

- TITULO: "El control del lenguaje en la producción de una base de datos: el tesauro de urbanismo ISOC".
- AUTORAS: Angeles Maldonado Martinez
Carmen Gavira Golpe
- INSTITUCION: ISOC (Instituto de Información y Documentación en Ciencias Sociales y Humanidades)

RESUMEN:

A la hora de crear una base de datos hay que tener en cuenta el tipo de documento a tratar, pues esto influye en la forma de indización. En la descripción del contenido de un documento puede utilizarse o no un lenguaje controlado. En el campo de las Ciencias Sociales y las Humanidades la recuperación en lenguaje libre no es aconsejable, dada la falta de sistematización de la terminología utilizada. Por este motivo el ISOC, creador de bases de datos documentales en Ciencias Sociales y Humanidades, consciente del mayor rendimiento de las bases de datos que utilizan un lenguaje controlado en la indización y recuperación de documentos, se encuentra en un proceso de creación de tesauros para cada una de las disciplinas que cubre.

La comunicación que presentamos versará sobre las dificultades metodológicas de la construcción de los vocabularios de indización, así como sobre la problemática que se suscita en otros aspectos de la producción de bases de datos, y su posterior explotación.

PALABRAS CLAVE: PRODUCCION DE BASES DE DATOS/ INDIZACION/ RECUPERACION DE LA INFORMACION/ RECUPERACION AUTOMATIZADA/ LENGUAJE DOCUMENTAL/ TESAURO.



2^o
Jornadas Españolas de Documentación Automatizada

"EL CONTROL DEL LENGUAJE EN LA PRODUCCION DE UNA BASE DE DATOS: EL TESAURO DE URBANISMO ISOC" por Carmen Gavira Golpe y Angeles Maldonado Martinez

INDICE

INTRODUCCION

I.- EL TESAURO DE URBANISMO ISOC

II.- METODOLOGIA DE ELABORACION

- 1) Selección de Fuentes
- 2) Recogida de Términos
- 3) Establecimiento de Grandes Familias y Agrupación de Términos
- 4) Selección de Términos
- 5) Establecimiento de relaciones de Equivalencia

III.- PRIMEROS RESULTADOS OBTENIDOS

IV.- EQUIPO DE TRABAJO

.....

INTRODUCCION

El Instituto de Información y Documentación en C. Sociales y Humanidades ISOC tiene por misión el tratamiento de las publicaciones periódicas españolas incluidas en estos campos. Con los documentos extraídos de las mencionadas fuentes, y tras su tratamiento documental e informático, está creando bases de datos accesibles por terminal. El Urbanismo cuenta con una base, en estos momentos en fase de experimentación, con más de 3.000 documentos.

Las bases de datos tienen la ventaja sobre los ficheros manuales, de que se puede manipular una gran cantidad de documentos en muy poco tiempo. Pero para que rindan todo lo deseado es necesario que los documentos hayan sido tratados convenientemente en el aspecto documental. La clave de una buena recuperación de docu-



2^o

ornadas Españolas de Documentación Automatizada

mentos, fin último de las bases de datos, se encuentra en una buena indización. Indizar un documento es analizarlo tanto en su aspecto formal (autor, título, revista...) como en su contenido. El contenido de un artículo se puede expresar, o bien a través de términos extraídos del lenguaje libre sin ningún tratamiento previo, o bien se puede optar por utilizar un lenguaje controlado. En el campo de las Ciencias Sociales y las Humanidades la recuperación en lenguaje libre no es aconsejable, dada la falta de sistematización de la terminología utilizada, por este motivo se tiende a emplear tesauros que garantizan una recuperación más precisa. El ISOC se encuentra en un proceso de creación de tesauros para cada una de las disciplinas que cubre, entre ellas el Urbanismo.

I.- EL TESAURO DE URBANISMO ISOC

El ISOC, llevado por la necesidad de sistematizar sus lenguajes de indización, solicitó del CSIC, organismo al que pertenece, en 1984 la aprobación de un Programa de Investigación, cuyo fin era la creación de tesauros sectoriales para cada una de las disciplinas que cubre, decidiendo comenzar por el de Urbanismo.

Pocas materias presentan tanta complejidad como el Urbanismo a la hora de delimitar su vocabulario científico, ya que en él se da, no sólo la concurrencia de múltiples disciplinas, sino también una gran diversidad en su ámbito de aplicación.

Así, si la enumeración de las distintas disciplinas básicas que aportan vocabulario es relativamente corta (7 u 8 campos científicos: Arquitectura, Geografía, Derecho, Ingeniería, Medio Ambiente, Sociología, Demografía, Economía, etc.), mucho más difícil es la delimitación de su ámbito de aplicación: Investigación (teórica o experimental), Planificación (sectorial y territorial en sus distintos niveles), Gestión (desde la tramitación del proyecto y su financiación, hasta el mantenimiento y administración de la obra), Docencia (en cada una de las disciplinas que lo integran) e Información (análisis, recupe-



2^o

ornadas Españolas de Documentación Automatizada

ración y difusión de los documentos de Urbanismo).

De esta forma llega a plantearse la pregunta de si existe un lenguaje específico de esta disciplina, y por tanto la pertinencia de un tesoro de Urbanismo distinto de la suma de tesoros de distintas disciplinas. De los 600 títulos de revistas, que actualmente analiza y procesa el ISOC, 26 se integran en el ámbito del Urbanismo (1). Su tratamiento por un especialista durante los últimos cinco años ha puesto de manifiesto la existencia de un vocabulario específico que va más allá de la selección de términos con alguna referencia especial a partir de un glosario general.

La incorporación en 1981 de un experto para la indización de las publicaciones periódicas de Urbanismo ha individualizado más la disciplina dentro del centro, y al cabo de tres años de trabajo se planteó la necesidad de contar con un vocabulario propio para el análisis de tal documentación. Si bien existían algunos catálogos, diccionarios e índices españoles que aportaban vocabulario desde áreas muy concretas (Derecho, Arquitectura, etc), faltaba la estructura general que facilita un tesoro. La inexistencia de un tesoro de Urbanismo en castellano que se acoplara a las necesidades del instituto, lleva al ISOC a comenzar su elaboración.

Un tesoro puede definirse como un conjunto de términos relacionados entre sí relativos a una disciplina. Las relaciones básicas que pueden establecerse entre los términos son las siguientes:

- Relaciones de equivalencia. En muchas ocasiones un concepto puede expresarse a través de varios términos. Establecer las relaciones de equivalencia consiste en agrupar todos los términos que tengan la misma significación, y elegir de entre ellos uno.
- Relaciones de jerarquía. Expresan las relaciones de superioridad y subordinación entre los conceptos. Los conceptos se encuentran unidos

(1) Gavira, C.: "Las Publicaciones Periódicas de Urbanismo en España". Ciudad y Territorio, nº 1, 1984, p. 7-12.



2^o

ornadas Españolas de Documentación Automatizada

por lazos de jerarquía, de tal forma que cada uno de ellos es específico de otro concepto más general, pero a su vez es genérico de otros más específicos.

- Relaciones de asociación. Definen las relaciones entre conceptos próximos, que no son propiamente sinónimos ni tampoco mantienen lazos de jerarquía, pero que son de uso alternativo, advirtiéndose que si utilizamos un término bien podríamos usar sus términos relacionados.

El mencionado Programa de Investigación prevee la recogida de términos y el establecimiento de las relaciones de equivalencia, jerarquía y asociación en sus tres años de duración (85/87). Pasado ya año y medio desde el comienzo de los trabajos, se han superado las fases de recogida de términos y establecimiento de las relaciones de equivalencia. El resto del trabajo será realizado en el tiempo que resta hasta la finalización del Programa. A continuación se detallará la metodología seguida, y los resultados obtenidos en lo ya realizado.

II.- METODOLOGIA DE ELABORACION

1) SELECCION DE FUENTES

En primer lugar se seleccionaron aquellas fuentes, españolas y extranjeras, que contuvieran una relación sistemática de términos urbanísticos: tesauros, diccionarios, catálogos, listados de términos, etc...

El Urbanismo ha ido adquiriendo su cuerpo teórico a partir de otras disciplinas, de las que ha tomado una parte de su terminología. En la actualidad es considerada como una disciplina autónoma con entidad propia, lo que le ha hecho configurar su propio vocabulario. Para extraer los términos se creyó conveniente partir de aquellas fuentes que tratan al Urbanismo como unidad, y dejar aquellas otras más específicas de cada una de las materias que han contribuido a su formación como disciplina.

Las fuentes seleccionadas fueron las siguientes:



2^ª

Asociación Española de Documentación Automatizada

a) Listado de Términos ISOC

El ISOC viene indizando desde 1976 las revistas españolas de Ciencias Sociales, entre ellas las más importantes especializadas en Urbanismo. Producto de esta indización es un listado de términos que reúne todos los utilizados desde entonces en las diversas disciplinas, y por lo tanto los procedentes de los artículos de temática urbana. De este listado se han entresacado todos los términos relacionados con el Urbanismo.

b) Otras Fuentes en Castellano

- "Tesaurus de Urbanismo". MOPU. 1984. Los términos contenidos en este tesaurus se ajustan únicamente a la terminología legal urbanística, pues fue pensado en función de una base de datos de normativa, creada por el Ministerio de Obras Públicas y Urbanismo (MOPU).
- "Catálogo de Términos Urbanísticos". Luis Chacón Ortega. Editorial Bayer Hnos. 1983. Este catálogo contiene aproximadamente 500 términos, extraídos en su gran mayoría, como en el caso anterior, de la normativa urbanística española.
- "Documentación Local y Urbanística". IEAL. 1971/1976. A partir de 1971, y hasta 1976, el Instituto de estudios de Administración Local (IEAL) publicó anualmente un volumen que recogía los documentos ingresados en su biblioteca, una de las más completas especializada en Administración Local y Urbanismo. Cada uno de estos volúmenes contenía un índice alfabético de materias con aproximadamente 500 términos.
- "Bibliografía sobre Derecho Urbanístico". Francisco Gonzalez Navarro. Editorial Montecorvo. 1981. De esta bibliografía se extrajeron los términos de su índice de materias.
- "Bibliografía Iberoamericana de Administración Local". Enrique Orduña Rebollo. IEAL. 1983. De esta bibliografía se han utilizado los términos urbanísticos extraídos de su índice de materias.



2^o

ornadas Españolas de Documentación Automatizada

- . "Glosario Internacional de Términos Técnicos Relacionados con la Vivienda y el Urbanismo". Dirección General de Arquitectura. 1942. Se trata de un glosario con términos y expresiones relativas a vivienda y urbanismo reproducidos en cinco idiomas, entre ellos el castellano, cuya traducción no se encontraba con facilidad en los diccionarios.
 - . "Enciclopedia de la Planificación Urbana". A. Whittick. IEAL. 1975. Se trata de una enciclopedia americana de la que se han seleccionado aquellos términos de interés.
 - . "Tesaurus DIC". Centro de Estudios y Experimentación de Obras Públicas. MOPU. 1982.
 - . "Diccionario Técnico Vial de la A.I.P.C.R.". Asociación Española la Permanente de los Congresos de Carreteras. Comité Nacional Español de la A.I.P.C.R. 1984.
- c) Tesoros en otras Lenguas

Es evidente que uno de los útiles terminológicos fundamentales para este tesoro son los propios tesoros de Urbanismo, elaborados a partir del tratamiento documental de las fuentes urbanísticas. Al no existir un tesoro español de Urbanismo que contemplara esta disciplina en su conjunto, se acudió a las existentes en otras lenguas. La práctica urbanística difiere de unos países a otros, pues parte de los condicionamientos económico-sociales y legales propios de cada uno de ellos. El urbanismo español ha sufrido las influencias de diversos países dependiendo del momento, pero quizá se encuentre más próximo del urbanismo francés. De los diferentes tesoros consultados se seleccionó el de la base de datos francesa URBAMET (3.800 términos), ya que fue considerada la más completa, pues aborda la disciplina urbanística en sus diferentes vertientes (sociológica, económica, espacial, legal, ...). Al no contar con una versión en castellano de este tesoro se procedió a su traducción. Una vez traducido se comprobó que, a pesar de la fuerte influencia francesa, el sistema organizativo de nuestro



urbanismo difiere en algunos aspectos del propio del sistema francés, por lo que no bastaba una mera traducción de él, sino que se imponía como necesaria una adaptación del mismo, lo que afianzaba nuestra decisión de elaborar un tesoro a la medida de nuestro urbanismo.

2) RECOGIDA DE TERMINOS

En primer lugar se consultaron las fuentes seleccionadas en castellano. Los términos extraídos de estas fuentes compusieron un fichero base, ordenado alfabéticamente, que fue completado más tarde con los procedentes del tesoro de Urbanismo de la base URBAMET. Las fuentes españolas aportaron primordialmente toda la terminología de nuestro sistema legal urbanístico. Del tesoro URBAMET se rechazaron los términos procedentes del derecho urbanístico francés, pero sin embargo se conservaron aquellos otros pertenecientes a otras parcelas (equipamientos, geografía, sociología, etc) que enriquecieron notablemente la terminología extraída de las fuentes españolas.

3) ESTABLECIMIENTO DE GRANDES FAMILIAS Y AGRUPACION DE TERMINOS

Una vez extraídos todos los términos y ordenados alfabéticamente, se pasó a establecer las grandes áreas o "familias" en que serian distribuidos posteriormente. En un principio el número de familias fue menor, pero estas se fueron subdividiendo hasta llegar a alcanzar la cifra de 15. Cada una de las familias se corresponde con uno de los grandes apartados o temas que configuran la disciplina urbanística. A continuación se reproduce el listado de las "familias" que se creyó oportuno considerar.

- Administración Pública - Colectividades Locales
- Arquitectura - Construcción - Habitat - Vivienda - Percepción del Espacio
- Derecho
- Actividad Económica



2°

Asociación Española de Documentación Automática

- Educación
- Equipamientos Colectivos
- Espacio
- Geografía Física
- Infraestructuras (excepto Transportes) - Obras Públicas
- Medio Ambiente - Paisaje - Recursos
- Medio Rural
- Metodología - Fuentes Documentales
- Ordenación
- Sociología - Psicología Social - Demografía - Salud - Deportes
- Transportes

Aunque las denominaciones de las familias sean genéricas, estas sólo contemplan términos utilizados en Urbanismo. Este no es un tesoro suma de otros tesoros sectoriales de aquellas materias que han contribuido a la configuración del Urbanismo como disciplina autónoma, sino un tesoro de Urbanismo, por eso únicamente han sido incluidos en él términos utilizados en el campo del Urbanismo propiamente dicho. Para evitar introducir términos que no pertenecen estrictamente al área urbana, se seleccionaron aquellas fuentes que tratan al Urbanismo como unidad, dejando a un lado todas aquellas otras específicas de cada una de las disciplinas base que le han servido de punto de partida.

Una vez establecidas definitivamente las grandes familias, se pasó a distribuir los términos entre ellas según su afinidad, de tal forma que cada uno de ellos quedó adscrito a una de las grandes áreas consideradas.

4) SELECCION DE TERMINOS

Una vez agrupados los términos por familias se eliminaron aquellos que a pesar de haber sido seleccionados en un primer momento, tras una reflexión más ponderada, se pensó que no tenían cabida en este tesoro, bien por estar mal empleados gramaticalmente, bien por representar conceptos que no se creyó oportuno incluir. A partir de este momento quedaron definitivamente seleccionados los términos que contendría el tesoro.



2^o

ornadas Españolas de Documentación Automatizada

5) ESTABLECIMIENTO DE RELACIONES DE EQUIVALENCIA

Con los términos seleccionados definitivamente se procedió a establecer las relaciones de equivalencia, es decir, a eliminar los sinónimos, para lo cual se agruparon los términos de una misma significación, y se eligió de entre ellos aquel que mejor representaba el concepto, quedando constituido como descriptor. El resto de los términos se conservó, remitiéndose a su correspondiente término elegido.

III.- PRIMEROS RESULTADOS OBTENIDOS

Una vez recogidos, seleccionados y distribuidos en familias los términos, ya se puede hacer balance de los resultados obtenidos. Primeramente diremos que en estos momentos el tesoro cuenta con 4348 términos en total, de los cuales 3493 son descriptores y 855 sinónimos. De los descriptores 589 tienen al menos un sinónimo, y de entre ellos alguno cuenta con más.

No todas las familias tienen el mismo peso específico dentro del conjunto, pues mientras hay apartados que cuentan con más de 500 descriptores, como es el caso de Sociología y Actividad Económica, otros no alcanzan los 100, vease el caso de Educación (61) y Medio Rural (69). En el Cuadro 1 se presentan las cifras para cada una de las familias.

Por último señalar que un tesoro no es un elemento estático, sino que está en continua evolución, por este motivo una vez cubiertas todas las etapas de elaboración deberá atravesar por un periodo de prueba que servirá para detectar, por una lado las lagunas existentes, incluyendo los términos necesarios, y por otro los términos superfluos que hayan sido incluidos en un primer momento, y eliminarlos. Además deberá someterse a revisiones periódicas con la misma finalidad.



2^ª

ornadas Españolas de Documentación Automatizada

AREAS O FAMILIAS	TERMINOS	DESCRIPTORES	DESCRIPTORES CON SINONIMO	SINONIMOS
Admon. Pública - Colectividades Locales	440	379	42	61
Arquitectura - Construcción - Habitat - Vivienda	441	368	48	73
Derecho	312	268	33	44
Economia	516	403	93	113
Educación	61	53	8	8
Equipamientos Colectivos	199	144	38	55
Espacio	174	115	37	59
Geografía Física	159	140	17	19
Infraestructuras - Obras Públicas	82	66	7	16
Medio Ambiente - Paisaje - Recursos	225	191	28	34
Medio Rural	69	56	10	13
Metodología - Fuentes Documentales	358	346	11	12
Ordenación	372	269	66	103
Sociología-Psicología Social-Demografía-Salud-Deportes	524	469	50	55
Transportes	414	226	101	188
TOTAL	4348	3493	589	855

CUADRO 1. - DISTRIBUCION DE TERMINOS POR FAMILIAS



2^ª

Asociación Española de Documentación Automatizada

IV.- EQUIPO DE TRABAJO

En los trabajos anteriores han colaborado un importante número de personas entre documentalistas y especialistas en Urbanismo, organizados en dos equipos de trabajo: uno compuesto por documentalistas del ISOC especialistas en temas legales, sociales, económicos y urbanos; y otro externo de apoyo compuesto por profesionales pertenecientes a varias instituciones, expertos en cada una de las grandes áreas.

Una vez comenzado el trabajo se vió la necesidad de reunir a expertos que pudiesen asesorar sobre los problemas que surgían en los campos específicos, estableciendo contacto con diferentes instituciones hasta crear un equipo de asesores que cumplieran las funciones de seguimiento y control del vocabulario en sus distintas áreas. A este equipo le ha correspondido la supervisión de los trabajos realizados por el equipo de Urbanismo del ISOC. En un primer momento hubieron de encargarse de revisar la parte de traducción del tesauro URBAMET correspondiente a su especialidad. La segunda tarea que les ha sido encomendada, y que en estos momentos están desarrollando, es la corrección del listado de términos seleccionados con sus correspondientes sinónimos del área relativa a su especialización. En estos listados habrán de eliminar los términos que, según su opinión, no debieran incluirse en su apartado; añadir aquellos otros términos que crean deban estar, y no hayan sido incluidos; y por último, verificar los sinónimos, es decir, comprobar que los términos considerados como sinónimos en realidad lo son, y que el elegido entre ellos, es el más apropiado.



2^o

ornadas Españolas de Documentación Automatizada