

Vínculos entre las Ontologías y la Biblioteconomía y Documentación

Vicente Guerrero Bote, Adolfo Lozano Tello
Facultad de Biblioteconomía y Documentación. Universidad de Extremadura.
{vicente, aloztel}@alcazaba.unex.es

Resumen

Las ontologías proporcionan un método más rico de organizar el conocimiento, permitiendo realizar deducciones mediante el uso de axiomas entre los conceptos formalizados. El conocimiento formalizado en ontologías tiene como característica fundamental su disposición para ser reutilizado, por lo que inducirán a un gran avance en el entendimiento de diversos sistemas de representación del conocimiento y llevará a fomentar el desarrollo de aplicaciones basados en estos sistemas. El artículo da una visión general de los sistemas basados en conocimientos, define los conceptos generales de ontologías y apunta las posibilidades que pueden ofrecer las ontologías en el campo de la Biblioteconomía y Documentación.

Palabras Clave: Ontologías, representación del conocimiento, reutilización del conocimiento.

1 Introducción

La palabra conocimiento tiene varias acepciones; en Biblioteconomía y Documentación (ByD) solemos utilizarla para referirnos a ese tercer mundo de Popper [1] del “conocimiento objetivo” recogido en los documentos. En ese sentido hablamos de la organización del conocimiento refiriéndonos a esa organización documental. Por otro lado, también es la noción del entorno, la ciencia o la sabiduría que nos permite realizar acciones intelectuales y que Brookes [2, 3, 4, 5] cuando tratando de modelar la interacción entre los tres mundos de Popper, lanza su ecuación fundamental del conocimiento, lo considera formado por una estructura de conceptos y las relaciones entre ellos. Ciertamente no se trata simplemente de un depósito de datos, ni siquiera de noticias, informaciones o conceptos, como dice H. Poincaré, “un montón de noticias, informaciones o datos no son conocimientos, del mismo modo que un montón de ladrillos y cemento no es un edificio”. Aludiendo a una frase típica actual, se trata más bien de un “amueblamiento”, donde no solo se almacenan objetos, sino que éstos se colocan de una determinada forma en relación a los demás. Para la Inteligencia Artificial, el concepto “conocimiento” es el soporte de los sistemas basados en conocimientos que se almacenan en forma de: definiciones descriptivas de términos específicos del dominio, de descripciones de objetos individuales del dominio y sus relaciones con otros objetos y con criterios para tomar decisiones.

En los años cuarenta cuando se comienzan a investigar sobre este tipo de sistemas, los problemas que se trataban de resolver eran suficientemente sencillos como para resolverse explorando todas las posibilidades, todo el espacio-problema, automatizadamente. Durante esta época se trabajó con toda una serie de algoritmos de exploración y búsqueda. No obstante, cuando se trataron de solucionar problemas cercanos a la realidad, estos métodos resultaron ser demasiado débiles ante la explosión combinatoria de la misma. Sin embargo, los humanos tenemos aprendidas ciertas reglas, ciertos conocimientos que nos permiten no tener que considerar todas las posibilidades. Su emulación fomentó la creación de los métodos fuertes que no son tan dependientes de la explosión combinatoria. El DENDRAL [6] o el MYCIN (que con un respaldo en las primeras versiones de 450 reglas, era capaz de hacer diagnósticos médicos parecidos a los de un experto) fueron los primeros sistemas que lo emplearon.

Esto dio lugar a los sistemas basados en conocimientos que suelen tener una estructura formada por los siguientes elementos:

- a) La Base de Conocimientos, es el esquema arquitectónico que contiene los datos (o hechos) y reglas (o heurísticas) declarados explícitamente.
- b) El Motor de Inferencias, es el programa que toma la decisión de aplicar las reglas acerca del problema que se intenta resolver.
- c) La Interfaz de Entrada/Salida mediante la que se comunican con el usuario para transmitirle resultados o captar nuevos conocimientos.

Si nos fijamos, esta estructura no dista mucho de la propuesta por Menou [7] intentando

representar la resolución (humana) de problemas utilizando la información. Este último indica que el motor de inferencias puede verse influido por la personalidad, experiencia, cultura, etc., que coincide incluso con algunas últimas tendencias, según las cuales se especifica en la misma base de conocimiento la forma de llevar a cabo las inferencias.

Con este tipo de sistemas se ha tratado también de rivalizar con los expertos humanos [8]; Smith [9] dice que:

"...preservan el conocimiento perecedero de los expertos humanos; distribuyen el que sea escaso; reduce costes derivados del pobre y mediocre rendimiento humano; y ayudan a los humanos intentando acceder a información y utilizando ordenadores..." (p. 51, se refería al caso concreto de los aplicados a la recuperación de la información)

Para ser adecuado para el empleo de este tipo de sistemas, un dominio debe tener claramente circunscritas las fronteras y contener un número limitado de entidades y relaciones [10, 11]. En ese sentido, a pesar de la existencia de mucha literatura sobre Sistemas Expertos en ByD, uno puede preguntarse si es apropiado dicho dominio para la aplicación de esta tecnología. Con respecto a sólo un subdominio de la ByD, la recuperación de la información, Brooks [12] afirmó:

"El mundo de la R. I. inteligente resulta ser muy amplio, incluyendo como deben ser usuarios y subproblemas, documentos y sus descripciones, y heurísticas de recuperación. En otras palabras el dominio no es reducido, ni discreto, ni homogéneo y contiene una gran cantidad de objetos y relaciones."

Sparck-Jones [13] manifiesta estar igualmente de acuerdo. Así no sorprende que ningún sistema basado en conocimientos cubra el rango de conocimientos y experiencia que un profesional de la ByD debería poseer. Entre los 139 contabilizados por Poulter et al. [14], que se han aplicado a este dominio, el 80% de los proyectos estaban destinados a subdominios de la ByD. Estos se pueden desglosar en un 32% para trabajos de referencia, un 28% para recuperación en línea, un 12% para indización, catalogación o clasificación, 7% para aplicaciones de gestión de bibliotecas, y el 1% para resumen. Dentro de cada subdominio, los sistemas tienden a restringirse a otros incluso más limitados para evitar problemas asociados con límites mal definidos, o con un gran número de entidades y relaciones.

Para la creación de sistemas basados en conocimientos se necesita almacenar esas reglas o ese conocimiento de forma que se pueda razonar con ellas, realizando la adquisición de conocimientos de diversas fuentes (expertos, manuales, publicaciones, etc.). Para la representación del mundo que se quiere mantener con un cierto nivel de detalle no basta con la lógica de proposiciones, sino que se utiliza la de predicados. Sin embargo, resulta difícil mediante la lógica tratar relaciones complejas, representar lo expresado mediante el lenguaje natural, y pasar de lo general a lo particular. Uno de los modelos que surgieron para aliviar el problema son las redes asociativas, dentro de las cuales podemos encontrar las redes semánticas, especialmente aptas para el proceso del lenguaje natural, o las redes bayesianas, para representar las relaciones causales, muy utilizadas en la diagnosis médica.

El último de los modelos surgidos es el de los marcos de Minsky [15], del inglés *frames*, con una larga tradición de uso en recuperación de la Información [16, 17]. En ellos se representan los distintos objetos que forman parte del dominio, sus propiedades y sus atributos. De forma similar han surgido los guiones para representar situaciones determinadas.

2 Definición y tipos de ontologías

La creación de estos sistemas basados en conocimientos depende demasiado de la informal recolección de conocimiento humano y de su representación en esquemas adecuados para el procesamiento simbólico [18]. Esto, unido a que a finales de los ochenta estaba aceptado que era necesaria una conceptualización del conocimiento del dominio [19] hizo que adquiriera gran notoriedad el concepto de ontología que ya utilizó Sowa en [10].

2.1 Definición de ontología.

El término ontología viene del campo de la filosofía y se define como la rama de la filosofía que se ocupa de la naturaleza y organización de la realidad. En el campo de la Inteligencia Artificial, existen diferentes autores que tienen varias definiciones declarativas y procedimentales de ontologías [20].

La definición declarativa más consolidada es la que describe una ontología como una especificación explícita de una conceptualización [21], es decir, que proporciona una estructura y contenidos de forma explícita que codifica las reglas implícitas de una parte de la realidad;

estas declaraciones explícitas son independientes del fin y del dominio de la aplicación en el que se usarán o reutilizarán sus definiciones.

Neches propuso la definición procedimental de ontología [22]: Una ontología define el vocabulario de un área mediante un conjunto de términos básicos y relaciones entre dichos términos, así como las reglas que combinan términos y relaciones que amplían las definiciones dadas en el vocabulario. En una ontología, además de las definiciones explícitas, existirán definiciones implícitas que podrán ser inferidas por el uso de axiomas.

2.2 Diferencias entre bases de conocimientos y ontologías.

Aunque bases de conocimientos y ontologías tienen en común en que ambas contienen conocimiento, existen varias diferencias entre ellas [23]:

Las características del lenguaje usado para codificar el conocimiento: las ontologías deben ser escritas en un lenguaje expresivo, declarativo, portable, independiente del dominio y legible por la máquina de forma bien definido semánticamente. No se puede asegurar que una base de conocimientos esté codificada con estas propiedades. Los lenguajes más usuales para implementar ontologías son Ontolingua [21], CycL [24] y LOOM [25].

El propósito de la codificación de su conocimiento: las ontologías son diseñadas con suficiente abstracción y generalidad como para compartir y reutilizar el conocimiento, pero las bases de conocimientos no.

La especificación de los requerimientos: la reutilización de las ontologías debe poder hacerse independientemente del comportamiento y dominio de la aplicación que la reutilice. Por tanto, los ontólogos deben especificar qué se espera de la ontología para un dominio dado.

De cualquier forma, tanto ontologías como bases de conocimientos tienen un problema base común, y es que es imposible capturar todo lo conocido acerca del mundo en una estructura finita. Así, uno de los más importantes problemas es garantizar la completitud, consistencia y concisión del conocimiento en los procesos de desarrollo de los sistemas.

2.3 Tipos de ontologías.

Van Heist [26], clasifica las ontologías de acuerdo a la cantidad y tipo de estructura de la conceptualización en:

- ◆ Ontologías terminológicas: especifican los términos que son usados para representar conocimiento en el universo del discurso. Suelen ser usadas para unificar vocabulario en un campo determinado.
- ◆ Ontologías de información: especifican la estructura de almacenamiento de bases de datos. Ofrecen un marco para el almacenamiento estandarizado de información.
- ◆ Ontologías de modelado del conocimiento: especifican conceptualizaciones del conocimiento. Contienen una rica estructura interna y suelen estar ajustadas al uso particular del conocimiento que describen.

En documentación se pueden considerar de especial interés la aplicación por un lado ontologías terminológicas que unificasen la terminología de cada concepto y las relaciones entre ellos; y por otro, las ontologías de información para unificar las estructuras de almacenamiento de forma que pudieran ser reutilizadas por varias aplicaciones informáticas que utilicen la misma fuente de información.

2.4 Uso de estándares ontológicos.

El conocimiento formalizado en ontologías tiene como característica fundamental su disposición para ser reutilizado. Mediante el uso de ontologías todos los problemas de formalización heterogénea del conocimiento son evitados, de hecho permiten el intercambio de conocimiento mediante lenguajes de representación del conocimiento que funcionan como interlenguas (o lenguajes de intercambio). Uno de los más extendidos es KIF (Knowledge Interchange Format) [27] desarrollado por el programa *ARPA Knowledge Sharing Effort* [28] por el *Interlingua Working Group*. Mediante la formalización del conocimiento realizada en KIF cualquier usuario puede reutilizar la información recopilada por otros. KIF es una versión prefija del cálculo de predicados de primer orden, con extensiones para mejorar su expresividad, con suficiente poder expresivo para representar el conocimiento declarativo contenido en bases de conocimientos de sistemas de información. Además, permite una traducción semiautomática desde otros lenguajes de representación que facilitaría el traspaso de información desde otras formas de representación a KIF.

Del mismo modo, existen diferentes protocolos de comunicación como por ejemplo KQML (*Knowledge Query Manipulation Language*) [28] que permiten a agentes interactuar haciendo consultas entre diferentes sistemas cooperativos que tienen como base ontologías.

3 Papel de las ontologías en los sistemas de representación del conocimiento.

El conocimiento documental tradicionalmente ha sido organizado utilizando mecanismos básicos de clasificación jerárquica de Dewey (o la CDU) y los encabezamientos de materia (clasificación asociativa) sobre los que también se han establecido un gran número de acuerdos. Actualmente, más orientado a la recuperación, se utilizan descriptores de un lenguaje controlado con el soporte de un tesoro. Dichos modos de almacenamiento y recuperación de información han jugado un papel fundamental en el uso generalizado y estandarizado de sistemas documentales permitiendo a los documentalistas tener unos modelos comunes y aceptados en la descripción y almacenamiento de información.

En los tesauros los descriptores se enlazan mediante tres tipos principales de relaciones muy simples formando un árbol taxonómico del sistema de información. Dichas relaciones no sólo se utilizan a la hora de la recuperación, sino que permiten tener una idea general del contenido de la base o navegar a través de los mismos descriptores.

La construcción de ontologías lleva implícito que cada término y cada relación entre términos se defina formalmente. Los conceptos se describen explícitamente para entender su significado, mediante acuerdos ontológicos. Con ello, un usuario que desee reutilizar una ontología desarrollada por otros, puede conseguir la información de todos los conceptos que soporta, su taxonomía y los axiomas. A diferencia de los tesauros, las ontologías poseen una mayor variedad de relaciones entre conceptos. Estas relaciones se crean a propósito del modelo conceptual existente en el dominio que se está formalizando. Por ejemplo, en la figura 1 se muestra algunas relaciones en el campo de la biología: entre los conceptos *Mamífero*, *Perro*, *Carnívoro*, y *Bulldog* pueden existir (entre otras) las relaciones *Es_de_raza*, *Es_de_especie* y *Tipo_de_alimentación* de forma que se puede realizar preguntas como “Dame razas de especies de mamíferos con alimentación carnívora” e inferir la respuesta a través del conocimiento contenido en la ontología. Como vemos, esta forma de representación del conocimiento permite hacer deducciones más complejas sobre el árbol taxonómico. Todo ello capacita a las ontologías ser utilizadas en sistemas documentales permitiendo mayores posibilidades de recuperación de la información.

De igual forma, el conocimiento almacenado en un sistema documental se puede organizar definiendo cada concepto (como: *Revista*, *Artículo*, *Libro*, *Autor*, etc.), y las relaciones que existen entre ellos de forma explícita (como: *Escrito_por*, *Aparece_en*, etc.). Con ello se ganará el establecimiento de un consenso tanto en el significado de cada término, como en una unificación formal del almacenamiento de esa información. Las ontologías se construyen siguiendo esta filosofía, por lo que pueden ser reutilizadas en diferentes dominios y con diferentes fines. Poniendo axiomas del tipo: si A es coautor de B y sabes que B escribió el libro L1 y L2, se sabe que A escribió alguno de estos dos libros, es decir, se pueden deducir conocimientos que no están explícitos en la ontología.

Un ejemplo que puede ser extrapolado al caso de reutilización de un sistema documental, es

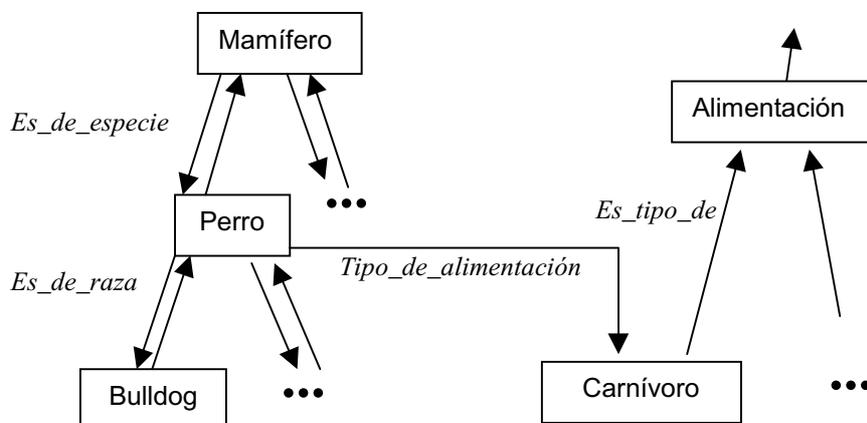


Figura 1. Ejemplo de relaciones entre conceptos.

la iniciativa del (KA)² [29], que modeliza a la Comunidad de Adquisición de Conocimientos (sus

investigadores, temas de investigación, publicaciones, etc.) en una ontología llamada $(KA)^2$ *Ontology* accesible a toda la comunidad a través del Ontology Server¹ [30]. En la figura 2 se muestran el árbol de clasificación de conceptos y las relaciones binarias “Ad-hoc” más representativas de $(KA)^2$. Arpírez y colegas [31] realizaron una extensión de la $(KA)^2$ *Ontology* para incorporarla a su ontología *Reference Ontology* que sirve como unas páginas amarillas para encontrar ontologías.

De modo análogo, con cualquier sistema documental formalizado mediante una ontología, se puede realizar la misma labor extendiéndolo o integrándolo para otros fines. Como ya se ha mencionado, todo esto es posible porque las ontologías describen de manera formal y explícita cada término y cada relación que formalizan.

Este último ejemplo podemos decir que encaja totalmente con las tareas tradicionales de la ByD. Como ya se ha dicho, las ontologías almacenan conocimiento en un formato que puede ser utilizado por sistemas automáticos para realizar deducciones y, además, su propósito de reutilización exige que sus conceptos estén descritos explícitamente, de modo que también son comprensibles y asimilables por el entendimiento humano. Además, conservan el conocimiento percedero de los expertos. Es decir, son un tipo particular de documento, no muy numeroso en la actualidad con relación a los sistemas documentales, pero que va adquiriendo cada vez más importancia. Por este motivo su almacenamiento y recuperación para la satisfacción de

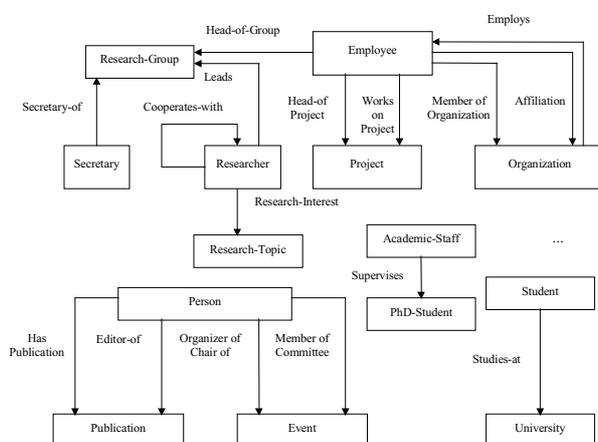


Figura 2. Diagrama de relaciones binarias “ad-hoc” en $(KA)^2$. necesidades de información le atañen en gran medida a esta disciplina.

4 Conclusiones

Acostumbrados a referirnos al conocimiento documental, las Ontologías permiten expresarlo en forma de conceptos y axiomas de modo que es manipulable por sistemas automáticos (incluso para hacer inferencias con el conocimiento almacenado). Se diferencian del resto de las bases de conocimientos (entre otras cualidades) por el objetivo de su construcción, que tiene como finalidad la reutilización y la interoperabilidad entre distintos sistemas.

Esto hace que la relación con nuestra disciplina se base en que:

- ♦ Se han aplicado sistemas basados en conocimientos, el desarrollo e interoperabilidad de los cuales se pueden optimizar si utilizan ontologías en lugar de bases de conocimiento particulares. Además, permiten preservar el conocimiento percedero de los expertos en cualquier campo de aplicación.
- ♦ Al describir los conceptos y sus relaciones, pueden ser empleadas de manera más general que los actuales tesauros, pudiendo establecer un mayor número de relaciones distintas (y tener definidas cada una de ellas).
- ♦ Visto que son un nuevo tipo de documento, se debe estudiar su almacenamiento y recuperación para la satisfacción de las necesidades que se puedan plantear.

¹ Accesible a través de la URL: <http://www-ksl.stanford.edu:5915>

5 Referencias.

- [1] Popper, K. El conocimiento objetivo. Madrid: Tecnos, 1992.
- [2] Brookes, B. C. The foundations of information science, part 1: philosophical aspects. *Journal of Information Science*, 2(3-4), 1980, 125-133.
- [3] Brookes, B. C. The foundations of information science, part 2: quantitative aspects: classes of things and the challenge of human individuality. *Journal of Information Science*, 2(5), 1980, 209-221.
- [4] Brookes, B. C. The foundations of information science. Part 3. Quantitative aspects: objective maps and subjective landscapes. *Journal of Information Science*, 2(6), 1980, 269-74.
- [5] Brookes, B. C. The foundations of information science. Part IV. Information science: the changing paradigm. *Journal of Information Science*, 3(1), 1981, 3-12.
- [6] Buchanan, B. G.; Sutherland, G. L. y Feigenbaum, E. A. Heuristic DENDRAL: a program for generating explanatory hypotheses in organic chemistry. En: Meltzer, B.; Michie, D. y Swann, M., ed. *Machine Intelligence 4*. Edinburgh, Scotland: Edinburgh University Press, 1969, 209-254.
- [7] Menou, M. J. The impact of Information-III Concepts of Information and its value. *Information Processing and Management*, 31(4), 1995, 479-90.
- [8] Wecker, J. y Cooper, C. Artificial Intelligence, Expert Systems and Librarianship: A Review of the Literature. *Australian Library Review*, 4(4), 1990, 281-299.
- [9] Smith, L. C. Artificial Intelligence and Information Retrieval. *Annual Review of Information Science and Technology*, 22, 1987, 41-77.
- [10] Sowa, J. F. *Conceptual structures: Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley, 1984.
- [11] Welbank, M. *A Review of Knowledge Acquisition Techniques for Expert Systems*. Martlesham Heath, England: Martlesham Consultancy Services, 1983.
- [12] Brooks, H. Expert Systems and Intelligent Information Retrieval. *Information Processing and Management*, 23(4), 1987, 367-382.
- [13] Sparck Jones, K. Information Retrieval. En: Shapiro, S.C., ed. *Encyclopedia of Artificial Intelligence*. New York, NY: Wiley, 1, 1987, 419-421.
- [14] Poulter, A.; Morris, A. y Dow, J. LIS Professionals as Knowledge Engineers. *Annual Review of Information Science and Technology*, 29, 1994, 305-350.
- [15] Minsky, M. "A Framework for Representing Knowledge" *The Psychology of Computer Vision*. De. P, Winston. McGraw-Hill, NY 1975.211-277.
- [16] Green, R. *Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference*, Frankfurt: Indeks Verlag, 1996.
- [17] Vickery, B. C. Conceptual relations in information systems. *Journal of Documentation* 52, 1996, 198-200.
- [18] Alexander, J. H. et al. Knowledge level engineering: ontological analysis. *Proceedings of AAAI 86*, 1986, 963-968.
- [19] Scott A. C. et al. *Practical Guide to Knowledge Acquisition*. Addison-Wesley, 1991.
- [20] Guarino N. and Giaretta P., "Ontologies and Knowledge Bases: Towards a Terminological Clarification", *Toward Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing*. IOS Press, 1995, pp.25-32.
- [21] Gruber T., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", *Technical Report KSL-93-04*, Knowledge Systems Laboratory, Stanford University, CA, 1993.
- [22] Neches R., Fikes R., Finin T., Gruber T., Patil R., Senator T., Swartout W., "Enabling Technology for Knowledge Sharing", *AI Magazine*, fall 1991, pp. 36-56 .
- [23] Gómez-Pérez A., "Knowledge Sharing and Reuse, *The Handbook of Applied Expert Systems*", Ed. J. Liebowitz, CRC Press, 1998.
- [24] Lenat, D.B., Guha, R.V., "Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project", Addison-Wesley Publishing Company, Inc. CA. 1990.
- [25] MacGregor R., "The Evolving Technology of Classification-based Knowledge Representation System". In J. Sowa Ed. *Principles of Semantic Networks: Explorations in the Representation of Knowledge*. San Mateo, CA. Morgan Kaufman, 1991.
- [26] Van Heist G., Schreiber A., Wielinga B., "Using Explicit Ontologies in KBS Development", *Int. J. Human-Computer Studies* , 1997, vol 45, pp 183-292.
- [27] Genesereth M., and Fikes R., "Knowledge Interchange Format. V3.0. Reference Manual." Report Logic-92-1. Computer Science Department. Stanford University. CA. 1992.
- [28] Finin T. and Wiederhold G. "An Overview of KQML: A Knowledge Query and Manipulation Language" Department of Computer Science. Stanford University. CA. 1991.
- [29] Benjamins R. and Fensel D., "Community is Knowledge! in (KA)²", *Knowledge Acquisition Workshop, KAW98*. Presented in FOIS 1998.
- [30] Faquhar A., Fikes R., Pratt W. and Rice J., "Collaborative Ontology Construction for Information Integration" *Technical Report KSL-95-69*, Knowledge Systems Laboratory, Stanford University, Stanford (USA) 1992.
- [31] Arpírez J., Gómez-Pérez A., Lozano A. and Pinto S. "(ONTO)2Agent: An ontology-based WWW broker to select ontologies", *Workshop on Applications of Ontologies and Problem-Solving Methods. ECAI'98*. Brighton, UK. pp.16-24.