



Número 5, desembre 2000

## Tipología documental citada en tesis doctorales de informática: bases empíricas para la gestión equilibrada de colecciones

[versió catalana]

[Cristóbal Urbano](#)

Facultat de Biblioteconomia i Documentació

Universitat de Barcelona

[urbano@fbd.ub.es](mailto:urbano@fbd.ub.es)

---

### Resumen

A partir del análisis de citas bibliográficas presentes en las tesis doctorales de informática de la Universitat Politècnica de Catalunya (UPC), leídas entre 1996-1998, se determina el uso de información según la tipología y el soporte documental. Los datos obtenidos son consistentes con las características apuntadas en la bibliografía sobre los investigadores del área de ciencias de la computación: una elevada obsolescencia de la bibliografía profesional, el control de la bibliografía mediante las citas en las lecturas efectuadas, la prioridad por las fuentes más próximas y accesibles, la consulta de buscadores de recursos web, la visita a sitios web de investigadores e instituciones de solvencia en su terreno, y el contacto con los colegas mediante la comunicación electrónica. Todo ello configura un uso por categorías documentales que prima los congresos y la literatura gris en mayor medida que otras áreas científico-técnicas.

### 1 Introducción

Con frecuencia, en el ámbito de las bibliotecas universitarias y de investigación, la gestión de colecciones se encuentra excesivamente focalizada en la selección de títulos de revistas. La indudable importancia de este tipo de documentos en el sistema de información de la ciencia y la tecnología, el plus de garantía de calidad que tradicionalmente se asigna a los trabajos publicados en revistas, el elevado coste de adquisición de gran parte de la edición científica de mayor impacto, así como la vinculación entre evaluación de la producción científica y visibilidad de los trabajos publicados en ellas, pueden llegar a ser factores que distorsionen el equilibrio presupuestario ideal entre las diversas categorías documentales a adquirir en una biblioteca.

La pregunta sobre a qué títulos de revistas se ha de suscribir una determinada biblioteca ha merecido, con mayor frecuencia, más atención que la pregunta previa sobre qué porcentaje del presupuesto se tenía que dedicar a cada categoría documental o, si se prefiere situar el tema más allá de las adquisiciones, qué atención había que dedicar al control bibliográfico de los diversos formatos y soportes documentales. Precisamente, el desarrollo de los instrumentos de control bibliográfico de las revistas científicas ha sido uno de los factores que puede haber contribuido en mayor medida a un posible sobredimensionamiento de esta clase de documentos en los debates profesionales sobre desarrollo de colecciones. Como prueba de ello, valga el siguiente párrafo en el que se debate entre adquisición de revistas y documentos secundarios para el control de las revistas:

"The director of a well known university library, when asked why he had not subscribed to the *Science citation index*, answered as follows. 'I think the *SCI* is a fine retrieval tool. But I asked the faculty members — which would you prefer, the *SCI* or 40 new journal subscriptions? We have finite budget and choices must be made.' In reply to this seemingly reasonable approach, I asked, 'Which 40 journals?'" (Garfield, 1968).

Ciertamente, la pregunta de Garfield era, y continúa siendo pertinente, si bien ante las grandes transformaciones que Internet está introduciendo en los mecanismos de comunicación científica

parece adecuado volver a reflexionar sobre la ratio revistas/monografías en los presupuestos de adquisiciones desde un contexto más amplio. Esto es, además de planificar la distribución del presupuesto de adquisiciones por categorías documentales y por títulos concretos de series, quizá haya que comenzar a plantearse los recursos económicos destinados a facilitar el control y el acceso a información, independientemente de si se trata de recursos a adquirir de forma permanente, o no.

Conocer qué usan realmente los investigadores de una determinada área de conocimiento en una universidad puede ser un dato de gran interés en el proceso de volver a definir las políticas de adquisiciones a la vista del nuevo entorno digital. La constatación empírica del uso de información desde un enfoque local, junto a otros elementos más globales sobre valoración de las fuentes de información, es una condición necesaria para llevar a cabo una gestión de colecciones centrada en los usuarios.

Es, en este sentido, que en el presente artículo se pretende conocer el uso de información realizado por investigadores en informática de la Universitat Politècnica de Catalunya con la finalidad de obtener información empírica de base con la que alimentar los procesos relacionados con la gestión de las colecciones de la biblioteca. Inicialmente, nuestro trabajo formaba parte de un estudio más amplio que analizaba con detalle el método del análisis de citas en publicaciones de usuarios como mecanismo para el estudio de usos y necesidades de información en bibliotecas universitarias (Urbano, 2000), y en el que se desarrollaba una aplicación práctica sobre tesis doctorales de informática de la UPC. Sin embargo, se ha considerado pertinente y de especial valor presentar ahora de forma independiente los datos obtenidos sobre la tipología documental utilizada por los usuarios.

## 2 Fundamentos y antecedentes

### 2.1 Análisis de citas en publicaciones de usuarios y gestión de colecciones

Los estudios de uso de información y los estudios de necesidades son de gran utilidad para la gestión de colecciones, y aunque se han de considerar de naturaleza diferente están íntimamente relacionados con el ámbito que nos ocupa. Los primeros tienen en esencia un carácter retrospectivo de descripción de lo sucedido; los segundos persiguen anticipar las posibles demandas de los usuarios. Sin embargo, son muchos los autores que vinculan uso y necesidades de información basándose en el principio de que "los usos efectuados en el pasado son una buena fuente de información para pronosticar el uso presente o futuro" (Baker y Lancaster, 1991, p. 82). Si bien puede ser considerado un principio discutible, lo que no ofrece duda alguna es que los usos efectuados sirven, como mínimo, para detectar si un determinado grupo de usuarios ha evolucionado en sus hábitos, ya sea a raíz de cambios en la oferta de servicios y recursos de la biblioteca, ya sea por cambios en sus características como investigadores. Así pues, la evolución del uso es de gran interés para el estudio de la evolución de las necesidades.

En este sentido, en el presente trabajo se ha optado por el análisis de citas en publicaciones de usuarios como método para determinar el uso de información de los mismos, teniendo en consideración que los métodos habitualmente utilizados en la medición del uso de colecciones están excesivamente limitados por la condición del acceso físico, y que difícilmente miden el uso que los lectores realizan con fuentes externas a la colección. En efecto, la mayoría de manuales y directrices sobre la evaluación bibliotecaria (Lancaster, 1996; Poll y Boekhost, 1998, entre otros) se basan en una definición de uso centrada en la colección de la biblioteca y en la medición de los usos que impliquen un movimiento físico de los documentos. Este enfoque es muy objetivo y, conscientemente, se limita a valorar el uso concreto de la colección, pero no entra en la investigación del uso de información que los usuarios locales realizan y que no necesariamente se manifiesta de forma única en el uso de la colección.

Por otra parte, si se desea realizar un estudio por métodos indirectos sobre la distribución del uso entre las diversas clases de documentos utilizadas por un grupo de usuarios, el análisis de citas en sus publicaciones es, sin duda, una de las opciones más eficaces. Pese a las discusiones que se plantean en torno a si citar un documento indica su lectura real, se entiende que, como método para medir el uso, el recuento de las citas recibidas por los diversos tipos de documentos es más adecuado que el recuento de los documentos movilizados en una sala de lectura cuando los usuarios estudiados son investigadores. En definitiva, se pretende obtener una imagen del uso de información que realiza el usuario y no necesariamente del uso realizado a partir de la colección.

En el caso concreto del recuento de referencias en las bibliografías de usuarios de bibliotecas, al tratarse de un método que estudia el uso de información citada, pero no necesariamente consultada en la biblioteca desde la que se realiza el estudio, podría decirse que se trata de un método que parcialmente permite estudiar tanto el uso como las necesidades porque las fuentes citadas y no disponibles en la colección de la biblioteca de referencia para el usuario

son un buen método para explorar las demandas no satisfechas.

De todos modos, conviene recordar que la validez del análisis de citas en los estudios de uso de información de usuarios de bibliotecas es objeto de un largo debate en el que se pueden distinguir tres posiciones. Para algunos autores los datos de citación de colectivos nacionales, o que afectan a toda la bibliografía circulante sobre una materia, son mucho más fáciles de obtener y normalmente igual, o más significativos, que los de los propios usuarios para establecer sus necesidades de información (Broadus, 1985a y 1985b; Lancaster, 1996). Sin embargo, las objeciones a la validez de datos externos frente a los de los usuarios locales son numerosas (Line, 1977 y 1985; Joswick y Stierman, 1997), por lo que se explica que, para muchos autores, el uso de publicaciones de los usuarios en combinación, o no, con otras fuentes de citas externas sea imprescindible para la aplicación del método del análisis de citas en estudios locales de necesidades y uso de fuentes de información bibliográficas<sup>1</sup> (McCain y Bobick, 1981; Buchanan y Herubel, 1994; Jiménez Contreras et al., 1994; Hughes, 1995; Sylvia y Leshner, 1995; Zipp, 1996; Hurd, Bleic y Vishwanathan, 1999; y otros). Finalmente, otros autores destacan que el análisis de citas, especialmente el que se basa en datos generales y no en documentos fuentes de los usuarios, mantiene un bajo nivel de correlación con el uso real medido directamente en las bibliotecas (Scales, 1976; Satariano, 1978; Line, 1985a) por lo que, comparativamente, defienden la medición directa del uso en sala o por medio de las estadísticas de circulación, o incluso el "juicio experto" de los bibliotecarios (Pan, 1978).

Ahora bien, como dice Broadus (1977, p. 309), existe una opinión general que considera que siempre que se trabaje con grandes cantidades de citas,<sup>2</sup> su recuento y análisis con finalidades bibliotecarias tiene un considerable valor y fiabilidad (Linda C. Smith, 1981, p. 84), especialmente si se tienen en cuenta datos locales como complemento de estudios de mayor alcance. Así, por ejemplo, Sylvia y Leshner (1995) reconocen la tentación de usar estudios previos y los datos generales de las grandes bases de datos de citas, pero advierten de sus limitaciones:

"Citation analysis has been done through nationally collected citation statistics, and proposals have been made to use these analyses for library collection and evaluation. The easiest approach is to simply look at previous studies and use the findings to govern one's own collection development strategies. There are, however, some very strong arguments for doing a local study. Maurice B. Line argued that every library has a differing clientele, and that clients have different information requirements so that national surveys may have little relevance for individual libraries." (Sylvia y Leshner, 1995, p. 314)

Por su parte, Zipp (1996), en un trabajo sobre la validez de las tesis doctorales como fuente para estudiar el uso de información del conjunto de los investigadores de una universidad, afirma al comparar los estudios de uso real por medio de estadísticas de préstamo o similares con el análisis de citas:

"Studies of circulation data and in-library use contribute to our understanding of what parts of a library collection are being used. The link between demand for research materials and use is most easily achieved by direct analysis of the scholarly communication process. Evaluation of the titles used in the process of discovery or problem identification yields valuable information. The titles in which researchers publish and those that they cite in their works also serve to elucidate real and potential demand on library collections." (Zipp, 1996, p. 335)

De todas formas los resultados de estudios de citas de alcance local también se han de leer con cautela. Así, por ejemplo, Lancaster (1996) los resume alertando fundamentalmente sobre tres aspectos: las diversas clases de documentos no son igualmente citables; las colecciones personales pueden distorsionar la imagen del uso real de la colección; y la mayor accesibilidad de la colección de la biblioteca respecto a publicaciones no localizadas en ella puede generar un sesgo de las citas hacia la colección.

En la discusión metodológica sobre el análisis de citas en publicaciones de usuarios de bibliotecas, otro de los puntos clave es la selección de los documentos fuente a analizar. En este sentido, las tesis doctorales se consideran en numerosos estudios un elemento muy significativo para analizar las necesidades de los investigadores en general. Numerosos autores consideran explícitamente que el análisis de tesis puede ser una fuente idónea para obtener datos del consumo de información del conjunto de investigadores de un departamento, estudiantes de tercer ciclo, becarios de investigación y profesorado ayudante incluido. Así lo establecen autores como McCain y Bobick (1981), Zipp (1996) y Barry (1999), quienes bien por métodos estadísticos, bien por medio de sólidas argumentaciones, afirman que el consumo de los estudiantes de doctorado es muy representativo de las necesidades y uso de información del conjunto de personal docente e investigador en un entorno universitario. En la misma línea Walcott (1994) observó que los estudiantes de tercer ciclo eran los que utilizaban más intensamente las publicaciones seriadas en la biblioteca de biología de la State University of

New York.

Por otra parte, en contraste con la tendencia de los artículos de revista a recoger preferentemente citas de otros artículos de revista, las tesis contienen una mayor variedad y cantidad de documentos citados, por lo que su utilización como documento fuente en análisis de citas permite un mejor reflejo de la variedad de fuentes consultadas por los usuarios de una biblioteca. El hecho de que una gran fracción de la bibliografía científica permanezca sin ser citada en las revistas que cubren los índices de citas del ISI no significa necesariamente que no sea leída, sino que por las limitaciones de número de páginas y de referencias que marcan las pautas de publicación de las revistas científicas, no son efectivamente citadas.

## 2.2 La ratio monografías/publicaciones en serie en la política de adquisiciones

Conocer la variedad y distribución de los documentos citados es una información muy útil para una biblioteca. El problema de las adquisiciones en las bibliotecas no se limita a una selección de títulos, sino que existe un problema previo que consiste en decidir el equilibrio que una colección ha de mantener entre libros, revistas, congresos, etc. Desde el punto de vista del servicio de información bibliográfica y de la adquisición de documentos secundarios, el interés de los usuarios por unos u otros tipos de documentos primarios, condicionará el uso de los instrumentos de control bibliográfico. Por otra parte, una de las formas tradicionales de caracterizar a los usuarios ha sido mediante el porcentaje de uso de cada categoría documental. Así, en ciencias aplicadas y en tecnología los porcentajes de citación de congresos, patentes o normas son mayores que en ciencia básica, en la que el dominio de la revista científica es mucho mayor que en las restantes áreas del conocimiento.

La evaluación de la colección también comporta la determinación de la ratio ideal entre el presupuesto dedicado a revistas y el dedicado a monografías. Es este uno de los ámbitos en el que los resultados del análisis de citas con enfoque local son más satisfactorios. Se trata de un problema clásico al que se han de enfrentar las bibliotecas universitarias y difícil de resolver por las implicaciones que tiene en la definición de las partidas presupuestarias de una biblioteca. Con frecuencia, se trasladan de forma mecánica porcentajes de citación de cada una de las tipologías documentales en estudios generales a porcentajes de participación en el presupuesto de adquisiciones, sin atender a la demanda del conjunto de la clientela de una biblioteca, ni a sus hábitos concretos y detallados.

El trabajo de Devin y Kellogg (1990) sobre la ratio revistas/monografías observada en un total de 70 estudios de uso de colecciones resulta muy clarificador. Los resultados obtenidos en los diversos estudios revisados eran muy variados en función de las materias, las instituciones, el periodo analizado, o la definición de la tipología documental<sup>3</sup> sobre la cual se realizaba el recuento. Algo similar se podría observar analizando la gran variedad que muestran las diversas instituciones españolas en relación con este indicador en las estadísticas anuales sobre bibliotecas universitarias que publica Rebiun (*Anuario estadístico...* 2000, p. 67).

Sin embargo, lo que parece fuera de toda duda en el desarrollo de una colección de apoyo a la investigación, concebida por tanto desde una lectura segregada de los presupuestos, es que ha de existir una cierta correlación entre los porcentajes de uso de los diversos tipos de materiales y los correspondientes a las partidas presupuestarias para cada tipo de documento. Así lo concluyen Devin y Kellogg:

"The authors believe that it is possible to determine guidelines for developing a serial/monograph ratio for research library budgeting based on objective criteria. We propose that, for a research library, the percentage of serials and monographs purchased should correlate with the characteristics of literature usage in each subject area. (...) In conclusion, we find that there is no one optimum serial/monograph ratio for all research libraries. Rather, the ratio for each library will vary depending on its own priorities and emphases. For each subject, however, the serial/monograph ratio should be based on the use of the literature by researchers in that subject area as determined by citation studies." (Devin y Kellogg, 1990, p. 54)

Para ello el mejor método del que se dispone es el análisis de las citas en trabajos de los usuarios y el contraste de los datos obtenidos con estudios de citas de alcance general, tal y como concluye Bowman (1991). Un ejemplo de reasignación presupuestaria en función de las referencias presentes en tesis de ingeniería se puede encontrar en el trabajo de Kriz (1978), quien observó un nivel de uso de los libros superior al inicialmente previsto y actuó incrementando la partida destinada a su adquisición en detrimento de las revistas.

### 2.3 Uso de información en el área de informática

Con la elección de la Universitat Politècnica de Catalunya y del área de informática para la realización del trabajo se pretende realizar una contribución al estudio de dos ámbitos en los que la bibliografía analizada muestra un número reducido de trabajos:<sup>4</sup> los informáticos como usuarios de bibliotecas y los estudios bibliométricos en el área de la informática.

Resulta en cierta forma sorprendente el poco interés que los investigadores de esta área han despertado entre los bibliotecarios-documentalistas, pues la bibliografía especializada en biblioteconomía y documentación no los ha considerado sino dentro de estudios genéricos de la ciencia y la tecnología. Si bien se pueden encontrar estudios de usuarios en los que los informáticos son estudiados junto a otros usuarios, son pocos los trabajos que se dedican a ellos de forma monográfica.

En general, todos los trabajos coinciden en dibujar un perfil similar de los investigadores del área: bajo nivel de uso de las bases de datos bibliográficas; preferencia por las colecciones personales de documentos frente a los fondos de una biblioteca; técnicas de interrogación a bases de datos basadas muy mayoritariamente en la combinación de palabras clave en lenguaje libre; control de la bibliografía mediante las citas en las lecturas efectuadas, hojeando<sup>5</sup> las fuentes más próximas, o recurriendo a medios electrónicos poco formalizados bibliográficamente como la consulta de buscadores de recursos web, la visita a sitios web de investigadores e instituciones de solvencia en su terreno, y el contacto con los colegas mediante la comunicación electrónica (correo electrónico, listas de distribución y grupos de noticias).

En definitiva, y tomando como referencia el trabajo de Cunningham y Connaway (1996), los informáticos, son muy selectivos en la cantidad de información "bibliográfica" que consumen, aunque no en términos de calidad y pertinencia, sino en términos de accesibilidad e integración con su entorno electrónico de trabajo. En cuanto a las fuentes documentales primarias más utilizadas, destaca el importante papel atribuido a las comunicaciones de congresos y los informes distribuidos vía web, no tanto por que se consideren fuentes de toda solvencia, sino por su gran actualidad frente a los lentos mecanismos de publicación de las revistas científicas.

Otro de los puntos que ha motivado nuestro interés, es la doble aplicación que el trabajo podía tener en el ámbito bibliotecario y en el de la evaluación de la producción científica. Tras la lectura de documentos referidos al sistema de evaluación de la UPC, hemos observado que los departamentos de las áreas de arquitectura de ordenadores y de lenguajes y sistemas informáticos otorgan una gran importancia a los congresos como forma de publicación relevante, y altamente visible, en un posible sistema de evaluación de la producción científica de la universidad. Por ello reclamaron de la Comisión de Publicaciones Notables del Vicerrectorado de Investigación que en dichas áreas se tuvieran en consideración tanto o más que ciertas revistas determinados congresos de prestigio. Dicha comisión es la responsable de la determinación de una lista de publicaciones complementaria al *Journal Citation Report* que permite una evaluación cualitativa y no meramente cuantitativa de las publicaciones generadas por cada investigador (UPC. Vicerectorat de Recerca 1997 y 1999), por lo que comprobar si realmente los congresos jugaban el papel que se les asignaba en los departamentos parecía un tema de gran interés.

## 3 Objetivos y metodología

Los objetivos concretos que se pretenden son:

1. Obtener la ratio correspondiente a la proporción entre las citas a publicaciones adquiridas por la biblioteca en forma de suscripción y las que reciben documentos adquiridos "como" monografías. La obtención de este dato se ha de realizar en función del modelo de adquisición que en la actualidad lleva a cabo la biblioteca, y no según la naturaleza de los diversos tipos de documentos.
2. Confirmar las aportaciones que los departamentos estudiados realizaron a la Comisión de Publicaciones Notables en el sentido de considerar mayor la visibilidad de los congresos en el ámbito de la informática en comparación con el resto de documentos. También se pretende comprobar si el uso de literatura gris por parte de este tipo de investigadores es tan elevado como la bibliografía acostumbra a describir.
3. Considerando que el colectivo analizado no presenta problemas significativos de acceso a los recursos disponibles en Internet, se pretende cuantificar la presencia explícita de recursos electrónicos y la presencia de determinados documentos citados en papel, pero que en realidad son documentos electrónicos —como es el caso de buena parte de la literatura gris citada—, o que poseen un potencial muy alto de migración a formato

electrónico.

Para ello, frente a la alternativa de realizar un estudio por métodos directos de entrevista o encuesta, el análisis bibliométrico de la bibliografía citada en tesis doctorales ha atraído nuestro interés por tratarse de un método que no interfiere en el comportamiento de los usuarios mientras se toman los datos y que permite una disponibilidad relativamente sencilla de los mismos ya que no se requiere la cooperación de un entrevistado o encuestado, ni la respuesta aparece contaminada por el proceso de estudio.

Se ha trabajado sobre el total de las tesis del periodo 1996-1998 presentadas en el Departamento de Arquitectura de Computadores y en el Departamento de Lenguajes y Sistemas Informáticos, sin realizar ningún tipo de muestreo. El total de tesis y referencias escritas ha sido de 54 y 6.807 respectivamente, los datos se han introducido en una base de datos gestionada por un programa relacional a partir de la cual se han generado resultados de tipo descriptivo y matrices de datos para su explotación con un paquete estadístico.

Se han contabilizado las citas a partir de las listas de referencias que figuran al final de las tesis en la forma de bibliografía citada. Se planteó la exclusión de las bibliografías que recogiesen bibliografía consultada pero no citada, pero no se observó ningún caso correspondiente a dichas características. También se planteó la exclusión de aquellas tesis que tuvieran la lista de referencias segregada y duplicada por capítulos, sin una lista unificada final de referencias citadas, pero tampoco se observó ningún caso en el conjunto analizado.

El recuento se ha realizado contabilizando una vez cada documento citado, entendiendo por documento en el caso de las series cada uno de los artículos o comunicaciones que forman parte de una publicación.

Las publicaciones se han considerado en función de la representación que de ellas se hacía por medio de la referencia bibliográfica consignada en la bibliografía. Así, las revistas se han distinguido por ir acompañadas de una designación de fascículo, o por coincidir con alguno de los títulos o abreviaturas que figuraban en la base de datos con la que se contaba en el momento de iniciar el recuento de citas.<sup>6</sup> Para los congresos se contó en un inicio con la lista de congresos notables UPC para alimentar la aplicación informática utilizada en el recuento, al tiempo que, por lo general, la designación cronológica y/o numérica es bien característica y diferenciada de las revistas. El punto más conflictivo ha sido la consideración que se había de dar a determinadas series monográficas que los autores en ocasiones citaban como serie, en otras ocasiones como congreso, y también en otras como monografía. Nuestro criterio ha sido considerarlas, siempre que fuera posible, monografías.

De todas formas, la falta de una definición clara y uniforme de las diferentes tipologías documentales, especialmente en el sentido del alcance dado al término publicación seriada, es uno de los principales problemas a la hora de contrastar resultados relativos a distribución de porcentajes por tipología documental entre los diversos trabajos publicados. Así, Walcott (1991) en su estudio sobre tesis doctorales de geología observa un 80% de publicaciones seriadas frente a un 20% de monografías, sin que quede claro dónde se contemplan los congresos citados. Por su parte, Glänzel y Schoepflin (1999), en un estudio monográfico sobre la proporción de citas a series en diversas áreas del conocimiento en el que procesaron con ordenador diez millones de referencias de las bases de datos del ISI correspondientes a 1993, consideraron publicación en serie todas las referencias en las que figuraba un número de volumen válido y la indicación de primera página de la parte componente, por lo que parece que los congresos no entraron en su definición automática de publicación seriada que, de facto, contemplaba únicamente revistas.

## 4 Resultados

Los departamentos seleccionados han generado a lo largo de los tres años considerados un total de 54 tesis (tabla 1). La contribución por parte de cada uno de ellos en porcentaje sobre el total es bastante equilibrada excepto en el año 1998, en el que la diferencia se eleva a más de 35 puntos, lo que ocasiona que la diferencia en el conjunto de años se eleve a 18,5 puntos. A lo largo de los años analizados el número de tesis leídas se mantiene uniforme en el Departamento de Arquitectura de Computadores (AC), mientras que en el Departamento de Lenguajes y Sistemas Informáticos (LSI) se produce un salto considerable entre el primero y los dos años siguientes.

Tabla 1. Tesis leídas por departamento y año de lectura

Departamentos	Total	%	1996	%	1997	%	1998	%
AC	22	40,74	7	50,00	9	42,86	6	31,58

<b>LSI</b>	32	59,26	7	50,00	12	57,14	13	68,42
<b>Total</b>	54	100,00	14	100,00	21	100,00	19	100,00

Tabla 2. Cifra total de referencias y promedio de referencias por tesis según departamento y año de lectura

Dep.	Total Ref.	%	Ref. por tesis	1996		1997		1998				
				Total Ref.	%	Total Ref.	%	Total Ref.	%			
<b>AC</b>	2485	36,50	113	1.002	51,65	143	1.032	41,20	115	451	19,10	75
<b>LSI</b>	4.322	63,50	135	938	48,35	134	1.473	58,80	123	1.911	80,90	147
<b>Total</b>	6.807	100,00	126	1.940	100,00	139	2.505	100,00	119	2.362	100,00	124

Sin embargo, si se atiende al número de referencias totales incluidas en las tesis (tabla 2) el Departamento LSI casi dobla al Departamento AC, con una diferencia porcentual de 27 puntos. Esta diferencia se ha de imputar tanto al mayor número de tesis, como al mayor número de citas por tesis en el Departamento LSI, y se encuentra en línea con en mayor consumo de información bibliográfica en el área de lenguajes.

#### 4.1 Tipología documental

La distribución de las fuentes citadas (fig. 1) presenta unos resultados que, a primera vista, pudieran parecer sorprendentes: un 31,9% de las citas corresponden a comunicaciones de congresos, conferencias o *workshops*, configurando este tipo de documento como el más citado por encima de los artículos de revista. En segundo lugar, figuran los artículos de revista con un 31,2% y, en tercero, los libros con el 13,3%, porcentaje que sumado al 6,5% de los capítulos de libro conforma una volumen de citas de este tipo de publicación muy elevado para lo que es habitual en ciencia y tecnología.

El porcentaje de documentos citados que la Biblioteca de la UPC adquiere por suscripción es de un 31,2%. Esta cifra se ajusta con mucha precisión a la distribución actual del presupuesto de adquisiciones de la UPC que, según los datos que figuran en las estadísticas Rebiun correspondientes al año 1999 (*Anuario estadístico... 2000: 67*), destina el 64% del gasto a la adquisición de monografías, entre las cuales se contemplan las actas de congresos.

Figura 1. Distribución de las citas según tipología documental

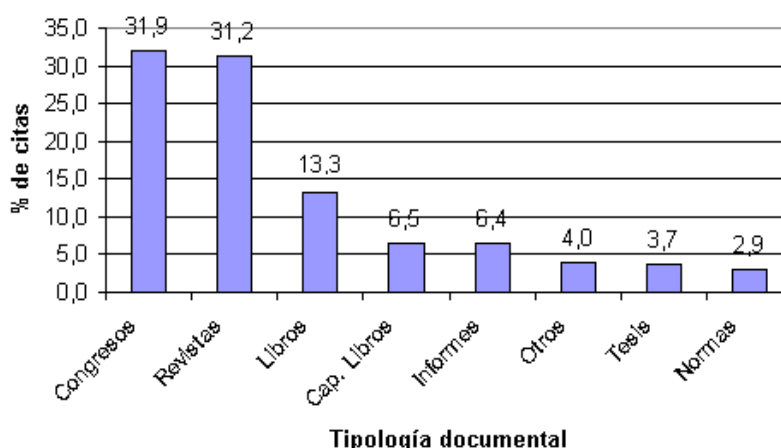


Figura 2. Comparación del uso de las diversas categorías documentales por parte de cada departamento una vez ponderado el peso de cada categoría en el total de citas de cada departamento

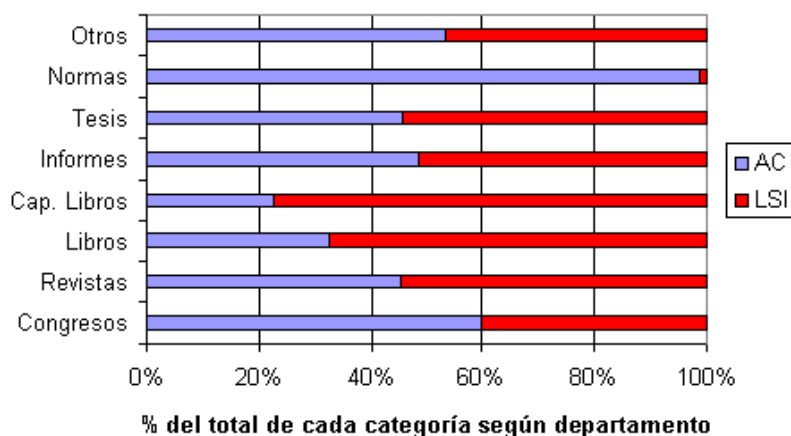


Tabla 3. Tipología de fuentes citadas y distribución según departamentos

Fuentes	Total	%	AC	%	LSI	%
Comunicaciones a congresos	2.171	31,89	1.000	40,24	1171	27,09
Artículos de revista	2.126	31,23	687	27,65	1439	33,29
Libros	908	13,34	197	7,93	711	16,45
Capítulos de libro	442	6,49	63	2,54	379	8,77
Informes	438	6,43	154	6,20	284	6,57
Tesis	252	3,70	82	3,30	170	3,93
Normas	196	2,88	192	7,73	4	0,09
Inf. comercial y de producto	79	1,16	54	2,17	25	0,58
Preprints y doc. de trabajo	76	1,12	16	0,64	60	1,39
Manuscritos	53	0,78	8	0,32	45	1,04
Notas técnicas y de aplicación	21	0,31	12	0,48	9	0,21
Proyectos final de carrera	6	0,09	3	0,12	3	0,07
Comunicaciones personales	4	0,06	3	0,12	1	0,02
Patentes	1	0,01	1	0,04	0	0,00
Sin determinar	34	0,50	13	0,52	21	0,49
<b>Total</b>	<b>6.807</b>	<b>100,00</b>	<b>2.485</b>	<b>100,00</b>	<b>4.322</b>	<b>100,00</b>

La distribución por tipo de documento es significativamente diferente según se trate de uno u otro departamento (fig. 2 y tabla 3). La mayor utilización de fuentes de información bibliográfica en el área de lenguajes informáticos concuerda con una utilización de fuentes más formalizadas, como las revistas científicas: con un 33,29% los artículos de revista son la categoría documental más citada en el Departamento LSI, frente a un 27,65% en el caso del Departamento AC.

En cambio, el dominio de los congresos en las tesis del Departamento AC con un 40,08% de las citas es muy destacable y concuerda con un área de conocimiento en la que la actualidad de la información prima sobre cualquier otra consideración (fig. 3).

No se observan grandes oscilaciones a lo largo del tiempo en el número de citas que recibe cada clase de documentos (tabla 4), excepto en el caso de las normas técnicas, con una punta fuera de lo normal en 1997, y en el de los *preprints*, en 1998, con un incremento importante pero dentro de unos límites normales. El caso de las normas se explica por la distorsión creada por una tesis en la que esta forma documental aparece citada de forma anormalmente dominante. Para explicar el incremento de los *preprints* en 1998 se podría especular con diversos factores, entre los que destaca la mayor y mejor distribución que ofrecen a medida que se consolida Internet, y en particular el web, como medio habitual de difusión de este tipo de fuentes.

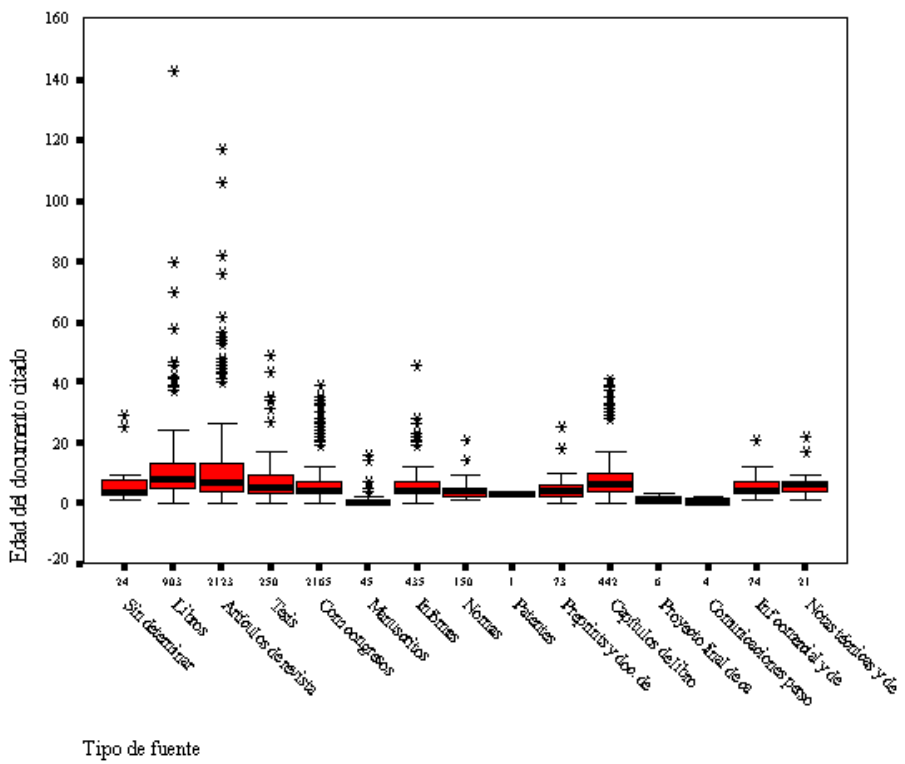


Tabla 4. Tipología de fuentes citadas y distribución por años de lectura de las tesis

Fuentes	Total	%	1996	%	1997	%	1998	%
Comunicaciones a congresos	2.171	31,89	641	33,04	747	29,82	783	33,15
Artículos de revista	2.126	31,23	622	32,06	781	31,18	723	30,61
Libros	908	13,34	292	15,05	283	11,30	333	14,10
Capítulos de libro	442	6,49	103	5,31	216	8,62	123	5,21
Informes	438	6,43	140	7,22	142	5,67	156	6,60
Tesis	252	3,70	73	3,76	85	3,39	94	3,98
Normas	196	2,88	4	0,21	169	6,75	23	0,97
Inf. Comercial y de producto	79	1,16	24	1,24	30	1,20	25	1,06
Preprints y doc. de trabajo	76	1,12	15	0,77	7	0,28	54	2,29
Manuscritos	53	0,78	15	0,77	24	0,96	14	0,59
Notas técnicas y de aplicación	21	0,31	4	0,21	5	0,20	12	0,51
Proyectos final de carrera	6	0,09	0	0,00	4	0,16	2	0,08
Comunicaciones personales	4	0,06	0	0,00	3	0,12	1	0,04
Patentes	1	0,01	1	0,05	0	0,00	0	0,00
Sin determinar	34	0,50	6	0,31	9	0,36	19	0,80
<b>Total</b>	<b>6.807</b>	<b>100,00</b>	<b>1.940</b>	<b>100,00</b>	<b>2505</b>	<b>100,00</b>	<b>2.362</b>	<b>100</b>

El análisis detallado de la actualidad de la bibliografía citada y del índice de envejecimiento para cada uno de los tipos de documentos considerados en el estudio se puede observar gráficamente por medio del diagrama de cajas que muestra la figura 3. El diagrama de cajas ofrece una visión resumida muy simple y clara para comparar el comportamiento de la variable edad entre los diversos tipos de documentos. Se trata de una comparación gráfica que permite representar múltiples características de cada una de las distribuciones.

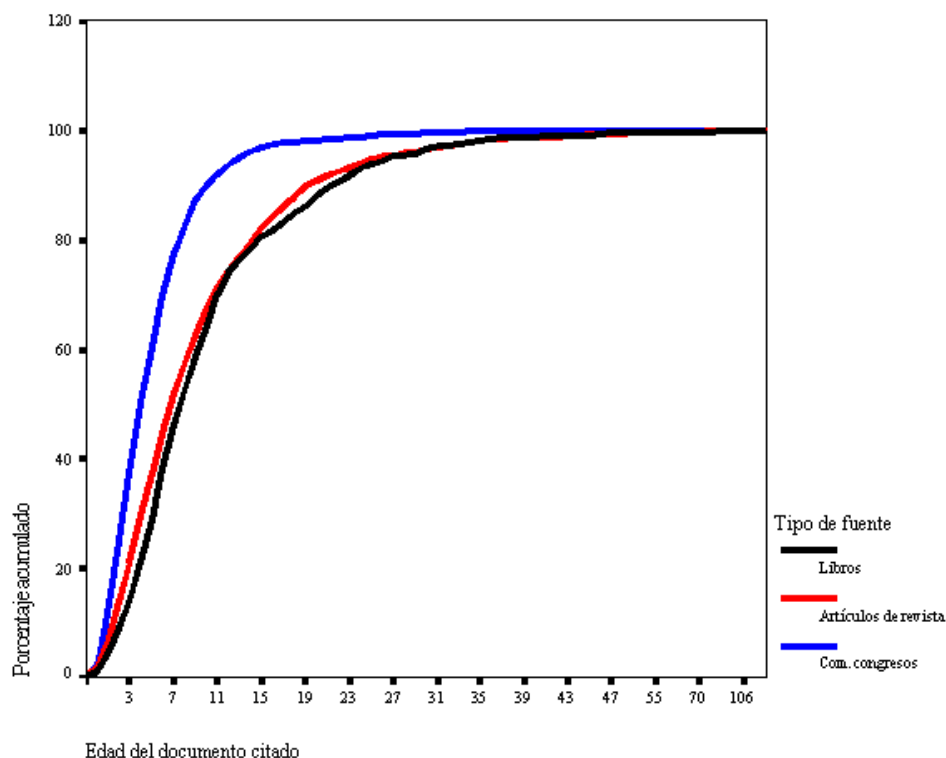
Figura 3. Diagrama de caja y bigote de la edad según los diversos tipos de documentos



Los límites de la caja roja vienen marcados por el primer y tercer cuartil de la distribución, por lo que el 50% de los casos están comprendidos dentro de la caja, mientras que la raya negra horizontal dentro de la misma representa la mediana. El valor más pequeño que esté dentro del primer cuartil menos 1,5 veces el rango intercuartil marca el límite inferior, representado por la raya horizontal que marca el inicio de la línea vertical que atraviesa la caja, mientras que el límite superior lo marca el valor más grande que esté dentro del tercer cuartil más 1,5 veces el rango intercuartil. Los valores que no llegan o superan dichos límites se muestran en forma de estrella.

Al observar la figura 3, se puede afirmar a simple vista que los libros, los capítulos de libro y las revistas son las publicaciones que muestran una mayor resistencia al envejecimiento, si bien el número de casos extremos y el rango de casos entre los límites es mayor para las revistas. El contraste entre la obsolescencia de los congresos y la de los libros y revistas es más que notable: el porcentaje de documentos citados que tienen una edad de cero a cinco años (índice de Price) para los primeros es del 60,23%, con un valor para la revistas del 38,81% y para los libros del 28,05%.

Figura 4. Porcentaje acumulado de citas según la edad para los tres tipos de documentos más numerosos



En este sentido, se puede establecer una relación entre el grado de actualidad de la bibliografía citada y el mayor o menor dominio de los congresos como fuente más citada. El departamento AC manifiesta una demanda de información ligeramente más actualizada, con una mediana de edad de los documentos citados de 5 años, por 6 del departamento LSI, y presenta un mayor porcentaje de congresos citados que el departamento LSI.

#### 4.2 Soporte documental

Se ha querido prestar especial atención a la observación del uso de fuentes electrónicas frente al papel, pero el análisis de citas se ha mostrado insuficiente por sí mismo. En la mayoría de los casos, la referencia bibliográfica del documento citado no permite discriminar con seguridad el soporte documental. Por esta razón se contabilizan los datos obtenidos del análisis de citas considerando como documentos electrónicos únicamente aquellos que aparecen explícitamente descritos como tales (tabla 5), pese a que se supone que su número real es mucho mayor.

Tabla 5. Soportes documentales citados y distribución según departamentos

Soporte documental	Total	%	AC	%	LSI	%
Papel	6.727	98,82	2.447	98,47	4.280	99,03
Electrónico explícito	47	0,69	26	1,05	21	0,49
- <i>Tangible (cd-rom, etc.)</i>	1	0,01	1	0,04	0	0,00
- <i>Internet FTP</i>	5	0,07	1	0,04	4	0,09
- <i>Internet web</i>	33	0,48	16	0,64	17	0,39
- <i>Internet sin determinar</i>	8	0,12	8	0,32	0	0,00
Sin determinar, o no aplicable	33	0,48	12	0,48	21	0,49
<b>Total</b>	6.807	100,00	2.485	100,00	4.322	100,00

Con los datos de citas en la mano, se observa un dominio abrumador del soporte papel o "asimilable", es decir, de documentos que quizás están disponibles en soporte electrónico pero que responden formalmente a la tradición del papel, o que conviven con una versión en papel a la que sirven de facsímil: revistas en formato pdf que tienen doble o triple presentación (papel, cd-rom, en línea), informes técnicos y de investigación, y notas técnicas. De todas formas, se ha de destacar que el porcentaje de documentos contabilizados como documentos impresos disminuye, aunque levemente, a lo largo del periodo analizado (tabla 6).

Tabla 6. Soportes documentales citados y distribución según años de lectura

Soporte documental	Total	%	1996	%	1997	%	1998	%
Papel	6.727	98,82	1.934	99,69	2.469	98,56	2.324	98,39
Electrónico explícito	47	0,69	2	0,10	26	1,04	19	0,80
<b>- Tangible (cd-rom, etc.)</b>	<b>1</b>	<b>0,01</b>	<b>0</b>	<b>0,00</b>	<b>1</b>	<b>0,04</b>	<b>0</b>	<b>0,00</b>
<b>- Internet FTP</b>	<b>5</b>	<b>0,07</b>	<b>1</b>	<b>0,05</b>	<b>1</b>	<b>0,04</b>	<b>3</b>	<b>0,13</b>
<b>- Internet web</b>	<b>33</b>	<b>0,48</b>	<b>1</b>	<b>0,05</b>	<b>16</b>	<b>0,64</b>	<b>16</b>	<b>0,68</b>
<b>- Internet sin determinar</b>	<b>8</b>	<b>0,12</b>	<b>0</b>	<b>0,00</b>	<b>8</b>	<b>0,32</b>	<b>0</b>	<b>0,00</b>
Sin determinar, o no aplicable	33	0,48	4	0,21	10	0,40	19	0,80
<b>Total</b>	<b>6.807</b>	<b>100,00</b>	<b>1.940</b>	<b>100,00</b>	<b>2505</b>	<b>100,00</b>	<b>2.362</b>	<b>100</b>

Tabla 7. Documentos en soporte electrónico según tipología documental

Tipología documental	Total Citas	Citas doc. electrónicos	% citas doc. electrónicos
Comunicaciones a congresos	2.171	4	0,18
Artículos de revista	2.126	2	0,09
Libros	908	0	0,00
Capítulos de libro	442	0	0,00
Informes	438	6	1,37
Tesis	252	0	0,00
Normas	196	11	5,61
Inf. comercial y de producto	79	10	12,66
Preprints y doc. De trabajo	76	9	11,84
Manuscritos	53	1	1,89
Notas técnicas y de aplicación	21	0	0,00
Proyectos final de carrera	6	0	0,00
Comunicaciones personales	4	0	0,00
Patentes	1	0	0,00
Sin determinar	34	4	11,76
<b>Total</b>	<b>6.807</b>	<b>47</b>	<b>0,69</b>

Más destacable es la vinculación de determinados documentos con la edición electrónica en Internet: el 5,61% de normas técnicas, el 11,84% de los preprints y el 12,66% de la información comercial y de producto son documentos citados explícitamente como documentos de Internet. Dado el periodo analizado (1996-1998), estas cifras no hacen otra cosa que indicar el inicio de una tendencia que se ha de intensificar en los próximos años.

## 5 Discusión

### 5.1 Tipología documental

La distribución observada según la tipología documental muestra un gran contraste con los datos de análisis de citas en tesis doctorales obtenidos en otros estudios y con los patrones observados en el uso de bibliografía científico-técnica en general.<sup>2</sup> Pero si se analiza desde el conocimiento de las características de la bibliografía informática y de los hábitos de consumo de los investigadores en esta área de conocimiento, la distribución aparece como bastante ajustada a lo que es paradigmático (Olander, 1992; Cunningham y Connaway, 1996).

En efecto, la primera posición de los congresos entre las categorías documentales citadas parece concordar con la importancia que los profesionales e investigadores de la informática otorgan a dicha fuente. Así, por ejemplo, el programa de publicaciones notables de la UPC (UPC. Vicerectorat de Recerca, 1997), que contempla tanto revistas como congresos, recoge en el área denominada "Tecnologías de la información y la comunicación y electrónica" un total de 132 títulos de congresos, frente a los 175 que figuran en el conjunto del listado de congresos notables formado por publicaciones de cinco áreas diferentes: los informáticos han otorgado más importancia a los congresos que los otros investigadores de la UPC en un proceso de evaluación de la investigación basado en las publicaciones.

Otra de las razones que explicaría el gran número de congresos citados se puede encontrar en el hecho de que se trata del tipo de documento en el que más han publicado previamente los autores de las tesis junto a sus directores y colegas de equipo de investigación.<sup>8</sup> Es evidente que siempre existe un sesgo hacia las fuentes en las que los autores y sus colegas más inmediatos publican. Este hecho unido al de la presencia de dichos documentos en las bibliotecas de los departamentos,<sup>9</sup> podría explicar también el mayor porcentaje de citas de esta categoría frente a las revistas.

Otro dato a tener en consideración es el hecho de que los congresos sean el tipo de documento más citado es el grado de obsolescencia que presentan las citas a los congresos en comparación con las revistas (fig. 3). Se puede observar que las publicaciones que presentan un mayor índice de obsolescencia son precisamente los congresos. Este dato es muy importante en un área de conocimiento en la que los cambios son muy rápidos, por lo que todo hace indicar que los congresos se adaptan mejor a las necesidades de comunicación de los informáticos.

Si se contempla el porcentaje acumulado de citas en función del incremento de edad de los documentos citados para las tres clases de fuentes más citadas (fig. 4) sorprende el comportamiento casi idéntico de libros y revistas, al tiempo que destaca la actualidad de la bibliografía correspondiente a los congresos. El hecho de que en el campo de la informática se necesite una información muy actualizada e inmediata explica el predominio de los congresos como forma documental frente al lento proceso de filtrado, revisión, edición y distribución que soportan las revistas. Más allá del predominio de una forma documental, el elevado número de congresos existentes en el ámbito de la informática y su gran especialización, explican por sí mismos el papel de la comunicación interpersonal por medio de reuniones científicas en esta área. Esta inmediatez, junto con la mayor accesibilidad de los congresos en bibliotecas de los departamentos o en colecciones personales, propias o de colegas,<sup>10</sup> explica en buena medida el elevado número de citas que recibe esta categoría documental.

También resulta destacable el alto porcentaje que acumulan las citas a libros y capítulos de libro conjuntamente, con un porcentaje nada despreciable del 19,83%. Esta cifra se ha de matizar en el apartado capítulos de libros, pues en ella se han considerado contribuciones individuales en volúmenes de series monográficas que con frecuencia recogen comunicaciones seleccionadas de reuniones y congresos. Aún así, las cifras de estas dos categorías resultan bastante elevadas para lo que es habitual en la bibliografía tecnológica y se podrían explicar considerando el enfoque más teórico y de fundamentación que los departamentos de la UPC quieren dar al doctorado y la tesis, frente a la orientación más aplicada de los proyectos final de carrera. Esta afirmación se ve reforzada por el bajo volumen de citas que reciben las patentes y los proyectos final de carrera.

En cuanto al estudio del volumen de literatura gris citada, el primer obstáculo que se presenta es que no existe una definición totalmente clara de lo que se entiende por literatura gris, ni tan sólo se dispone de un consenso en castellano respecto al término que la designa. A nivel operativo, en este trabajo en el que se analizan citas y no documentos, se ha optado por considerar dentro del apartado literatura gris fuentes de información que muestran un control bibliográfico débil tales como informes técnicos y de investigación, tesis, información comercial y de producto, *preprints* y documentos de trabajo, manuscritos, notas técnicas y de aplicación, proyectos final de carrera, y comunicaciones personales.

Realizadas estas consideraciones, el volumen de literatura gris se puede considerar más que notable porque representa un 13,65% de las citas lo que equivale a 929 citas sobre un total de 6.807. Esta cifra podría ser mayor si se contemplasen dentro de este apartado algunas contribuciones a congresos, que habitualmente no alcanzan una publicación y distribución editorial "normal" y que se quedan formalmente como *papers*. Este alto volumen de citas de publicaciones de literatura gris concuerda con las características del uso de información observadas entre los tecnólogos en general (Terrada, 1994, p. 94). El volumen de consumo de este tipo de fuentes se puede decir que es directamente proporcional a la distancia que tiene una determinada área de investigación de enfoques teóricos, o de lo que se conoce como investigación básica, esto es, cuanto más aplicada es una disciplina científica, más numerosa es la presencia de la literatura gris en su bibliografía.

En este sentido, una lectura atenta de la figura 2 permite matizar el menor número de referencias por tesis en el departamento AC con la distribución de las mismas según la tipología documental. En dicho gráfico podemos ver que el consumo de información es menor, pero respecto al total de citas recibidas por congresos, informes, tesis, normas y otros, la cuota de participación del departamento AC es mayor, tanto en valor absoluto en algunos casos como las normas, como en valor relativo una vez descontado el menor número de referencias generadas por las tesis de dicho departamento. Por tanto, no sólo se da un menor número de referencias por tesis en el departamento AC, sino que el consumo de información se basa en mucha mayor medida en fuentes de la denominada literatura gris.

En estudios locales de usuarios como el presente, es de gran interés el contraste de los resultados obtenidos con los datos disponibles para el ámbito nacional o internacional. Para ello, las ciencias de la computación cuentan con un instrumento de control bibliográfico de primer orden que puede ser de gran utilidad para analizar si la distribución de citas observadas según tipología documental concuerda con lo que sucede en el conjunto de la bibliografía informática. Se trata de la bibliografía de bibliografías de informática titulada *The collection of computer science bibliographies* (Achilles, 1999), a la cual se puede acceder libremente por Internet. Además de actuar como una bibliografía de bibliografías clásica, con una potencia de recuperación de información más que notable, también permite obtener en tiempo real las estadísticas de todas y cada una de las bibliografías indizadas, con datos sobre la tipología documental y las fechas de publicación. Así mismo ofrece las estadísticas agregadas de todo el conjunto, por lo que se puede obtener una imagen bastante interesante entre lo que se publica sobre informática y lo que se publicita por medio de bibliografías selectivas y bibliografías temáticas. Los datos representan un análisis de nada menos que un millón de referencias bibliográficas, algunas de las cuales, por supuesto, están duplicadas.

Al comparar los datos observados en el presente estudio con los observados en bibliografías cubiertas por la base de datos de Achilles (1999) que tienen perfiles temáticos similares a cada uno de los departamentos estudiados, se observa una gran similitud con las tipologías documentales citadas en las tesis. Por su parte, el resumen general del conjunto de la bibliografía de bibliografías de Achilles ofrece un dato revelador sobre la estructura de la comunicación científica en estas áreas de conocimiento: sobre un total de 1.038.616 noticias,<sup>11</sup> los artículos de revistas representaban un 46,70%, las comunicaciones a congresos un 29,18%, y los informes técnicos un 10,37%.

Estos datos, obtenidos de bibliografías de propósito informativo, y no de citas en trabajos, informan sobre la producción editorial y no sobre el uso, al tiempo que priman el valor permanente de la información. Por ello, en los datos agregados de conjunto, los artículos están sobrerrepresentados. De todas formas lo que es relevante es que los congresos representen casi un 30% de la bibliografía circulante, algo que no es habitual en otras disciplinas científicas.

A grandes trazos, la tipología de fuentes citadas en las tesis y su distribución porcentual es consistente con lo observado en la bibliografía que estudia el uso de información por métodos directos (Takayama, 1986; Olander, 1992; Cunningham y Connaway, 1996), y en la que se confirma la imagen obtenida en el presente estudio según la cual una parte muy significativa de la comunicación científica se realiza por medio de los congresos y de literatura gris. Se ha de destacar que, en conjunto, las revistas tan sólo representan un 31,2% del total de citas. Resulta éste un dato muy significativo, pues por lo general se asumen para la ciencia y la tecnología porcentajes de uso más altos para esta categoría documental, porcentajes que con frecuencia se aplican de forma muy mecánica en el desarrollo de colecciones.

En efecto, los datos obtenidos son realmente relevantes en la evaluación de políticas de adquisición. Si bien los congresos se podrían considerar dentro del capítulo de publicaciones seriadas, la verdad es que, en general, su adquisición no se realiza de forma consolidada como publicación en serie, imputándose su adquisición en la mayoría de las ocasiones a la partida presupuestaria de monografías. Tanto por esta razón, como por el significativo porcentaje de libros y capítulos de libro citados, los datos que se han obtenido en este apartado pueden ser motivo de reflexión en el terreno de la ratio entre monografías y revistas del presupuesto de adquisiciones.

## 5.2 Soporte documental

La cifra de documentos electrónicos accesibles en línea por Internet, dentro de su reducido nivel, es bien significativa si se compara con estudios anteriores sobre el uso de recursos Internet por parte de investigadores en ciencia y tecnología basados en el análisis de citas (Urbano, 1997) y en los que se constataba un número aún más bajo de citas explícitas a documentos electrónicos. En este caso, se observa un porcentaje del 0,66% sobre el total de citas, con una tendencia al incremento a lo largo del tiempo (tabla 6) de los recursos web que podría relacionar con el papel creciente de la edición no-tradicional vinculada a Internet. Comparado con el 1,9% de citas que recibían los recursos electrónicos en el estudio de Harter y

Kim (1996b) realizado a partir de artículos publicados en revistas electrónicas, el 0,66% en las tesis de la UPC es ciertamente relevante.

Como puede observarse en la tabla 7, la presencia de documentos en soporte electrónico no es uniforme, ya que es mucho más notable en el ámbito de la literatura gris. Se ha de tener en consideración que la revisión y el arbitraje como condición para publicar en ciencia y tecnología está mucho más extendida y es más exigente que en muchas disciplinas de las humanidades y las ciencias sociales. Este hecho provoca que se otorgue mucho más valor a fuentes reconocidas, asentadas tradicionalmente en modelos de publicación pre-Internet. Así, como revista electrónica de acceso en línea y de distribución gratuita tan sólo se ha identificado *NHSE review* con dos citas. Ninguna de las revistas electrónicas de materias afines a los usuarios estudiados y con mayor número de citas en del estudio de Harter y Kim (1996b) figura en la lista de las citadas en las tesis de la UPC.<sup>12</sup>

No es sorprendente, que en el momento en que las condiciones de acceso y de capacitación en el uso de Internet de los investigadores en ciencias "blandas" y humanidades se igualan a los de ciencias "duras" y tecnología, el nivel de recursos Internet citados sea mayor en las áreas de conocimiento con mecanismos de comunicación científica menos estrictos. Así, un estudio de las citas en publicaciones de distribución normalizada, como los libros y las revistas, mostraría que se pueden encontrar numerosas citas a documentos electrónicos en trabajos que estudian Internet en relación con aspectos sociales y de aplicación, pero en menor medida en trabajos de investigación tecnológica reconocida y de vanguardia, pues la red es una buena fuente de información sobre sí misma en términos de normalización y difusión de su tecnología, así como para el estudio de su impacto social. Esto explica que los estudios sobre el uso de Internet en términos de análisis sociológico o, como en caso de los bibliotecarios-documentalistas, sobre la radical transformación de una práctica profesional a causa de la red, son los que muestran un mayor número de citas a recursos Internet (Masip, 1997; Estivill y Urbano, 1998).

En cualquier caso, a partir de las características de la colección, de las observaciones del personal bibliotecario encargado de las publicaciones periódicas, de las suscripciones electrónicas que a nivel personal o institucional mantienen determinados investigadores sin estar consolidadas en el catálogo de la biblioteca, y del análisis de los servidores web de los departamentos estudiados, se podría llegar a concluir lo que aquí se formula sólo como suposición, es decir, que un buen número de las referencias bibliográficas, a pesar de que no se pueda establecer con certeza por la descripción, corresponde a consultas de documentos sobre soportes diferentes al papel (microformas, cd-rom, recursos en línea comerciales y recursos en línea gratuitos).

En este terreno, el análisis de citas se ha mostrado insuficiente y serían necesarias otras técnicas para evaluar el uso de recursos electrónicos como, por ejemplo, el análisis de conexiones a servidores locales o remotos vinculados al suministro de documentos primarios. Este tema es especialmente interesante pues existen al menos dos factores recientes que pueden modificar el comportamiento futuro en el uso de la colección. De una parte, a partir del año 1999 tanto el Servei de Biblioteques UPC como el Consorci de Biblioteques Universitàries de Catalunya ofrecen acceso a revistas electrónicas desde ordenadores con IP de la universidad. De otra, la Biblioteca Rector Gabriel Ferraté, central del Campus Nord de la UPC, dispone de la base de datos *IEL* (IEE/IEEE Electronic Library)<sup>13</sup> desde el año 1997 en cd-rom con acceso local, lo que ha comportado una progresiva cancelación de suscripciones a revistas en papel del IEEL, cancelaciones que se ha acentuado claramente en 1999.

Si los datos sobre citas de los doctorandos de la UPC se rectificasen tan sólo en lo que se refiere a literatura de informes y *preprints* posteriores a 1990,<sup>14</sup> considerándolos recursos electrónicos accesibles por Internet, se contarían un total de 404 citas, cifra que representaría casi un 6% sobre el total de las 6.807 contabilizadas en el estudio. Los datos sobre uso de documentos disponibles vía Internet que se obtendrían realizando una proyección como la enunciada concuerdan mejor con los datos de las estadísticas ofrecidas por Achilles (1999) en *The Collection of computer science bibliographies*, en la que en total, sobre el millón de noticias analizadas, 92.543 contienen el URL en la referencia bibliográfica como elemento de descripción documental, por lo que no es aventurado afirmar, que en la actualidad un 9% de la bibliografía circulante sobre informática a nivel mundial pertenece al ámbito de la edición no comercial en red.

El gran volumen de literatura gris citada, ha de ser tenido en consideración tanto desde el servicio de información bibliográfica, como desde la política de adquisiciones y de catalogación de recursos electrónicos. El debate sobre cómo se ha de actuar con relación a los recursos electrónicos disponibles en Internet es especialmente relevante en el terreno de este grupo de documentos. Se han de identificar los sitios web que se dedican a repertoriar este tipo de documentos; se ha de tener conocimiento de los sitios web más productivos en la publicación de informes y *preprints*, pero especialmente los más utilizados por los usuarios de la biblioteca; se ha de analizar el grado de cobertura con el que son contemplados por los sistemas clásicos de control bibliográfico; y, por último, se ha de decidir si la biblioteca ha de ofrecer algún

mecanismo de acceso a los mismos, y si es este mecanismo pasa por el catálogo, o por repertorios independientes de recursos web.<sup>15</sup>

En cualquier caso, los datos obtenidos demuestran que en el ámbito de las ciencias de la computación el papel que tiene la comunicación científica entre colegas, frente a la que canalizan las grandes editoriales y las revistas científicas, es realmente significativo. Por otra parte, tanto los instrumentos de control bibliográfico en la red, como las colecciones digitales relevantes, ya sean de libre acceso o de pago, tienen un nivel de cobertura y de calidad muy destacable.<sup>16</sup> Como ejemplos reseñables de recursos gratuitos se pueden citar: *ResearchIndex* (Lawrence, Giles, Bollacker, 1999a y 1999b), índice de citas electrónico de publicaciones en Internet; *The collection of computer science bibliographies* (Achilles, 1999), bibliografía de bibliografías de informática en formato electrónico; y *Networked Computer Science Technical Reference Library* (NCSTRL 1999), repertorio de informes técnicos y de investigación. Por parte de la edición comercial, los dos editores más importantes en el campo de la informática — confirmados a la luz de lo observado en el presente estudio en función del número de citas que reciben sus publicaciones— disponen de sedes web destinadas a distribuir todas sus publicaciones: la ACM (Association of Computing Machinery) con la *ACM Digital Library* (ACM, 1999) y el IEEE (Institute of Electrical and Electronics Engineers) con la *IEEE/IEE Electronic Library Online* (*IEL online...* 1999).

En definitiva, las posibilidades de trabajar sólo con recursos en línea son reales, por lo que estos datos tendrán que ser tenidos en consideración por las bibliotecas que quieran añadir valor a estos documentos, único método para ser valoradas como un mecanismo eficiente de acceso a la información en el nuevo escenario de edición en la red que se dibuja tanto para los canales no-comerciales como para los comerciales.

## 6 Conclusiones

El tipo de documento más citado corresponde a las comunicaciones de congresos, confirmándose el papel clave otorgado a esta categoría documental en el área de la informática por parte de la Comisión de Publicaciones Notables a requerimiento de los dos departamentos estudiados. Este dominio en conjunto de los congresos, junto a la presencia de un importante volumen de literatura gris, así como de libros y capítulos de libros, ha de ser tenida muy en consideración para no sobrevalorar en la política de adquisiciones sectorial de la informática la importancia de la revista científica.

La bibliografía citada muestra una elevada obsolescencia. De todas formas, al haberse tomado tesis doctorales como fuente para el análisis de citas, se han recogido menciones a documentos de una cierta antigüedad ya que una de las características de la investigación llevada a cabo por los doctorandos acostumbra a ser la revisión de la bibliografía sobre un tema considerando sus fundamentos y los primeros desarrollos sobre el mismo.

El método utilizado no se ha demostrado válido en la determinación del uso de documentos electrónicos ya que, en muchas ocasiones, las referencias no se han realizado consignando la naturaleza del soporte. Sin embargo, el volumen de citas explícitas a documentos electrónicos accesibles por Internet es muy destacable si tenemos en consideración el periodo cubierto (1996-1998) y tomando como referencia lo que es habitual en publicaciones científicas de un cierto nivel del ámbito de las ciencias "duras" y tecnológicas.

## 7 Referencias

Achilles, Alf-Christian (1999). *The collection of computer science bibliographies* [en línea]. [Alemania]: el autor, 1995–1999. <<http://iinwww.ira.uka.de/bibliography>> [consulta: 11 nov. 2000].

ACM [Association for Computing Machinery] (2000). *ACM digital library* [en línea]. New York: ACM. <<http://www.acm.org/dl>> [consulta: 13 nov. 2000].

*Anuario estadístico de las bibliotecas universitarias y científicas españolas 1999* [en línea]. (2000). Rebiun. Madrid: Universidad Autónoma de Madrid: CRUE (Conferencia de Rectores de las Universidades Españolas). 78 p. <<http://www.uma.es/rebiun/informeEstadistico99.pdf>> [consulta: 12 nov. 2000].

Baker, Sharon L.; Lancaster, F.W. (1991). *The measurement and evaluation of library services*. 2<sup>nd</sup> ed. Arlington, Virginia: Information Resources Press. XVIII, 411 p.

Bandyopadhyay, A.K. (1996). "Citation analysis of doctoral dissertations in mathematics using



- dBase 3 plus", *Annals of library science and documentation*, vol. 43, no. 3 (1996), p. 81–107.
- Barry, Christine A. (1999). "Las habilidades de información en un mundo electrónico: la formación investigadora de los estudiantes de doctorado", *Anales de documentación*, nº 2 (1999), p. 237–258. [Traducción del artículo publicado en 1997: "Information skills for an electronic world: training doctoral research students", *Journal of information science*, vol. 23, no. 3 (1997), p. 225–238.]
- Bowman, Michael. (1991). "Format citation patterns and their implications for collection development in research libraries", *Collection building*, vol.11, no. 1 (1991), p. 2–8.
- Broadus, Robert N. (1977). "The applications of citation analyses to library collection building", *Advances in librarianship*, no.7 (1977), p. 299–335.
- Broadus, Robert N. (1985a). "A proposed method for eliminating titles from periodical subscription lists" *College & research libraries*, vol. 46, no. 1 (1985), p. 30–35.
- Broadus, Robert N. (1985b). "On citations, uses, and informed guesswork: a response to Line", *College & research libraries*, vol. 46, no. 1 (1985), p. 38–39.
- Buchanan, Anne L.; Herubel, Jean Pierre V.M. (1994). "Profiling PhD dissertation bibliographies: serials collection development in political science", *Behavioral and social sciences librarian*, vol. 13, no. 1 (1994), p. 1–10.
- Cunningham, S.J.; Connaway, L.S. (1996). "Information searching preferences and practices of computer science researchers". En: *Sixth Australian Conference on Computer-Human Interaction*. [Australia]: IEEE. p. 294–299.
- Deshpande, M.; Rajyalakshmi, D. (1997) "Citation study of dissertations in library and information science", *Annals of library science and documentation*, vol. 44, no. 2 (1997), p. 41–53.
- Devin, Robin B.; Kellogg, Martha (1990). "The serial/monograph ratio in research libraries: budgeting in light of citation studies", *College & research libraries*, vol. 51, no. 1 (1990), p. 46–54.
- Estivill, Assumpció; Urbano, Cristóbal (1998). "Cómo citar recursos electrónicos: del uso a la norma y de la norma al uso: análisis de prácticas de citación en revistas españolas de biblioteconomía". En: *VI Jornadas Españolas de Documentación*. Valencia: Fesabid, p. 245–256.
- Garfield, E. (1968). "Decision making in libraries acquisitions: indexes or journals?", *Current contents*, (April 9<sup>th</sup> 1968).
- Glänzel, Wolfgang; Schoepflin, Urs (1999). "A bibliometric study of reference literature in the sciences and social sciences", *Information processing & management*, 35 (1999), p. 31–44.
- Harter, Stephen P.; Kim, Hak Joon (1996). *Electronic journals and scholarly communication* [en línea]: a citation and reference study. A paper delivered at the Midyear Meeting of the American Society for Information Science, San Diego, CA, May 20–22, 1996.  
<<http://php.indiana.edu/~harter/harter-asis96midtxt.html>> [consulta: 11 nov. 2000].
- Hughes, Janet (1995). "Use of faculty publication lists and ISI citation data to identify a core list of journals with local importance", *Library acquisitions: practice and theory*, vol.19, no. 4 (1995), p. 403–413.
- Hurd, Julie M.; Bleic, Deborah D.; Vishwanatham, Rama (1999). "Information use by molecular biologist: implications for library collections and services", *College & research libraries*, vol. 60, no. 1 (1999), p. 31–43.
- IEL online* [en línea]: *IEEE/IEE electronic library online*. (1999). Information Handling Services. <<http://iel.ihs.com>> [consulta: 11 nov. 2000].
- Jiménez Contreras, E.; et al. (1994). "Determinación de las colecciones básicas de publicaciones periódicas en hemerotecas científicas". En: *IV Jornadas españolas de documentación automatizada*. Oviedo: Universidad de Oviedo, p. 216–221.
- Joswick, Kathleen E.; Stierman, Jeanne K. (1997). "The core list mirage: a comparison of the journals frequently consulted by faculty and students", *College & research libraries*, vol. 58, no.

1 (1997), p. 48–55.

Kriz, Harry M. (1978). "Subscriptions vs. books in a constant dollar budget", *College & research libraries*, vol. 39, no. 2 (1978), p. 105–109.

Lancaster, F.W (1996). *Evaluación de la biblioteca*. Traducción por Ramón Abad y Belén Altuna. Madrid: ANABAD, 1996. 374 p. [Traducción de: *If you want to evaluate your library...* 2<sup>nd</sup> ed. (Londres: The Library Association, 1993)].

Lawrence, Steve; Giles, C. Lee; Bollacker, Kurt (1999a). "Digital libraries and Autonomous Citation Indexing", *IEEE computer*, vol.32, no. 6 (1999), p. 67–71.

Lawrence, Steve; Giles, C. Lee; Bollacker, Kurt (1999b). *CiteSeer* [en línea]: *autonomous citation indexing*. NEC Research Institute. <<http://www.neci.nec.com/~lawrence/citeseer.html>> [consulta: 11 nov. 2000].

Line, Maurice B. (1977). "On the irrelevance of citation analyses to practical librarianship". En: *EURIM II: an European conference on the application of research in information services in libraries*. London: Aslib, p. 51-53.

Line, Maurice B. (1985). "Use of citation data for periodicals control in libraries: a response to Broadus", *College & research libraries*, vol. 46, no. 1 (1985), p. 36–37.

Masip, Pere (1997). "L'ús de informació digital en ciències socials". En: *6es Jornades Catalanes de Documentació*. Barcelona: Socadi: Col·legi Oficial de Bibliotecaris-Documentalistes, p. 439–446.

McCain, Katherine W.; Bobick, James E. (1981). "Patterns of journal use in a departmental library: a citation analysis", *Journal of the American Society for Information Science*, vol. 32, no. 4 (1981), p. 257–267.

Mubeen, M. A. (1996). "Citation analysis of doctoral dissertations in chemistry", *Annals of library science and documentation*, vol. 43, no. 2 (1996), p. 48–58.

NCSTRL. (1999). *Networked Computer Science Technical Reference Library* [en línea]. NCSTRL. <<http://www.ncstrl.org>> [consulta: 11 nov. 2000]

Nederhof, A. J.; Van Raan, A.J. (1987). "Citation theory and the Ortega Hypothesis", *Scientometrics*, vol.12, no. 5–6 (1987), p. 325-328.

Nisonger, Thomas E. (1998). *Management of serials in libraries*. Englewood, Colorado: Libraries Unlimited. XXVIII, 433 p.

Olander, Anna Birgitta (1992). *Personal information management in computer science research*. Ph. D. thesis University of Toronto. 227 p.

Pan, E. (1978). "Journal citation as a predictor of journal usage in libraries", *Collection management*, vol. 2, no. 1 (1978), p. 29–38.

Poll, Roswitha; Boekhorst, Peter (1998). *Medición de la calidad: directrices internacionales para la medición del rendimiento en bibliotecas universitarias*. Madrid: ANABAD, 1998. 209 p.

Satariano, W. A. (1978). "Journal use in sociology: citation analysis versus readership patterns", *Library quarterly*, vol. 48, no. 3 (1978), p. 293–300.

Scales, P.A. (1976). "Citation analyses as indicators of the use of serials: a comparison of ranked title lists produced by citation counting and from use data", *Journal of documentation*, vol.32, no. 1 (1976), p. 17–25.

Smith, Linda C. (1981). "Citation analysis", *Library trends*, vol. 30, no. 1 (1981), p. 83–106.

Sylvia, Margaret; Leshner, Marcella (1995). "What journals do psychology graduate students need?: a citation analysis of thesis references", *College & research libraries*, vol. 56, no. 4 (1995), p. 313–318.

Takayama, Masaya (1986). "Analysis of technological information transfer among Japanese computer scientists at a research front", *Information services & use*, vol. 6, no. 1 (1986), p. 9–25.

Terrada, María Luz (1994). "Documentos, citas y referencias en la comunicación científica". En: Pérez-Álvarez Ossorio, J.R; Goedegebuure (ed.). *New worlds in information and documentation* Amsterdam: Elsevier, p. 89–97.

UPC. Vicerrectorat de Recerca (1997). *Publicacions notables* [en línea]. <<http://inopia.upc.es:80/sia/cat/butlleti-upc/documents/76-1997.html>> [consulta: 11 nov. 2000].

UPC. Vicerrectorat de Recerca (1999). "Sistema d'indicadors de l'activitat de recerca de la UPC." Documento accesible en línea en la aplicación *Fenix* de producción científica de la UPC.

Urbano, Cristóbal (1997). "Information seeking behaviour of doctoral students in Barcelona since the Internet explosion: groundwork for a project of research". En: *5<sup>th</sup> International Symposium BOBCATSSS 1997, Budapest*. Amsterdam: Hogeschool van Amsterdam.

Urbano, Cristóbal (2000). *El Análisis de citas en publicaciones de usuarios de bibliotecas universitarias [en línea]: estudio de las tesis doctorales en informática de la Universidad Politècnica de Catalunya, 1996-1998*. Director: Ernest Abadal i Falgueras. 529 p. Tesis doctoral, Universitat de Barcelona, 2000. <<http://www.tdcat.cesca.es/TDCat-0706100-091717>> [consulta: 11 nov. 2000].

Walcott, Rosalind (1991). "Characteristics of citations in geoscience doctoral dissertations accepted at United States academic institutions 1981-1985", *Science and technology libraries*, vol. 12, no. 2 (1991), p. 5–16.

Walcott, Rosalind (1994). "Local citation studies: a shortcut to local knowledge", *Science and technology libraries*, vol. 14, no. 3 (1994), p. 1–14.

Zipp, Louise S. (1996). "Thesis and dissertation citations as indicators of faculty research use of university library journal collections", *Library resources and technical services*, vol. 40, no. 4 (1996), p. 335–342.

## Notas

<sup>1</sup> Se hace referencia al uso por parte del usuario, y no de la colección: el análisis de citas como método de evaluación del uso de la colección está más discutido en comparación con otros sistemas de medición (Baker y Lancaster, 1991).

<sup>2</sup> Así se pueden diluir algunos de los sesgos y distorsiones. Nederhof y Van Raan (1987) en este sentido afirman que si se consideran grandes cantidades de documentos fuente de forma agregada, las influencias omitidas y las citadas pero no reales tienden a diluirse y a contrarrestarse.

<sup>3</sup> Nisonger (1998, p. 145) subraya que los diversos estudios dedicados a establecer la ratio publicaciones seriadadas/monografías no parten de una definición uniforme de lo que es una publicación seriada.

<sup>4</sup> La informática como disciplina en términos bibliométricos y los informáticos como usuarios de bibliotecas han recibido poca atención tanto a nivel español como en el ámbito internacional (Cunningham y Connaway, 1996; Urbano 2000).

<sup>5</sup> Se traduce *to browse* por *hojear*, si bien en los diferentes contextos en los que se puede utilizar la expresión inglesa se podría ajustar su traducción a un término específico: *explorar, curiosar, leer superficialmente, buscar aleatoriamente*, etc.

<sup>6</sup> Antes de comenzar el recuento de citas se construyó una base de datos relacional en la que se volcaron datos del Journal Citation Report del ISI, y en la que se introdujeron los congresos notables que figuraban en la documentación del Programa de Publicaciones Notables de la UPC.

<sup>7</sup> Como ejemplo del dominio de las citas a revistas en otros estudios se pueden mencionar los siguientes porcentajes de revistas citadas: 80,01% en matemáticas (Bandyopadhyay, 1996), 91,3% en biología molecular (Hurd, Bleicic y Vishwanathan, 1999), 68,74% en biblioteconomía y documentación (Deshpande y Rajyalakshmi, 1997), 73% en química (Mubeen, 1996), o 62,2% en ingeniería electrónica (Glänzel y Schoepflin, 1999). Todos ellos contabilizaban de forma separada congresos y revistas, algo que no queda claro cuando se utiliza el concepto genérico de publicación seriada, que incluye a ambas categorías.

<sup>8</sup> Si bien el dato de las autocitas no se ha contabilizado, sí que se ha observado esta práctica hacia los trabajos publicados en congresos.

<sup>9</sup> Un primer análisis exploratorio realizado con una pequeña muestra, permite observar una elevada presencia de este tipo de documentos en las bibliotecas departamentales de la UPC.

<sup>10</sup> El volumen de autocitas a trabajos propios o de otros miembros del departamento o grupo de investigación se concentra fundamentalmente en informes de investigación y en comunicaciones de congresos, hecho que hace suponer como cierta la afirmación en relación a las colecciones personales.

<sup>11</sup> A fecha de 14 de diciembre de 1999.

<sup>12</sup> En el estudio llevado a cabo por Harter y Kim (1996b) se identificaron en el directorio de revistas electrónicas de la Association of Research Libraries un total de 131 revistas científicas electrónicas que disponían de sistema *peer review*, de las cuales 28 pertenecían a las ciencias, 34 a las ciencias sociales, 31 a las humanidades y 38 a una categoría miscelánea. Educación y literatura con 13 y 12 eran las áreas concretas con más títulos. Dicho directorio en las fechas del estudio recogía principalmente revistas electrónicas de nueva creación y de acceso libre.

<sup>13</sup> La *IEL* en cd-rom es una base de datos mensual en texto completo e imagen que reúne las publicaciones (revistas, congresos y normas) del IEEE (Institute of Electrical and Electronics Engineers) y de la IEE (Institution of Electrical Engineers). Contempla tanto las publicaciones en curso como las retrospectivas desde 1988. La Biblioteca Rector Gabriel Ferraté de la UPC dispone de una colección que cubre desde 1991 hasta la actualidad. Actualmente también se comercializa la versión web (*IEL online...* 1999).

<sup>14</sup> Por medio del análisis de sedes web de distintos departamentos universitarios del área, se ha observado que 1990 podría ser una fecha frontera convencional a partir de la cual buena parte de los informes de investigación y preprints son disponibles en formato electrónico.

<sup>15</sup> Las bibliotecas de la UPC parece que ha optado, de momento, por la creación de una base de datos independiente del catálogo para controlar los recursos Internet (Beumala et al., 1998).

<sup>16</sup> Los números de abril de 1995 y de 1998 de la revista *Communications of the ACM* y el número de febrero de 1999 de *IEEE computer* son monográficos sobre las bibliotecas digitales, especialmente en el área de informática. Son muchas las páginas web de bibliotecas o páginas web personales dedicadas a repertoriar selectivamente este tipo de recursos; una selección de ellas se puede encontrar en Achilles (1999).