

Prueba de Concepto de Expansión de Consultas basada en Ontologías de Dominio Financiero

Proof of Concept of Ontology-based Query Expansion on Financial Domain

Julián Moreno Schneider

Grupo de Bases de Datos Avanzadas
Departamento de Informática
Universidad Carlos III de Madrid
Avda. Universidad, 30, 28911
Leganés, Madrid, España
jmschnei@inf.uc3m.es

Thierry Declerck

Language Technology Lab
German Research Center for Artificial
Intelligence, DFKI GmbH
Stuhlsatzenhausweg, 3, 66123
Saarbrücken, Alemania
declerck@dfki.de

José Luis Martínez Fernández

DAEDALUS – Data, Decisions and
Language S.A.
Avda. de la Albufera 321, 28031
Madrid, Spain
jmartinez@daedalus.es

Paloma Martínez Fernández

Grupo de Bases de Datos Avanzadas
Departamento de Informática
Universidad Carlos III de Madrid
Avda. Universidad, 30, 28911
Leganés, Madrid, España
pmf@inf.uc3m.es

Resumen: Este trabajo presenta el uso de una ontología en el dominio financiero para la expansión de consultas con el fin de mejorar los resultados de un sistema de recuperación de información (RI) financiera. Este sistema está compuesto por una ontología y un índice de Lucene que permite recuperación de conceptos identificados mediante procesamiento de lenguaje natural. Se ha llevado a cabo una evaluación con un conjunto limitado de consultas y los resultados indican que la ambigüedad sigue siendo un problema al expandir la consulta. En ocasiones, la elección de las entidades adecuadas a la hora de expandir las consultas (filtrando por sector, empresa, etc.) permite resolver esa ambigüedad.

Palabras clave: Búsqueda Semántica, Recuperación de Información, Ontología, Expansión de Consulta

Abstract: This paper explains the application of ontologies in financial domains to a query expansion process. The final goal is to improve financial information retrieval effectiveness. The system is composed of an ontology and a Lucene index that stores and retrieves natural language concepts. An initial evaluation with a limited number of queries has been performed. Obtained results show that ambiguity remains a problem when expanding a query. The filtering of entities in the expansion process by selecting only companies or references to markets helps in the reduction of ambiguity.

Keywords: Semantic Search, Information Retrieval, Ontology, Query Expansion

1 Introducción

La cantidad de información semántica existente en formato electrónico es cada vez mayor. Cualquier consulta que se realice para acceder a estos vastos volúmenes de información semántica proveerá grandes cantidades de resultados, imposibles de procesar

manualmente. Por otro lado, si dicha consulta no se ha planteado adecuadamente, la respuesta puede no ser la esperada. Esto nos lleva a pensar que la utilización de técnicas externas que ayuden al usuario a concretar o enriquecer la consulta realizada pueden ayudar a gestionar grandes cantidades de información.

En este sentido, la utilización de elementos que aporten información semántica a la consulta pueden ayudar a completar su formulación para obtener mayor cantidad de información a la vez que más diversa. Incluso puede llegar a resolver problemas de ambigüedad Martínez-Fernández et al. (2008).

El dominio contemplado en esta investigación es el financiero, concretamente noticias sobre economía e información de mercados bursátiles. Este dominio presenta la particularidad de que es altamente cambiante, lo que complica la recuperación de información semántica y requiere además que la información recuperada sea actual y precisa si se va a utilizar en procesos de toma de decisiones.

Como ejemplo intuitivo podríamos mencionar a una persona que quiere invertir en bolsa y ha oído que Inditex es una buena opción. Si el usuario busca el nombre de esta empresa en cualquier sistema de RI obtendrá documentos en los que se mencione la propia empresa. Aunque esta puede ser información interesante para el usuario, está claro que no es suficiente. El interés que tiene el usuario en esta empresa podría extenderse a sus filiales, empresas asociadas, información relacionada con sus directivos o con empresas del mismo sector.

El usuario no tiene todos esos datos antes de realizar la consulta, por lo que ayudarle a enriquecer su consulta puede ser una buena forma de mejorar la calidad del proceso de recuperación.

2 Trabajo Relacionado

La expansión de consultas es un tema que ha sido ampliamente estudiado en diversos dominios como describen Bhogal, Macfarlane y Smith (2007). Carstens (2011) describe tres modos de realizar la expansión de una consulta (EC): manual, automático e interactivo. La EC manual es aquella que realiza el usuario directamente; en la EC interactiva es el sistema propone una expansión que el usuario debe confirmar, mientras que en la automática el sistema no necesita confirmación. Este último tipo de expansión es el que vamos a describir en mayor profundidad.

Carpineto et al. (2012) realizan un estudio en profundidad de la expansión automática de consultas. Mencionan aplicaciones como la búsqueda de respuestas (en inglés, *question answering*) o la

recuperación de información en entornos multilingües y describen las técnicas empleadas en la expansión, desde la utilización de ontologías hasta el análisis de *logs* de consultas.

Dentro de la investigación de expansión automática, se debe distinguir entre trabajos que realizan la expansión de la consulta con información obtenida de la colección de documentos y aquellos que expanden la consulta con información externa.

Enmarcado en el primer tipo, Dragoni, Pereira y Tettamanzi (2012) describen un sistema de RI basado en ontologías, pero sin hacer una expansión de la consulta. En su lugar plantean la extracción de los conceptos de la ontología presentes en la consulta y en los documentos y calculan la concordancia en función de estos conceptos. Wollersheim y Rahayu (2005) presentan un trabajo similar

En lo que se refiere a trabajos independientes de la colección, la propuesta de Díaz-Galiano, Martín-Valdivia y Ureña-López (2009) utiliza una ontología para expandir consultas en el dominio biomédico y trabaja sobre un sistema multimodal. La ontología ayuda a expandir la consulta con términos relacionados que aparecen en otros documentos médicos. Farhoodi et al. (2009) describen un trabajo similar, en el que la expansión se realiza utilizando una ontología basada en Wikipedia. Intentan focalizar las necesidades de información del usuario haciendo desambiguación a la vez que expanden la consulta. Utilizan un sistema de vectores para expandir la consulta si algún término de esta se encuentra en la ontología. Utilizan Google como sistema de RI para evaluar la expansión.

Otros trabajos que realizan expansión de la consulta son: Schweighofer y Geist (2007) trabajan en el dominio legal y utilizan una ontología para adecuar las consultas a la terminología legal. Emplean la ontología para añadir términos semánticamente similares; Wu, Wu y Fu (2007) procesan términos de una consulta en función de un razonador lógico y de la frecuencia de aparición de los términos, para posteriormente expandirlos contra una ontología; y, por su parte, Liu y Li (2011) hacen expansión de consultas basándose en una ontología de tres capas.

Mandala, Tokunaga y Tanaka (2000) realizan una expansión de consulta basada en tesauros, utilizando una combinación de la relevancia de cada término de la consulta en

cada tesoro del que disponen para determinar si el término se expandirá o no.

El trabajo más cercano al presentado en este artículo es el de Dey et al (2005). En él se utilizan ontologías para mejorar las expresiones de búsqueda, seleccionando primero los términos de la consulta que son adecuados para ser expandidos y posteriormente realizando la expansión utilizando la ontología.

El trabajo de Fu, Jones y Abdelmoty (2005) realiza expansión de la información geográfica de la consulta en función de una ontología específica de dominio y una ontología geográfica. La expansión de la consulta en la recuperación de imágenes se analiza y se describe en el trabajo (Gulati y Sharma, 2010). Las técnicas de expansión de consulta son similares pero la consulta posteriormente se aplica a recuperación de imágenes en lugar de textos.

El sistema (Tuominen et al, 2009) realiza una expansión de consulta textual basada en las ontologías publicadas en el servicio ONKI Ontology Service¹. Lo destacable de este sistema no es su funcionamiento sino que es un *widget* que se puede integrar en cualquier sistema, como páginas web o sistemas de recuperación. Ha sido evaluado con ontologías específicas de dominio y generales.

Para finalizar, en CLEF² también se han hecho muchos experimentos sobre expansión de consultas. En concreto, (Martínez-Fernández et al, 2008) concluye que si no se resuelve la ambigüedad de algún modo, la expansión no funciona (introduce más ruido) y empeora los resultados de precisión y cobertura.

En resumen, los trabajos actuales han utilizado ampliamente ontologías para realizar expansión de consultas y han demostrado ser una herramienta adecuada para ello. Por eso, este trabajo persigue explotar esta técnica para aplicarla al dominio financiero.

3 Propuesta

Este trabajo plantea el uso de una ontología para mejorar los resultados de la búsqueda en un dominio particular.

La Figura 1 muestra la arquitectura completa de la propuesta que consta de dos partes: una parte de expansión de la consulta y otra de acceso a una ontología.

La búsqueda basada en ontología se describe cómo la obtención de información semánticamente relacionada con una consulta planteada. Considerando un funcionamiento típico de la ontología, una búsqueda debería tomar como punto de partida un concepto (individuo o clase) de la propia ontología. En este trabajo se ha utilizado una versión más amplia y se ha implementado una búsqueda textual sobre la ontología, permitiendo la obtención de conceptos (individuos y clases) de la ontología en función de una búsqueda expresada en lenguaje natural.

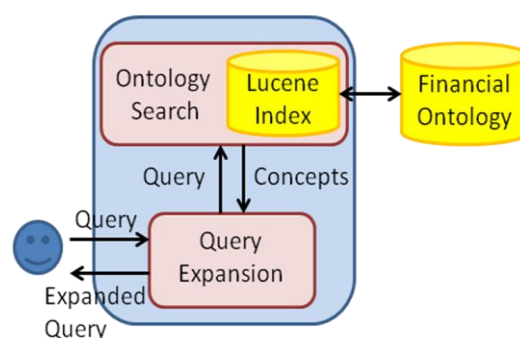


Figura 1. Arquitectura de la propuesta

Aunque no se cubre en este trabajo, esta estructura favorece la recuperación multilingüe, tal como explican Declerck y Gromann (2012). Esto se debe, entre otras cosas, a que si la ontología es multilingüe, almacena relaciones entre conceptos de distintos idiomas sin ambigüedad. Por ejemplo, *salt* en inglés estará relacionada con el concepto *sal* (*condimento*) en castellano y no con la forma imperativa del verbo *salir*.

3.1 Descripción de la Ontología de Doble Capa

La ontología desarrollada en este trabajo se centra en el dominio financiero y para ser más concretos, en mercados bursátiles y en las empresas que operan en estos mercados. El formato de esta ontología está basado en el utilizado por Liu y Fi (2011) aunque sin dividir el contenido en varias ontologías. Únicamente disponemos de una ontología en la que podemos identificar dos ‘capas’: la primera que relaciona todas las entidades presentes en un mercado bursátil y la segunda que asigna metadatos a cada una de estas entidades.

La primera de las capas está compuesta por cuatro clases diferentes:

¹ <http://onki.fi/>

² <http://www.clef-initiative.eu/>

- Mercados: representación semántica de los mercados bursátiles. Contiene empresas que operan en ellos.
- Sectores: elemento perteneciente a la clasificación de los diferentes ámbitos en los que las empresas pueden operar.
- Empresas: empresa que opera en mercados bursátiles y en sectores. Está relacionada con las personas que trabajan en ella (junta directiva, presidentes, etc.).
- Personas: personas físicas que forman parte o representan a las empresas.

Además de las propias clases, las relaciones semánticas entre ellas (representadas por propiedades de objeto) también se engloban en la primera capa. Estas relaciones son las que ofrecen una completitud semántica a las entidades. La lista completa de las relaciones existentes se muestra en la Tabla 1. El dominio y el rango pueden tomar los siguientes valores: Mercado (M), Empresa (E), Sector (S) o Persona (P).

La segunda capa es la que aporta información complementaria a cada una de las entidades. Esto se realiza a través de la asignación de anotaciones a las entidades. Estas anotaciones son de dos tipos (nombre y descripción) y pueden aparecer en cualquier idioma. En nuestro caso nos hemos limitado a español e inglés. Como ejemplo se muestran las anotaciones del concepto *Indra* en formato OWL en la Tabla 1.

Propiedad	Dominio	Rango
hasCeo	E	P
isParticipatingIn	E	M
hasActivityIn	E	S
hasCompany	S	E
hasCompanyIn	M	E
hasPresident	E	P
isCEOIn	P	E
isPresidentOf	P	E
isSubSectorOf	S	S
isSuperSectorOf	S	S

Tabla 1: Propiedades de Objetos de la Ontología

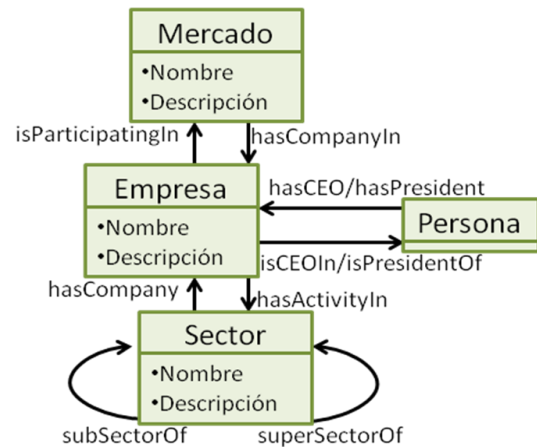


Figura 2. Diagrama de la Ontología

Empresa

Nombre (esp.): INDRA SISTEMAS, S.A.,SERIE A
 Nombre (ing.): INDRA SISTEMAS, S.A.,SERIE A
 Descripción (español): Indra es la multinacional de tecnologías de la Información número 1 en España y una de las principales de Europa y Latinoamérica. Es la segunda compañía europea por capitalización bursátil de su sector y es también la segunda empresa española que más invierte en I+D. En 2008 sus ventas alcanzaron los 2.380 M , de los que un tercio procedieron del mercado internacional. Cuenta con más de 29.000 profesionales y con clientes en más de 100 países. Indra es líder en soluciones y servicios de alto valor añadido para los sectores de Seguridad y Defensa, Transporte y Tráfico, energía e Industria, Servicios Financieros, Sanidad y Administraciones Públicas, Telecom y Media.

Sector

Nombre (esp.): Tecnología y telecomunicaciones
 Nombre (ing.): Technology and telecommunication
 Descripción (español): Este sector engloba aquellas actividades relacionadas con las telecomunicaciones tales como la telefonía (tanto básica como móvil), y el diseño, instalación, gestión y mantenimiento de redes e infraestructura de comunicaciones. Además, se incluyen todas aquellas actividades de electrónica y software así como las empresas dedicadas a la fabricación y distribución de hardware tecnológico y equipamiento.

Tabla 2: Ejemplo de metadatos asociados a empresa y sector.

3.2 Construcción de la Ontología

La ontología ha sido creada automáticamente a partir de un *web crawler* sobre la página de la Bolsa de Madrid³.

Mediante este *crawler* se obtuvo tanto información de las empresas como de los sectores. En total se obtuvieron 41 sectores que están agrupados en 6 sectores generales que a su vez se dividen en subsectores. En cuanto a las empresas, se han obtenido 34 empresas que cotizan en el mercado bursátil de Madrid.

La información que se obtuvo de las empresas está compuesta por: nombre, descripción, dirección, últimas cotizaciones, información bursátil, etc. Para la aplicación que se ha desarrollado, solo se ha utilizado el nombre y la descripción.

Con respecto a los sectores, que también fueron obtenidos con un *web crawler* sobre la página de la Bolsa de Madrid, se obtuvo el nombre de cada sector y su descripción (ambos en español y en inglés). Además, también se almacenó la jerarquía de sectores, para poder determinar aquellos que pertenecen a otros. En el caso de los sectores, toda la información obtenida (en español) se utiliza en la ontología.

Una vez que se dispone de toda esta información, es necesario poblar la ontología. Los sectores tienen distintas realizaciones lingüísticas en las descripciones de las empresas. Por ejemplo, en el caso de *Bankia* el identificador del sector es “*Serv.Financieros e Inmob. - Bancos/Cajas Ahorro*”, mientras que el nombre del sector que le corresponde es “*Servicios financieros e inmobiliarias*”. Por este motivo se ha aplicado una normalización para hacer corresponder la etiqueta de sector asociada a la empresa con su correspondiente sector en la clasificación.

La distancia utilizada para comparar los nombres es la distancia de Jaro-Winckler (Winckler, 1990). Se valoró la utilización de la distancia de Levenshtein (Levenshtein, 1965), pero dado que los sectores en las empresas aparecían con abreviaturas, se descartó esa posibilidad.

3.3 Búsqueda de Lenguaje Natural

El sistema de búsqueda en lenguaje natural se basa en la utilización conjunta de la ontología y un índice de Lucene⁴ creado sobre la misma. Este índice tiene tres campos: el nombre de los

conceptos (individuos), su descripción y la URL que los identifica. La consulta textual se lanza sobre el nombre del concepto y su descripción y se recuperan las URLs de los conceptos como identificador semántico. El peso que se otorga al consultar cada campo del índice es el mismo. Tanto para la indexación de las descripciones como para la consulta se utiliza el analizador Snowball⁵ de Lucene. Además, el peso que se otorga a todos los campos del índice es el mismo.

El resultado obtenido de consultar este índice son los identificadores únicos (URLs) de los conceptos almacenados en la ontología.

3.4 Enriquecimiento de Consulta

El enriquecimiento de la consulta se realiza con los nombres de los conceptos que devuelve la búsqueda en lenguaje natural sobre la ontología.

La expansión de la consulta se ha abordado desde dos puntos de vista diferentes: lanzando la consulta completa del usuario a la búsqueda de lenguaje natural y el análisis semántico de la consulta para expandir únicamente las entidades.

El análisis semántico de la consulta se ha realizado utilizando Textalytics⁶, un conjunto de APIs de procesamiento lingüístico en distintos idiomas, incluyendo español e inglés. Esta herramienta ha permitido el reconocimiento de las entidades de la consulta, obteniendo además etiquetas semánticas que permiten distinguir el tipo de entidad, es decir, si se trata de una persona, una organización, un lugar, etc.

En ambos casos, la consulta planteada por el usuario se mantiene invariable mientras que la información extraída de la ontología se añade directamente al final de la propia consulta. La información obtenida puede ser de cinco tipos diferentes: (i) nombre de la empresa (+Nom); (ii) texto completo de la descripción de la empresa; (iii) nombre de la empresa y texto completo de la descripción de la empresa; (iv) entidades extraídas de la descripción de la empresa;(+Ents.Descr.) y (v) nombre de la empresa y entidades extraídas de la descripción de la empresa.

⁵

http://lucene.apache.org/core/old_versioned_docs/versions/3_5_0/api/all/org/apache/lucene/analysis/snowball/SnowballAnalyzer.html

⁶ <http://www.textalytics.com>

³ <http://www.bolsamadrid.es>

⁴ <http://lucene.apache.org/core/>

En una segunda aproximación, se ha utilizado la ontología para extraer información relacionada (sectores en los que operan las empresas) con los conceptos recuperados mediante la búsqueda textual. Esto añade dos tipos nuevos de información a añadir: (vi) nombre del sector en el que operan las empresas (+Sect.); (vii) nombre de la empresa y del sector; y permite añadir el sector en los tipos (iii) y (v) previamente definidos.

4 Evaluación preliminar

Este sistema está trabajando en el dominio financiero sobre una ontología poblada con información en español del mercado bursátil de Madrid (IBEX35).

La evaluación del sistema se realiza de manera incremental, evaluando en primera instancia la búsqueda en la ontología para posteriormente evaluar la expansión de la consulta. La finalidad de dividir la evaluación radica en comprobar la influencia que tiene en la expansión de la consulta los posibles fallos producidos por la eficacia de la búsqueda en la ontología.

4.1 Búsqueda en la Ontología

Para la evaluación de la búsqueda en la ontología se ha diseñado una prueba de Cranfield. Para cada una de las consultas planteadas se ha definido un *gold standard* (definido por los expertos que han colaborado en la evaluación) de conceptos que deben recuperarse, por lo que la evaluación de esta búsqueda se hará mediante la precisión y el *recall* o cobertura que ofrece la propia recuperación. En la Figura 3 se muestra la precisión y cobertura de la búsqueda con dos números diferentes de conceptos recuperados (10 y 1) para cada consulta. Al calcular la media se observa que la precisión aumenta (de 36% a 50%) y que la cobertura baja (de 45,5% a 35,4%).

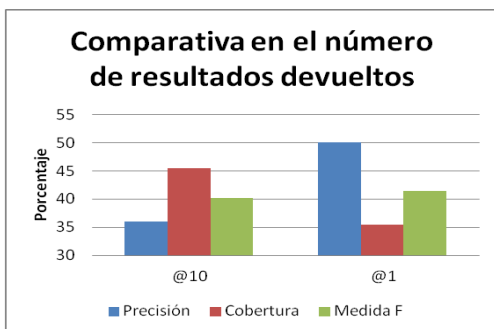


Figura 3: Comparativa de la Recuperación en función del número de Conceptos devueltos.

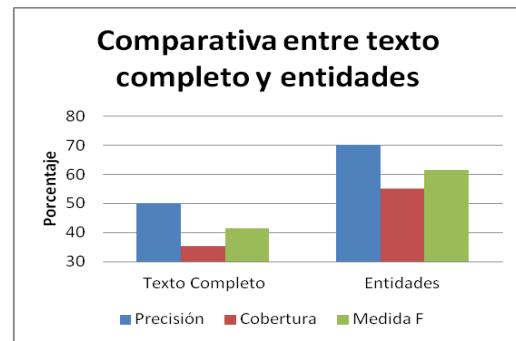


Figura 4: Comparativa numérica de la búsqueda en ontología con texto completo o entidades.

La segunda evaluación que se consideró es la comparación entre la recuperación cuando se lanza el texto completo contra la ontología y cuando se lanzan únicamente las entidades (Figura 4).

Los resultados de precisión y cobertura son muy parecidos en ambos casos, siendo mejores en el caso en el que solo se utilizan las entidades para consultar la búsqueda en la ontología: 70% precisión y 55% cobertura (buscando el texto completo tenemos 50% y 34,4%).

4.2 Expansión de Consulta

Para realizar esta prueba se han utilizado 20 consultas textuales del dominio financiero que se han sido generadas por un licenciado en Economía y en Administración de Empresas. Algunos ejemplos de consulta son: acciones de Telefónica, ¿Cuál ha sido la última cotización de Bankia?, empresas eléctricas que suben, liquidez de Repsol.

Cada una de estas consultas ha sido enriquecida utilizando la búsqueda en ontología, y posteriormente el resultado ha sido evaluado manualmente por dos expertos diferentes, que han valorado la consulta de tres maneras:

- Completamente Expandida (CE): la consulta se ha expandido y toda la información que se ha añadido es relevante para la búsqueda que queremos hacer.
- Parcialmente Expandida (PE): la consulta se ha expandido, pero lo que se ha añadido no es suficiente o tiene una parte correcta y otra no.

- Sin expansión (SE): no se ha expandido o la información que se ha añadido no es relevante para lo que buscamos.

Los expertos han presentado valoraciones dispares a la hora de valorar los resultados, lo que introduce una variable de subjetividad en el proceso. Para calcular el porcentaje de acuerdo que existe entre ellos se ha procedido a calcular el coeficiente Kappa (Cohen, 1960). Este coeficiente se ha calculado en los siete casos diferentes de expansión que se quieren probar (Figura 5).

Según los coeficientes obtenidos, existe una gran disparidad en el acuerdo entre los evaluadores. Las únicas informaciones añadidas a la consulta que son útiles son el nombre de la empresa ($k=0,69$), el sector en el que opera la empresa ($k=0,7$) y las entidades presentes en la descripción de la empresa ($k=0,75$).

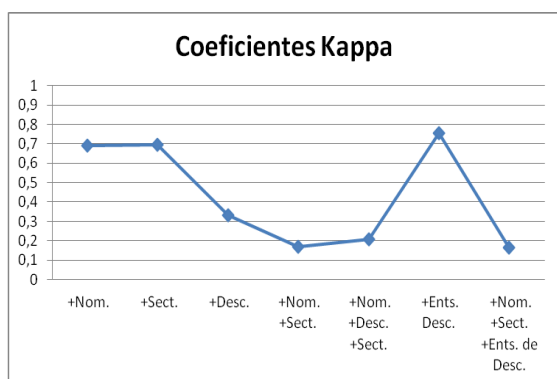


Figura 5: Coeficiente Kappa para evaluar el porcentaje de acuerdo entre los evaluadores.

En la Figura 6 se muestra el porcentaje (media de ambos expertos) de consultas que se han considerado de cada tipo para cada información añadida.

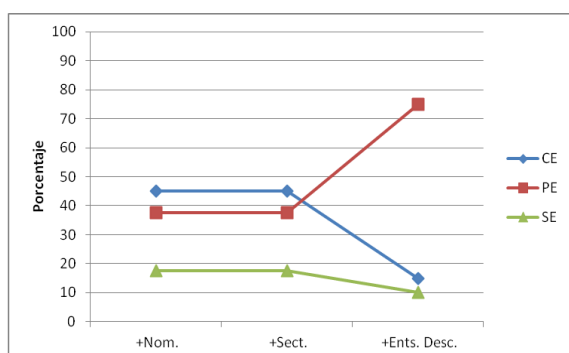


Figura 6: Porcentajes de consultas de cada tipo al realizar las expansiones.

Al añadir el nombre o el sector de la empresa, el porcentaje de consultas que son evaluadas como CE es mucho mayor (45%) que en el caso de añadir las entidades de la descripción (15%). Por el contrario, el porcentaje de consultas consideradas como PE es mucho mayor en el último caso (75%) mientras que en los anteriores era mucho más bajo (37,5%). Estas cifras llevan a que el número de consultas valoradas como SE es menor en el último caso (10% por el 17,5% de los dos primeros casos).

5 Conclusiones y Líneas Futuras

Se ha mostrado una prueba de concepto de expansión de consulta usando una ontología pero los resultados no permiten asegurar la utilidad del enfoque, para ello sería necesario escalar el tamaño de la ontología y aplicarlo a un problema de recuperación de información financiera.

La expansión de consulta presenta unos porcentajes bajos en el caso de consultas sin expansión: un 17,5% para +Nom y +Sect. y un 10% en +Ents.Descr.

En los casos de expansión +Nom y +Sect. se añaden pocos términos, mientras que en el caso +Ents.Descr se añaden muchos términos (porque la descripción de una empresa hace referencia a muchas entidades). En ambas situaciones hay ambigüedad, introducida por la recuperación de conceptos que no son relevantes para la consulta, y que en una primera aproximación se ha eliminado limitando los resultados al primero que devuelve el índice de Lucene. Sin embargo, este es un enfoque que habría que mejorar introduciendo algún algoritmo que haga uso de las relaciones semánticas de la ontología para resolver la ambigüedad y añadir a la consulta aquellos términos relacionados entre sí.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto Trendminer (EU FP7-ICT287863), el proyecto Monnet (EU FP7-ICT 247176) y MA2VICMR (S2009/TIC-1542).

Bibliografía

Bhagal, J., A. Macfarlane y P. Smith. 2007. A review of ontology based query expansion. En *Information Processing and Management*, 43, 4, pp: 866-886.

- Carpineto, C. y G. Romano. 2012. A survey of automatic query expansion in information retrieval. *ACM, Comput. Surv.* 44, 1, Article 1 (January 2012), 50 pages, DOI = 10.1145/2071389.2071390
- Carstens, C. 2011. Ontology Based Query Expansion-Retrieval Support for the Domain of Educational Research.
- Cohen, J. 1960. A coefficient of agreement for nominal scale. En *Educ. Psychol. Measure*, 20, pp: 37-46.
- Declerck, T. y D. Gromann. 2012. Towards the Generation of Semantically Enriched Multilingual Components of Ontology Labels. En: *Proceedings of the 3rd International Workshop on the Multilingual Semantic Web (MSW3) volume 936, CEUR Workshop Proceedings*, pp: 11-23.
- Dey, L., S. Singh, R. Rai y S. Gupta. 2005. Ontology Aided Query Expansion for Retrieving Relevant Texts. En *Advances in Web Intelligence*, pp: 988-991.
- Díaz-Galiano, M. C., M. T. Martín-Valdivia y L. A. Ureña-López. 2009. Query expansion with a medical ontology to improve a multimodal information retrieval system. En *Computers in Biology and Medicine*, 39, 4 pp: 396-403.
- Dragoni, M., C. C. Pereira y A. G. B. Tettamanzi. 2012. A conceptual representation of documents and queries for information retrieval systems by using light ontologies. En *Expert Systems with Applications* 39, 12, pp: 10376-10388.
- Farhoodi, M., M. Mahmoudi, A. M. Zare Bidoki, A. Yari y M. Azadnia. 2009. Query Expansion Using Persian Ontology Derived from Wikipedia. En *World Applied Sciences Journal* 7 (4), pp: 410-417.
- Fu, G., C. B. Jones y A. I. Abdelmoty. 2005. Ontology-based Spatial Query Expansion in Information Retrieval. En *OTM Conferences 2005*, pp: 1466-1482.
- Gulati, P. y A. K. Sharma. 2010. Ontology Driven Query Expansion for Better Image Retrieval. En *International Journal of Computer Applications* 5(10), pp: 33-37.
- Levenshtein, V.I. Binary codes capable of correcting deletions, insertions, and reversals. 1966. *Doklady Akademii Nauk SSSR*, 163(4):845-848, 1965 (Ruso). Traducción a inglés en *Soviet Physics Doklady*, 10(8):707-710.
- Liu, L. y F. Li. 2011. 3-layer Ontology Based Query Expansion for Searching. En *ISNN 2011, Part III, LNCS 6677*, pp: 621-628.
- Mandala, R., T. Tokunaga y H. Tanaka. 2000. Query expansion using heterogeneous thesauri. En *Information Processing and Management* 36, pp: 361-378.
- Martínez-Fernández, J. L., A. M. García-Serrano, J. Villena-Román y P. Martínez. 2008. Expanding Queries Through Word Sense Disambiguation, Evaluation of Multilingual and Multi-modal Information Retrieval; LNCS Vol. 4730, pp: 613-616.
- Schweighofer, E. y A. Geist. 2007. Legal Query Expansion using Ontologies and Relevance Feedback. En *LOAIT 2007*, pp: 149-160.
- Tuominen, J., T. Kauppinen, K. Viljanen, y E. Hyvonen. 2009. Ontology-Based Query Expansion Widget for Information Retrieval. En *Proceedings of the 5th Workshop on Scripting and Development for the Semantic Web (SFSW 2009), 6th European Semantic Web Conference (ESWC 2009)*.
- Winkler, W. E. (1990). String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. *Proceedings of the Section on Survey Research Methods (American Statistical Association)*: 354-359.
- Wollersheim, D. y W. J. Rahayu. 2005. Ontology Based Query Expansion Framework for Use in Medical Information Systems., En *IJWIS 2005, 1, (2)*, pp: 101-115.
- Wu, F., G. Wu y X. Fu. 2007. Design and Implementation of Ontology-Based Query Expansion for Information Retrieval. En *IFIP - The International Federation for Information Processing 254*, pp: 293-298.