica, por lo que no tienen ni ISBN, ni ISSN, ni NIPO (número de identificación para publicaciones oficiales) ni DL (depósito legal), no pudiéndose adquirir por los canales normales de distribución, como por ejemplo, una empresa editora. (Currás, 1995).

Textos de / para conferencias o charlas
AU: Autor
AC: Autor corporativo
TI: Título
TM: Tema
LG: Lugar físico
FE: Fecha
SP: Entidad organizadora
PC: Palabras clave

Fig. 1. Formato de descripción para textos de/para conferencias o charlas

2. La literatura gris se puede obtener públicamente, no siendo convencional en su contenido, no estando bien controlada su publicación y no siendo accesible por los canales normales de distribución, lo que la hace difícil de localizar y obtener. Además son documentos de muy variado tipo, yendo desde las pre-publicaciones, no contrastadas, hasta documentos de contenido muy concreto. (Currás, 1990 ; Henrichs y Currás, 1994).
3. Los documentos de literatura gris son documentos fugitivos, transparentes (que no se ven en los catálogos de editoriales, librerías, bibliotecas, etc.), de difícil localización, que, sin embargo en la mayoría de los casos, contienen datos relevantes e importantes. (Aguiar, 1992 ; Henrichs y Currás, 1994).
4. Pertenecen a la “literatura gris” aquellos documentos con información primordial de formato no convencional, que no tienen ISBN, ni ISSN ni NIPO (número de identificación para publicaciones oficiales), ni DL (depósito legal); pero sí pueden tener otro número ISDN. No se adquieren a través de los canales normales de difusión. (Tomada del propio disco: Henrich y Currás, 1998)

Para conocer la historia de la evolución de la “literatura gris” se aconseja consultar la bibliografía. Destaca un artículo firmado por la Prof. Dinah Aguiar de Población (1992, nota 5; 1993), además de algún libro (Auger, 1989; Knowles, 1981) y las actas de los Congresos Grey Literature (G), editadas estos últimos años (1).

3. Breve descripción histórica de este proyecto

La idea de realizar este trabajo surgió como consecuencia del problema práctico que planteaba el almacenamiento creciente de documentos de literatura gris en el Gabinete de Documentación Científica de la Universidad Autónoma de Madrid, que me exigía ordenarlos, indizarlos, etc. para formar una base de datos. Base de datos que no conseguía formar como un todo, sino que se dispersaba en varias colecciones, según los temas de que trataban. Era importante, también, saber de dónde procedían los documentos y qué instituciones los originaban.

3.1. Trabajo desarrollado en UCLA-USA

Estábamos en el año 1990 cuando los sistemas expertos y sobre todo los sistemas hipertexto empezaban a tomar realidad y se difundía su utilidad. Tuve noticias de que compañeros y buenos amigos de mis relaciones con la FID, el Prof. Harold Borko y el Prof. William Richardson de la UCLA (Universidad de California. Los Ángeles), estaban trabajando en un sistema experto para localizar documentos oficiales en el Congreso de los Estados Unidos de América. Con una beca del Programa Sectorial de Promoción General del Conocimiento del Ministerio de Educación y Ciencia, fui para allá. (Currás, 19, nota 2).

El programa que estaban desarrollando, en Knowledge Pro, se limitaba muy concretamente a los documentos de origen y daba lugar a pocas extrapolaciones. Sin embargo, me permitió familiarizarme con el planteamiento teórico de los sistemas hipertexto—basados en condicionales ‘si... entonces’ y en la formación de “árboles” anidados en los que se iba contestando “sí” y “no” a las sucesivas preguntas resultantes de aquella condicionalidad—. A veces se llega a una gran complejidad, pues hay que ir continuamente volviendo al “sí” o al “no” para continuar la extensión del entramado arborescente.

Con estas investigaciones no conseguía mis objetivos, por eso se imponía seguir investigando.

3.2. Trabajo realizado en la Universidad Hebrea de Jerusalén

Asimismo, por las actividades realizadas en el Comité FID/ET, tenía una buena amistad con la Prof. Bluma Pérez de la Universidad Hebrea de Jerusalén, quien me informó sobre unas investigaciones realizadas en la Facultad de Química para desarrollar un sistema experto con el cual poder llegar a conseguir que todos los resúmenes de un mismo artículo sobre cualquier tema de química fuesen iguales. A mi modo de ver, la cuestión parece bastante utópica y no conozco el resultado de aquellas investigaciones. Igualmente, en este caso, obtuve una beca del Programa Sectorial de Promoción General del Conocimiento, del Ministerio de Educación y Ciencia.

Allí me familiaricé con la sistemática de los sistemas expertos y su posible conexión con los sistemas hipermedia; pero seguía sin conseguir mi objetivo. Por esta causa, pensé que se debería desarrollar un programa específico y propio para la finalidad que perseguía.

3.3. Trabajo realizado en la Universidad de Düsseldorf

Dada, igualmente mi relación colegial con el Prof. Norbert Henrichs, gran experto, tanto en teoría de la Ciencia de la Información, como en Informática, le propuse realizar este trabajo conjuntamente. Aceptó muy gustoso y acordamos realizarlo en su Departamento “*Informationswissenschaft Abteilung*”. Él realizaría el programa informático y yo la parte teórica, es decir, la preparación del árbol de condicionalidad, las conexiones entre los diferentes datos, las listas de los posibles documentos de literatura gris, sus definiciones en español, etc.

En realidad trabajamos en colaboración y se fueron ampliando los contenidos (Henrich y Currás, 1993; 1994; 1995; 1998), según se iba avanzando en las investigaciones. El Prof. Henrichs decidió unir al sistema experto otro hipermedia, realizado en Knowledge Pro, que posteriormente fue adaptado al programa Windows 95. (Currás, 1990) Así surgió este “sistema experto-hipermedia para reconocimiento, definición, indización y recuperación de documentos de literatura gris, en conexión con las entidades que los producen”, presentado mediante ventanas multicolor, muy vistosas, donde se pasa de unos datos a otros, por la conexión entre ellas (Henrichs y Currás, 1998) (2).

4. Estructura del programa

El programa se estructura siguiendo el procedimiento de condicionalidad, haciendo preguntas basadas en el fórmula (3) “si ... entonces ...”, y a las que se debe contestar “sí” o “no”.

Tanto en un caso, como en otro, se pasa a la siguiente pregunta, donde se vuelve a contestar “sí” o “no” y así sucesivamente hasta obtener un resultado, respuesta, la cual no admite más preguntas. Con esto se construye un árbol, más o menos, como el del siguiente ejemplo esquematizado:

- ¿Tiene el documento ISBN ?
Sí (Se trata de un documento que no pertenece de partida a la literatura gris)/No
- ¿Tiene una clasificación?
Sí/No
- ¿Tiene un formato determinado ?
Sí/No

- ¿Figura una universidad en el título?

Sí/No

Cada vez que se responde a una pregunta se puede acceder a definiciones, tanto de cuestiones generales, como de los propios documentos, donde se han caracterizado hasta 63 clases diferentes. Éstos se ordenan tanto alfabéticamente como por tipos, agrupando los que pertenecen al mismo tipo. (Fig. 2).

En una segunda fase del programa se hacen constar definiciones de los documentos en cuestión, donde figuran además las empresas e instituciones que los producen.(Fig. 3).

La siguiente fase del programa ,dedicada a la indización y recuperación de la literatura gris, indica qué datos se deben tomar de cada clase de documentos para su correcta indización. (Fig. 4). Presentamos también otro conjunto de ejemplos referentes a los “documentos relacionados con la docencia”. (Fig. 5, 6 y 7).

En otra siguiente fase del programa figuran las entidades que producen literatura gris, con envío a los lugares, donde se detallan qué documentos son esos. Asimismo se establecen conexiones con los datos necesarios para una correcta indización de cada uno de aquéllos.

El programa permite pasar de un dato a otro, retroceder al principio, continuar, etc., haciendo click donde se indica. La variación de colores, hace muy comprensible y agradable el manejo del programa.

5. Presentación del disco

Por tratarse de un programa experto-hipermedia, el disco tiene estructura de ventanas, donde figuran las instrucciones precisas para su manejo.

Al iniciarse el programa aparece la ventana I (Fig.8) (Currás, 1995, nota 1), donde figura el título y contenido.

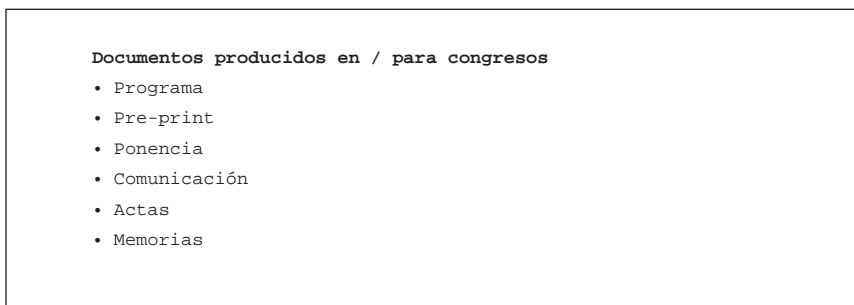


Fig. 2. Tipos de textos de/para conferencias o charlas

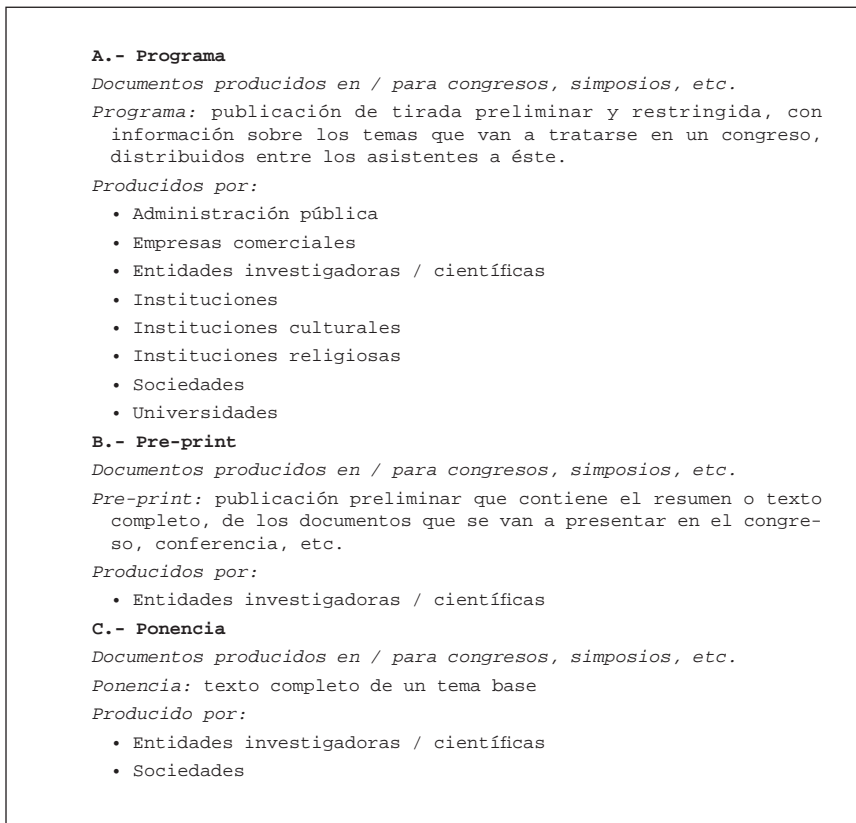


Fig. 3. Descripción de los textos de/para conferencias o charlas

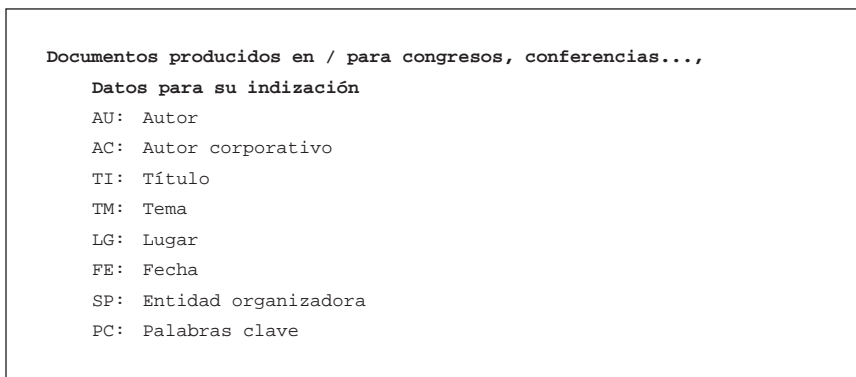


Fig. 3. Formato de descripción de los documentos producidos en / para congresos

Documentos relacionados con la docencia

- Apuntes de clase
- Textos de o para exámenes
- Memorias de examen
- Reglamentos de exámenes

Fig. 5. Tipos de literatura gris relacionados con la docencia

A.- Apuntes de clase

Documentos relacionados con la docencia

Apuntes de clase: Documentos manuscritos o mecanografiados, o presentados en cualquier forma telemática, que contienen los textos originales de las lecciones dadas por el docente a sus alumnos.

Producidos por:

Universidades

B.- Textos de o para exámenes

Documentos relacionados con la docencia

Textos de o para exámenes: Prueba escrita que realiza un individuo para comprobar su nivel de conocimiento o facultades en una determinada materia.

Producidos por:

• Universidades

Fig. 6. Descripción de algunos tipos de literatura gris relacionados con la docencia

Documentos relacionados con la docencia

Datos para su indización

AU: Autor

AC: Autor corporativo

TI: Título

DT: Tesis doctoral

TM: Tema

DI: Departamento

UN: Universidad

LG: Lugar físico

FE: Fecha

SP: Entidad organizadora

PC: Palabras clave

Fig. 7. Formato de descripción de los documentos relacionados con la docencia



Fig. 8. Ventana I.

Si se hace pulsa en “organigrama” se visualiza la ventana II (Fig.9) (Currás, 1995, nota 1), que contiene un esquema simple del método arborescente utiliza-

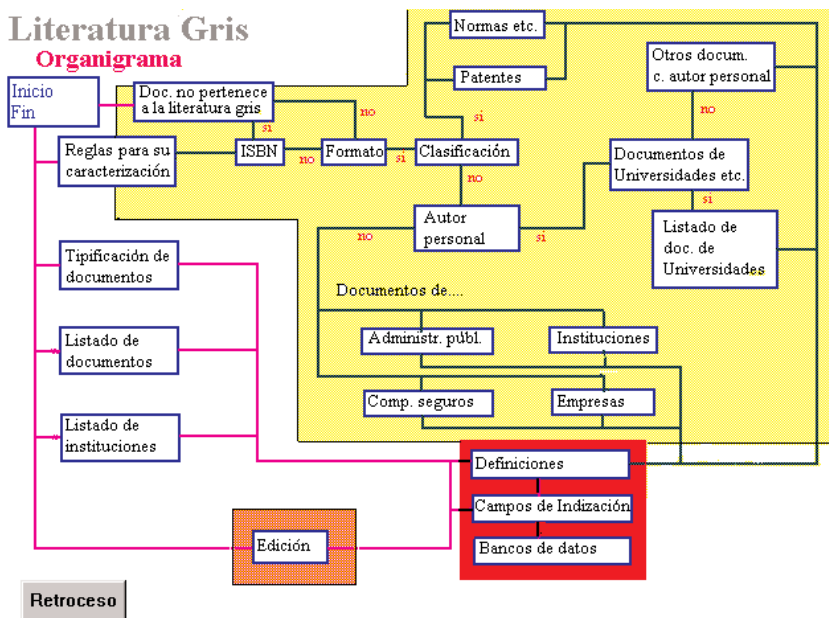


Fig. 9. Ventana II.

do, con la indicación de las conexiones que se pueden establecer. Esta ventana solo permite volver a la ventana I.

Cuando en la ventana I se presiona “info” aparece otra ventana con la definición dada aquí para “literatura gris”, anotada en páginas anteriores.

Volviendo a la ventana I, cuando se presiona en “Reglas para su caracterización”, aparece la primera pregunta, a la que hay que contestar “sí” o “no”. Se muestra en la ventana III (Fig. 10).

Siguiendo sucesivamente las indicaciones aparecidas en las subsiguientes ventanas, se accede a todos los datos del programa, como normalmente se hace en estos casos, sin dificultad ninguna. Sin dificultad ninguna, una vez construido; pero el Prof. Henrichs empleó muchas horas para conseguir esas conexiones y relaciones precisas.

6. Instituciones españolas que tratan literatura gris

La principal institución oficial, con carácter nacional, que trata y almacena este tipo de documentos, es el CINDOC, en su sección ISOC. Allí se construye

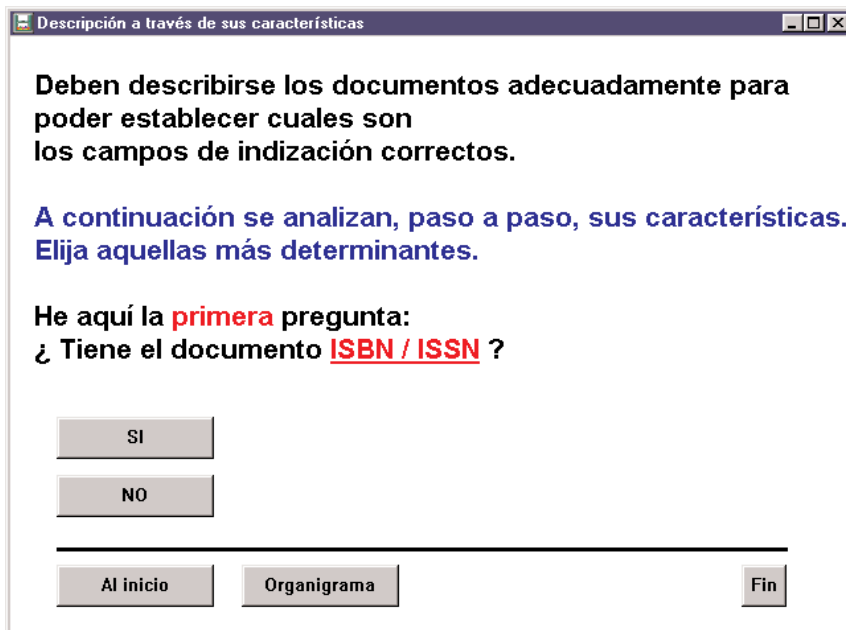


Fig. 10. Ventana III.

una base de datos, relacionada con las entidades internacionales y nacionales de otros países. (Aguilar, 1994, referencias)

En los Ministerios, Congreso de los Diputados, Senado y demás instituciones oficiales, tanto a nivel nacional, como autonómico, también se lleva un control de la literatura gris, principalmente la propia.

Entre las entidades privadas, con más actividades en estos temas, se pueden citar DOC 6 y la sección española de SISGLE.

7. Acciones internacionales

7.1. Bases de Datos

De unos años a esta parte se han incrementado las acciones internacionales, comenzando por el establecimiento de una base de datos convencional, donde se recogen los resúmenes y originales de documentos de literatura gris. Esto a nivel internacional, con oficinas en muchos países a ellos conectados. Se trata de SISGLE, distribuido por Silverplatte. Naturalmente funciona con un programa en línea conectado por Internet.

7.2. Congresos

En lo que atañe a los Congresos, es la empresa Transatlantic la que más destaca, organizando una primera "International Conference on Grey Literature, GL'93" (4) teniendo en cuenta el Informe Weinberg 2000 (5), que ha sido seguida de GL'95 y GL'97.

De las acciones a nivel nacional, pero de ámbito internacional, merecen destacarse las cubanas radicadas en sus institutos nacionales de información y documentación, en su sección PROINFO, que asimismo organiza congresos con amplias sesiones dedicadas a estos temas.

Se viene observando que, los congresos nacionales, sectoriales o internacionales ofrecen, igualmente, cada vez en mayor número, sesiones sobre literatura gris.

7.3. Publicaciones

Entre las publicaciones cabe destacar las actas de los congresos (Farace, 1993), los catálogos editados por los constructores de bases de datos, algunos libros (Auger, 1987; Knowles, 1981) y artículos aparecidos en revistas. Se citan algunos en el apartado de referencias.

8. Conclusión

Para comprender la importancia de este trabajo de investigación se debe desplegar el disco e ir visualizando las distintas ventanas, realizando las intercon-

xiones necesarias. Explicar su funcionamiento y manejo en un papel escrito no resulta fácil, pues para dar una idea completa de su ámbito y utilidad se necesitarían muchas páginas, quizá poco comprensibles y fatigosas de leer, con referencias hacia delante y hacia atrás. Aquí se ha tratado de dar una visión de conjunto, sencilla, intentando hacer comprensible el programa y su posible uso.

Se hace hincapié en la utilidad de este programa realizado en Knowledge Pro, adaptado para Windows 95, no sólo para usuarios y profesionales en estas cuestiones, sino como elemento didáctico, pues, por un lado, permite la explicación de lo referente a la literatura gris, y por el otro, permite enseñar la estructura y el manejo de los sistemas expertos, hipertexto e hipermedia. Igualmente resulta muy intuitivo poder hacer ejercicios y prácticas sobre el propio disco.

La realización de este trabajo de investigación, disco incluido (6), ha sido laboriosa y compleja, pero, dada su utilidad, parece que ha merecido la pena.

9. Notas

- (1) First International Conference on Grey Literature. Weinberg Report 2000, 1993 y años sucesivos, GL' 95 y GL' 96. Asimismo, véase el artículo de Schoch (1997).
- (2) Este trabajo se comenzó en el año 1991, con una beca, asimismo, del Programa Sectorial de Promoción General del Conocimiento del MEC. Continuó en años sucesivos, con idéntica beca menos en 1997 y 1998 que recibió una bolsa de viaje de la Universidad Autónoma de Madrid. Se trabajaba solamente pocos días, pocas semanas, en período de vacaciones, de ahí lo dilatado del tiempo, además de sucesivas ampliaciones anuales hasta llegar al disco final.
- (3) Véase la nota 1 y el trabajo de Henrichs y Currás (1998).
- (4) Véase la nota 1, así como Farace (1993 y las publicaciones de la empresa Transatlantic —*GreyNet, International guide to persons and organisations in grey literature*—, así como las revistas de la Asociación Transatlantic: *Grey Literature Network Service* y *News Brief News*.
- (5) Véase la nota 1.
- (6) Se dispone del disco completo y puede ser enviada una copia a quien lo solicite.

10. Referencias

- Aguiar de Población, D. (1992). Literatura cinzenta ou nao convencional: un desafio a ser enfrentado. // *Ciencia de la Informação*. 3 : 21 (setembro-dezembro ,1992) 243-246.
- Aguiar de Población, D. (1993). Biblioteca virtual amplia o "status" da literatura cinzenta: Relatório da etapa de 1993 da pesquisa de Pós-graduação que esta sendo realizada na Universidad Autónoma de Madrid. [Trabajo de investigación]. Madrid : Universidad Autónoma, 1993. Inédito.
- Aguiar de Población, D. ; Pires Noronha, D. ; Currás, E. (1996). Literatura cinzenta versus literatura branca: transição dos autores das comunicações dos eventos para produtores de artigos. // *Ci. Inf., Brasília*. 25: 2 (maio-agosto 1996) 228-248.

- Alberani, V. (1987). Introduzione alla letteratura grigia: definizione, tipologia, caratteristiche e controllo bibliografico. // *Bolletino d' Informazioni*. AIB. 27 : 3-4 (1987) 307-324.
- Alberani, V. (1992). *La letteratura grigia. Guida per le biblioteche speciali e i servizi d'informazione*. Roma: La Nuova Italia Scientifica, 1992.
- Auger, C. P. (1987). *Information sources in grey literature*. London: Bowker-Saur, 1989 (2ª ed. y ediciones sucesivas).
- Bolletino d' Informazioni*. AIB. 27: 3-4 (1987) (número monográfico dedicado a la literatura gris).
- Currás, E. (1990). Estructuración de un programa experto para literatura gris: memoria del trabajo de investigación realizado en la Universidad de California. Los Ángeles : UCLA ; Madrid, septiembre 1990. Inédito
- Currás, E. (1995). Instrucciones para reconocimiento, catalogación e indización de documentos de literatura gris. Proyecto de investigación entre el Gabinete de Documentación Científica de la UAM y la Universidad de Düsseldorf, 1995. Inédito.
- Farace, D. J. (1993). GL' 93 First International Conference on Grey Literature. // *FID News Bulletin*. 43: 4 (April 1993) 84-85.
- Helmut, M. (1994). Wird die Graue Literatur endlich zum Forschungsgegenstand?: Nachdenken über eine Konferenz in Amsterdam. // *NFD*. 45 : (1994) 112-120.
- Henrichs, N. ; Currás, E. (1993). Proyecto de sistema experto para reconocimiento e indización de documentos de literatura gris. Memoria de investigación inédita.
- Henrichs, . N; Currás, E. (1994). Estructuración de un programa de sistema experto para reconocimiento y localización de ciertos documentos de literatura gris. Proyecto de investigación entre el Gabinete de Documentación Científica de la UAM y la Universidad de Düsseldorf, 1994.
- Henrichs, N. ; Currás, E. (1995). Estructura de un programa de sistema experto e hipertexto para reconocimiento y localización automatizada de documentos. Memoria de investigación inédita.
- Henrichs, N.; Currás, E. (1998). Instrucciones para reconocimiento, catalogación e indización de documentos de literatura gris. [Disco de ordenador]. Trabajo de investigación realizado entre el Gabinete de Documentación Científica de la UAM y el Departamento Informationswissenschaft de la Universidad de Düsseldorf.
- Knowles, Ch.M. (1981). *The bibliographic presentation of grey literature*. Luxemburg : Commission of The European Communities, 1981.
- La ciencia española se conecta al Banco Europeo de Literatura Gris (1993). // *ABC* (7 febrero 1993) 65.
- Lara Guitard, A. (1985). La literatura gris como medio de comunicación científica: algunos resultados teóricos y empíricos de la investigación científica. // *Revista Española de Documentación Científica*. 8 : 2 (1985) 187-189.
- Pujol, R. (1995). La literatura gris en expansión. // *Information World en Español*. 33 (abril, 1995) 14.
- SISGLE (1992). *Online user guide*. Den Haag : EAGLE Secretariat, 1992.

- Schoch, N.A.; Bowes, L.; Lagasse, P. (1997). Highlights of Grey Works 96: The Second U.S. Workshop on Grey Literature. // FID News Bulletin. 47 : 3 (March, 1997) 92-96.
- Torres, I. de (1997). La llamada "literatura gris" : Reflexiones sobre su naturaleza y desarrollo e instrumentos bibliográficos para identificación y localización (II). // Boletín de la Asociación Andaluza de Bibliotecarios. 10 : 35 (Junio, 1994) 9-26.