

# **UNIVERSIDAD NACIONAL DE ROSARIO**

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Licenciatura en Ciencias de la Computación

Tesina

## **Búsqueda Semántica en Repositorios de Objetos de Aprendizaje**

Marcos Emanuel Belén

Directoras de Tesina:

Cristina Bender

Claudia Deco

2010

## Tabla de Contenidos

Resumen .....	4
Introducción .....	5
1. Conceptos Teóricos .....	7
1.1 Recursos Lingüísticos .....	7
1.1.1 WordNet .....	7
1.1.2 Otros Recursos .....	10
1.2 Objetos de Aprendizaje (OA).....	13
1.2.1 Definición.....	13
1.2.2 Atributos de los OA.....	14
1.2.3 Granularidad.....	14
1.2.4 Reutilización de los OA .....	14
1.2.5 Metadatos .....	15
1.3 Repositorios de Objetos de Aprendizajes (ROA).....	16
1.3.1 Definición.....	16
1.3.2 Tipos y Características Principales.....	16
1.3.3 Iniciativas de ROA .....	18
1.3.4 Ejemplos [Repositorios Utilizados].....	20
1.4 Estándares y Especificaciones de los ROA.....	22
1.4.1 Ventajas del Uso de Estándares.....	22
1.4.2 IMS.....	23
1.4.3 SCORM.....	24
2. Trabajos Relacionados .....	25
3. Propuesta.....	27
3.1 Descripción general.....	27
3.2 Procesamiento de documentos recuperados.....	34
4. Implementación.....	36
4.1 Tecnologías Utilizadas .....	36

4.2 Repositorios Utilizados .....	36
4.3 Prototipo.....	37
5. Experimentación .....	39
5.1 Primer análisis.....	39
5.2 Segundo análisis.....	41
5.3 Tablas Comparativas.....	42
6. Discusiones .....	48
6.1 Repositorio .....	48
6.2 Pruebas de la Implementación.....	48
6.3 Limitaciones.....	48
6.4 Trabajo futuro.....	49
7. Bibliografía .....	50
8. Apéndice .....	54
8.1 Clases Utilizadas.....	54
8.2 OA recuperados en la experimentación.....	59
8.2.1 Primer análisis.....	60
8.2.2 Segundo análisis.....	61

## Resumen

Debido al incremento exponencial de información contenida en la Web y a la necesidad de tener acceso al enorme caudal de datos, han surgido diversos problemas relacionados con la recuperación de los mismos. La mayoría de los usuarios son inexpertos y no aplican técnicas de búsqueda o las aplican incorrectamente por lo que se generan dificultades a la hora de conseguir información valiosa de manera eficiente. Con una búsqueda de este tipo la mayoría de los documentos devueltos tienen poco o nada que ver con los intereses del usuario. En cambio muchos documentos relevantes no se tienen en cuenta poniendo así de manifiesto la importancia de encontrar una solución a este problema.

Este problema también sucede en el escenario de la recuperación de información aplicada a los repositorios educacionales. Sin embargo aquí no sólo se busca la documentación correcta sino que también se exige que la misma presente ciertas características de interés o de necesidad para el usuario. Los elementos a recuperar se llaman objetos de aprendizaje (OA), los cuales son elementos para la instrucción, aprendizaje o enseñanza basada en computadora. Los OA se encuentran almacenados en contenedores conocidos como Repositorios de Objetos de Aprendizaje (ROA). El objetivo principal de los ROA es organizar los OA y mantenerlos disponibles para diferentes usos.

En este trabajo se describe un método que permite expandir la consulta del usuario de forma automática mediante recursos lingüísticos, buscando incrementar las posibilidades de recuperar la información deseada. Además de las palabras claves usadas para expresar la consulta, también se tienen en cuenta características y preferencias individuales del usuario mediante el uso de metadatos. Debido a que todos los OA se describen mediante metadatos, el uso de los mismos nos permitirá recuperar los OA que mejor satisfagan las necesidades de información del usuario.

## Introducción

La representación y organización de los documentos debería proveer al usuario un fácil acceso a la información en la cual está interesado. Desafortunadamente, la caracterización de esta necesidad de información no es un problema simple. Dada la consulta del usuario, el objetivo principal de un sistema de IR (Information Retrieval - *Recuperación de Información*) es recuperar información que podría ser útil o relevante para el usuario.

La recuperación de datos en el contexto de un sistema de IR consiste principalmente en determinar qué documentos de una colección contienen las palabras claves de la consulta las cuales, generalmente, no son suficientes para satisfacer las necesidades de información del usuario.

En realidad el usuario de un sistema de IR *está vinculado mas con la recuperación de información sobre un tema* que con la recuperación de datos que satisfacen una consulta determinada. Para un sistema de recuperación de información, los objetos recuperados podrían ser (y lo son la mayoría de las veces) imprecisos. La razón principal es que IR usualmente trata con texto en lenguaje natural el cual no siempre está bien estructurado y muchas veces es semánticamente ambiguo.

Para ser efectivos en este intento de satisfacer las necesidades de información del usuario, el sistema de IR debe saber *cómo* interpretar el contenido de los documentos y clasificarlos de acuerdo un grado de relevancia respecto a la consulta. Esta interpretación del contenido implica extraer información sintáctica y semántica del texto del documento y usarla para recuperar información relevante. La dificultad no es solo saber cómo extraer esta información sino también como usarla para decidir relevancia. Por lo tanto la noción de relevancia esta en el centro de la recuperación de información.

La Web se ha convertido en un repositorio universal de cultura y conocimiento humano el cual ha permitido compartir de una manera sin precedentes las ideas e información en una escala nunca antes vista. Este universo sin fronteras ha atraído la atención de millones de personas desde el comienzo. Más aun, ha causado una revolución en la forma en que las personas usan las computadoras y realizan sus tareas diarias. Por ejemplo, las compras on-line desde casa (home shopping) y los servicios bancarios disponibles desde la comodidad del hogar (home banking) son muy populares y han generado varios cientos de millones de dólares en ingresos. Como el ingreso por publicidad de los buscadores y portales está fuertemente vinculado con el volumen de acceso del público, el incremento de la demanda para un sitio es un problema de negocio extremadamente serio. Indudablemente, este incentivo financiero genera uno de los mayores impulsos para la enorme cantidad de estudios e investigaciones sobre la recuperación de información basada en la Web.

A pesar de tanto éxito, existen problemas complejos. Encontrar información útil es frecuentemente una tarea difícil y tediosa. El principal obstáculo es la ausencia de un modelo de datos subyacente bien definido para la Web, lo cual implica que la definición y estructura de la información es frecuentemente de baja calidad.

Los sistemas de recuperación que emplean técnicas de indexación automática para crear términos representativos de documentos en lenguaje natural deben tratar con problemas de polisemia y sinonimia. La polisemia (palabras con más de un significado) degrada la precisión causando falsas coincidencias, mientras que la sinonimia (múltiples palabras teniendo el mismo

significado) degrada la proporción de documentos relevantes que son recuperados, causando la pérdida de verdaderas coincidencias conceptuales. En principio, la polisemia y sinonimia se pueden manejar o controlar asignando a diferentes sentidos de una palabra diferentes identificadores del concepto, y asignando el mismo identificador de concepto a los sinónimos. En la práctica esto requiere procedimientos que sean capaces de reconocer sinónimos, y que puedan no solo detectar los usos de diferentes sentidos de una palabra sino que también puedan resolver qué significado se desea en cada caso.

Más allá del desorden de los recursos en la Web, la complejidad de la búsqueda de información y muchas veces la inseguridad de las fuentes, han surgido nuevas iniciativas y tecnologías para la organización de estos recursos que tienen como objetivo precisamente que los recursos disponibles en la Web sean verdaderamente explotables. Específicamente para el sector educativo han emergido sistemas de organización llamados repositorios de objetos de aprendizaje. Un usuario que busca información en un repositorio, *describe* de manera breve el documento que quiere obtener como respuesta. El resultado es que el usuario *navega* en el repositorio realizando consultas y refinando resultados hasta que queda satisfecho. Aquí, los criterios de búsqueda deben considerar bastante más que títulos, autores o palabras claves. Sin embargo, aún en un entorno como éste el usuario necesita ayuda para conseguir la información que necesita y evitarse, o por lo menos alivianar, la tarea de filtrar lo que no es considerado relevante a su criterio.

En este trabajo se presenta un método automático para construir una estrategia de búsqueda adecuada con el objetivo de ayudar al usuario a encontrar información útil de manera eficiente. Por un lado, se busca ampliar la consulta con sinónimos definidos en WordNet para lidiar con el problema de la sinonimia (diferentes palabras o expresiones representan un mismo concepto). Por otro, se busca el sentido correcto de cada sustantivo o frase hallada en la consulta mediante un método de des-ambigüedad que se basa en las relaciones existentes entre los términos utilizando el contenido semántico de WordNet. Esta estrategia de búsqueda servirá de base para la búsqueda en la Web en general y también para la búsqueda de objetos de aprendizaje en repositorios educacionales o ROA.

El resto del trabajo se organiza de la siguiente forma: en el Capítulo 1 se presentan con cierto detalle los *actores* involucrados en el tema como lo son los objetos de aprendizaje y los repositorios; el Capítulo 2 presenta trabajos relacionados; el Capítulo 3 describe el método propuesto; el Capítulo 4 presenta la implementación del método para la búsqueda en ROA; experimentaciones se describen en el Capítulo 5; finalmente las conclusiones y trabajos futuros se abordan en el Capítulo 6.

# 1. Conceptos Teóricos

## 1.1 Recursos Lingüísticos

### 1.1.1 WordNet

WordNet es un sistema léxico construido manualmente por George Miller y sus colegas en el laboratorio de Ciencias Cognitivas de la Universidad de Princeton [WordNet]. Originario de un proyecto cuyo objetivo era construir un diccionario en el que se pudiera buscar manera conceptual en lugar de solo alfabéticamente, WordNet evoluciono en un sistema que refleja las actuales teorías psicolingüísticas acerca de cómo los seres humanos organizan sus recuerdos léxicos. El objeto básico en WordNet es un conjunto de sinónimos exactos llamados synsets. Por definición, cada synset en el cual una palabra aparece es un sentido o significado diferente de dicha palabra, es decir, cada synset representa un concepto.

Hay cuatro divisiones principales en WordNet, para sustantivos, verbos, adjetivos y adverbios. Hay un rico conjunto de enlaces que permite la relación entre palabras, palabras y synsets, y también entre synsets. Dentro de una división, los synsets están organizados por las relaciones léxicas definidas sobre ellos. Para los sustantivos, la única división de WordNet utilizada en este trabajo, las relaciones léxicas incluidas son: antonimia, hiperonimia/hiponimia (relación IS-A), y tres diferentes relaciones de meronimia/holonimia (PART-OF). La relación IS-A es la relación dominante, y organiza a los synsets en un conjunto de aproximadamente diez jerarquías<sup>1</sup>. Como un ejemplo, la Figura 1.1 muestra la jerarquía IS-A relacionando los ocho sentidos diferentes del sustantivo 'board'. Los synsets con borde grueso son los sentidos de 'board', y los synsets restantes son ancestros o descendientes de uno de los sentidos. Otras relaciones, tal como 'director' (un sentido del mismo) es MIEMBRO-DE el sentido de committe de 'board', están incluidas en WordNet pero no se muestran en la figura. Los synsets {group, grouping} y {entity, thing} son ejemplos de cabeceras de jerarquías. Otras cabeceras incluyen {act, human\_action, human\_activity}, {abstraction}, {possession} and {psychological\_feature}.

La versión 3.0 contiene un número de 117798 sustantivos, organizados en 82115 synsets como se observa en la siguiente tabla (Tabla 1.1).

#### Number of words, synsets, and senses

POS	Unique Synsets		Total
	Strings		Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

Tabla 1.1 - Número de palabras y conceptos para la división de sustantivos en WordNet 3.0.

<sup>1</sup> La estructura actual no es una jerarquía porque algunos synsets tienen más de un padre.

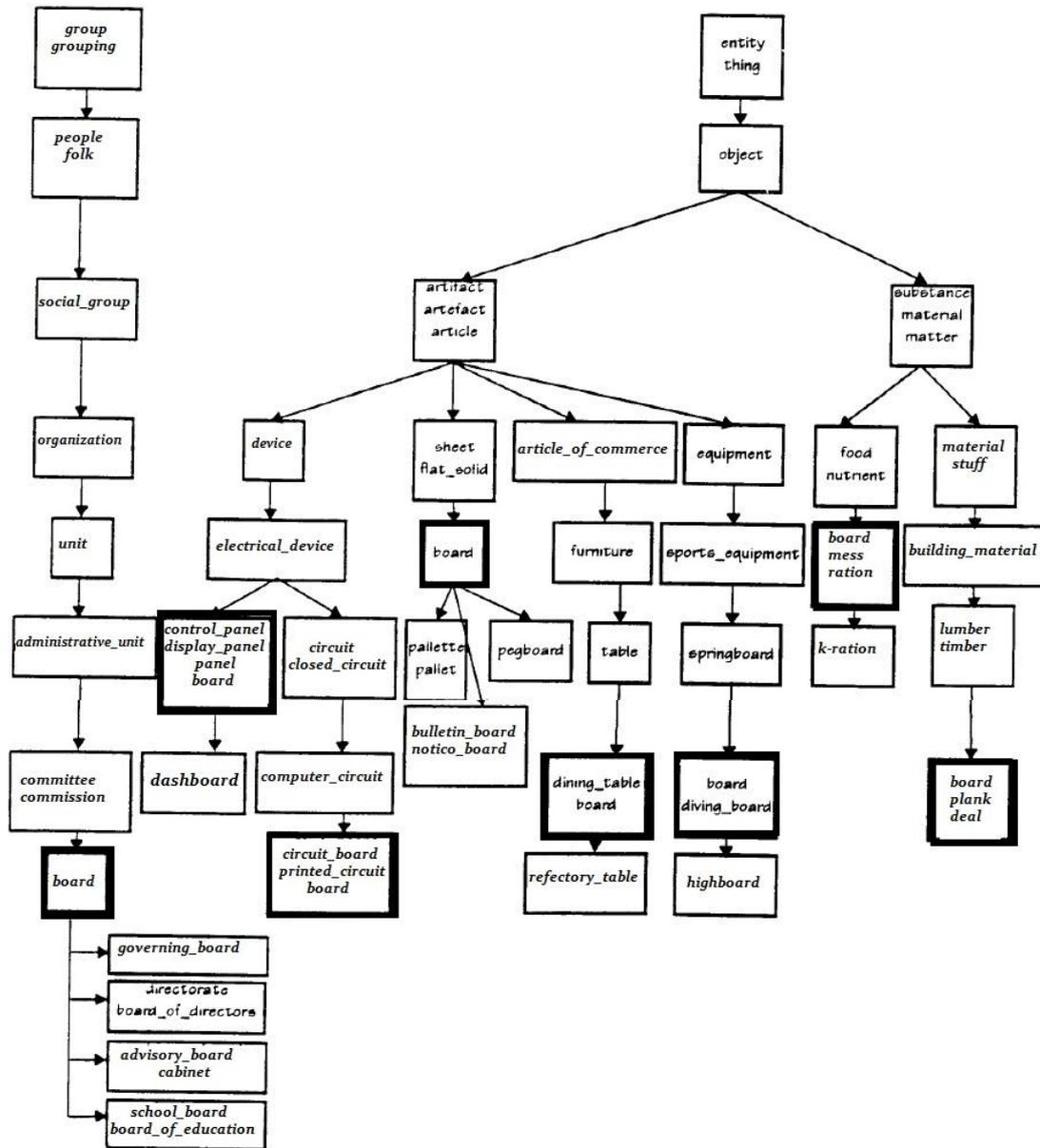


Figura 1.1 – La jerarquía IS-A para ocho sentidos diferentes del sustantivo **board**.  
Figura extraída de [Voorhees].

Para el proceso de desambiguación de los términos que componen la consulta y posteriormente la búsqueda de los sinónimos y homónimos para la expansión se ha utilizado WordNet. Su integración se ha realizado mediante una conversión de parte de la red semántica al formato XML (solo las relaciones de sinonimia, hiperonimia e hiponimia). En la Figura 1.2 se puede observar parte del archivo. Dentro del archivo los synsets se distinguen por la etiqueta <synset>. La etiqueta <id> describe el identificador del synset, <num\_words> indica el número de términos que componen el synset, <synonyms> contiene los términos sinónimos del synset los cuales se distinguen por la etiqueta <term>, <hyperonyms> contiene los identificadores de los synsets “padres”(contienen términos más generales), en <hyponyms> se

encuentran los identificadores de los synsets “hijos” (contienen términos más específicos), y finalmente la etiqueta **<descripción>** contiene una descripción del concepto que representa ese synset.

```

- <synset>
  <id>02139199</id>
  <num_words>02</num_words>
- <synonyms>
  <term>bat</term>
  <term>chiropteran</term>
</synonyms>
- <hyperonyms>
  <term>01886756</term>
</hyperonyms>
- <hyponyms>
  <term>02139671</term>
  <term>02141306</term>
</hyponyms>
<descripcion>nocturnal mouselike mammal with forelimbs modified to form
membranous wings and anatomical adaptations
for echolocation by which they navigate</descripcion>
</synset>

```

Figura 1.2 – Parte de la base de datos de WordNet. En este caso se muestra uno de los synsets correspondiente al término bat.

Para poder encontrar los distintos sentidos de una palabra o frase existe otro archivo en WordNet donde se asocia cada término con sus respectivos synsets. Este archivo también ha sido convertido a XML y se puede ver un ejemplo del mismo a continuación (Figura 1.3).

```

- <noun>
  <id>bat</id>
  <synset>02139199</synset>
  <synset>00458456</synset>
  <synset>04292414</synset>
  <synset>03132076</synset>
  <synset>02806379</synset>
</noun>
- <noun>
  <id>bat_boy</id>
  <synset>09843443</synset>
</noun>
- <noun>
  <id>bat_mitzvah</id>
  <synset>07454196</synset>
</noun>

```

Figura 1.3 – Ejemplo del XML que contiene los términos existentes en WordNet y los synsets correspondientes.

Se remarca en amarillo el identificador del synset descrito en la figura anterior. Cada sustantivo junto con todos los synsets que tenga asociado se encuentran entre las etiquetas **<noun></noun>**. Dicho sustantivo se distingue por la etiqueta **<id>** y cada uno de los sentidos del mismo es identificado por **<synset>**.

## 1.1.2 Otros Recursos

Otros recursos que se pueden utilizar en la preparación de la estrategia de búsqueda son tesauros y ontologías.

### Tesauros

La flexibilidad y variedad del lenguaje natural crea serias dificultades para el manejo automatizado de la información. Para solucionar este problema, surgen los tesauros, que permiten el control del vocabulario para representar en forma unívoca cada concepto.

Según la definición de la Unesco, un tesauro es un instrumento de control terminológico utilizado para traducir a un lenguaje más estricto el idioma natural empleado en los documentos y por los indizadores, que son las personas que asignan las palabras claves a cada documento. Por su estructura, es un vocabulario controlado y dinámico de términos relacionados semántica y genéricamente, los cuales cubren un dominio específico del conocimiento [Tesauros&Ontolog.].

El tesauro está estructurado formalmente para hacer explícitas las relaciones entre conceptos. Estas relaciones pueden ser jerárquicas, de afinidad, y preferenciales.

Las relaciones jerárquicas indican términos más amplios ó más específicos de cada concepto. Las de afinidad muestran términos relacionados conceptualmente, pero que no están ni preferencial ni jerárquicamente relacionados. Las relaciones preferenciales se utilizan para indicar cuál es el término preferido o descriptor entre un grupo de sinónimos; y para la calificación de homónimos a fin de diferenciar su significado, eligiendo un significado preferido para cada término.

A diferencia de un diccionario, donde todos los sinónimos de un concepto son representativos y tratados por igual, en un tesauro se tiene una palabra clave preferida y representativa del conjunto de sinónimos para cada concepto. Son obras en las que se relacionan numerosas palabras que guardan una relación más o menos directa con la palabra objeto de consulta. En cambio en los diccionarios se incluyen únicamente palabras con un significado similar y equivalente.

Los tesauros generalmente se limitan a un área específica, y desde un cierto ángulo en particular, por lo que existen innumerables tesauros específicos. Como ejemplo de tesauros podemos citar el MeSH o encabezamientos temáticos médicos en lengua inglesa. En la siguiente Figura 1.4 se muestra como ejemplo el resultado de una consulta realizada con el término “lung cancer”. Al ser éste un término prohibido para esta base de datos, el tesauro nos redirige en forma automática a su término permitido: “lung neoplasms”, mostrando además los conceptos relacionados con este último. Como se observa en la Figura 1.5, un término puede aparecer en varias jerarquías conceptuales, y el usuario puede moverse por estas jerarquías para ubicar el concepto más cercano a su necesidad.



The screenshot displays the MeSH search interface. At the top, there are logos for NCBI and MeSH, along with the text 'A service of the National Library of Medicine and the National Institutes of Health'. Below the logos, there are navigation tabs for 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', 'Structure', 'OMIM', and 'PMC'. The search bar contains 'MeSH' in a dropdown menu and 'lung cancer' in the input field. There are 'Go', 'Clear', and 'Save Search' buttons. Below the search bar, there are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The display settings show 'Summary' selected, 'Show 20' items, and a 'Send to' dropdown menu. The results section shows 'All: 8' items. The first result is '1: Lung Neoplasms' with a description 'Tumors or cancer of the LUNG.' and a link to 'DLEC1 protein, human [Substance Name]'. The second result is '2: DLEC1 protein, human [Substance Name]' with a link to 'DLEC1 protein, human [Substance Name]'.

Figura 1.4 – Búsqueda en MeSH del término “lung cáncer”

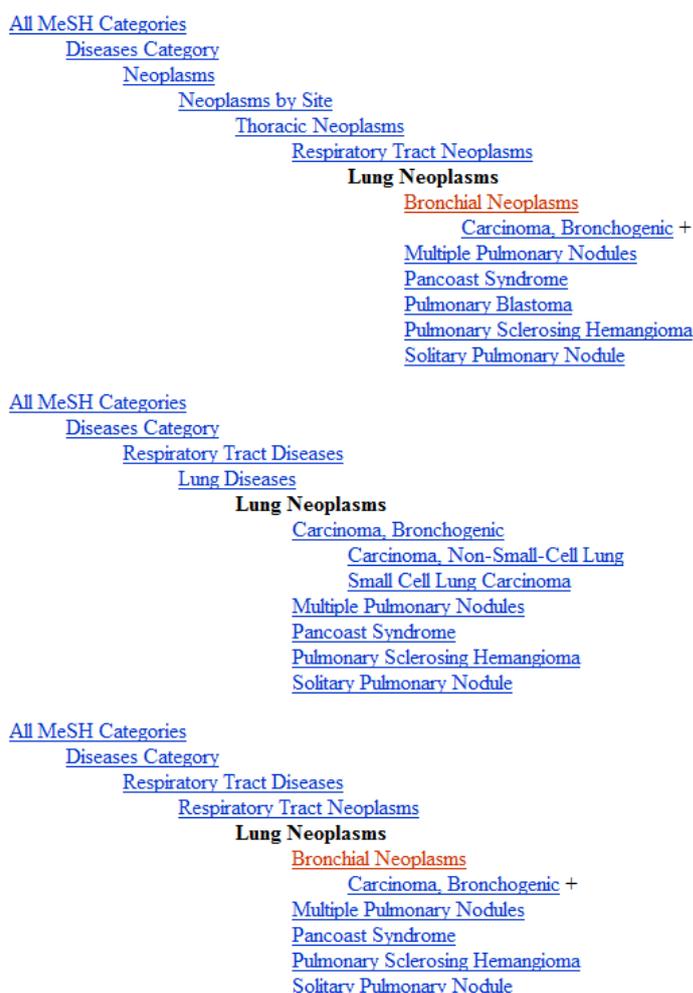


Figura 1.5 – Jerarquías conceptuales en las que aparece *lung neoplasm*.

## Ontologías

Proporcionan una vía para representar el conocimiento. La definición más consolidada es la que la describe como "un especificación explícita de y formal de una conceptualización compartida" [Gruber][Studer]. Consisten de términos, sus definiciones y axiomas que los relacionan con otros términos. Los términos están organizados en una taxonomía y los axiomas permiten realizar búsquedas con inferencias.

Para poder darle semántica a la *web*, se necesitan lenguajes de marcado apropiados que representen el conocimiento. El lenguaje XML (*eXtensible Markup Language*) con sus respectivos DTD (*Document Type Definition*) no es suficiente para esto [Broekstra et al. (2002)]. Existen otros lenguajes de marcado como ser RDF (*Resource Description Framework*), que permite representar algunos aspectos sobre conceptos de un dominio y, mediante relaciones taxonómicas, crear una jerarquía de conceptos. RDF es recomendado por el consorcio W3C [W3C] como estándar para los metadatos. Un lenguaje con gran capacidad expresiva que está emergiendo como un estándar para representar ontologías en la *web* es OWL (*Ontology Web Language*), y es desarrollado por el consorcio W3C.

A diferencia de los tesauros y de los diccionarios, en las ontologías se agregan axiomas que permiten realizar inferencias sobre los conceptos.

Las ontologías catalogan y definen los tipos de cosas que existen en un cierto dominio, así como sus relaciones y propiedades. Por ejemplo, una ontología del mundo empresarial usará conceptos como Venta, Compra, Transferencia, Pago entre otros, y relaciones como “Una Transferencia corresponde a una Venta o a una Compra”, “Un Pago corresponde a una o varias Transferencias”, etc.

Existen ontologías específicas (de términos médicos, empresariales, aeronáuticos, etc.) y ontologías de carácter general (proporcionan terminologías útiles para varios campos):

- En <http://www.snomed.org/> se puede encontrar una ontología médica (SNOMED).
- En <http://www.unspsc.org/> se describe una ontología de carácter general desarrollada por las Naciones Unidas (es un vocabulario para servicios y productos).

Un ejemplo de uso de ontologías en el área de Recuperación de Información se puede ver en [Navigli, Velardi]. Aquí se emplean las ontologías de forma experimental para extraer el dominio semántico de una palabra y expandir la consulta agregando términos que coocurren con las palabras de la consulta (por ejemplo *car* y *driver*) argumentando que el uso de los mismos es más efectivo que la expansión con sinónimos e hiperónimos.

## 1.2 Objetos de Aprendizaje (OA)

### 1.2.1 Definición

Los OA son elementos para la instrucción, aprendizaje o enseñanza basada en la computadora. No son realmente una tecnología, sino una filosofía, que según [Wiley (2000)] se fundamenta en la corriente de las ciencias de la computación conocida como *orientación a objetos* [Coad & Jill (1993)].

La orientación a objetos se basa en la creación de entidades con la intención de que puedan ser reutilizadas en múltiples aplicaciones. Componentes pequeños que pueden ser reutilizados en diferentes aplicaciones educativas [Wiley (2000)].

Formalmente no hay una única definición del concepto de objeto de aprendizaje y las definiciones son muy amplias. El Comité de Estandarización de Tecnología Educativa [IEEE (2001)], dice que los objetos de aprendizaje son “una entidad, digital o no digital, que puede ser utilizada, reutilizada y referenciada durante el aprendizaje apoyado con tecnología”. Según [Wiley (2000)] son “cualquier recurso digital que puede ser reutilizado para apoyar el aprendizaje”.

En [Morales & García (2005)] se definen a los OA como una unidad de aprendizaje independiente y autónomo que está predispuesta a su reutilización en diversos contextos. Por otra parte, [JORUM+ Project (2004)] dice que “un OA es cualquier recurso que puede ser utilizado para facilitar la enseñanza y el aprendizaje y que ha sido descrito utilizando metadatos”.

Sin embargo podemos considerar como una definición más acertada a la definida en [Guzmán] donde dice que “cualquier recurso con una intención formativa, compuesto de uno o varios elementos digitales, descrito con metadatos, que pueda ser utilizado y reutilizado” puede ser un OA.

Se dan como ejemplos de objetos de aprendizaje los contenidos multimedia, el contenido para instruir, los objetivos de aprendizaje, software de enseñanza, personas, organizaciones o eventos referenciados durante el aprendizaje basado en tecnología [IEEE (2001)]. Otros autores son menos específicos en cuanto a recursos del campo educativo, como [González (2005)] que considera como OA a archivos de texto, ilustraciones, vídeos, fotografías, animaciones y otros tipos de recursos digitales. Por su parte, el [JORUM+ Project (2004)] dice que como ejemplos se puede incluir una imagen, un mapa, una pieza de texto, una pieza de audio, una evaluación o más de uno de estos recursos.

Los beneficios que los objetos de aprendizaje pueden tener en un contexto educativo son:

- **flexibilidad**, ya que el mismo recurso puede utilizarse en distinto contextos;
- **administración del contenido**, que se facilita porque los recursos están descritos con metadatos que permiten su control;
- **adaptabilidad**, que facilita al diseñador poder seleccionar y componer recursos según la aplicación; y
- **código abierto** que elimina los problemas de incompatibilidad entre plataformas [Lowerison, Gallart & Boyd, (2003)].

Un OA podrá utilizarse y/o reutilizarse en la medida en que su diseño haya sido el adecuado, es decir, que sus objetivos estén claros y que sea fácilmente integrable a diversas aplicaciones, tanto por su contenido como por la descripción que se haga de él y que le permita ser identificado adecuadamente, de esto se encargan su granularidad (sección 1.2.3) y sus metadatos (sección 1.2.5).

## 1.2.2 Atributos de los OA

En su concepción debe pensarse que sean recursos con atributos específicos, fáciles de localizar, utilizar, almacenar y compartir. Para ello, estos recursos deben ser [Rehak & Mason, (2003)]:

- *Reutilizables*. El recurso debe ser modular para servir como base o componente de otro recurso. También debe tener una tecnología, una estructura y los componentes necesarios para ser incluido en diversas aplicaciones.
- *Accesibles*. Pueden ser indexados para una localización y recuperación más eficiente, utilizando esquemas estándares de metadatos.
- *Interoperables*. Pueden operar entre diferentes plataformas de hardware y software.
- *Portables*. Pueden moverse y albergarse en diferentes plataformas de manera transparente, sin cambio alguno en estructura o contenido.
- *Durables*. Deben permanecer intactos a las actualizaciones de software y hardware.

La creación de OA no es sencilla, pero los esfuerzos y costos de producción se equilibran con las veces que el recurso pueda reutilizarse.

## 1.2.3 Granularidad

Aunque se menciona que un OA es “una pieza pequeña” o un recurso “modular” no se puede especificar una dimensión precisa. El tamaño de un OA es variable y esto se conoce como granularidad.

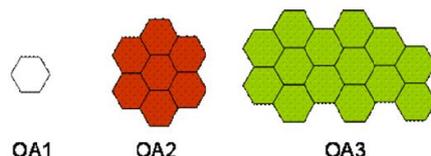


Figura 1.7 – OA de distinta granularidad.

En la Figura 1.7 se han representando, a través de hexágonos, lo que pueden ser unidades de contenido o elementos que componen al OA. Por ejemplo, el OA1 podría ser una imagen y el OA2 podría ser una página web que incluye texto e imágenes. El OA3 puede ser un recurso multimedia en el que se incluyen más unidades de contenido que en los objetos anteriores.

No es posible definir la cantidad de información o elementos que un OA debe contener, esto dependerá de las necesidades y habilidades del autor para trabajar y conceptuar trozos de contenidos que irán formando un curso. Sin embargo, el mejor criterio para definir la granularidad de un objeto es por sus propósitos u objetivos [Duncan (2003)].

Se considera una buena práctica que los OA cubran un único objetivo de aprendizaje y para lograrlo deben mantener independencia del contexto y no requerir de otros recursos, es decir, que sean autosuficientes.

## 1.2.4 Reutilización de los OA

La característica más notable en las diferentes definiciones de los OA es la reutilización. Lograr la reutilización requiere de tener un diseño, un desarrollo y una documentación que aseguren un alto nivel en la calidad del producto y pueda éste trabajar de forma sencilla con otros. La reutilización de un contenido aumenta su valor y produce ahorro, en diferentes sentidos, a nivel institucional o individual. El gran potencial de la reutilización de los OA es poder aprovechar los contenidos que han desarrollado otros para formar nuevos recursos.

En la Figura 1.8 se ejemplifica cómo a partir de tres OA independientes se genera otro nuevo.

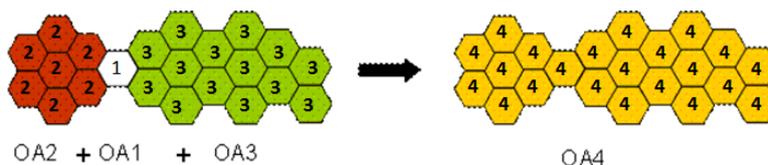


Figura 1.8 – Creación de un nuevo OA a partir de la utilización de otros.

Para la reutilización, así como para lograr los otros atributos descritos, es necesario que el objeto de aprendizaje cuente con los metadatos que le permitan ser identificado, organizado y recuperado, entre otros aspectos como la categorización y calificación pedagógica del objeto, pero lo más importante es que esos metadatos estén basados en un estándar, a fin de asegurar su compatibilidad e interoperabilidad con los sistemas que puedan reutilizarlos [López, García & Pernías, (2005)].

### 1.2.5 Metadatos

Los metadatos son un conjunto de atributos necesarios para describir un recurso. A través de los metadatos se tiene un primer acercamiento con el objeto, conociendo rápidamente sus principales características. Son especialmente útiles en los recursos que no son textuales y en los que su contenido no puede ser indexado por sistemas automáticos, por ejemplo, los archivos multimedia o de audio.

En [Olmos & Ballesteros] se define a los metadatos como “*datos estructurados que proporcionan un resumen corto sobre cualquier recurso de información, impresa o electrónica, y facilita la ubicación, identificación o descubrimiento de ese recurso*”.

Los metadatos no sólo son descriptivos, también pueden ser administrativos y de estructura [Caplan (2003)]:

- *Metadatos descriptivos*: tienen propósito de **descubrimiento** (cómo se encuentra un recurso), **identificación** (cómo un recurso puede distinguirse de otro), y **selección** (cómo determinar que un recurso cubre una necesidad particular).
- *Metadatos administrativos*: es información que facilita la administración de los recursos. Incluyen información sobre cuándo y cómo fue creado el recurso.
- *Metadatos estructurales*: sirven para identificar cada una de las partes que componen al recurso, definen la estructura que le da forma.

Para describir los recursos educativos, se requieren elementos específicos, además de los metadatos bibliográficos comunes.

Para capturar estos elementos, se han desarrollado **estándares de metadatos educativos**. Los siguientes son dos estándares desarrollados para este propósito:

- El estándar **Metadatos de Objetos de Aprendizaje (LOM)** tiene toda una sección **dedicada a las** características educativas de un recurso de información. Éstas incluyen: tipo de interactividad, tipo del recurso de aprendizaje, nivel de interactividad, densidad semántica, rol intentado del usuario final, contexto, rango típico de edad, dificultad y tiempo típico de aprendizaje.
- **El esquema de metadatos educativos Dublin Core, desarrollado por el Grupo de trabajo en Educación DCMI, propone dos elementos:** “audiencia” y “normas”; además de los quince elementos básicos de Dublin Core. El grupo también ha propuesto la adaptación de tres elementos del estándar LOM: tipo y niveles de interactividad, y el tiempo típico de aprendizaje.

En el año 2002 se emite el estándar 1484.12.1 [IEEE, (2002)] que acredita al modelo de datos LOM como el estándar de metadatos para OA.

Para organizar los metadatos, LOM los agrupa en nueve categorías: general, ciclo de vida, meta-metadatos, técnica, educativa, derechos, relación, anotación y clasificación.

Con esto se busca, además de la organización, la reutilización de recursos y la interoperabilidad entre los sistemas involucrados con el uso de contenidos. Para hacerlo realmente posible es necesario que los metadatos estén representados a través de XML [Bray, Paoli, Sperberg-MacQueen, Maler & Yergeau, (2004)], ya que se considera “que los metadatos basados en tecnología XML son un elemento clave para la administración de repositorios digitales, con esta alianza se puede llevar a cabo el intercambio de información y de contenidos, entre plataformas y entre repositorios, de forma transparente para el usuario” [Ahmed, Ayers, Birbeck, Cousins, Dodds, Lubell, et al., (2001)].

### **1.3 Repositorios de Objetos de Aprendizajes (ROA)**

Los ROA comienzan a posicionarse como importantes herramientas que tienen como función resguardar los recursos, hacerlos disponibles para diversos usos y para compartirlos con otras aplicaciones, facilitando con esto el flujo de contenidos y la expansión de servicios. A continuación se dará su definición, los tipos y características principales, también se explicará el esquema de metadatos que se recomienda utilizar y se hará mención también de las iniciativas más relevantes de los proyectos de este tipo de repositorios.

#### **1.3.1 Definición**

En el apartado sobre OA se ha insistido en que la reutilización debe ser una de sus principales bondades, es entonces casi intrínseco que se tenga un lugar destinado para su almacenamiento y clasificación para facilitar posteriormente su mantenimiento, localización y, posiblemente, también compartir ese OA con otros sistemas para aplicaciones diversas. Este almacén es lo que se conoce como Repositorio de Objetos de Aprendizaje. Los sistemas de repositorios son la infraestructura clave para el desarrollo, almacenamiento, administración, localización y recuperación de todo tipo de contenido digital [ADL, (2002)].

El [JORUM+ Project (2004)] adopta la siguiente definición: “Un ROA es una colección de OA que tienen información (metadatos) detallada que es accesible vía Internet. Además de alojar los OA los ROA pueden almacenar las ubicaciones de aquellos objetos almacenados en otros sitios, tanto en línea como en ubicaciones locales”.

Se puede decir que los ROA son un tipo de bibliotecas digitales especializadas en recursos educativos que utilizan los estándares de metadatos que han desarrollado los organismos encargados de la estandarización del e-learning, preparadas tecnológicamente para interoperar con otros repositorios y con otras aplicaciones de los entornos e-learning.

#### **1.3.2 Tipos y Características Principales**

Por la forma en la que se concentran los recursos, principalmente se identifican dos tipos de ROA [Downes (2004)] [Rehak & Mason, (2003)]:

- los que contienen los objetos de aprendizaje y sus metadatos, en éstos los objetos y sus descriptores se encuentran dentro de un mismo sistema e incluso dentro de un mismo servidor (Figura 1.9a), y
- los que contienen sólo los metadatos, en este caso el repositorio contiene sólo los descriptores y se accede al objeto a través de una referencia a su ubicación física que se encuentra en otro sistema o repositorio de objetos (Figura 1.9b).

También es común encontrar repositorios mixtos, en los que se hace una combinación de estos dos tipos mencionados.

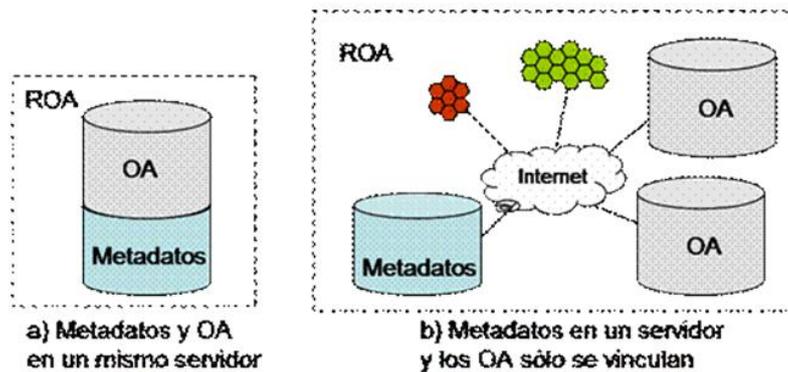


Figura 1.9 – Tipos de ROA por la distribución de los OA.

Los ROA más conocidos (que se menciona en el apartado 1.3.3) comúnmente funcionan de forma independiente (*stand-alone*). Son aplicaciones con una interfaz web, un mecanismo de búsqueda y listados con algún tipo de clasificación. Otra clase de ROA operan sólo como módulos adicionales a otros productos (LMS *Sistemas de Administración de Aprendizaje* o LCMS *Sistemas de Administración de Contenidos de Aprendizaje*) que utilizan los contenidos de forma exclusiva y sin que el usuario tenga acceso directo al repositorio. Lo deseable es que los ROA tengan ambas capacidades, tanto ofrecer una interfaz web, para que los usuarios humanos puedan acceder a la colección, así como la capacidad de comunicarse directamente con las plataformas de aprendizaje y hacer posible la interoperabilidad entre sistemas de diferente naturaleza.

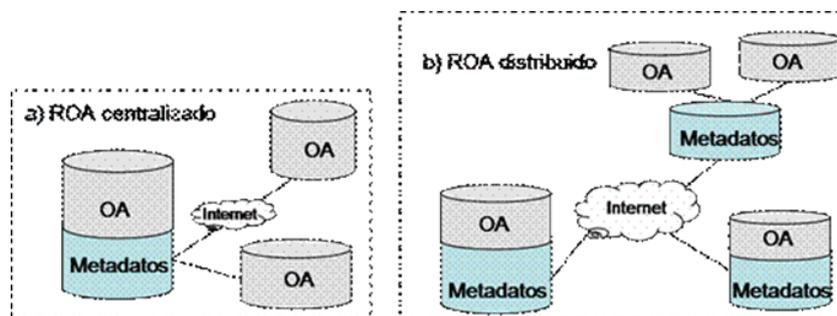


Figura 1.10 – Tipos de ROA por la distribución de los metadatos.

Por la forma en la que los catálogos de metadatos se organizan, se diferencian dos modelos de ROA [Downes (2004)]: centralizados y distribuidos. Los más comunes son los centralizados (Figura 1.10a), en los cuales los metadatos de los OA están contenidos en un mismo servidor, aunque el objeto esté localizado en algún otro. El modelo distribuido (Figura 1.10b) opera a través de varios servidores, cada uno contiene diferentes grupos de metadatos y se comunican entre ellos para intercambiarlos.

[ADL, (2002)] propone un conjunto básico de funciones que los repositorios deben proveer a fin de dar acceso a los objetos de aprendizaje en un ambiente seguro. Estas funciones son:

- *Buscar/encontrar*. Es la habilidad para localizar un objeto de aprendizaje apropiado. Esto incluye la habilidad para su despliegue.
- *Pedir*. Un objeto de aprendizaje que ha sido localizado.
- *Recuperar*. Recibir un objeto de aprendizaje que ha sido pedido.
- *Enviar*. Entregar a un repositorio un objeto de aprendizaje para ser almacenado.
- *Almacenar*. Poner dentro de un registro de datos un objeto, con un identificador único que le permita ser localizado.

- *Colectar*. Obtener metadatos de los objetos de otros repositorios por búsquedas federadas.
- *Publicar*. Proveer metadatos a otros repositorios.

Además de estas funciones también deben considerarse el manejo de los derechos de copia o DRM (*Digital Rights Management*).

### 1.3.3 Iniciativas de ROA

La creación de ROA es relativamente reciente, las iniciativas de desarrollo se han dado inicio a finales de la década pasada y a principios de ésta. [Leslie, Landon, Lamb, & Poulin, (2004)], en su análisis sobre software para ROA, afirman que el mercado de software para estas aplicaciones es todavía inmaduro. A pesar de ello, el crecimiento ha sido rápido y los resultados se pueden ver en los repositorios ya disponibles en la Web, con decenas de miles de objetos de aprendizaje recopilados. La Tabla 1.2 muestra la lista de algunos de los principales proveedores y de las instituciones académicas involucradas en el desarrollo de repositorios.

También hay iniciativas que están trabajando en propuestas para la interoperabilidad entre repositorios, con la finalidad de formar redes de sistemas distribuidos que permitan búsquedas federadas [Hatala, Richards, Eap & Willms, (2004)].

<i>Proveedores</i>	<i>Instituciones</i>
Artesia	Cornell University
IBM	Nacional Science Foundation
Sun Microsystems	Old Dominion University
EMC	Simon Frasier University
Learning Object Network	University of Alberta
Microsoft Corporation	University of Calgary
Digital Concept, Inc.	University of Wisconsin

Tabla 1.2 – proveedores e Instituciones mayormente conocidas en el desarrollo de ROA (ADL, 2002)

A continuación se mencionan a las iniciativas más conocidas, tanto de repositorios como de propuestas de redes interoperables:

- **MERLOT (*Multimedia Educational Resource for Learning and Online Teaching*)**, sin duda es el repositorio más conocido y reconocido, es quizá el que está marcando la pauta para el desarrollo y tendencia de los ROA. Es un repositorio centralizado que contiene sólo los metadatos y apunta a los objetos ubicados en sitios remotos. Es independiente y funciona como un portal de objetos de aprendizaje. Provee búsquedas y otros servicios como personalización, importación y exportación de objetos. Cualquier usuario puede tener acceso a todos los objetos contenidos en MERLOT y sólo los miembros contribuyen agregando objetos, pero para ser miembro no se requiere más que inscribirse y no se adquiere ninguna responsabilidad. La revisión por pares es una actividad que MERLOT utiliza para evaluar la calidad de los objetos agregados. Disponible en <http://www.merlot.org/>.
- **CAREO (*Campus Alberta Repository of Educational Objects*)**, es un repositorio centralizado de objetos de aprendizaje multidisciplinarios de profesores de Alberta (Canada). Es un repositorio independiente que da acceso a objetos remotos y locales a través de los metadatos contenidos en su colección. Cualquier usuario puede tener acceso a los objetos, pero los miembros tienen servicios adicionales, al igual que MERLOT ser miembro es gratis y abierto a cualquier persona. Disponible en <http://www.careo.org/>.
- **CLOE (*Co-operative Learning Object Exchange*)**, es un modelo cooperativo para el desarrollo, uso y reutilización de objetos de aprendizaje. Es un proyecto de la Universidad de Waterloo en el que participan 17 universidades de Ontario y se tiene

acceso a su colección sólo siendo miembro de dichas universidades. Disponible en <http://cloe.on.ca/>.

- **SMETE ( *Science, Mathematics, Engineering and Technology Education* )**, es un repositorio distribuído, que se presenta como una biblioteca digital que integra de forma federada las colecciones de varias bibliotecas de recursos educativos. El acceso es libre para la consulta. Disponible en <http://www.smete.org/smete/>.
- **GEM ( *Gateway to Educational Materials* )**, es un proyecto del Departamento de Educación de los EEUU, originalmente conocido como *National Library in Education Advisory Task Force* . La colección GEM está orientada a la interoperabilidad entre múltiples bases de datos a través del uso de módulos que extraen los metadatos de los objetos en su formato GEM. Disponible en <http://www.thegateway.org/>.
- **POOL ( *Portals for Online Objects in Learning* )**, es un consorcio de organizaciones educativas privadas y públicas, que pretende crear un gran repositorio distribuido de objetos de aprendizaje, desarrollando y distribuyendo herramientas para crear repositorios conectados. Disponible en <http://www.edusplash.net/>.
- **CeLeBraTe ( *Context eLearning with Broadband Technologies* )**, es un proyecto desarrollado para los ambientes de aprendizaje virtual de la *European Learning Network*, con la finalidad de que se intercambien los recursos digitales educativos de sus miembros. Se plantea un repositorio centralizado pero cada miembro tiene la opción de conservar, total o parcialmente, la administración local de los metadatos de su colección. Las búsquedas se realizan tanto en el sistema central como en los repositorios locales. Disponible en [http://celebrate.eun.org/eun.org2/eun/en/index\\_celebrate.cfm/](http://celebrate.eun.org/eun.org2/eun/en/index_celebrate.cfm/).
- **ELENA/Edutella**, es un proyecto europeo que propone mediadores de servicios educativos que llama *Smart Spaces*, que permiten la integración de servicios heterogéneos de aprendizaje como herramientas de tutoría, LMS, sistemas de videoconferencia y repositorios. ELENA es una capa de la infraestructura propuesta por Edutella, en la cual se conectan aplicaciones con tipos diferentes de repositorios, modelos de búsqueda y diferentes esquemas de metadatos. Disponible en <http://www.elena-project.org/>.
- **eduSourceCanada**, este proyecto es una propuesta para crear una red de ROA en Canadá, uniendo los principales repositorios creados en este país con una infraestructura abierta e interoperable. La infraestructura soportará una amplia variedad de servicios y promete sistemas fáciles de usar y comunicar. Disponible en <http://www.edusource.ca/>.

En la Tabla 1.3 se presenta un resumen de las características principales de los ROA y de los proyectos para crear redes interoperables de ROA. Se destaca el nivel escolar al que están dirigidos, la organización que los promueve y el país asociado, el tipo de acceso y la forma en la que organizan tanto los metadatos como los OA.

Repositorio	Nivel	Organización/ País	Acceso	Metadatos	OA
MERLOT	Superior	Internacional	Abierto	Locales	Distribuidos
CAREO	Superior	Universida de Calgary/ Canadá	Abierto	Locales	Distribuidos
CLOE	Superior	Cooperative Learning Object Exchange/ Canadá	Cerrado	Locales	Locales y distribuidos
SMETE	k-12	SMETE Open Federation/ EEUU	Abierto	Distribuidos	Distribuidos
GEM	Todos	GEM Consortium / EEUU	Abierto	Distribuidos	Distribuidos
POOL	Todos	Varias/ Internacional	Abierto	Distribuidos	Distribuidos
CeLeBraTe	Todos	European Learning Network/ Europa	En desarrollo	Distribuidos	Distribuidos
ELENA/Edutella	Todos	Edutella/ Europa	Interope- rabilidad	Distribuidos	Distribuidos
eduSource- Canada	Todos	EduSource/ Canadá	Interope- rabilidad	Distribuidos	Distribuidos

Tabla 1.3 – Resumen de las características de los ROA y proyectos asociados

### 1.3.4 Ejemplos [Repositorios Utilizados]

Un problema crucial que se le presenta a la comunidad de e-learning es cómo producir y distribuir contenido de calidad para el aprendizaje on-line. Mejorando la reutilización e interoperabilidad de los objetos, el costo de crearlos se minimiza. El primer problema a concentrarse es el de encontrar el objeto de aprendizaje más útil para desarrollar la tarea educativa requerida. Para este propósito, los objetos de aprendizaje están almacenados en repositorios que pueden ser accedidos mediante consultas y motores de búsqueda.

Dentro de los repositorios los objetos de aprendizaje se encuentran usualmente descritos por metadatos. Por lo tanto en el acceso a los mismos se puede tomar ventaja si las consultas se realizan sobre estos metadatos.

**Ariadne** [Ariadne] fue uno de los repositorios que se tuvo en cuenta para realizar las pruebas de consulta y recuperación de objetos de aprendizaje. Es un proyecto de investigación y de desarrollo tecnológico de telemática para la educación y el entrenamiento, patrocinado por la Unión Europea. El proyecto se enfoca en el desarrollo de herramientas y metodologías para producir, administrar y reutilizar elementos pedagógicos basados en computadora, así como el plan de estudios de entrenamiento a distancia.

Para proveer a los miembros de Ariadne acceso a otros repositorios, se ha desarrollado un motor de búsqueda federada. Este motor ofrece una interfaz SQI (Simple Query Interfaz) [SQI] síncrona. SILO, la herramienta de búsqueda e indexación de Ariadne, utiliza este target para consultar un conjunto de repositorios por ejemplo. Luego el motor reenvía la consulta a diferentes repositorios SQI habilitados.

Actualmente, Ariadne requiere que cada uno de esos repositorios devuelva un mínimo de campos en los metadatos, codificados en LOM XML: una URL, un identificador, un título y un identificador del repositorio de origen. Otros campos en los metadatos son opcionales.

El núcleo de la infraestructura ARIADNE es una red distribuida de repositorios de objetos de aprendizaje. En un sentido amplio, la arquitectura de búsqueda federada de Ariadne consiste en 3 capas:

- *Search Clients*: pueden ser clientes de búsqueda tales como la herramienta de consulta e indexación de Ariadne (SILO - Search & Index Learning Objects), portales Web tales como el sitio Web prolearn o integraciones VLE tal como el plugin moodle.

- *Middleware*: el motor de búsqueda federado es responsable de distribuir consultas que recibe de search clients y recuperar y ordenar el resultado, antes de enviarlo de vuelta al search client. El registro UDDI mantiene metadatos de diferentes repositorios. El motor de búsqueda federado utiliza estos registros para cargar la lista de repositorios.

- *Repositorios de Objetos de Aprendizaje* que pueden ser usados para la búsqueda a través de SQI.

Además de Ariadne, también fueron considerados otros repositorios para llevar a cabo las consultas de prueba, sin embargo fueron posteriormente descartados por diferentes motivos. Se describen brevemente a continuación.

**BIOE** (Banco Internacional Objetos Educativos): El Banco Internacional de Objetos Educativos es un repositorio creado en 2008 por el Ministerio de Educación de Brasil, que cuenta con la cooperación del Ministerio de Ciencia y Tecnología, de la Red Latinoamericana de Portales Educativos – RELPE, de la Organización de los Estados Iberoamericanos – OEI, entre otros. Este Banco Internacional tiene por objetivo mantener y compartir recursos educativos digitales de acceso libre, producidos en distintos formatos - como audio, video, animación, simulación, software educativo – además de imagen, mapa y hipertexto, considerados relevantes y adecuados a la realidad de la comunidad educativa local, teniendo en cuenta las diferencias de lengua y culturas regionales.

En este momento BIOE cuenta con alrededor de 9948 objetos disponibles y 2838 más están siendo evaluados o están esperando la autorización del autor para ser publicados.

Para conectarse a este repositorio se debe utilizar el protocolo SQI mediante web services.

**OER Commons**: públicamente puesto en marcha en febrero de 2007, su misión es dar apoyo y construir una base de conocimiento entorno al uso y aprovechamiento de los recursos educativos abiertos (OER por sus siglas en inglés). Ha forjado alianzas con más de 120 socios para proporcionar un punto único de acceso a través del cual los profesionales de la enseñanza y los estudiantes pueden buscar a través de colecciones para acceder a más de 24.000 artículos, encontrar y proporcionar información descriptiva sobre cada recurso, y recuperar los que ellos necesitan. Al ser "abierto", estos recursos están disponibles al público para uso de todos.

Este repositorio se basa en el estándar OAI-PMH [OAI] para la recuperación de objetos mediante consultas en la URL.

**arXiv**: Acceso abierto a más de 606.452 documentos sobre Física, Matemáticas, Ciencias de la Computación, Biología Cuantitativa, Finanzas Cuantitativas y Estadísticas. arXiv es propiedad de Universidad de Cornell y está operado por la misma, una entidad privada

sin fines de lucro institución educativa. arXiv es un proyecto financiado por la Cornell University Library.

## **1.4 Estándares y Especificaciones de los ROA**

Al manejar diferentes tipos de recursos para distintas aplicaciones y con diferentes tecnologías la estandarización se vuelve un punto clave para seguir operando e incluso permitir crecer a las aplicaciones actuales.

La estandarización se requiere a distintos niveles, primero, cuando los recursos son creados deben considerarse tecnologías, políticas y formatos compatibles con lo común en el sector; segundo, cuando esos recursos son incluidos en un repositorio, y deben ser descritos se utilizarán esquemas que aseguren su fácil localización y compatibilidad con otros sistemas de metadatos; tercero, cuando esos recursos sean utilizados y tengan que incorporarse a diferentes servicios, repositorios, plataformas y aplicaciones en un contexto dado; y cuarto, cuando los sistemas involucrados en un entorno tengan que interactuar con otros para cumplir sus funciones o ampliar sus capacidades.

El reto de los estándares es acordar de qué forma compartir, comunicar o desarrollar modelos y sistemas con la finalidad de lograr la interoperabilidad entre los diversos componentes.

En el ámbito de la educación en línea, los estándares se ven como necesarios ahora más que antes, dado el alcance global que tienen las aplicaciones *e-learning* por el uso de los sistemas de telecomunicación y al creciente interés de los individuos en la autoformación y en el aprendizaje a lo largo de toda la vida, que está marcando un mayor uso de los modelos de aprendizaje en línea que crecen de forma dispersa. Contar con aplicaciones estandarizadas marcará un mejor y mayor aprovechamiento de los esfuerzos hasta ahora invertidos en este campo.

La creación de estándares globales es una tarea compleja. Sería muy difícil llegar a un consenso que cubra las necesidades de todos o de una gran mayoría para ser adoptados de forma genérica, sin embargo, diferentes grupos están trabajando en el desarrollo tanto de especificaciones como de estándares en los diferentes niveles que se requieren, para poder establecer entornos *e-learning* integrados e interoperables.

### **1.4.1 Ventajas del Uso de Estándares**

Los estándares han sido un elemento indispensable para la masificación de cualquier tecnología ya que favorecen el crecimiento, la expansión y la generalización. En el campo del *e-learning*, los estándares están en proceso de comprensión y de adopción [López (2004)], por lo que los beneficios aún no son fácilmente perceptibles. Conforme los estándares se vayan introduciendo cada vez más, el sector se estará beneficiando en interoperabilidad, reutilización, manejabilidad, accesibilidad, durabilidad, escalabilidad y confiabilidad [Masie Center Learning Consortium (2003)], tanto en contenidos como en infraestructura y funcionalidad.

Muchas de las empresas y organizaciones que están haciendo desarrollos o impartiendo programas de educación a distancia ven estos beneficios reflejados como:

- Contenidos reutilizables.
- Consistencia en la descripción de los contenidos.
- Normalización en la organización de sus recursos.
- Acceso a más contenidos y más fácilmente localizables.
- Persistencia de sus beneficios.
- Migración sencilla de sus sistemas a nuevas versiones, e incluso a una nueva plataforma.
- Comunicación e intercambio de información con otros sistemas.

- Administración de la información apropiada tanto del recurso como del estudiante.
- Extensión de los servicios y de las capacidades de las plataformas.
- La inversión en la infraestructura se asegura por mayor tiempo.

La utilización de estándares amplía las opciones de los usuarios finales, reduciendo las restricciones de los sistemas propietarios y de soluciones aisladas [CETIS (2003)]. Las instituciones, los docentes, los estudiantes y los proveedores se verán beneficiados al contar con contenidos flexibles, plataformas homogéneas y bases de datos compartidas y distribuidas.

Las propuestas más ampliamente adoptadas en gran número de herramientas han sido las propuestas de IMS *Learning Consortium* y la de ADL [(CETIS, s.f.)], que han tenido como resultado las especificaciones IMS y el modelo SCORM, respectivamente. Las mismas se describen brevemente a continuación.

## 1.4.2 IMS

Las especificaciones IMS (2004) son el resultado de una activa iniciativa que está desarrollando y proponiendo especificaciones basadas en tecnologías abiertas (XML) para facilitar las actividades de aprendizaje sobre tecnología web, principalmente para el intercambio de contenidos y de información sobre los estudiantes. Es una propuesta ambiciosa que cubre, entre otros rubros, accesibilidad y adaptación del estudiante, la definición de competencias, el empaquetamiento de contenidos, información de agentes del proceso educativo, el diseño del aprendizaje a través de un lenguaje para expresar diferentes modelos pedagógicos, así como la formación de repositorios de contenidos digitales. Algunas de las especificaciones disponibles son:

- **IMS Resource List Interoperability** (IMS RLI, <http://www.imsglobal.org/rli/index.html>). Detalla como los metadatos estructurados pueden intercambiarse entre sistemas que almacenan y proveen recursos para la creación de listados y para aquellos que reúnen y organizan esos listados para fines educativos o de capacitación.
- **IMS Learning Resources Meta-Data** (IMS LRM, <http://www.imsglobal.org/metadata/index.html>). Esta especificación hace más eficiente el proceso de búsqueda y uso de los recursos, ya que proporciona una estructura para los elementos (metadatos) que describen o catalogan los recursos de aprendizaje, incluye también cómo los elementos deben ser usados, representados y organizados. La especificación se basa en la aplicación de LOM.
- **IMS Digital Repositories Interoperability** (IMS DRI, <http://www.imsglobal.org/digitalrepositories/index.html>). Esta especificación provee recomendaciones para la interoperabilidad de las funciones más comunes entre repositorios. En el nivel más general, define los repositorios digitales como colecciones de recursos con acceso a través de una red, sin conocimiento previo de la estructura de la colección. Los repositorios pueden contener los objetos o los metadatos que los describen y no importa si los objetos y los metadatos se encuentran en diferentes repositorios.
- **IMS Content Packaging** (IMS CP, <http://www.imsglobal.org/content/packaging/index.html>). Provee la funcionalidad para describir y empaquetar materiales de aprendizaje, tales como cursos individuales o una colección de cursos, en paquetes interoperables y distribuibles. Esta especificación direcciona la descripción, estructura y ubicación de materiales de aprendizaje en línea, así como la definición de algunos tipos específicos de contenidos. Los proveedores y desarrolladores de contenidos utilizan este formato para asegurar que sus productos serán compatibles e importables/exportables con cualquier herramienta que soporte esta especificación.

### 1.4.3 SCORM

El modelo SCORM [ADL (2005)] es un conjunto de estándares y especificaciones para compartir, reutilizar, importar y exportar OA. Este modelo describe cómo las unidades de contenidos se relacionan unas con otras a diferentes niveles de granularidad, cómo se comunican los contenidos con el LMS (*Learning Management System*), define cómo empaquetar los contenidos para importarse y exportarse entre plataformas, y describe las reglas que un LMS debe seguir a fin de presentar un aprendizaje específico. SCORM es expandible e incluye a trabajos de IEEE, AICC y de IMS para algunas de sus funciones. Maneja las unidades de contenido con el nombre de SCO (*Sharable Content Object*) que son simplemente objetos de aprendizaje que cumplen con la especificación SCORM.

Los detalles de la especificación se encuentran en cuatro documentos a los que se da mantenimiento de manera independiente. La versión 1.3 es la más reciente y se conoce como SCORM 2004, los documentos que la componen son:

- **SCORM Overview** (Thropp, 2004a). Este libro describe la historia y los objetivos de la Iniciativa ADL y de SCORM.
- **SCORM Content Aggregation Model** (CAM) (Thropp, 2004b). Describe los componentes utilizados en el aprendizaje, cómo empaquetar esos componentes para el intercambio entre sistemas, cómo describir esos componentes para permitir la búsqueda y la recuperación, y cómo definir las reglas de secuencia de los componentes.
- **SCORM Run-Time Environment** (RTE) (Thropp, 2004c). Este libro describe el medio para interoperar contenidos de aprendizaje basados en SCO y los LMS.
- **SCORM Sequencing and Navigation** (SN) (Thropp, 2004d). El documento SN describe las reglas que un LMS debe seguir a fin de presentar un aprendizaje específico.

Aunque SCORM ha ido actualizando versiones y ha extendido sus funciones, su alcance es limitado y en su modelo sólo cubre el empaquetamiento y la comunicación del recurso con el LMS, lo que hace su entendimiento e implementación mucho más sencilla que la de IMS, quizá este sea el motivo por el que es el más ampliamente utilizado hoy día para el intercambio de paquetes entre plataformas. Sin embargo, para la creación de repositorios todavía no tiene un desarrollo específico, pero las funciones hasta ahora disponibles, con el uso de metadatos y la creación de paquetes para mover recursos entre sistemas, pueden jugar un papel importante para facilitar las funciones de los ROA [ADL, (2002)].

Las cuatro especificaciones IMS nombradas anteriormente son aquellas que están relacionadas con la estandarización de los ROA: *IMS Digital Repositories Interoperability*, *IMS Content Packaging*, *IMS Learning Resource Metadata* e *IMS Resource List Interoperability*. Con estas especificaciones se pueden formar repositorios interoperables que consideren la importación y exportación de contenidos como paquetes identificados con metadatos. Con la aparición de IMS RLI, se incluyen también como contenidos a las listas de recursos que se transforman a OA. IMS RLI abre un vínculo importante para la comunicación entre las bibliotecas automatizadas y los ROA.

Por su parte, SCORM no tiene en su modelo contemplado el desarrollo de repositorios, pero sí puede comunicarse con ellos ya que se ha apoyado en IMS CP para el intercambio de paquetes, lo que lo hace compatible con las aplicaciones que se hagan sobre IMS para los ROA. Además, dado que ambas especificaciones utilizan un derivado de LOM para la definición de su esquema de metadatos, el intercambio entre ambas especificaciones es posible.

## 2. Trabajos Relacionados

En [Loupy & El-Bèze] se realizan experimentos de recuperación de documentos usando conocimiento semántico. En un primer conjunto de experimentos, sinónimos e hipónimos obtenidos de WordNet se utilizan para enriquecer las consultas. En la segunda parte de experimentación se utiliza un word sense disambiguation (WSD) system para hacerle frente al problema de la polisemia. Finalmente también se experimenta con tesauros especializados. Los resultados conducen a que el uso de sinónimos no necesariamente decrece la precisión, que el uso de WSD no necesariamente decrece el recall y que el uso de recursos especializados puede ser útil para mejorar el rendimiento.

En [Resnik] se presenta un método para WSD automático de sustantivos dentro de un conjunto de sustantivos relacionados. La esencia del algoritmo de desambiguación es el cálculo de lo que llama similitud semántica o “semantic similarity” usando la taxonomía WordNet restringiéndose a la relación “IS-A” para sustantivos y evaluándose en base a la información compartida por los ítems siendo comparados. Esta similitud se basa en la observación de que cuando dos palabras polisémicas son similares sus synsets ancestros más “informativos” proveen información sobre cuál es el sentido más relevante de cada palabra. Se remarca también la diferencia entre similitud y relación entre dos palabras: similitud es una relación más especializada que relación o asociación, *doctors* y *sickness* pueden estar altamente relacionadas pero no se distinguirían por ser particularmente similares.

[Mandala et. al] analiza el porqué de la pérdida de performance cuando se utiliza WordNet para mejorar la recuperación de información. Indica como razón principal que la mayoría de las relaciones entre términos no se encuentran en WordNet y que algunos términos como los nombres propios tampoco son incluidos. Como solución se propone un método para enriquecer WordNet mediante tesauros construidos de forma automática.

En [Fernandez & Serrano] se muestran aspectos relacionados con la integración de la tecnología disponible para el tratamiento del lenguaje natural en el desarrollo de un metabuscador que alcance un mayor grado de acierto en la recuperación de información realizada por un buscador tradicional así como también en el tratamiento posterior de los documentos recuperados. En particular se describe el proceso realizado para la extensión de las consultas de los usuarios mediante información lingüística empleando dos recursos léxicos para el castellano: ARIES para el tratamiento de la morfología y EuroWordNet para el tratamiento de la semántica. Sin embargo este último recurso plantea algunos problemas en el campo de la recuperación de la información tal como se describe en [Gonzalo et al. 1998] y [Mihalcea]: por ejemplo EWN no incluye relaciones semánticas con referencias cruzadas entre categorías gramaticales y las distinciones semánticas son muy sutiles en algunos casos (el grado de granularidad es bajo).

En [Voorhees] se describe un procedimiento de indexación automático que utiliza la relación “IS-A” de WordNet junto con un conjunto de sustantivos contenidos en un texto para seleccionar un sentido para cada sustantivo polisémico en el texto. Los resultados no fueron buenos debido a la dificultad del método utilizado para desambiguar el sentido correcto de las

palabras, principalmente en consultas cortas; con poco contexto el procedimiento falla en la desambiguación.

En [Moldovan & Mihalcea] se describe un sistema que se enfoca en dos problemas: el primero es la traducción de una pregunta o sentencia en lenguaje natural a una consulta usando WordNet, y el segundo es la extracción de párrafos con información relevante contenidos en los documentos recuperados por motores de búsqueda. Con el primer paso se busca aumentar el *recall* y con el segundo la *precisión*.

La diferencia principal con estos trabajos es que las experimentaciones no se realizan sobre repositorios educacionales. En la mayoría se utilizan bases de datos especialmente diseñadas para este tipo de experimentación: son ambiente controlados que proporcionan un conjunto de consultas predefinidas asociadas cada una a un conjunto determinado de documentos permitiendo de esta forma calcular con facilidad la precisión y en especial el *recall*. En [Loupy & El-Bèze] se utiliza TREC-6 (*the 6th Text Retrieval Conference*) como base de datos para la recuperación de documentos en las pruebas realizadas y también se realizan experimentos utilizando un tesoro especializado. Aparte de sinónimos también se utilizan hipónimos.

En las pruebas efectuadas en [Resnik] se ignoran aquellos sustantivos que no se encuentran en WordNet y no se diferencia entre términos plurales y singulares. Las únicas pruebas realizadas corresponden a la efectividad del algoritmo de WSD.

En [Fernandez & Serrano] se usan recursos lingüísticos en español y la pruebas se realizan sobre el contenido de un sitio. Además para mejorar la presentación de los resultados se utiliza una ontología de dominio específico para deducir temas relacionados con las páginas devueltas en los resultados y de esta forma poder mejorar la interacción con el usuario.

En [Moldovan & Mihalcea] se utilizan nuevos operadores léxicos que reemplazan los operadores booleanos clásicos utilizados por los motores de búsqueda. Las pruebas se realizan utilizando la colección de tópicos provista en TREC-6. En un principio se pensó la implementación en nuestro trabajo de este tipo de operadores (por ej. un operador de filtrado que medía la distancia entre las palabras claves en el texto de los documentos) pero el único texto a aplicar en los OA es en su descripción, la cual es un texto muy reducido en comparación al que se encuentra en los documentos recuperados en la Web. Por este motivo se decidió la implementación de otro filtro que actúa como un ranking de objetos de aprendizaje.

### 3. Propuesta

#### 3.1 Descripción general

En este trabajo se propone expandir semánticamente la consulta ingresada por el usuario utilizando sinónimos (e hiperónimos) de la base de datos léxica WordNet. La expansión se realiza de forma automática en base a un método de des-ambigüedad de términos descrito en la Figura 3.2. EL algoritmo de WSD es el propuesto en [Liu, Liu, Yu, Meng] con una ligera modificación.

Incluyendo términos conceptualmente equivalentes y jerárquicamente relacionados se espera mejorar la precisión así como también aumentar la cantidad de documentos relevantes recuperados.

El proceso de expansión se presenta en la Figura 3.1.

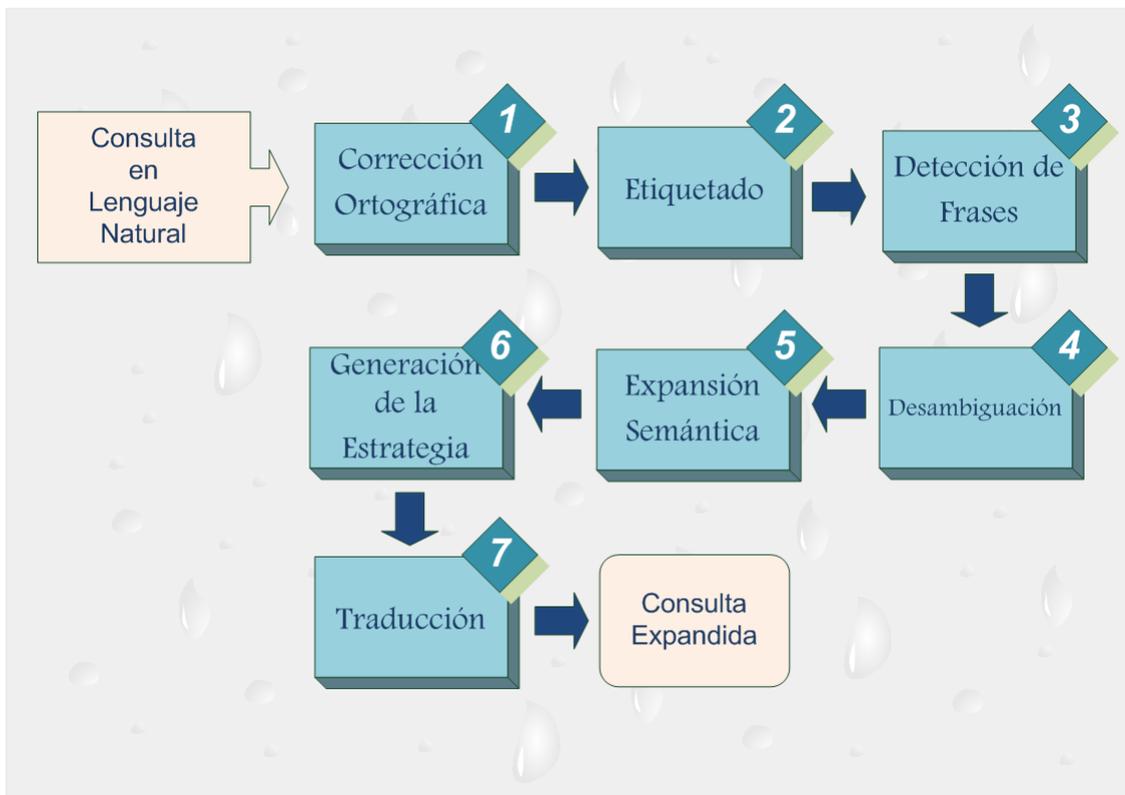


Figura 3.1 – Proceso de expansión semántica de una consulta de usuario.

En la figura anterior se observan una serie de módulos que se comunican para llevar a cabo la tarea de expandir de manera conceptual una consulta ingresada por el usuario el lenguaje natural. La descripción de cada uno de estos módulos es la siguiente:

- Dada una consulta  $Q = [t_1, t_2, .., t_n]$  el módulo 1 verifica que todos los términos se hayan escrito de manera correcta. La salida de este módulo es un conjunto  $Q' = [t'_1, t'_2, .., t'_n]$  donde cada elemento  $t'_i$  es la corrección ortográfica de su contraparte en  $Q$ .

Debido a que todo el proceso de expansión se realiza de forma automática los resultados de este modulo solo se utilizaran como sugerencia al usuario.

- El segundo módulo se encarga del etiquetado morfológico de los términos. Aquí se detallan los rasgos de género y número de cada  $t_i$ . Para esto se utiliza el etiquetador sintáctico (Part Of Speech tagger) de Eric Brill [Brill].
- La consulta resultante del etiquetado es procesada en el tercer modulo, el cual está encargado del reconocimiento de nombres propios, frases de diccionario y stop-words.
- Una vez reconocidos los términos y frases que componen la consulta, el siguiente modulo se ocupará de hallar el sentido correcto de las palabras en base a WordNet. Sea entonces  $[w_1, w_2, w_3, \dots, w_k]$  una lista de sustantivos resultante del modulo anterior, para encontrar el sentido correcto de  $w_i$  se efectuarán los siguientes pasos hasta obtener un resultado:
  - 1) Si  $w_j$  o uno de sus sinónimos se encuentra en la definición de un synset de  $w_i$  digamos S, entonces S será el sentido de  $w_i$ .
  - 2) Para cada sentido de  $w_i$  se recuperan los sinónimos de aquellos synsets que son ancestro común con otros sentidos. El conjunto de sinónimos recibe el nombre de  $SC_i$ . Dada la definición de  $w_i$  se eliminan todas las ocurrencias de las palabras en  $SC_i$ . Se procede de igual forma para  $w_j$ .  
La definición de cada synset de  $w_i$  y  $w_j$  son comparados entre sí. La combinación que tenga el número máximo de palabras contenidas (no stop-words) en común será el sentido para  $w_i$  y también para  $w_j$ .
  - 3) Si  $w_j$  o uno de sus sinónimos aparece en la definición de un synset S conteniendo un hipónimo de  $w_i$ , entonces el sentido de  $w_i$  será el synset  $S_1$  el cual contiene a  $w_i$  y tiene al descendiente S.
  - 4) Sea  $w_i$  en un synset  $S_1$ .  $S_1$  tiene un synset hipónimo U que contiene un término h. Si h aparece en la definición de un synset  $S_2$  que contiene a  $w_j$  entonces el sentido de  $w_i$  será  $S_1$ , y  $S_2$  será el sentido de  $w_j$ .
  - 5) Si no se obtiene un resultado se deja al usuario la decisión.

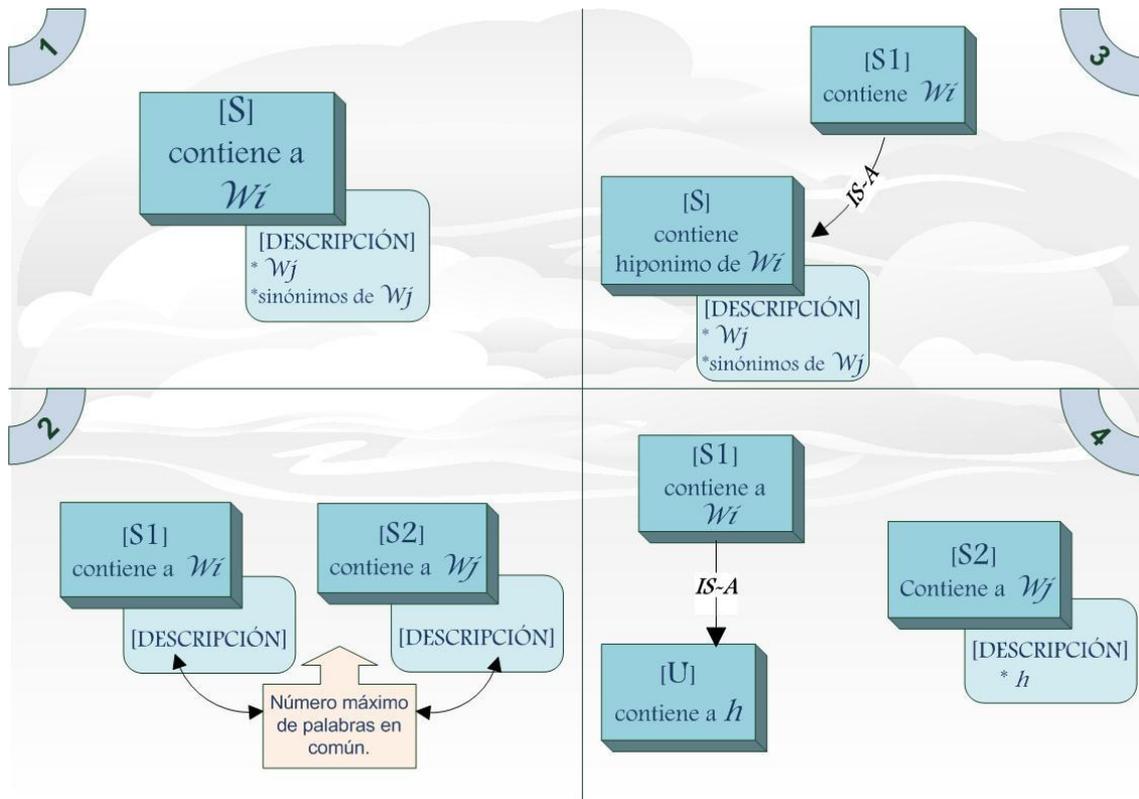


Figura 3.2 – Proceso de desambiguación de un término  $w_i$ .

En la Figura 3.2 se observa gráficamente el proceso de desambiguación de  $w_i$ , donde  $i=1 \dots k$  y  $w_j$  es elegido en base a su proximidad con  $w_i$  en la consulta suponiendo que mientras más relacionados estén dos términos entonces más cerca estarán uno de otro en la consulta. La Figura 3.3 ilustra esta idea.

- El quinto módulo tiene como objetivo expandir la consulta original para la recuperación de documentos que son relevantes pero que no contienen los términos exactos ingresados por el usuario. Son documentos que contienen términos conceptualmente equivalentes a los originales.

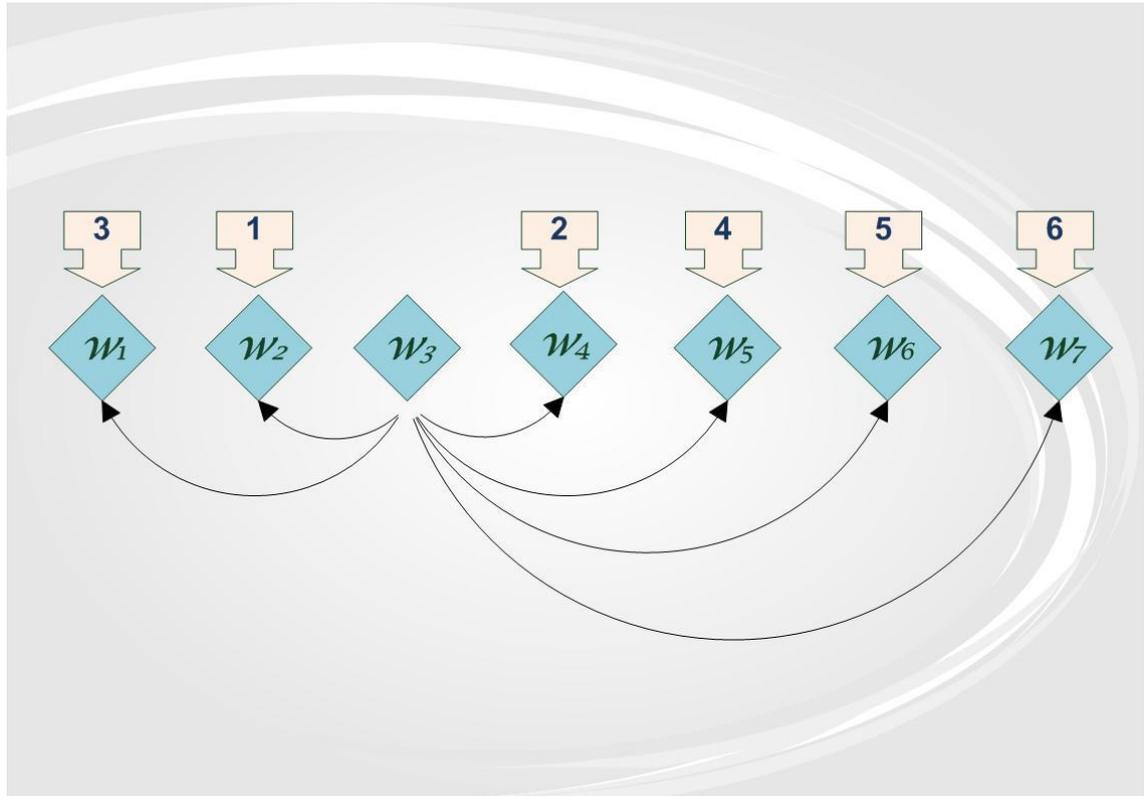


Figura 3.3 – Orden de selección de  $w_j$  para hallar el sentido de  $w_3$ .

- El módulo número seis produce la estrategia de búsqueda en XML. Consiste en la disyunción de las expansiones de cada concepto y luego la conjunción de estas expansiones. Esta estrategia es trasladada a la sintaxis de algún motor de búsqueda en particular en el siguiente módulo.
- El séptimo módulo convierte la estrategia a la sintaxis del motor de búsqueda. Cada repositorio soporta por lo menos un lenguaje de consulta. Ejemplos de los mismos pueden ser PLQL, PLQL1, Lucene [Lucene].

Para ejemplificar todo este proceso consideremos una consulta de usuario sin errores ortográficos en este caso.

- Consulta:

Usar conceptos mas generales

- Etiquetamos los términos en la consulta: [cancer NN], [disease NN]
- Analizamos en busca de frases: no hay frases dentro de la consulta.
- Desambiguación:
  - [cancer]:

Búsqueda de synsets: 5 synsets encontrados

- S11: **<term>**Cancer**</term>**  
**<term>**genus\_Cancer**</term>**  
**<descripcion>**  
type genus of the family Cancridae  
**</descripcion>**
- S12: **<term>**Cancer**</term>**  
**<term>**Cancer\_the\_Crab**</term>**  
**<term>**Crab**</term>**  
**<descripcion>**  
the fourth sign of the zodiac; the sun is in this sign  
from about June 21 to July 22  
**</descripcion>**
- S13: **<term>**Cancer**</term>**  
**<descripcion>**  
a small zodiacal constellation in  
the northern hemisphere; between Leo and Gemini  
**</descripcion>**
- S14: **<term>**Cancer**</term>**  
**<term>**Crab**</term>**  
**<descripcion>**  
(astrology) a person who is born while the sun is in Cancer  
**</descripcion>**
- S15: **<term>**cancer**</term>**  
**<term>**malignant\_neoplastic\_disease**</term>**  
**<descripcion>**  
any malignant growth or tumor caused by abnormal and  
uncontrolled cell division; it may spread to other parts of the  
body through the lymphatic system or the blood stream  
**</descripcion>**

[[disease](#)]:

Busqueda de synsets: 1 synset encontrado.

- S21: **<term>**disease**</term>**  
**<descripcion>**  
an impairment of health or a condition of abnormal functioning  
**</descripcion>**

Desambiguamos [[cancer](#)]:

Paso1: **Descripción S11:** type genus of the family Cancridae  
[disease](#) no se encuentra en la descripción. Pasamos a S2.

**Descripción S12:** the fourth sign of the zodiac; the sun is in this sign  
from about June 21 to July 22

[disease](#) no se encuentra en la descripción. Pasamos a S3.

**Descripcion S13:** a small zodiacal constellation in the northern hemisphere; between Leo and Gemini

**disease** no se encuentra en la descripción. Pasamos a S4.

**Descripcion S14:** (astrology) a person who is born while the sun is in Cancer

**disease** no se encuentra en la descripción. Pasamos a S5.

**Descripcion S15:** any malignant growth or tumor caused by abnormal and uncontrolled cell division; it may spread to other parts of the body through the lymphatic system or the blood stream

**disease** no se encuentra en la descripción. Pasamos al Paso 2.

Paso2: Comparamos

**S11:** type genus of the family Cancriidae

**S21:** an impairment of health or a condition of abnormal functioning

Numero de palabras en comun: 0

Comparamos

**S12:** the fourth sign of the zodiac; the sun is in this sign from about June 21 to July 22

**S21:** an impairment of health or a condition of abnormal functioning

Numero de palabras en comun: 0

Comparamos

**S13:** a small zodiacal constellation in the northern hemisphere; between Leo and Gemini

**S21:** an impairment of health or a condition of abnormal functioning

Numero de palabras en comun: 0

Comparamos

**S14:** (astrology) a person who is born while the sun is in Cancer

**S21:** an impairment of health or a condition of abnormal functioning

Numero de palabras en comun: 0

Comparamos

**S15:** any malignant growth or tumor caused by abnormal and uncontrolled cell division; it may spread to other parts of the body through the lymphatic system or the blood stream

**S21:** an impairment of health or a condition of abnormal functioning

Numero de palabras en comun: 1

Por lo tanto **S15** es el Synset elegido para **cancer**.

Desambiguamos [**disease**]:

Contiene un unico synset **S21**.

- Sinonimos en **S15**: <term>cancer</term>

`<term>malignant_neoplastic_disease</term>`

Sinonimos en **S21**: `<term>disease</term>`

- Generación de la consulta:

`("cancer" OR "malignant neoplastic disease") AND ("disease")`

- Convertimos la consulta a la sintaxis del motor de búsqueda del repositorio **BIOS** en este caso:

```
<simpleQuery>  
  <term>  
    ("cancer" OR "malignant neoplastic disease") AND ("disease")  
  </term>  
</simpleQuery>
```

### 3.2 Procesamiento de documentos recuperados

Debido a que a veces muchos documentos son devueltos y pocos son realmente importantes para el usuario, sería interesante que la colección de documentos devuelta por el motor de búsqueda sea filtrada por un módulo. Su tarea sería determinar cuáles podrían ser relevantes y cuáles no. Con la incorporación de este proceso de análisis de los resultados, es decir un pos-procesamiento de los documentos recuperados, se espera aumentar aún más la precisión de la búsqueda.

Basándonos en el estándar LOM, los objetos recuperados contienen metadatos llamados *keyword* (Figura 3.4) con las palabras claves que definen el objeto de aprendizaje. Sin embargo el motor de búsqueda de los repositorios no priorizan estos metadatos a la hora de realizar las búsquedas, es decir, no hay distinción si una palabra aparece en la descripción, en un *keyword* o en otro metadato. Por esto muchas veces los objetos que son relevantes no se encuentran entre los primeros recuperados. Más aun, existen casos en los cuales los objetos devueltos fueron recuperados en base a aproximaciones de los términos de las consultas (por ejemplo objetos conteniendo *program* en lugar de *programming*) perdiendo de esta forma precisión.

```
<keyword>
  <string language="en-US">cell signaling</string>
</keyword>
<keyword>
  <string language="en-US">stem cells</string>
</keyword>
<keyword>
  <string language="en-US">human diseases</string>
</keyword>
<keyword>
  <string language="en-US">genetics</string>
</keyword>
<keyword>
  <string language="en-US">protein localization</string>
</keyword>
<keyword>
  <string language="en-US">biology</string>
</keyword>
<keyword>
  <string language="en-US">Biology</string>
</keyword>
<keyword>
  <string language="en-US">cell cycle</string>
</keyword>
<keyword>
  <string language="en-US">gene regulation</string>
</keyword>
```

Figura 3.4 – Parte del XML devuelto en una consulta al repositorio Ariadne. Los mismos describen las palabras claves de uno de los objetos recuperados.

El procesamiento de los resultados consistirá en ordenar los objetos comparando los términos de la consulta con las palabras en los *keywords* y las palabras que conforman la descripción del objeto.

Se sumaran puntos por cada coincidencia, sin embargo los términos que ocurran en el título y en los *keywords* sumaran más que los términos hallados en la descripción. Más aún, los puntos también dependerán del tipo de ocurrencia del término dentro del OA.

Por ejemplo un OA cuyo título coincide con un término de la consulta suma 2.5 puntos. Si el término es una frase y se encuentra contenida en el título entonces el OA suma 2 puntos. Y si el término no es una frase pero se encuentra contenido en el título, el OA sumará 1.5 puntos. Caso contrario a todo esto, el OA no recibirá puntos por este término en particular. De manera similar se realizan comparaciones sobre los *keywords* y la descripción.

Además los objetos que no hayan sido recuperados mediante los términos exactos de la consulta serán posicionados en los últimos lugares.

## 4. Implementación

### 4.1 Tecnologías Utilizadas

Para el desarrollo del presente trabajo se ha utilizado tecnología Microsoft, puntualmente Visual Studio 2008 .NET Framework 3.5. El contenido de WordNet fue representado mediante tres archivos de formato XML.

### 4.2 Repositorios Utilizados

La siguiente lista muestra los repositorios que se tomaron en cuenta para realizar las pruebas sin embargo sólo Ariadne resultó ser el más conveniente. Una descripción más detallada de los mismos puede verse en la sección 1.4:

- i) **BIOE** (Banco Internacional Objetos Educativos): Actualmente BIOE permite realizar consultas simples sin posibilidad de consultar por lenguaje o cantidad de elementos que se deseen recuperar.  
Cabe recordar que SQI se encarga de enviar y recibir las consultas, no de la estructura de las mismas por lo que distintos repositorios que utilicen a SQI para su interoperabilidad podrán tener lenguajes de consulta con diferente grado de expresividad.
- ii) **OER Commons**: Este repositorio se basa en el estándar OAI-PMH [OAI] para la recuperación de objetos mediante consultas en la URL. El problema es que las consultas realizadas con este estándar no permiten la búsqueda mediante palabras claves. Las búsquedas se pueden realizar por rango de fecha de creación del objeto, identificador del objeto en el repositorio o conjunto al que pertenece dentro del mismo repositorio entre otras alternativas.
- iii) **arXiv**: Este repositorio también se basa en el estándar OAI-PMH por lo que conlleva las mismas limitaciones que OER Commons.
- iv) **Ariadne**: Este repositorio permite realizar las consultas vía REST o Soap, y es el elegido para realizar las pruebas debido a que ninguno de los otros repositorios disponibles presentaba una interfaz adecuada para las consultas (OER Commons, arXiv) o contenía una cantidad apreciable de objetos de aprendizaje en inglés (BIOE).

### 4.3 Prototipo

Para la implementación del prototipo se ha pensado una arquitectura de capas según se observa en la Figura 4.1. Estas capas están conformadas por clases las cuales cumplen diferentes funciones e interactúan entre sí. Una descripción más detallada de las clases y sus funcionalidades se puede encontrar en la sección 8.1.

La primera capa está conformada por tres clases. Una de las clases, llamada XMLFiles, es la encargada de recuperar la información de WordNet contenida en los archivos XML. Esta clase cuenta con funciones que realizan acciones sobre los archivos Index.xml y Data.xml, dos archivos que contienen los synsets de los sustantivos y las definiciones de los synsets respectivamente. Otra de las clases es ALMACENFrase que permite determinar si una determinada frase forma parte del conjunto de sustantivos que abarca WordNet. Finalmente ALMACENTags recupera el contenido dentro del archivo para determinar el Part Of Speech de los términos de las consultas.

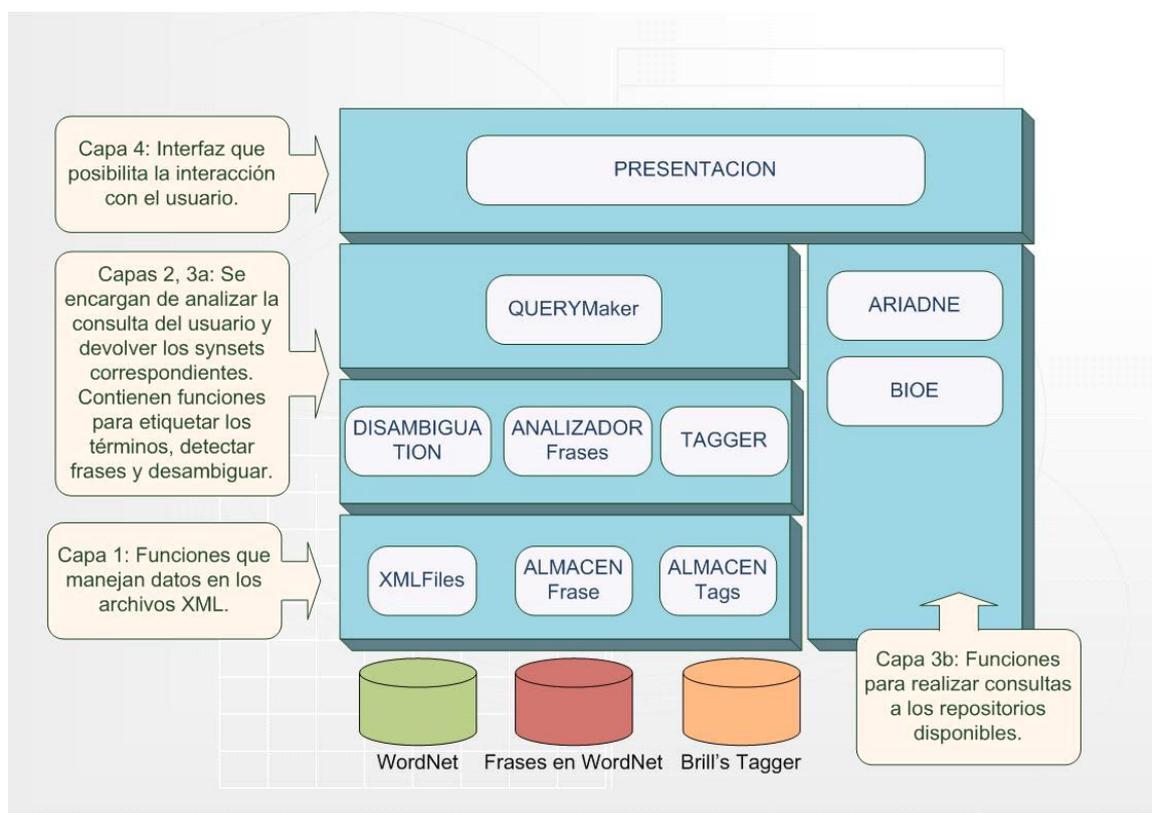


Figura 4.1 – Esquema donde se observa la arquitectura del prototipo.

La segunda capa también está conformada por tres clases. Estas clases utilizan los servicios que proveen las clases de la primera capa. Una de ellas, llamada DISAMBIGUATION, es la que contiene las funciones encargadas de desambiguar los términos de las consultas en base a la información de WordNet. Otra de las clases, ANALIZADORFrases, es la encargada de la detección de frases pertenecientes a WordNet en la consulta. La última clase, TAGGER, contiene funciones para etiquetar semánticamente los términos de las consultas.

La capa 3a está compuesta por una sola clase cuyo objetivo es dividir la consulta en frases y términos, etiquetar las mismas y recuperar los synsets desambiguando aquellos términos que tengan más de un sentido.

La capa 3b está conformada dos clases que corresponden a los repositorios educativos Ariadne y BIOE. Las mismas contienen funciones que permiten realizar búsquedas en dichos repositorios utilizando los web services provistos por esas entidades.

La última capa es la de presentación, una interfaz con la que el usuario puede ingresar las consultas y obtener los resultados de las mismas además de información sobre los términos utilizados en la expansión de la consulta.

En el Apéndice 8.1 se puede observar un resumen de todas las clases involucradas.

## 5. Experimentación

Se ha realizado experimentación ejecutando diferentes consultas sobre el repositorio Ariadne. Las consultas que se utilizaron fueron del tipo simple (solo con palabras ingresadas por el usuario), con sinónimos (compuesta con los sinónimos de las palabras claves) y con hiperónimos (además de los sinónimos, completando la consulta con hiperónimos de las palabras claves).

El análisis se divide de acuerdo a los términos ingresados por el usuario que componen la consulta.

En una primera parte se analizarán los resultados de dos consultas en particular (secciones 5.1 y 5.2). La primera contiene los términos “mathematics” y “calculus”. La segunda consulta está compuesta por “artificial intelligence” y “programming”.

Finalmente en la sección 5.3 se muestran dos tablas comparativas donde se analizan en total 12 consultas, cada una con tres tipos de resultados finales: simples (cuyo resultado citaremos como **Sin Expansión**), expandidas (cuyo resultado llamaremos **Con Expansión**) y, ordenadas y filtradas (cuyo resultado sólo, por abreviar, se mencionará como **Ordenados**).

Para aquellas consultas con más de 20 objetos recuperados se hará un análisis teniendo en cuenta los primeros 10, 15 y 20 elementos recuperados.

Las consultas simples o ‘sin expansión’ son aquellas que no contienen términos adicionales más que aquellos ingresados por el usuario. En las pruebas que se llevarán a cabo supondremos que en dichas consultas se distingue un término simple de una frase. Las consultas se generan en base a este supuesto y las frases pertenecen al conjunto de frases existentes en WordNet. Por ejemplo los términos *Artificial Intelligence Programming* corresponderán a la consulta (*Artificial Intelligence AND Programming*).

### 5.1 Primer análisis

Consulta: *Mathematics Calculus*

01. Sin expansión

*mathematics AND calculus*

La primera consulta recupero una lista de 78 objetos que contienen las palabras claves "mathematics" y "calculus" con distintos grados de relevancia. El cálculo de la misma se basa en la cantidad de palabras claves que encuentra en la definición de cada objeto. A medida que se desciende en la lista la relevancia de los ítems es menor.

Por ejemplo algunos de los primeros ítems recuperados tienen que ver con archivos de audio o lecciones de cálculo matemático en formato de video. En cambio objetos que se encuentran cerca del final de la lista tienen una relación más "distante" al cálculo matemático. Tal es el caso de “**8.022 Physics II: Electricity and Magnetism, Fall 2004 (MIT)**” que es un curso online del MIT sobre electricidad y magnetismo y que requiere conocimientos de cálculo vectorial entre otros. Parte de la lista de objetos recuperados puede verse en Apéndice 8.2.1.

02. Con expansión

*("mathematics" OR "math" OR "maths") AND ("calculus" OR "infinitesimal calculus")*

En la segunda búsqueda se recuperaron 79 objetos. Expandimos los términos de la primera mediante el uso de sinónimos. El uso de los sinónimos "math" y "maths" permitieron recuperar un objeto más que había sido ignorado en la búsqueda anterior. El objeto es un curso online del

MIT sobre física que trata temas como fuerza, gravedad, energía, etc. dirigido a estudiantes con conocimientos de cálculo. Es un objeto de relevancia media lo que condice con la posición en la que aparece en la lista de objetos recuperados. Parte de la lista de objetos recuperados puede verse en Apéndice 8.2.1.

03. Con hiperónimos

*("mathematics" OR "math" OR "maths") AND ("science" OR "scientific discipline") AND ("calculus" OR "infinitesimal calculus") AND ("pure mathematics")*

La tercera búsqueda no logró ningún resultado. Los hiperónimos, añadidos como una conjunción de disyunciones, hacen que la consulta sea más restrictiva. Si los hiperónimos son términos poco comunes existe la posibilidad de que una consulta de esta forma no recupere ningún elemento, como sucede en este caso.

04. Con hiperónimos 2

*("mathematics" OR "math" OR "maths" OR "science" OR "scientific discipline") AND ("calculus" OR "infinitesimal calculus" OR "pure mathematics")*

La cuarta consulta recuperó 96 objetos. En la misma se utilizan los hiperónimos desde otra perspectiva. Agregándolos mediante una disyunción junto con los sinónimos con los cuales están relacionados se genera una consulta más permisiva donde los resultados están relacionados con la consulta original pero abarcando temas más generales. Por ejemplo se pueden observar resultados en los cuales se aplica el cálculo pero en otras áreas de la ciencia tales como la física (Describing motion along a line) o la química.

05. Con hiperónimos 3

*("mathematics" OR "math" OR "maths") AND ("science" OR "scientific discipline") AND ("calculus" OR "infinitesimal calculus")*

En la consulta número cinco se eliminó el componente más restrictivo el cual provocaba que no se recuperaran ítems.

La cantidad de ítems recuperados fue de 17, es decir, hubo 61 objetos que no fueron devueltos porque no cumplieron con la restricción ("science" OR "scientific discipline").

Sin embargo hubo elementos con una relevancia importante que no fueron recuperados. Tal es el caso de los siguientes objetos **"Calculus Podcasts"**, **"Sequences and limits"** o **"Teach yourself limits"**. Y también hubo objetos no relevantes que sí fueron devueltos, como por ejemplo **"14.12 Economic Applications of Game Theory, Fall 2005 (MIT)"**.

Se puede concluir que la consulta número 2 es la más apropiada. Recupera más elementos que la consulta 1 gracias a la inclusión de sinónimos. No presenta el problema de hiperónimos restrictivos que dan como resultado una devolución nula de objetos como en la consulta 3. La consulta 4 no es de utilidad ya que recupera muchos objetos no relevantes lo cual provoca una disminución de la precisión. Por otro lado la consulta 5 supone saber cuál es el termino hiperónimo que restringe una consulta de tipo 3 por lo que sería práctica.

## 5.2 Segundo análisis

Consulta: *Artificial Intelligence Programming*

01. Sin expansión

*"artificial intelligence" AND "programming"*

La primera consulta recupero una lista de 53 objetos que contienen las palabras claves "artificial intelligence" y "programming" con distintos grados de relevancia. Nuevamente a medida que se desciende en la lista la relevancia de los ítems es menor.

Por ejemplo algunos de los primeros ítems recuperados son "**Perspective on Artificial Intelligence Programming**" y "**6.871 Knowledge-Based Applications Systems, Spring 2005 (MIT)**". El primero de ellos aborda el tema del uso de lenguajes especializados, entornos y herramientas basadas en el conocimiento para la programación de sistemas para aplicaciones que utilizan técnicas de inteligencia artificial, mientras que el segundo es un curso del MIT sobre el desarrollo de programas que contienen una cantidad apreciable de conocimiento sobre el dominio de la aplicación, es decir, sistemas expertos.

En cambio objetos que se encuentran cerca del final de la lista contienen información poco relevante o poco relacionada al objetivo de la búsqueda.

Tal es el caso de **Collage, Technology, and Creative Process** el cual poco tiene que ver con la inteligencia artificial y la programación y más con el arte. Parte de la lista de objetos recuperados puede verse en Apéndice 8.2.2.

02. Con expansion

*("artificial intelligence" OR "AI") AND ("programming" OR "programing" OR "computer programming" OR "computer programing")*

En la segunda búsqueda expandimos los términos de la primera mediante el uso de sinónimos. Principalmente con el término "AI" el número de objetos relevantes recuperados aumentó considerablemente. Por ejemplo "**The Logical Basis for Computer Programming : Deductive Reasoning**" y "**Logic Programming**" son dos objetos recuperados con una relevancia importante. El primero es una introducción a las nociones básicas de lógica de predicados y proposicional, mientras que el segundo es un curso sobre programación lógica. (Sin embargo hubo elementos recuperados que no estaban relacionados con el tema. Estos elementos fueron filtrados en la etapa de pos procesamientos ya que los mismos no verificaban la segunda cláusula de la consulta). Parte de la lista de objetos recuperados puede verse en Apéndice 8.2.2.

03. Con hiperónimos

*("artificial intelligence" OR "AI") AND ("computer science" OR "computing") AND ("programming" OR "programing" OR "computer programming" OR "computer programing") AND ("creating by mental acts")*

La tercera búsqueda no logro ningún resultado. Los hiperónimos, añadidos como una conjunción de disyunciones, hacen que la consulta sea más restrictiva. Si los hiperónimos son términos poco comunes existe la posibilidad de que una consulta de esta forma no recupere ningún elemento, como sucede en este caso.

04. Con hiperónimos 2

*("artificial intelligence" OR "AI" OR "computer science" OR "computing") AND ("programming" OR "programing" OR "computer programming" OR "computer programing" OR "creating by mental acts")*

Claramente esta última consulta es la que más recupera elementos indeseados. Una gran cantidad está relacionada con la programación en general como es en el caso de "**6.189 A Gentle Introduction to Programming Using Python, January (IAP) 2008 (MIT)**" o "**Programming in PASCAL**".

05. Con hiperónimos 3

*("artificial intelligence" OR "AI") AND ("computer science" OR "computing") AND ("programming" OR "programing" OR "computer programming" OR "computer programing")*

En la consulta número cinco se eliminó el componente más restrictivo el cual provocaba que no se recuperaran ítems.

La cantidad de ítems recuperados fue de 55, es decir, hubo 25 objetos que no fueron devueltos (en relación con la búsqueda 02) porque no cumplieron con la restricción ("science" OR "scientific discipline").

Sin embargo hubo elementos con una relevancia importante que no fueron recuperados. Tal es el caso de los siguientes objetos "**The Handbook of Artificial Intelligence**" o "**Logic Programming**".

Asimismo varios de los objetos devueltos tenían poca relevancia. Por ejemplo "**21W.765J Interactive and Non-Linear Narrative: Theory and Practice, Spring 2004 (MIT)**" es un curso del MIT que tiene que ver más que nada con la literatura, la narración de historias y novelas. Parte de la description se muestra a continuación:

Analysis of the structure and evaluation of the literary qualities of computer-based narratives including hypertexts, adventure games, and classic **artificial intelligence** programs like "Eliza".... Knowledge of **programming** helpful but not necessary...

Nuevamente vemos que la consulta número 2 es la más conveniente. Recupera más elementos relevantes que la consulta 1 gracias a la inclusión de sinónimos y no tiene las limitaciones de la consulta 3, la cual una vez más no recupera ningún elemento. La consulta nuevamente retorna muchos elementos irrelevantes y provoca una disminución de la precisión.

### **5.3 Tablas Comparativas**

En las siguientes tablas se lista una serie de consultas y valores correspondientes al número de objetos recuperados, la relevancia de los mismos, la precisión y el número de documentos recuperados por cada tipo de consulta (reemplazando el *recall* ya que no se sabe cuántos documentos relevantes existen para una consulta dada).

Debido a que el número de elementos devueltos por cada consulta varía demasiado, se presentan dos tipos de tablas para realizar las comparaciones de los resultados.

La primera tabla contiene los resultados de menos de 20 elementos recuperados y se calcula la precisión sobre el total de elementos obtenidos.

La segunda tabla contiene aquellas consultas que recuperaron más de 20 objetos. En esta tabla la precisión se calcula sobre los primeros 10, 15 y 20 elementos recuperados.

Posteriormente se realiza un análisis de los resultados.

Relevantes Sin Expansión = RSE  
 Relevantes Con Expansión = RCE  
 Relevantes Ordenados = RO

Precisión Sin Expansión = PSE  
 Precisión Con Expansión = PCE  
 Precisión Ordenados = PO

Número	A	B	C	D	E	F	G
Tema	plate tectonics	encryption AND algorithm	"computer science" AND "game theory"	"mathematical logic" AND "computer science"	cancer AND disease AND treatment	"solar energy" AND "photovoltaic cell"	"communications protocol" AND tcp
Consulta Expandida	("plate tectonics" OR "plate tectonic theory")	("encryption" OR "encoding") AND ("algorithm" OR "algorithmic rule" OR "algorithmic program")	("computer science" OR "computing") AND ("game theory" OR "theory of games")	("mathematical logic" OR "symbolic logic" OR "formal logic") AND ("computer science" OR "computing")	("cancer" OR "malignant neoplastic disease") AND ("disease") AND ("treatment" OR "medical care" OR "medical aid")	("solar energy" OR "solar power") AND ("photovoltaic cell" OR "solar cell")	("communications protocol" OR "protocol") AND ("tcp" OR "transmission control protocol" OR "TCP")
Recuperados Sin Expansión	10	2	2	1	9	1	2
Recuperados Con Expansión	10	22	7	9	9	9	12
Total Ordenados	10	10	3	2	9	9	8
ReISE	8	1	1	9	9	1	1
ReICE	8	1	3	9	9	9	5
ReIO	8	1	2	9	9	9	4
PreSE	0,8	0,5	0,5	1	1	1	0,5
PreCE	0,8	0,05	0,428	1	1	1	0,416
PreO	0,8	0,1	0,667	1	1	1	0,5

Tabla 5.1: muestra los valores de precisión calculados sobre un máximo de 20 elementos recuperados.

En la consulta A, al realizar el procesamiento de la consulta simple se obtuvieron como sinónimos de "plate tectonics" los términos "tectonics" y "plate tectonic theory". Una vez recuperados los elementos y procesados para su ordenamiento resultó que el número de elementos relevantes en Ordenados era menor que en Recuperados Sin Expansión y Recuperados Con Expansión.

El motivo por el cual Ordenados disminuía era debido al sinónimo "tectonics" que fue añadido a la consulta al ser parte de los sinónimos del synset (06118370) correspondiente a "plate tectonics". Sin embargo "tectonics" también pertenece a otro synset (06124395) relacionado con la ciencia de la arquitectura. Por este motivo objetos con contenido de este tipo (relacionado a la arquitectura) eran etiquetados con un importante nivel de relevancia por Ariadne. Este problema surge principalmente en consultas cortas donde la misma consta de un solo término o frase en este caso. Una solución a este problema es la eliminación de términos sinónimos pertenecientes a más de un synset cuando la consulta está compuesta por un solo término o frase.

La consulta expandida original era ("tectonics" OR "plate tectonics" OR "plate tectonic theory").

En B, la consulta expandida no tuvo mayor éxito que la simple. El motor de búsqueda del repositorio recupero objetos que no satisfacían la consulta. La causa de la baja precisión en los resultados de la consulta expandida se debe en parte al motor de búsqueda del repositorio ya que el mismo recupera objetos que no verifican la consulta realizada. Este inconveniente es solucionado con la etapa de **Ordenamiento y Filtrado** al descartar aquellos objetos que no verifican la clausula (encryption AND encoding).

Sin embargo todavía quedan ocho elementos que no son relevantes pero verifican la consulta. Es posible que parte del motivo sea que *encoding* no es un término tan relacionado a *algorithm* como lo es *encryption*. Estos objetos son recuperados por la clausula (*encoding* AND (*algorithm* OR ...)).

Por otro lado hay que tener en cuenta la cantidad de elementos relevantes vinculados con la consulta. Es probable que los dos elementos recuperados en la consulta simple sean los únicos relevantes en la base de datos. Por lo tanto aumenta la posibilidad de que una consulta expandida recupere objetos indeseados.

En E los sinónimos no recuperaron ningún objeto adicional. Especialmente los sinónimos de *treatment* ("*medical care*" OR "*medical aid*") no tuvieron impacto en la recuperación.

En la consulta F el sinónimo de *photovoltaic cell* (*solar cell*) hizo la diferencia en la recuperación de objetos relevantes. Hubo ocho elementos que fueron recuperados por el uso del sinónimo *solar cell*.

En G, la consulta simple solo devolvió dos elementos. Y la inclusión del sinónimo *protocol* de *communication protocols* permitió la recuperación de diez elementos adicionales.

La escasa cantidad de elementos recuperados en la consultas se debe en parte a que las mismas están compuestas por frases lo cual genera que las búsqueda sean más acotadas.

En G por ejemplo la consulta simple, compuesta por '*communication protocols*' arrojó solamente dos resultados mientras que una consulta sin frases de la forma '*communication protocols tcp*' devuelve más de 6800 elementos.

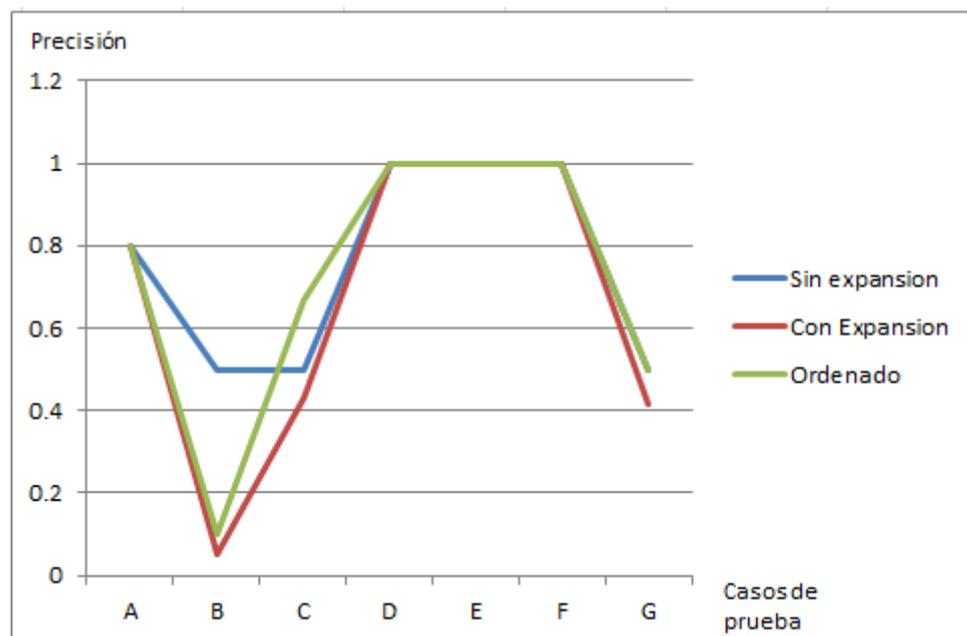


Figura 5.1: Precisión de menos de 20 documentos recuperados para las 12 consultas.

En la Figura 5.1 se muestra una gráfica donde se comparan los valores de precisión de la Tabla 5.1. En general los valores de **Con Expansión** parecen estar por debajo de los de **Sin Expansión**.

Por otro lado **Ordenado** está por arriba de **Sin Expansión** salvo en consultas donde la cantidad de elementos recuperados es muy pequeña como en el caso de la consulta número 2.

Para realizar una comparación más detallada de la precisión realizamos un análisis de aquellas que devuelven al menos 20 elementos y realizamos el cálculo sobre 10, 15 y 20 elementos recuperados. El resultado de esta aproximación se muestra en la siguiente tabla.

Número	1	2	3	4	5	
Tema	biology AND evolution	biology AND ecology	"artificial intelligence" AND programming	mathematics AND calculus	"operating system" AND unix	
Consulta Expandida	("biology" OR "biological science") AND ("evolution" OR "organic evolution" OR "phylogeny" OR "phylogenesis")	("biology" OR "biological science") AND ("ecology" OR "bionomics" OR "environmental science")	("artificial intelligence" OR "AI") AND ("programming" OR "programing" OR "computer programming" OR "computer programing")	("mathematics" OR "math" OR "maths") AND ("calculus" OR "infinitesimal calculus")	("operating system" OR "OS") AND ("UNIX" OR "UNIX system" OR "UNIX operating system")	
Recuperados Sin Expansión	91	250	53	78	37	
Recuperados Con Expansión	94	255	80	79	37	
Total Ordenados	35	35	34	37	26	
10	RelSE	6	4	8	10	5
	RelCE	6	6	9	10	6
	RelO	9	8	10	10	8
	PreSE	0,6	0,4	0,8	1	0,5
	PreCE	0,6	0,6	0,9	1	0,6
	PreO	0,9	0,8	1	1	0,8
15	RelSE	10	5	13	15	9
	RelCE	8	9	14	15	7
	RelO	12	10	13	15	11
	PreSE	0,67	0,33	0,87	1	0,6
	PreCE	0,53	0,6	0,93	1	0,47
	PreO	0,8	0,67	0,87	1	0,73
20	RelSE	14	9	16	20	10
	RelCE	13	11	18	18	10
	RelO	16	13	15	18	13
	PreSE	0,7	0,45	0,8	1	0,5
	PreCE	0,65	0,55	0,9	0,9	0,5
	PreO	0,8	0,65	0,75	0,9	0,65

Tabla 5.2: muestra los valores de precisión obtenidos para aquellas consultas con más de 20 elementos recuperados.

En la consulta 1 aparece un sinónimo clave de *evolution* en la descripción de uno de los objetos: *phylogeny*. Este objeto es desechado en la etapa de ordenamiento debido a que no contiene ocurrencias de ninguno de los términos que conforman la primer clausula (“*biology*” OR “*biological science*”).

Si tomamos la consulta 3 como ejemplo vemos que con la consulta expandida se recuperaron 24 elementos más que con la consulta simple (27 si contamos los repetidos).

De esos elementos, 2 no tenían relación con el tema. Los mismos cuentan con descripciones en varios idiomas, incluido el italiano, y al ser “*ai*” una palabra muy común en ese idioma los objetos fueron devueltos por el motor de búsqueda. Sin embargo al realizar el post procesamiento de los resultados se analiza la descripción en ingles la cual no contenía ningún término perteneciente a la consulta por lo que fueron desechados.

Luego 15 de esos elementos contenían información relacionada pero no eran relevantes; la mayoría trataba temas relacionados con la IA pero no desde el punto de vista de la programación.

En general el ordenamiento de estos objetos provoco que varios elementos con relevancia fueran posicionados más cerca de los primeros 10. Tal es el caso de “Logic Programming” que pasó de la posición 51 a la 5, “Concurrent Prolog: A Progress Report” que pasó de la posición 21 a la 12, o “Rule- Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project” que paso de la posición 72 a la 17.

El objeto “The Logical Basis for Computer Programming: Deductive Reasoning” ya se encontraban dentro de los 10 primeros y mediante el ordenamiento se acercó aún más a la cima.

En la consulta 4 varios de los 10 primeros elementos recuperados en la consulta simple y también en la consulta sin ordenar contienen material en otro idioma o no son suficientemente descriptivos. Muchos no contienen el termino *mathematics*, sin embargo por la naturaleza misma de la consulta, matemática es un término tan relacionado con calculo que resulta de poca importancia si el mismo aparece o no. Sin embargo esto nos lleva a otro punto; mas allá de que todos contienen material sobre cálculo matemático, no todos tienen la misma importancia o calidad de información. Los elementos en RO (relevantes ordenados) tienen información más detallada e interesante que aquellos encontrados en RSE (relevantes sin expansión) y RCE (relevantes con expansión).

Si observamos los valores de la precisión de los primeros 10 elementos vemos que en la mayoría de los casos aumenta al realizar la expansión de la consulta y especialmente al ordenarla. También vemos que los valores de precisión para **Con Expansión y Ordenada** superan en más de 10% y 26 % respectivamente a **Sin Expansión**.

La siguiente tabla muestra la cantidad promedio de objetos recuperados y la precisión promedio de las 5 consultas de la tabla, calculadas para cada tipo de consulta:

		Sin Expansión	Con Expansión	Ordenada
Cantidad promedio de documentos recuperados		101,8	109	33,4
Precisión promedio (%)	10 >	0,66	0,74	0,9
	15 >	0,694	0,706	0,814
	20 >	0,69	0,7	0,75

Tabla 5.3: muestra los valores promedios de elementos recuperados y la precisión de cada tipo de consulta por cantidad de elementos recuperados.

Si comparamos las precisiones resultantes en la tabla anterior vemos que **Con Expansión y Ordenada** disminuyen en un 4,59% y 9,55% respectivamente al pasar de 10 a 15 el número de documentos recuperados, mientras que **Sin Expansión** aumentó en más de un 5% su precisión.

En las siguientes figuras se pueden apreciar mejor estos resultados.

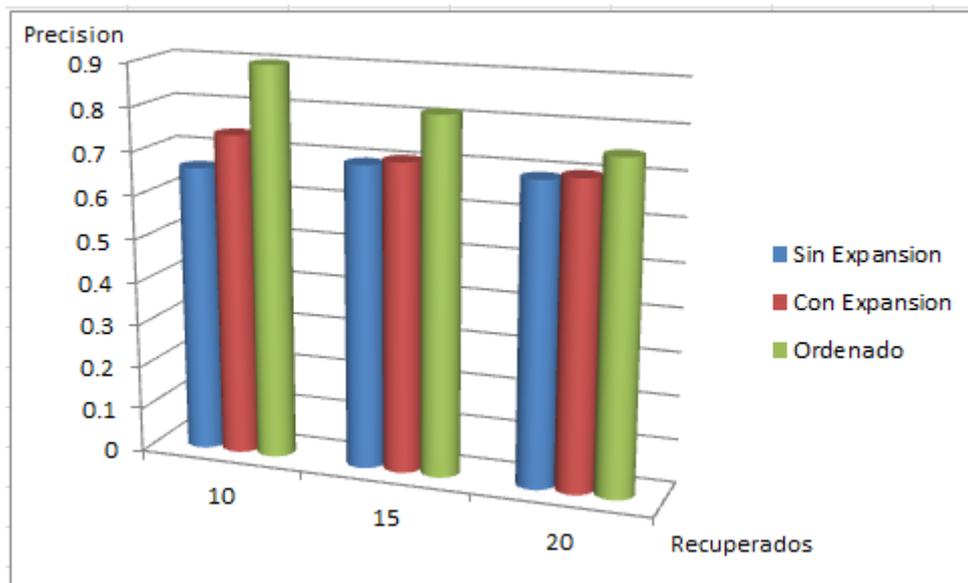


Figura 5.2: Comparación en barras de las precisiones de los 10, 15 y 20 elementos recuperados

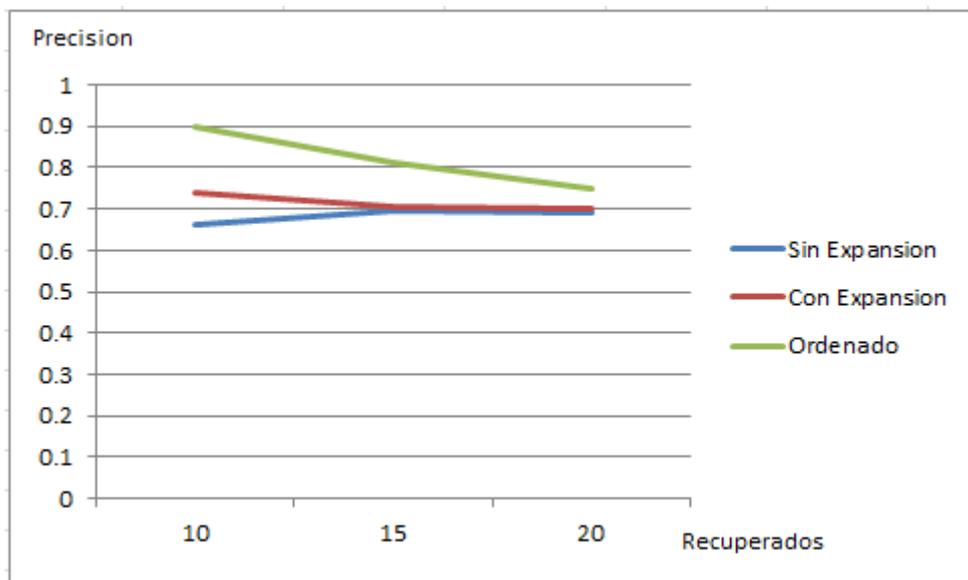


Figura 5.3: Comparación en líneas de las precisiones de los 10, 15 y 20 elementos recuperados

Se observa que a pesar de haber una disminución importante en la precisión de **Ordenado** y **Con Expansión** en los últimos 20 objetos recuperados, los valores siguen siendo mayores en un 1,43 y 8 % respectivamente que la precisión en **Sin Expansión**. En particular se advierte una mayor diferencia con los resultados de **Ordenado** donde, además de tener una mayor precisión, la calidad de los objetos relevantes recuperados es superior a sus pares devueltos por **Sin Expansión** (entendiendo por calidad al nivel de utilidad de la información contenida en el objeto).

## **6. Discusiones**

En este trabajo se ha presentado la propuesta de un sistema de búsqueda semántica de objetos de aprendizaje en repositorios educacionales. Como recurso lingüístico se ha utilizado WordNet, una base de datos léxica en inglés, y como repositorio se ha elegido Ariadne, que cuenta con un número de objetos en su mayoría en inglés y bajo el estándar lom. La implementación se ha desarrollado bajo el framework 3.5 de .Net y se han utilizado los web services que provee el repositorio para hacer la conexión con el mismo.

En las siguientes secciones se discutirán las conclusiones resultantes.

### **6.1 Repositorio**

De todos los repositorios analizados, Ariadne fue el más conveniente para la realización de las consultas. Tiene el mayor número de objetos (la mayoría en inglés), mayor cantidad de metadatos bajo el estándar lom en comparación con otros repositorios, y su uso es gratuito.

Sin embargo hubo inconvenientes durante las pruebas realizadas con Ariadne. Encontramos que varios de los objetos de aprendizaje de los repositorios no reciben mantenimiento, existen duplicados, links rotos o modificados, y otros que ni siquiera poseen una dirección que los identifique.

Otro dato no menor es que al realizar las consultas en varias ocasiones, hubo resultados que no verificaban alguna de las cláusulas que componían a la misma.

### **6.2 Pruebas de la Implementación**

Ya sea de forma automática o semiautomática, el proceso de expansión de la consulta generó resultados más precisos dándole también al usuario, mediante el ranking, una idea de la relevancia de los objetos.

Los hiperónimos presentaron ventajas en ciertos casos. Algunas veces restringieron de forma moderada los resultados obtenidos incrementando la precisión. Sin embargo otras veces condicionaron las consultas a tal punto que los resultados fueron nulos.

Se observó que la precisión de los resultados aumentaba en mayor medida en aquellos casos donde la consulta recuperaba más de 20 elementos. En los casos donde el número de resultados era mínimo en general la precisión era la misma tanto para las consultas simples como para las expandidas.

### **6.3 Limitaciones**

Las limitaciones pueden surgir por parte de los recursos lingüísticos como se discutió anteriormente. En este caso WordNet presenta falencias al momento de vincular términos altamente relacionados. Al observar la estructura de WordNet y las relaciones entre los elementos que la componen se verifican los vacíos existentes entre esos términos (por ejemplo *cancer* y *doctor* son palabras relacionadas en el ámbito de la medicina sin embargo es difícil encontrar una relación dentro de la estructura de WordNet).

Asimismo aunque WordNet contiene una gran variedad de palabras comunes, no cubre el vocabulario específico de cada dominio.

Además el reconocimiento de las frases ingresadas por el usuario está limitado a aquellas contenidas en WordNet.

También se notó escasa disponibilidad de servicios gratuitos prestados por los repositorios educacionales para realizar las pruebas.

#### ***6.4 Trabajo futuro***

- Sería interesante la inclusión de otros recursos lingüísticos tales como ontologías o tesauros. Contar con una ontología de dominio general nos permitiría inferir nuevo conocimiento y llenar los vacíos que deja WordNet.
- Además de los recursos se podría pensar en la conexión con más repositorios educacionales. El objetivo sería brindar al usuario más opciones de búsqueda y una mayor calidad en los resultados.

## 7. Bibliografía

[ADL, (2002)] ADL. (2002). Emerging and Enabling Technologies for the design of Learning Object Repositories Report. Advanced Distributed Learning Initiative. Recuperado el 15 de marzo de 2005, de <http://xml.coverpages.org/ADLRepositoryTIR.pdf>.

[ADL (2005)] ADL. (2005). Sharable Content Object Reference Model. Recuperado el 10 de marzo de 2005, de <http://www.adlnet.org/scorm/index.cfm>.

[Ahmed, Ayers, Birbeck, Cousins, Dodds, Lubell, et al., (2001)] Ahmed, K., Ayers, D., Birbeck, M., Cousins, J., Dodds, D., Lubell, J., Nic, M., Rivers-Moore, D., Watt, A., Worden, R. & Wringhtson, A. (2001). *Professional XML Meta Data*. Bimingham: UK Press.

[Ariadne] Ariadne (Alliance of Remote Instructional Authoring and Distribution Networks for Europe), <http://www.ariadne-eu.org>

[arXiv] <http://arxiv.org/>

[BIOE] <http://objetoseducacionais2.mec.gov.br/>

[Bray, Paoli, Sperberg-MacQueen, Maler & Yergeau, (2004)] Bray, T., Paoli, J., Sperberg-MacQueen, C. M., Maler, E., & Yergeau, F. (Eds.) . (2004). *Extensible Markup Language (XML) 1.0* . 3a. edición. W3C. Consultado el 28 de marzo de 2005, en <http://www.w3.org/TR/REC-xml>.

[Brill] [http://en.wikipedia.org/wiki/Brill\\_tagger](http://en.wikipedia.org/wiki/Brill_tagger)

[Broekstra et al. (2002)] Broekstra, J. Klein, M., Decker, S., van Harmelen, F., Horrocks, I. Enabling knowledge representation on the Web by extending RDF Schema. En *Computer Networks* 39, pp. 609-634, 2002.

[Caplan (2003)] Caplan, P. (2003). *Metadata Fundamentals for All Librarians*. Chicago : American Library Association.

[CETIS (2003)] CETIS. (2003). *Why are Learning Standards so Important?* Recuperado el 3 de abril de 2005, de <http://metadata.cetis.ac.uk/static/why.html>.

[(CETIS, s.f.)] CETIS. (s.f.). *Standards-Compliant Products Directory* . Recuperado el 5 de abril de 2005, de <http://www.cetis.ac.uk/directory>.

[Coad & Jill (1993)] Coad, P. & Jill, N. (1993). *Object Oriented Programming* . New Jersey: Yourdon Press.

[Downes (2004)] Downes, S. (2004). *The Learning Marketplace. Meaning, Metadata and Content Syndication in the Learning Object Economy*. Moncton, New Brunswick: el autor. Recuperado el 19 de abril de 2005, de <http://www.downes.ca/files/book3.htm>.

- [**Duncan (2003)**] **Duncan, C.** (2003). Granularization. En A. Littlejohn (Ed.), *Reusing Online Resources: A Sustainable Approach to E-learning* (pp 12-19). Open & Flexible Learning Series. London , Sterling : VA Taylor & Francis
- [**Fernandez & Serrano**] Fernandez, Serrano. *Utilizando Recursos Lingüísticos para La Mejora de la Recuperacion de Informacion en la Web.*
- [**García (2000)**] **García, F. J.** (2000). *Modelo de Reutilización Soportado por Estructuras Complejas de Reutilización Denominadas Mecanos* . Colecciones Vitor; 53. Salamanca: Ediciones Universidad de Salamanca.
- [**González (2005)**] **González, M.** (2005). Cómo Desarrollar Contenidos para la Formación on Line Basados en Objetos de Aprendizaje. *Revista de Educación a Distancia* , IV, Monográfico III. Recuperado el 4 de abril de 2005, de <http://www.um.es/ead/red/M3/>.
- [**Gonzalo et al. 1998**] J. Gonzalo, M. F. Verdejo, I. Chugur, F. Lopez, A. Peñas. “Extraccion de relaciones semánticas entre nombres y verbos en EuroWordNet”. Actas de SEPLN98, 1998.
- [**Gruber**] **Gruber T.,** *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. Technical Report KSL-93-04, Knowledge System Laboratory, Standford University, CA, 1993.
- [**Guzmán**] **Clara Guzmán.** *LORs como soporte para entornos e-learning.* <http://www.biblioweb.dgsca.unam.mx/libros/repositorios/contenido.htm>
- [**Hatala, Richards, Eap & Willms, (2004)**] **Hatala, M., Richards, G., Eap., T. & Willms, J.** (2004). Interoperability of Learning Object Repositories and Services: Standards, Implementations and Lessons Learned. *Proceedings of the 13 th WWW Conference* . Recuperado el 22 de mayo de 2005, de <http://www.www2004.org/proceedings/docs/2p19.pdf>.
- [**IEEE (2001)**] **IEEE.** (2001). *Learning Object Metadata Working Group* . Recuperado el 14 de diciembre de 2004, de <http://ltsc.ieee.org/wg12/index.html>.
- [**IEEE, (2002)**] **IEEE.** (2002). *1484.12.1 Standard for Learning Object Metadata* . ANSI/IEEE.
- [**JORUM+ Project (2004)**] **JORUM+ Project.** (2004). *The JISC Online Repository for [learning and teaching] Materials* . Recuperado el 5 de marzo de 2005, de [http://www.jorum.ac.uk/docs/Vol1\\_Fin.pdf](http://www.jorum.ac.uk/docs/Vol1_Fin.pdf).
- [**Leslie, Landond, Lamb, & Poulin, (2004)**] **Leslie, S., Landond, B., Lamb, B. & Poulin., R.** (2004). *Learning Object Repository Software* . EduTools. Recuperado el 21 de abril de 2005, de [http://www.edutools.info/documents/lor/2004/final\\_report.doc](http://www.edutools.info/documents/lor/2004/final_report.doc).
- [**Liu, Liu, Yu, Meng**] **Shuang Liu, Fang Liu, Clement Yu, Weiyi Meng** (2004). *An Effective Approach to Document Retrieval via Utilizing WordNet and Recognizing Phrases.*

[**López (2004)**] **López, C.** (2004) *Reflexiones sobre los Estándares en e-Learning* . eCampus UNIACC. Recuperado el 10 de mayo de 2005, de [http://webct.uniacc.cl/Textos/tecnologia/clara\\_lopez/claralopez.pdf](http://webct.uniacc.cl/Textos/tecnologia/clara_lopez/claralopez.pdf).

[**López, García & Pernías, (2005)**] **López C., García, F. J. & Pernías, P.** (2005). Desarrollo de Repositorios de Objetos de Aprendizaje a través de la Reutilización de los Metadatos de una Colección Digital: de Dublin Core a IMS. *RED Revista de Educación a Distancia* , año IV, monográfico II. <http://www.um.es/ead/red/M2/>.

[**Loupy & El-Bèze**] Claude de Loupy, Marc El-Bèze. Managing Synonymy and Polysemy in a Document Retrieval System Using WordNet.

[**Lowerison, Gallart & Boyd, (2003)**] **Lowerison G., Gallant, G. & Boyd, G.** (2003). Learning Objects in Distance Education: Addressing issues of Quality, Learner Control and Accessibility. *Proceedings of the 2003 CADE-ACED Conference* . Recuperado el 25 de agosto de 2004, de [http://www.cade-aced2003.ca/conference\\_proceedings/Gallant.pdf](http://www.cade-aced2003.ca/conference_proceedings/Gallant.pdf)

[**Lucene**] [http://lucene.apache.org/java/2\\_0\\_0/queryparsersyntax.html](http://lucene.apache.org/java/2_0_0/queryparsersyntax.html)

[**Mandala et. al**] Mandala, Takenobu, Hozumi. *The Use of WordNet In Information Retrieval*.

[**Masie Center Learning Consortium (2003)**] **Masie Center Learning Consortium.** (2003). *Making Sense of Learning Specifications & Standards: A Decision Maker's Guide to their Adoption* . (2ª Ed). Recuperado el 12 de marzo de 2004, de [http://www.staffs.ac.uk/COSE/cosenew/s3\\_guide.pdf](http://www.staffs.ac.uk/COSE/cosenew/s3_guide.pdf).

[**Mihalcea**] **R. Mihalcea**, *Word Sense Disambiguation and its application to Internet search*, Master Thesis School of Engineering and applied Science, Southern Methodist University, 1999.

[**Moldovan & Mihalcea**] Moldovan, Mihalcea. *Improving the Search on the Internet by Using WordNet and Lexical Operators*.

[**Morales & García (2005)**] **Morales, E. & Garcia F.** (2005). Quality Content Management for e-Learning: General Sigues for a Decisión Support System. *7 th International Conference on Enterprise Information Systems* . In Press

[**Navigli, Velardi**] **Navigli R., Velardi P.**, *An Analysis of Ontology-based Query Expansion Strategies*. Dipartimento di Informatica. Università di Roma "La Sapienza".

[**OAI**] <http://www.openarchives.org/OAI/openarchivesprotocol.html>

[**Olmos & Ballesteros**] **Olmos, K. & Ballesteros, B.** *Clasificación y Organización de los OA*. <http://www.slideboom.com/presentations/127717/Clasificación-y-organización-de-los-objetos-de-aprendizaje>

[**Rehak & Mason, (2003)**] **Rehak, D. & Mason, R.** (2003). Keeping the Learning in Learning Objects. En A. Littlejohn (Ed), *Reusing Online Resources: A Sustainable Approach to E-learning* (pp 20-34). London: Sterling, VA Taylor & Francis.

[**Resnik**] Philip Resnik. *Disambiguating Noun Groups With Respect To WordNet Sense*.

[**SQI**] <http://es.wikipedia.org/wiki/SQI>

[**Studer**] **Studer S., Benjamin R., Fensel D.**, *Knowledge Engineering: Principles and Methods*. Data and Knowledge Engineering, 1998. 25: pp. 161 - 197.

[**Tesauros&Ontolog.**] *Tesauros y Ontologías*.

<http://personales.upv.es/ccarrasc/doc/2003-2004/TesaurosOnto/principal.html>

[**Toshniwal & Agrawal, (2004)**] **Toshniwal, R. & Agrawal, D. P.** (2004). Tracing the Roots of Markup Languages. *Communications of the ACM* . Vol 47, No. 5, 95-98.

[**Voorhees**] **Ellen Voorhees**. *Using WordNet to disambiguate word senses for text retrieval*. <http://www.cs.nott.ac.uk/~bsl/G52HPA/articles/Voorhees:93a.pdf>

[**W3C**] [www.w3c.org](http://www.w3c.org)

[**Wiley (2000)**] **Wiley, D. A.** (2000). Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy. En D.A. Wiley (Ed.) *The Instructional Use of Learning Objects* . Recuperado el 14 de diciembre de 2004, de <http://reusability.org/read/chapters/wiley.doc>.

[**WordNet**] George A. Miller (1995). WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.

Christiane Fellbaum (1998, ed.) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Princeton University "About WordNet." WordNet. Princeton University. 2010. <http://wordnet.princeton.edu>

## 8. Apéndice

### 8.1 Clases Utilizadas.

En esta sección se describen brevemente las clases utilizadas en el prototipo y sus funciones principales.

#### CAPA DATOS (Nivel 0)

##### Clase XMLFiles

- Función **ObtenerIndices**: **Entrada** → List<string>, List<string>  
**Salida** → List<string> / Null

Las entradas de esta función son dos listas de strings que representan la forma plural y singular de una frase. Dichas listas también pueden representar un solo término por lo que en ese caso ambas listas contendrán un solo elemento.

La salida será una lista de strings que representa una lista de synsets o un valor nulo en caso de que no se encuentre synsets para la entrada.

- Función **ObtenerData**: **Entrada** → List<string>  
**Salida** → List<Data>

Dado un conjunto de identificadores de synsets se realiza una búsqueda de los synsets correspondientes. Estos identificadores son válidos ya que provienen de **ObtenerIndices**. La salida es una lista de todos los synsets los cuales son representados mediante el tipo Data (Data = {identificador, sinónimos, hiperónimos, hipónimos, definición}).

- Función **ObtenerSinonimos**: **Entrada** → string  
**Salida** → List<string>

Dado el identificador de un synset buscamos y devolvemos el conjunto de sinónimos de dicho synset.

##### Clase ALMACENFrase

- Función **BuscarFrase**: **Entrada** → string []  
**Salida** → bool

Dado un conjunto de strings que conforman una frase me devuelve True o False dependiendo si la frase se encuentra en la lista de frases de WordNet. La búsqueda se realiza sin tener en cuenta el orden en el cual se encuentran las palabras dentro de la frase.

##### Clase ALMACENTags

- Función **GetStream**: **Salida** → Stream

Devuelve el archivo que utiliza el tagger.

## CAPA NEGOCIOS (Nivel 1)

### Clase DISAMBIGUATION

- Función **Buscar**: **Entrada** → (List<string>, List<string>)  
**Salida** → List<string>

Dada una frase y su tag correspondiente, diferencia entre los nouns de la frase que son plurales y los que son singulares. Si es plural busca su singular y viceversa. De esta manera hay dos grupos que contienen los términos de la frase tanto en su forma plural como singular. Esta función llama a **XMLFiles.ObtenerIndices** para encontrar los índices.

- Función **BusquedaEnIndexWN**: **Entrada** → (List<string>, List<string>)  
**Salida** → List<string>

Utiliza la función anterior **Buscar** para encontrar los identificadores de los synsets (en Index.xml) de una frase dada y su tag. En el caso de las frases comienza buscando la frase entera, si no la encuentra sigue un método de reducción de la frase y renueva la búsqueda. El método se describe en [].

- Función **BusquedaEnDataWN**: **Entrada** → (List<string>, List<string>)  
**Salida** → List<string>

Toma como entrada dos conjuntos de identificadores para encontrar el sentido del primero con ayuda del segundo.

Llama dos veces a **XMLFiles.ObtenerData** para obtener los synsets de cada grupo de identificadores. Los synsets del segundo grupo servirán para encontrar al synset correcto en el primer grupo a través de cuatro pasos. Si no se encuentra se devuelve un valor nulo.

- Función **Desambiguar**: **Entrada** → (List<QTerm>, int)  
**Salida** → List<string>

Un QTerm representa una frase o un término junto con su POS (Part Of Speech) y el synset asociado. Dada entonces una lista de QTerm que será la consulta y el índice que indica el QTerm a desambiguar, se utiliza **BusquedaEnIndexWN** para devolver el conjunto de identificadores de los synsets que representan todos los sentidos posibles de dicho QTerm. Si no existen synsets se devuelve un valor nulo. Si existe un synset se llama a **XMLFiles.ObtenerSinonimos** con el único identificador y se devuelven los sinónimos correspondientes. Si hay más de un synset se procede como se describe en el **Capítulo 3**. Si en este caso tampoco se encuentra el sentido se devuelve un valor nulo y la búsqueda se realiza solo con el término ingresado por el usuario permitiéndole luego a éste elegir el sentido entre todos los disponibles en WordNet. En la **Capa de Presentación** se puede observar un ejemplo.

### Clase ANALIZADORFrasas

- Función **DeteccionFrasas**: **Entrada** → string  
**Salida** → bool

Dada una frase de tipo string, esta función separa los términos que la componen y llama a **ALMACENFrases.BuscarFrase** que toma un valor de tipo string[] y devuelve un booleano.

- Función **AnalizarConsulta**: **Entrada** → List<string>  
**Salida** → List<string>

Dado un string el cual representa a la consulta misma del usuario sin procesar, esta función es la encargada de dividirla en términos y frases con la ayuda de **DeteccionFrases**. Devuelve la consulta dividida.

#### Clase TAGGER

- Función **ObtenerTags**: **Entrada** → List<string>  
**Salida** → List<string>

Recibe un conjunto de strings que representan la frase o el término al cual se necesita etiquetar.

Llama a **Tag**, función que realiza todo el trabajo del etiquetado, y devuelve el resultado de la misma también.

- Función **Tag**: **Entrada** → List<string>  
**Salida** → List<string>

Calcula el etiquetado o POS (Part Of Speech) de una frase o término.

### CAPA SOLUCIONES (Nivel 2)

#### Clase QUERYMaker

- Función **ArmarConsulta**: **Entrada** → string  
**Salida** → List<Resultado>

Dada una string que será la consulta del usuario:

- Se divide la consulta en frases y términos usando la función **ANALIZADORFrases.AnalizarConsulta**.
- Se etiqueta cada frase y término usando **TAGGER.ObtenerTags**.
- Armamos la lista de QTerm.
- Llamamos a **DISAMBIGUATION.Desambiguar** por cada uno de los QTerm que componen la consulta.
- Mostramos los resultados.
- Resultado = {string término, List<Data> sinónimos }

### CAPA BUSQUEDA (Nivel 3)

#### Clase Ariadne

- Función **ConsultarSimple**: **Entrada** → string  
**Salida** → string

Función que toma la consulta del usuario representada como una string sin expandir y realiza una búsqueda en el repositorio Ariadne. Principalmente se utiliza para comparar los resultados de búsquedas simples y expandidas. El resultado obtenido es una cadena de texto en formato XML donde figuran los objetos devueltos.

- Función **Consultar**: **Entrada** → List<Resultado>, bool  
**Salida** → string

La entrada es una lista del tipo Resultado donde se encuentran los sinónimos e hipónimos de los synsets que representan los sentidos de los términos ingresados por el usuario. Una vez procesada la consulta se realiza la búsqueda utilizando el web service correspondiente el cual devuelve el resultado en formato XML.

El segundo parámetro determina si los hipónimos son agregados a la consulta.

#### CAPA PRESENTACION (Nivel 4)

Esta última capa contiene una clase que consta de una interfaz con la cual el usuario puede interactuar.

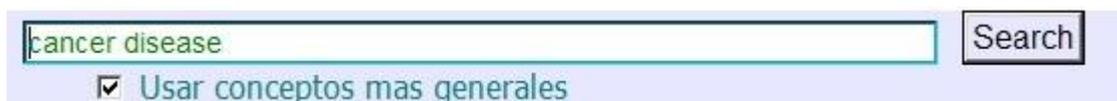


Figura 8.1.1: Interfaz de usuario para la búsqueda.

El botón de búsqueda **Search** llama a la función **Ariadne.Consultar** (que se encuentra en la capa anterior **CAPA DE BUSQUEDA**) para realizar la consulta especificada por el usuario al repositorio Ariadne. La consulta puede ser también expandida utilizando conceptos más generales vinculados a los términos buscados (hiperónimos).

Con Costo:	No
Duración Máxima:	No def.
Lenguaje:	No def.
Dificultad:	media

Figura 8.1.2: Opciones de búsqueda para los OA.

Otras opciones de búsqueda abarcan el costo de los objetos de aprendizaje y la duración de los mismos tal como se ve en la figura anterior así como también otras opciones que incluirían el idioma y el nivel de conocimiento del usuario dependiendo de las posibilidades que ofrezca el repositorio.

Una vez procesada la consulta, la misma se muestra al usuario mediante una estructura de árbol. El nodo raíz se llama **Consulta**, y sus nodos hijos son todos los términos hallados en la consulta. Cada uno de estos nodos contiene a su vez un nodo hijo llamado **Sinónimos** cuyos hijos serán todos los sinónimos pertenecientes al synset elegido para el término, y otro nodo llamado **Hiperónimos** que contendrá a todos los hiperónimos pertenecientes al synset antes mencionado. Éste último es opcional y aparece siempre que el usuario opte por utilizar conceptos más generales.

Se puede ver un ejemplo de esto en la siguiente figura. En la misma se muestra el resultado del procesamiento de la consulta “**cancer disease**” con los términos y sus respectivos y sus respectivos sinónimos e hiperónimos.



Figura 8.1.3: Resultado de la expansión de los términos ingresados por el usuario.

Los resultados de la búsqueda se presentan al usuario de la siguiente forma:

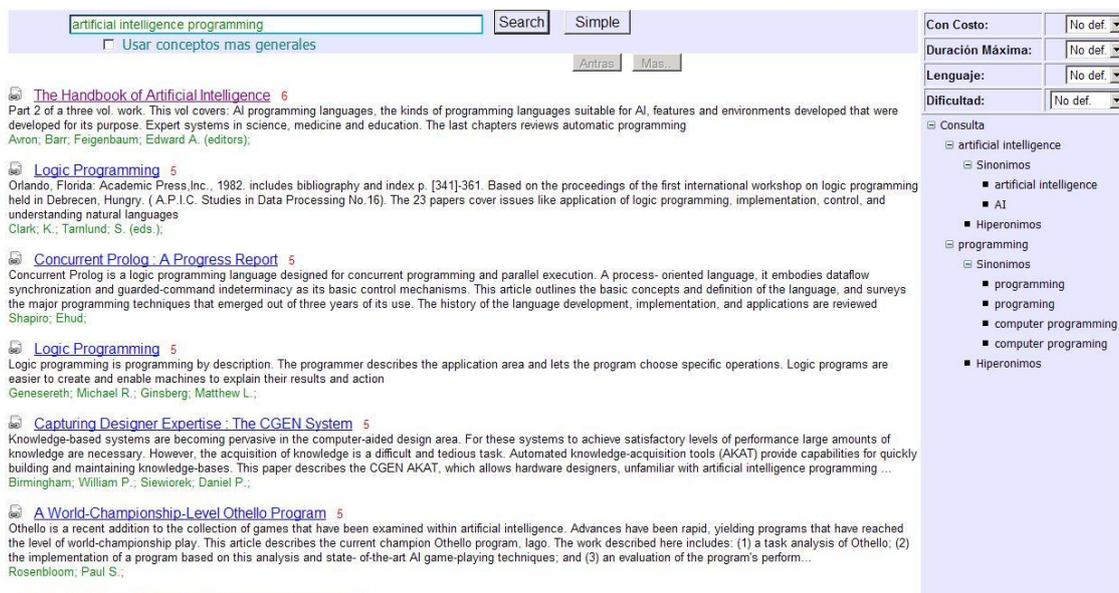


Figura 8.1.4: Resultado de una búsqueda.

Cada resultado consta de cuatro partes; el título del objeto (el mismo es un hyperlink que apunta hacia donde se encuentra el objeto), su descripción, el/los autor/es y un icono antes del título que indica si ese objeto se encuentra o no disponible.

 [HST.512 Genomic Medicine, Spring 2004 \(MIT\)](http://ocw.mit.edu)  
 An online course from the MIT OpenCourseWare website (<http://ocw.mit.edu>) use who accept the terms and conditions of use. The description of this course technologies and computational approaches that are driving advances in progn  
 Alma Taawo;

Figura 8.1.5: Componentes de un resultado.

En el caso en que no haya sinonimos asociados a un término debido a que no pudo hayarse un sentido para el mismo, primero se realizará sólo la búsqueda con el termino ingresado por el usuario y luego se presentará al mismo las opciones disponibles en WordNet tal como se muestra en la siguiente figura.

| **Cancer | genus Cancer**  
 type genus of the family Cancridae

| **Cancer | Cancer the Crab | Crab**  
 the fourth sign of the zodiac; the sun is in this sign from about June 21 to July 22

| **Cancer**  
 a small zodiacal constellation in the northern hemisphere; between Leo and Gemini

| **Cancer | Crab**  
 (astrology) a person who is born while the sun is in Cancer

| **cancer | malignant neoplastic disease**  
 any malignant growth or tumor caused by abnormal and uncontrolled cell division; it may spread to other parts of the body through the lymphatic system or the blood stream

---

| **doctor**  
 children take the roles of physician or patient or nurse and pretend they are at the physician's office; "the children explored each other's bodies by playing the game of doctor"

| **doctor | doc | physician | MD | Dr. | medico**  
 a licensed medical practitioner; "I felt so bad I went to see my doctor"

| **doctor | Dr.**  
 a person who holds Ph.D. degree (or the equivalent) from an academic institution; "she is a doctor of philosophy in physics"

| **Doctor of the Church | Doctor**  
 (Roman Catholic Church) a title conferred on 33 saints who distinguished themselves through the orthodoxy of their theological teaching; "the Doctors of the Church greatly influenced Christian thought down to the late Middle Ages"

Usar conceptos mas generales

 [HST.512 Genomic Medicine, Spring 2004 \(MIT\)](http://ocw.mit.edu) -1  
 An online course from the MIT OpenCourseWare website (<http://ocw.mit.edu>) operated by Massachusetts Institute of Technology. This course material is free for anyone to use who accept the terms and conditions of use. The description of this course, published on the MIT OpenCourseWare website, is: "This course reviews the key genomic technologies and computational approaches that are driving advances in prognostics, diagnostics, and treatment. Throughout the semester, emphasis will return to issue...  
 Alma Taawo;

Figura 8.1.6: Cuadro de opciones que se muestra al usuario en el caso en que no se pudo hallar el sentido de algún término.

### 8.2.1 Primer análisis

#### Sin Expansión

Calculus Podcasts  
Video Lessons in Calculus (course 2)  
Combining exponential and sinus function  
Video Lessons in Calculus  
Stokes theorem for mathematicians, physicists and engineers  
Riemann Integration  
Exponential function  
Differentiator  
Teach yourself limits  
Sequences and limits  
Limits of sequences  
Limits of Series  
Exponential function - real and imaginary  
Online Single Variable Calculus  
Cálculo con soporte interactivo en Moodle (Calculus with interactive support over Moodle)  
Applets sobre Equações Recursivas e Diferenciais  
Video Lessons in Number Theory  
18.013A Calculus with Applications, Spring 2005 (MIT)  
18.02 Multivariable Calculus, Fall 2007 (MIT)  
18.024 Calculus with Theory II, Spring 2003 (MIT)  
18.014 Calculus with Theory I, Fall 2002 (MIT)  
18.075 Advanced Calculus for Engineers, Fall 2004 (MIT)  
Achilles, Turtle, and the Geometric Series  
Diagnostic Testing at University of Helsinki  
An Operator Calculus for Surface and Volume Modeling

#### Con Expansión

Differential equations  
First-order differential equations  
First-order differential equations  
Differential equations  
Wortel TU/e , Math Learning Environment  
18.01 Single Variable Calculus, Fall 2006 (MIT)  
18.01 Single Variable Calculus, Fall 2005 (MIT)  
18.02 Multivariable Calculus, Spring 2006 (MIT)  
18.014 Calculus with Theory I, Fall 2002 (MIT)  
18.024 Calculus with Theory II, Spring 2003 (MIT)  
18.02 Multivariable Calculus, Fall 2007 (MIT)  
18.013A Calculus with Applications, Spring 2005 (MIT)  
Modelling with differential equations: oscillations  
2.035 Special Topics in Mathematics with Applications: Linear Algebra and the Calculus of Variations, Spring 2007 (MIT)  
18.075 Advanced Calculus for Engineers, Fall 2004 (MIT)  
Modelling with differential equations: oscillations  
18.994 Seminar in Geometry, Fall 2004 (MIT)  
3.016 Mathematics for Materials Scientists and Engineers, Fall 2005 (MIT)  
18.307 Integral Equations, Spring 2006 (MIT)  
24.242 Logic II, Spring 2004 (MIT)

18.238 Geometry and Quantum Field Theory, Fall 2002 (MIT)  
18.104 Seminar in Analysis: Applications to Number Theory, Fall 2006 (MIT)  
18.022 Calculus, Fall 2005 (MIT)  
14.12 Economic Applications of Game Theory, Fall 2005 (MIT)  
DISTANCE TEACHING OF MATHEMATICS FOR STUDENTS OF ARCHITECTURE: IS IT POSSIBLE?

## 8.2.2 Segundo análisis

### Sin Expansión

6.871 Knowledge-Based Applications Systems, Spring 2005 (MIT)  
Perspective on Artificial Intelligence Programming  
A Project Management Perspective ? utilization of artificial intelligence techniques  
Robot Manipulators : Mathematics, Programming and Control  
SOME EPISTEMOLOGICAL CONCERNS REGARDING ARTIFICIAL INTELLIGENCE AND KNOWLEDGE-BASED APPROACHES TO ARCHITECTURAL DESIGN - A RENEWED AGENDA  
A New Design Studio: Intelligent Objects and Personal Agents  
Artificial intelligence and networking in integrated building management systems  
A World-Championship-Level Othello Program  
SP.293 Lego Robotics, Spring 2007 (MIT)  
The Handbook of Artificial Intelligence  
16.413 Principles of Autonomy and Decision Making, Fall 2005 (MIT)  
21W.765J Theory and Practice of Non-linear and Interactive Narrative, Spring 2003 (MIT)  
A NEW APPROACH TO THE INTERACTIVE RESOLUTION OF CONFIGURATION PROBLEMS IN VIRTUAL ENVIRONMENTS  
Generative Design in an Evolutionary Procedure - An approach of genetic programming  
6.370 Robocraft Programming Competition, January (IAP) 2005 (MIT)  
6.186 Mobile Autonomous Systems Laboratory, January (IAP) 2005 (MIT)  
Perceptrons  
Knowledge Systems  
HST.951J Medical Decision Support, Fall 2005 (MIT)  
HST.951J Medical Decision Support, Spring 2003 (MIT)  
Communicating with Databases in Natural Languages  
Steps toward artificial intelligence  
21W.765J Interactive and Non-Linear Narrative: Theory and Practice, Spring 2006 (MIT)  
Architectural Tool Building: Introduction to Pascal for Architects and Designers Using Graphics on the IBM PC and Macintosh  
Capturing Designer Expertise : The CGEN System  
Spike : A Generic Design Standards Processing Expert System  
An AI-based shell for linking thermal and form-making considerations  
The Knowledge Level  
Knowledge-Aided Design System for Intelligent Building Design  
A knowledge based system for linking information to support decision making in construction

### Con Expansión

Perspective on Artificial Intelligence Programming 9.75  
The Logical Basis for Computer Programming : Deductive Reasoning 7.5  
An Evaluation of Programming Languages and Language Features for Engineering 7 Software Development  
The Handbook of Artificial Intelligence 7  
Logic Programming 6.5

Logic Programming 6.5  
 Capturing Designer Expertise : The CGEN System 6.25  
 16.413 Principles of Autonomy and Decision Making, Fall 2005 (MIT) 5.75  
 6.370 Robocraft Programming Competition, January (IAP) 2005 (MIT) 5.5  
 Robot Manipulators : Mathematics, Programming and Control 5.5  
 A World-Championship-Level Othello Program 5.25  
 Concurrent Prolog : A Progress Report 5  
 A NEW APPROACH TO THE INTERACTIVE RESOLUTION OF CONFIGURATION  
 PROBLEMS IN VIRTUAL ENVIRONMENTS 4.5  
 A New Design Studio: Intelligent Objects and Personal Agents 4.25  
 Computer Power and Human Reason : From Judgement to Calculation 4  
 21W.765J Theory and Practice of Non-linear and Interactive Narrative, Spring 2003 (MIT)  
 3.75  
 Rule- Based Expert Systems : The MYCIN Experiments of the Stanford Heuristic Programming  
 Project 3.5  
 A Theory of the Learnable 3  
 SP.293 Lego Robotics, Spring 2007 (MIT) 2.5  
 Developing Digital Design Techniques Investigations on Creative Design Computing The 25th  
 International Symposium on Automation and Robotics in Construction. ISARC 2008 2.25  
 21W.765J Interactive and Non-Linear Narrative: Theory and Practice, Spring 2004 (MIT) 2.25  
 Perception Aspects in Underground Spaces using Intelligent Knowledge Modeling Generic  
 Design Standards Processing in a Knowledge-based expert system Environment 2.25  
 BIMGame: Integrating Building Information Modeling and Games to Enhance Sustainable  
 Design and Education 2.25  
 Spike : A Generic Design Standards Processing Expert System 2.25  
 Architectural Tool Building: Introduction to Pascal for Architects and Designers Using  
 Graphics on the IBM PC and Macintosh 2.25  
 21W.765J Interactive and Non-Linear Narrative: Theory and Practice, Spring 2006 (MIT) 2.25  
 Knowledge Systems 2.25  
 Perceptrons 2.25  
 Special Section on Architectures for Knowledge-Based Systems 2.25  
 Possible Futures in Computer Art Generation 2.25  
 Logic Models of Design 2.25  
 The Emperor's New Mind. Concerning Computers Minds, and the Laws of Physics. 2.25