

Búsqueda de documentos utilizando criterios semánticos

Procesamiento de Lenguaje Natural - Proyecto de Grado 2012

Autores:

Santiago COTELO
Alejandro MAKOWSKI

Tutores:

Luis CHIRUZZO
Dina WONSEVER

Resumen

En la actualidad la búsqueda de contenidos en sistemas de recuperación de información, como los buscadores web, se realiza a través del criterio de palabras claves donde se buscan documentos que contengan una o varias de esas palabras y luego son ordenados según algún criterio propio de cada buscador. En este contexto no se puede a partir de una búsqueda, asociar las palabras de modo que estas reflejen la semántica deseada y que de esta forma, los documentos recuperados sean más relevantes para el usuario, según el propósito de su búsqueda. La búsqueda semántica es un concepto que está siendo estudiado en la actualidad pero no existen sistemas de recuperación de información que la resuelvan de una manera eficiente.

Como aporte del proyecto se plantea una solución que permita a un usuario de un sistema recuperación de información, plantear una consulta con mayor que la de la simple presencia de palabras, y que los algoritmos que recuperan los documentos puedan utilizar esa información para interpretarla y recuperar los documentos que mejor se adapten a ella.

En el proyecto se definió un lenguaje de consulta que permite extender las habilidades usuales de los buscadores considerando vínculos entre palabras, expansión por sinónimos e interpretación de las expresiones de fecha. Se definió además un método de ranqueo de los documentos recuperados que toma en cuenta los aspectos mencionados.

Se construyó un prototipo y se realizaron pruebas sobre un conjunto de documentos definido especialmente para el proyecto, lo que permitió mostrar las características del sistema construido y verificar en qué aspectos puede funcionar mejor que otros sistemas de recuperación ya existentes.

Contenido

Resumen	2
1. Introducción.....	5
1.1. Objetivos	5
1.2. Resultados Esperados	6
1.3. Organización del documento	6
2. Herramientas para búsquedas semánticas.....	7
2.1. Buscadores	7
2.2. Lenguajes de Consulta	12
3. Funcionalidades del buscador semántico.....	15
3.1. Requerimientos.....	15
3.1.1. Identificación de objetos y atributos relacionados en una consulta	15
3.1.2. Definición por parte del usuario del significado de palabras polisémicas.....	15
3.1.3. Semántica temporal en consultas.....	16
3.1.4. Expandir consulta por sinónimos.....	16
3.1.5. Incluir operadores en consulta	16
3.2. Evaluación de requerimientos en buscadores actuales.....	18
3.2.1. Casos de Prueba	18
3.2.2. Resultados.....	20
4. Solución propuesta	23
4.1. Diseño de la solución	23
4.1.1. Arquitectura del sistema.....	23
4.1.2. SemQL - Lenguaje de consulta semántico	24
4.1.3. Análisis de dependencias	26
4.1.4. Motor de búsqueda	31
4.2. Implementación	38
4.2.1. Módulos externos	38
4.2.2. Módulos internos.....	39
5. Validación de la solución.....	44
5.1. Corpus para validación de datos.....	44
5.1.1. Corpus Time	44
5.1.2. Corpus Cranfield.....	46

5.1.3.	Corpus BusSem-2012	47
5.2.	Resultados y evidencias	49
5.2.1.	Resultados por funcionalidad	49
5.2.2.	Comparación de resultados con The Times	52
6.	Conclusiones y trabajo a futuro	57
6.1.	Conclusiones	57
6.2.	Trabajo a futuro	58
7.	Referencias.....	59
ANEXO I – Consultas Capturadas		62
ANEXO II - Modelos de recuperación de información.		82
ANEXO III – Herramientas evaluadas		84

1. Introducción

En la actualidad la búsqueda de contenidos en sistemas de recuperación de información como son los buscadores web, se realiza a través del método de palabras claves donde se buscan documentos que contengan una o varias de esas palabras y son ordenados según diversos criterios (cantidad de ocurrencias, proximidad de palabras, popularidad de cada documento).

Como este método en su forma básica, implica que el usuario ingrese palabras en cualquier orden sin importar la relación entre palabra y palabra, la interpretación semántica, es decir, comprender que es lo que el usuario está buscando no es posible. En este contexto es probable que muchos de los documentos recuperados no le sean útiles al usuario.

Dada esta situación, se plantea que permitiendo al usuario agregar información semántica a sus búsquedas, se pueden ordenar los mismos documentos que serían devueltos por un buscador normal. Estos documentos serían ordenados dependiendo de la coincidencia en las palabras pero también en su significado y por lo tanto, es posible que el usuario visualice los documentos más relevantes en primer lugar.

De esta forma los sistemas de recuperación de información tendrían un insumo más para mejorar la calidad de las búsquedas, ya que del universo de documentos, los primeros resultados serían los más relevantes para el usuario en toda búsqueda.

1.1. Objetivos

El objetivo principal del proyecto consiste en definir estrategias que mejoren la calidad de los resultados en sistemas de recuperación de información a partir de la interpretación semántica de los documentos a recuperar.

La estrategia posible para cumplir este objetivo es definir un lenguaje de consulta que pueda ser utilizado para recibir consultas con información semántica y un algoritmo de ordenamiento que permita ordenar los documentos según el significado de la consulta. Dentro de las características deseables del lenguaje de consulta se encuentran:

- identificar objetos y atributos de los mismos, por ejemplo diferenciar “perro negro y gato blanco” de “perro blanco y gato negro”
- permitir al usuario indicar el significado de una palabra polisémica
- incluir semántica temporal en las consultas, por ejemplo “escuelas construidas a principios del siglo XX”
- expandir la consulta con sinónimos
- permitir operadores sobre los predicados, por ejemplo “perro –blanco”

Otro objetivo del proyecto es relevar el estado de arte de diversos buscadores para detectar si alguno de los estudiados cuenta con las características antes mencionadas. Los buscadores que se estudian son los más populares como **Google** y **Bing** además de otros buscadores que manejan conceptos de semántica como **Hakia** y **Kosmix** por ejemplo.

Además se desea contar con un conjunto de documentos recopilados que se denomina **corpus**, que permita realizar pruebas sobre el algoritmo implementado.

1.2. Resultados Esperados

- Estado del arte acerca de los buscadores que incorporan información semántica.
- Diseño del lenguaje de consultas.
- Implementación de un algoritmo de búsqueda que obtenga los documentos especificados por consultas del lenguaje.
- Desarrollo de un corpus de prueba de consultas y documentos para la evaluación de desempeño del algoritmo.
- Implementación de una interfaz web que permita visualizar los resultados en línea.

1.3. Organización del documento

A partir de aquí el documento está presentado de la siguiente forma:

En el capítulo 2 se estudian las distintas herramientas existentes para resolver los problemas de las búsquedas semánticas.

En el capítulo 3 se detallan los requerimientos que deberían cumplirse en un buscador semántico y pruebas realizadas sobre distintos buscadores para verificar si cumplen esos requerimientos o no.

En el capítulo 4 se propone una solución para implementar un sistema de recuperación que cumpla satisfactoriamente algunos de los requerimientos planteados

En el capítulo 5 se valida la solución propuesta definiendo dónde y cómo se realizan las pruebas

El capítulo 6 expone las conclusiones y las posibles mejoras o nuevas funcionalidades a implementar en el sistema propuesto.

Por último se encuentran diferentes anexos que complementan la información del proyecto.

2. Herramientas para búsquedas semánticas

Existen muchos centros de investigación o empresas de software que desarrollan distintos métodos de recuperación de documentos procurando mejorar los resultados introduciendo los conceptos de búsqueda semántica. Se estudiaron distintos buscadores y lenguajes de consulta para ver cómo se resuelven este tipo de búsquedas y qué nivel de madurez existe sobre esta problemática.

2.1. Buscadores

A continuación se detallan los diversos motores de búsqueda en este marco con el objetivo de encontrar pautas, problemas y soluciones a nuestro cometido de realizar búsquedas semánticas en la web.

Freebase

Es un repositorio de información del mundo que se construye gracias al aporte de la comunidad web donde todos pueden contribuir a crear la información y todos pueden accederla posteriormente mediante búsquedas.

Freebase [01] [02] maneja el concepto de entidades que pueden estar conectadas entre sí dependiendo de la relación entre las mismas, formando un gran grafo relacional. Las entidades además pueden tener diversos significados como por ejemplo “Boston” podría ser una ciudad, un equipo de Basketball o un grupo musical, donde para el Freebase lo importante es la entidad “Boston”. Esto implica que el resultado de buscar “Boston” en Freebase, sería un tema unificado como centro de información sobre la entidad “Boston” donde se despliega información sobre todos los distintos significados que puede tener asociados, no se busca según significado de la palabra.

A partir del manejo de entidades, entendiendo que la web se construye en base a cierto grupo de entidades, este buscador está más alineado a la **Web Semántica** que a la búsqueda semántica.

Powerset

Es un motor de búsqueda que en un principio tenía como objetivo el de procesar el lenguaje natural, entender la búsqueda completa y no enfocarse solo en palabras claves.

La idea de **Powerset** [03] [04] [05] se basa en interpretar la web indexando los contenidos según su semántica, recibir una consulta e interpretarla semánticamente, buscar en el repositorio de contenidos y devolver los resultados que tengan la misma semántica.

A pesar de que se hace una búsqueda semántica, se hace especial hincapié en el almacenamiento de los documentos indexados con una estructura lingüística que permita comparar luego semánticamente con las búsquedas. Muchos buscadores intentan interpretar la consulta, lo cual tiene menor costo a realizar un *crawl* de la web indexando semánticamente los documentos y esa diferencia entre tener los documentos con cierta estructura lingüística y no tenerlos, puede provocar resultados muy dispares.

Al realizar la interpretación de los documentos en la web, el motor puede contestar preguntas ya que está capacitado para entender la pregunta y encontrar respuestas dentro de su repositorio.

En la actualidad el motor de búsqueda Powerset ha sido adquirido por Microsoft para incorporarse al buscador Bing.

Duck Duck Go (DDG)

Es un motor de búsqueda que utiliza la información de sitios de origen público como Wikipedia con el objetivo de aumentar los resultados tradicionales y mejorar la relevancia de ciertos documentos.

El motor de **Duck Duck Go** [06] [07] utiliza al igual que **Google**, la técnica de búsqueda por palabras claves pero utiliza ciertas técnicas que permiten mejorar los resultados por ejemplo restringiendo la búsqueda en un sitio, en una zona geográfica u ordenando los resultados dentro de otras posibilidades.

En las consultas se permite agregar funciones lógicas como AND, OR, NOT para mejorar las búsquedas, agrupando las palabras claves según las funciones lógicas y filtrando los resultados con dicha consulta.

Si bien esto es sumamente útil, una de las funcionalidades más importantes de **DDG** es que brinda la posibilidad de crear *plugins* que al ser incorporados al motor de búsqueda interpretan la consulta del usuario y devuelve un resultado más exacto, como por ejemplo si se hace una pregunta o un cálculo matemático.

Kosmix

Es un buscador semántico cuyo objetivo es proporcionar información sobre un tema de búsqueda ofreciendo una colección de links a contenidos, descripciones, videos, enlaces y otros objetos de interés relacionados con la búsqueda [08].

Según los creadores de **Kosmix** existe una diferencia entre buscar y explorar la web donde buscar permite encontrar información específica y navegar webs que son conocidas para quien realiza la búsqueda, mientras que explorar se descubre lo que hay en la web sobre el tema de búsqueda [09]. El enfoque del desarrollo de **Kosmix** fue justamente permitir a los usuarios explorar la web, ayudando al cliente a organizar mejor sus resultados.

El motor primero explora y reúne el contenido de toda la web mediante *crawling* creando un catálogo enciclopédico multimedia. Esta exploración va generando una clasificación completa de categorías temáticas, incluyendo conexiones entre todas ellas [10].

Una vez indexado el contenido de toda la web, al realizar una búsqueda se detecta la temática buscada y mediante un algoritmo propietario al que no se tiene acceso, se recuperan los contenidos de mayor ranking según los temas que abarquen a veces hasta sin importar las palabras claves de búsqueda. De esta forma es un buscador que no intenta conocer la consulta semánticamente sino que la categoriza y luego explora en la web qué resultados encuentra sobre la temática.

Los desarrolladores de **Kosmix**, si bien tuvieron interés en competir con **Google**, manifiestan que es una herramienta distinta ya que **Google** es el mejor para buscar cuando se sabe con precisión qué es lo que se está buscando, realizando una búsqueda de navegación mientras que con **Kosmix** sería una búsqueda de información [11].

En la actualidad los laboratorios de *Wal-Mart* han adquirido el producto **Kosmix** para utilizar su diseño y algoritmos de búsqueda para productos comerciales evolucionando el producto a lo que se conoce como *Social Genome* que permite a las empresas comerciales tener un mejor relacionamiento con el cliente.

Hakia

El proyecto **Hakia Web Search** [12] es un proyecto de la compañía **Hakia**, que se dedica exclusivamente a mejorar la experiencia de los usuarios al realizar búsquedas basándose en la búsqueda semántica.

Dentro de la compañía existen tres proyectos distintos al de *Web Search* en los que aplican las técnicas de búsqueda semántica para buscar sobre contenidos de temáticas específicas.

Los tres proyectos son:

- AeroSpace – Informaciones aéreas
- Finance – Información financiera
- Medical Research (PubMed) – Información médica

En el proyecto de *Web Search* existen dos tecnologías utilizadas para procesamiento de segmentos. Primero se indexan los contenidos de la web con tecnología propia de **Hakia**, llamada **QDEX** y luego para obtener los resultados se usa tecnología de ranking semántico de plataformas de terceras partes.

QDEX

Este sistema es una manera de analizar y almacenar el contenido de un artículo web interpretando y entendiendo semánticamente qué quiere decir el mismo [13].

Se analiza el contenido entero y luego mediante un algoritmo se extraen todas las posibles consultas que se podrían hacer sobre dicho contenido y estas posibles consultas son el nexo para vincularlos con la búsqueda. El análisis no se hace en tiempo real sino que primero se indexan los contenidos de un sitio y luego de actualizado el índice de contenidos se pueden realizar búsquedas sobre los mismos.

El punto crítico en este sistema es la habilidad de descomponer oraciones de un documento en segmentos significativos sin perder el significado al quitar ciertas palabras y que los segmentos resultantes sean relevantes. En una oración que por ejemplo tiene ocho palabras claves, se pueden encontrar billones de combinaciones pero que forman segmentos que no tienen sentido para el humano, pues es clave entonces ignorarlos. Para resolver este problema **Hakia** utiliza la tecnología de Ontología Comercial.

Ranking Semántico

El ranking semántico consiste en que luego de obtener un conjunto de párrafos relevantes a una consulta específica, párrafos que provienen del sistema QDEX, se determina la relevancia de los documentos mediante un algoritmo que se basa en un análisis profundo de la oración y de concepto entre la consulta y la mejor oración y de cada párrafo del conjunto [14].

Swoogle

Es un motor de búsqueda para la **Web Semántica** que indexa, recupera y organiza la información de la misma. Actualmente **Swoogle** [15] [16] no indexa contenido de la World Wide Web y por lo tanto no realiza búsquedas sobre la misma.

Google

Google [17] [18], es el motor de búsqueda más utilizado del mundo y tiene acceso a miles de millones de sitios web de todo tipo. Su objetivo principal es el de la búsqueda de texto en páginas web, pero también se pueden buscar específicamente libros, blogs, imágenes, videos, documentos académicos, entre otras cosas.

Entre las principales características de búsqueda se incluyen sinónimos, clima, zonas horarias, cotizaciones, mapas, cartelera de cine, aeropuertos y resultados deportivos. También hay funciones especiales para trabajar con datos numéricos, incluyendo intervalos, precios, temperaturas, conversiones de dinero o unidad, cálculos y traducción de páginas mostradas.

Una nueva característica es la de mostrar de forma dinámica los resultados de las búsquedas a medida que se escribe en la caja del buscador. De esta forma, cada vez que se escribe una nueva palabra o frase, el conjunto de resultados es actualizado mostrando información relevante y relacionada sin necesidad incluso de tener que recurrir a hacer clic sobre el botón de "búsqueda".

Para el *crawling*, se segmenta la web de determinada manera y se hacen actualizaciones sobre cada segmento independientemente, y dependiendo del nivel jerárquico en que se encuentre el documento, será actualizado con mayor o menor frecuencia.

La indexación de contenidos se realiza con la técnica de *índice invertido* [19] donde, a partir de las palabras claves, se indexan las palabras claves y con cada palabra la lista de documentos en los que aparece. Para mejorar la búsqueda, no se indexan palabras comunes como "es", "en", "como" entre otras, palabras que comúnmente se denominan *stopwords*.

El orden de los resultados de búsqueda se basa en un algoritmo llamado *PageRank* [20] el cual ayuda a clasificar las páginas web que coincidan con una cadena de búsqueda. Este algoritmo se basa en cientos de factores, por ejemplo si se buscan dos palabras juntas, que ambas aparezcan juntas en la página y no separadas, o que una página tenga buena reputación o muchos links hacia ella y muchas otras combinaciones.

Google además provee para las consultas operadores de búsqueda como ser el "OR", el símbolo "-" para excluir una palabra, el símbolo "+" para forzar la aparición de una palabra, el "*" comúnmente utilizado como comodín y otros tipos de metadatos para las búsquedas como por ejemplo palabras claves como "site" para restringir los resultados a un dominio dado, o la palabra "Allinurl" el cual buscará solo en las direcciones URL de las páginas y no en sus contenidos.

Ofrece búsquedas avanzadas pudiendo restringir los resultados por idioma, región, última actualización de las paginas, formato de los archivos, excluir páginas de un determinado sitio y mucho más.

Hoy en día **Google** apunta a mejorar sus búsquedas y se dice que mejorará sus algoritmos para que pongan más énfasis en una búsqueda semántica en vez de una enumeración de palabras claves; una búsqueda de entidades o objetos y no solo textos, utilizando un grafo de conocimiento donde se almacenan todos esas entidades y objetos de interés [21]

Otras herramientas relacionadas

En la investigación realizada se encontraron otros buscadores o herramientas derivadas de los buscadores con algunas particularidades interesantes que no estaban previstos para ser estudiados, pero de todas formas tienen características importantes como para ser referenciados aquí.

Wolfram Alpha

Wolfram Alpha [22] es un servicio en línea que responde a las preguntas mediante el procesamiento de la respuesta extraída de una base de datos estructurados, pero no es un motor de búsqueda.

A diferencia de **Google** y otros buscadores tradicionales, éste le agrega más análisis semántico a la oración que se quiere buscar. Además obtiene los resultados no solo de la web pública sino también de la “web profunda”, la cual son fuentes de datos que requieren una suscripción, por lo que se podría decir que los datos se obtienen de páginas más estructuradas, robustas y confiables.

Ask

Ask [23], fue el primer buscador comercial que respondía preguntas del usuario en la World Wide Web realizadas en lenguaje natural. Permite una amplia variedad de consultas de preguntas, así como también busca por palabras claves y su intención es hacer que las búsquedas sean más intuitivas que otros resultados.

Siri

Siri [24] es una aplicación para **iOS** que procesa el lenguaje natural para responder preguntas, hacer recomendaciones y realizar acciones mediante la delegación de solicitudes a un conjunto de servicios web que va en aumento.

2.2. Lenguajes de Consulta

El análisis realizado para ver el funcionamiento de los buscadores permitió distinguir algunas pautas sobre el lenguaje de consulta que cada buscador utiliza.

El lenguaje de consulta permite al usuario utilizar todas las características que el buscador ofrece y además es una forma de ordenar o simplificar el análisis de la consulta por el motor de búsqueda.

Existen diversos lenguajes de consulta, algunos sencillos que funcionan únicamente con palabras claves y operadores lógicos comunes, como es el caso de Google y otros más complejos que requieren de un estudio para ser utilizados por el usuario.

El objetivo de los lenguajes de consulta es acortar las distancias entre lo que el usuario desea buscar y como está organizada la información sobre la que se va a consultar, de manera de optimizar los resultados. Tienen como ventaja ofrecerle al usuario brindar más detalle sobre la consulta, pero lo que en realidad están haciendo es ayudando al motor de búsqueda a buscar mejor luego de tener etiquetada la consulta.

Si bien existen muchos lenguajes de consulta, donde cada buscador podría definirse uno propietario, se detallan algunos de ellos para introducir el tema ya que luego se definirá el lenguaje de consulta utilizado para el proyecto de búsqueda semántica en la web.

Freebase - MQL (Lenguaje de Consultas MetaWeb)

Como se ha mencionado anteriormente, el buscador Freebase está más alineado a la **Web Semántica** que a la búsqueda semántica y por lo tanto, las búsquedas son sobre entidades y sus propiedades, y el lenguaje de consulta utilizado, denominado **MQL** [25], es bastante particular.

Esto se debe a que el conjunto de documentos sobre el que se realiza la recuperación está modelado de cierta forma de facilitar la categorización de los contenidos según las entidades y sus propiedades, de lo contrario no sería posible definir dichas búsquedas.

Un ejemplo del uso de MQL sería para buscar las películas que dirigió Sofia Coppola y tendría el siguiente formato de búsqueda:

```
{
  "type" : "/film/director",
  "name" : "Sofia Coppola",
  "film" : []
}
```

Analizando el cuerpo de la consulta se puede visualizar que con el tag *type* se define el dominio donde se desea buscar, es decir sobre qué entidades, mientras que el tag *name* define lo que se desea buscar dentro del dominio y el tag vacío *film* indica dónde se quieren obtener los resultados.

Este lenguaje de consulta no pretende ser utilizado por el usuario sino la idea es que un usuario redacte su búsqueda en lenguaje natural, se procese para generar una consulta en **MQL** y luego esta es ejecutada en el motor de búsqueda para obtener los resultados.

Al ser un lenguaje de consulta basado en entidades, si bien permite contemplar algunos aspectos de los requeridos para el proyecto, requiere de una estructuración del corpus de búsqueda y a su vez las características de la búsqueda dependen muy fuertemente de las entidades buscadas.

CQL - Lenguaje de Consultas Contextual

Es un lenguaje formal que es utilizado para sistemas de recuperación de información como motores de búsqueda, catálogos bibliográficos, colecciones de información sobre un cierto tema específico y otros [26].

Además de la búsqueda intuitiva que ofrece con los operadores booleanos AND, OR, NOT, ofrece búsqueda exacta introduciendo caracteres especiales, operadores de proximidad para permitir distancia entre palabras, índices de búsqueda que permiten especificar los resultados aún más dependiendo el contexto donde se desea buscar. A su vez incluye modificadores para la búsqueda de palabras que son parecidas fonéticamente, palabras relevantes o semejantes a la búsqueda, conjugaciones y otras características.

Al igual que **MQL**, la utilización del lenguaje de consultas contextual parece brindar al usuario la posibilidad de mejorar sus criterios de búsqueda agregando información sobre la misma, restringiendo o expandiendo los resultados según los criterios que el usuario seleccione y de esta forma generar varias búsquedas a partir de una sola consulta para devolver los resultados que cumplan con todas las características deseadas.

Este lenguaje puede ser utilizado por los usuarios o por los buscadores. Los usuarios pueden definir ciertas características de su búsqueda. Por otro lado los buscadores pueden a partir de una consulta en lenguaje natural creada por el usuario, expandir o restringir la consulta para mejorar los resultados según las características del buscador.

Google - Lenguaje de consulta

A pesar de que **Google** utiliza una búsqueda por palabras claves muy eficiente existen ciertos operadores y opciones de consulta complementarias que permiten especificar la búsqueda para obtener mejores resultados [27]. Este tipo de búsqueda avanzada, si bien no hemos encontrado datos que demuestren que son poco utilizadas, según hemos evaluado en distintos ámbitos como ser oficinas de trabajo, casas de familia entre otros, no es algo que el usuario del buscador conozca con exactitud y utilice frecuentemente.

A la búsqueda común, por aparición de palabras, se le pueden agregar operadores algebraicos (“-” para excluir una palabra, “+” para incluir forzosamente una palabra, “*” como comodín), el operador booleano OR y otras opciones para restringir tanto la colección de documentos donde se desea buscar (por ejemplo en un sitio determinado) o para restringir dependiendo de las propiedades de los documentos como por ejemplo la fecha de actualización.

Es un lenguaje simple que es muy poco utilizado ya que en la búsqueda por coincidencia de palabras generalmente no se necesita especificar más que las palabras ya que los algoritmos de ranking de los documentos por aparición de palabras parecen bastante precisos.

El problema con este lenguaje se da cuando la búsqueda no es por aparición de palabras, sino que se está realizando una búsqueda que tiene un significado mayor, Google no brinda posibilidades de agregar semántica a las búsquedas. A modo de ejemplo, en el caso de una búsqueda con un operador para excluir una palabra, “*cat -white*”, no traería ninguna página que diga *White*, aunque ese adjetivo en ninguna parte del documento esté relacionado con un gato.

BingQL - Lenguaje de Consulta de BING

El buscador de BING [28] permite introducir operadores avanzados en una búsqueda para afinar la consulta y obtener mejores resultados.

En el sitio de referencia se encuentra el detalle de los distintos operadores que se pueden utilizar para refinar las búsquedas pero realizando un análisis de los mismos el lenguaje ofrece las mismas operaciones que el CQL y que el lenguaje de Google. Ofrece operadores booleanos, filtros por atributos del documento o de la ubicación de los mismos.

3. Funcionalidades del buscador semántico

En este capítulo se pretende analizar los requerimientos de un buscador semántico para que pueda resolver los objetivos planteados en el proyecto y se realizan pruebas sobre los buscadores estudiados anteriormente para determinar qué tan eficientes son según cada característica deseada.

3.1. Requerimientos

Se presentan los distintos requerimientos que son deseables que un buscador semántico pueda resolver y un breve análisis para evaluar cómo podrían resolverse.

3.1.1. Identificación de objetos y atributos relacionados en una consulta

En los buscadores relevados, si bien se pueden asociar términos con operadores de manera de que en la consulta se contemplen todas las palabras, no es posible asociar palabras entre sí para marcar un atributo corresponde a un determinado objeto (y no a otro) de la cadena de consulta.

Un ejemplo es si se estuviera realizando una búsqueda sobre documentos que hablen sobre los soldados americanos y los civiles afganos. En un buscador de los tradicionales se podría escribir la consulta de diversas formas, la clásica es como una secuencia de palabras: **“american soldier afghan civilian”**.

Como se aprecia en esta consulta no se asocian los atributos **american**, y **afghan** a los objetos **soldier** y **civilian** respectivamente, por lo que en los documentos recuperados podría ocurrir que primero contengan los *objeto-atributo* cambiados, o ni siquiera los contengan.

Es por eso que como una ayuda para el buscador semántico, se plantea la idea de redactar la consulta asociando los objetos-atributos entre ellos para que luego puedan ordenarse los documentos dependiendo de estas coincidencias.

3.1.2. Definición por parte del usuario del significado de palabras polisémicas

Las palabras polisémicas son aquellas que tienen más de un significado, que dependen del contexto donde se esté utilizando y son una fuente de ambigüedad que requiere poderse resolver con el objetivo de entender una consulta y devolver los documentos relevantes.

En consultas realizadas en un contexto determinado, una palabra puede tener un significado y por lo tanto se deberían poder recuperar los documentos relevantes para ese significado, mientras que si cambia el contexto, cambia el significado de la palabra y se deberían obtener documentos diferentes, relacionados con el nuevo contexto de búsqueda. Por ejemplo, buscar por “apple” la compañía o “apple” la fruta.

Esta funcionalidad requeriría que el usuario que está ingresando la consulta especifique el significado de las palabras que componen la consulta y que los documentos en los que se busca, estén desambiguados, teniendo claro para cada palabra polisémica qué significado tiene.

La desambiguación de las palabras de una oración es un tema complejo en el procesamiento de lenguaje natural y si bien existen algunos clasificadores para poder intentar desambiguar las palabras, la creación de oraciones de entrenamiento para que los clasificadores puedan desambiguar y tener un correcto

comportamiento para luego aplicarlo a los documentos donde se desea buscar, está por fuera del alcance del proyecto. De esta forma resulta difícil poder implementar la búsqueda contemplando las palabras polisémicas.

3.1.3. Semántica temporal en consultas

En la actualidad existe una limitante en las búsquedas que es la de introducir expresiones temporales en las mismas para obtener documentos que estén asociados a las fechas buscadas. Es deseable entonces, permitir al usuario marcar un evento y su fecha de ocurrencia o el rango en el que interesa que haya ocurrido, para luego poder devolver los documentos que cumplen con ese criterio.

Algunos ejemplos de búsquedas con expresiones temporales podrían ser **“XVI century astronomers”** o **“world cup stadiums at 1934”** o **“meat exported in last ten years”**. Estas búsquedas además de contar con entidades a buscar, como son los astrónomos, o los estadios de fútbol, o la cantidad de carne exportada, indican un tiempo en el que se debe contextualizar la búsqueda. Por lo tanto, no interesa cualquier astrónomo sino los del siglo XVI, no interesa cualquier estadio de fútbol, sino los de la copa del mundo de 1934 y no interesa cuanta carne se exportó ayer, sino en los últimos diez años.

Por lo tanto se plantea la posibilidad de resolver este tipo de consultas brindando al sistema de recuperación de información la capacidad de poder resolver que, si en un documento aparecen expresiones temporales relacionadas a los eventos o entidades buscadas, estas expresiones deben estar en el rango de tiempo que se introdujo en la búsqueda.

3.1.4. Expandir consulta por sinónimos

La búsqueda por palabras claves tiene una limitante en caso de que el motor de búsqueda del sistema de recuperación no implemente la resolución de búsqueda por sinónimos porque podría pasar que existan documentos muy relevantes para una búsqueda, pero como las palabras ingresadas no aparecen, el documento no es recuperado.

De esta forma, una funcionalidad casi obligatoria de un buscador semántico es la de resolver las búsquedas de palabras claves por las propias palabras y sus sinónimos, con el objetivo de recuperar más documentos que contengan cualquier conjunto de esas palabras.

Un ejemplo podría ser al realizar una búsqueda del estilo **“giant dog”**, si en el conjunto de documentos donde se realiza la búsqueda, no aparece la palabra **giant**, el documento no sería recuperado, a pesar de por ejemplo contar con palabras que son sinónimos, como podría ser **“huge dog”**.

La inclusión de este requerimiento es sencilla dado que lo que se requiere es expandir la consulta para no solo buscar por las palabras claves ingresadas, sino por los sinónimos de cada palabra.

3.1.5. Incluir operadores en consulta

Otra funcionalidad que se desea cubrir es la posibilidad de incluir operadores para marcar expresiones semánticas adicionales, como por ejemplo sería la negación de una palabra y que el motor de búsqueda interprete tal operador para puntuar con más o menos relevancia un documento.

El objetivo es que por ejemplo si se realiza una búsqueda del estilo **“dog not white”**, los documentos recuperados con más relevancia, tengan en su contenido alguna oración donde sea explícito que el perro no es blanco, luego aparecerían los documentos que no dicen si es blanco, es decir no se sabe con certeza si se cumple o no la negación y luego aparecerían los documentos que afirman que es blanco, que está claro que son menos relevantes.

Esta funcionalidad está relacionada con la de objeto-atributo, pues los operadores, en principio el de negación, se podría utilizar para cualquier atributo de forma de poder especificar mejor la búsqueda deseada.

3.2. Evaluación de requerimientos en buscadores actuales

3.2.1. Casos de Prueba

Con el objetivo de poder medir los buscadores mencionados anteriormente para ver si alguno cumple con las características requeridas del proyecto, se definen ciertos casos de prueba para testear cada uno y de esta forma poder concluir si sirven para nuestros objetivos o qué camino se podría tomar para lograr desarrollar un buscador que sí cumpla con todas las características.

3.2.1.1. *Objetos y atributos*

Se deben realizar consultas que tengan dos o más objetos y dos o más atributos, donde al menos un atributo aplica para cada objeto.

Las pruebas fueron realizadas con las siguientes consultas

- Black dog White cat
- Black cat White dog
- Us civilian Afghanistan soldier
- High blood pressure low cholesterol
- High temperature low humidity
- Think fast move slow

3.2.1.2. *Palabras polisémicas*

Se realizan consultas con palabras que tienen más de un significado esperando que el buscador pregunte por el significado de esa palabra o muestre todos los resultados para todos los significados posibles.

Las pruebas fueron realizadas con las siguientes palabras polisémicas

- Watch: Mirar, Reloj
- Nut: Loco, Nuez
- Kind: Tipo, Amable
- Book: Libro, Alquilar
- Take - Tomar, Llevar
- Look - Mirar, Aspecto
- Apple - Manzana, La compañía
- Windows - Ventana, La compañía

3.2.1.3. *Semántica temporal*

Se realizan consultas con frases que contengan marcas temporales con el objetivo de que los buscadores sepan resolver los tiempos de los sucesos independientemente de cómo este redactado el mismo.

Las pruebas fueron realizadas con las siguientes consultas

- Maradona goals in the decade of 80

- XVI century astronomers
- 1934 world cup stadiums
- meat exported between 2002 to 2012
- meat exported in the last ten years

3.2.1.4. Expansión por sinónimos

Se realizan diversas consultas con palabras que tienen distintos sinónimos esperando que los buscadores retornen resultados con la palabra original o con los sinónimos de la misma.

- Giant dog - Huge dog - Enormous dog
- Deep hole - Profound hole
- Dusk - Nightfall - Sundown
- Shiny light - Bright light - Brilliant light

3.2.1.5. Operadores sobre predicados

Las pruebas consistirían en marcar de alguna forma por ejemplo una negación a un predicado y que las consultas no devuelvan resultados donde el predicado se cumpla.

- Dog not white
- Cat not black

3.2.2. Resultados

Sobre los buscadores que se han mencionado que realizaban búsquedas semánticas, se realizaron pruebas para ver cuál de ellos cumplía con alguna de las características mencionadas anteriormente, deseables para el proyecto.

Para realizar dicho estudio se seleccionaron los cuatro buscadores disponibles, donde dos son los buscadores más utilizados por los usuarios de internet y los otros dos son reconocidos por sus extensiones y mejoras para la búsqueda semántica.

Se presenta el cuadro con los resultados obtenidos:

CARACTERÍSTICA	GOOGLE	BING	DUCK DUCK GO	HAKIA
Objetos y atributos	0	1	0	0
Palabras polisémicas	0	0	1	0
Expansión por sinónimos	2	0	0	1
Semántica temporal	0	0	1	1
Operadores sobre predicados	0	0	0	0

Puntuación: 0 – No cumple requisito | 1 – Se cumple requisito en ciertas condiciones | 2 – Se cumple requisito | 3 – Supera expectativas

Como conclusión sobre el estudio de los cuatro buscadores se puede decir que ninguna cumple con los requerimientos del proyecto en la fecha de realizado el estudio y si bien **Google** resuelve correctamente el problema de Sinónimos, el resto de los problemas o no son resueltos o son resueltos en contextos de búsquedas muy particulares.

A continuación se presenta el detalle por cada característica probada y en el *ANEXO I – Consultas Capturadas*, se pueden ver las capturas de algunas pruebas realizadas.

3.2.2.1. *Objetos y atributos*

Se pretende mostrar la diferencia que hay en los resultados al intercambiar los atributos de ambos objetos y ver si se respetan las asociaciones planteadas.

En **Google** al cambiar el orden de los objetos y atributos los resultados de las consultas son muy similares, por lo que hace pensar que el orden de los factores no altera el resultado. Si bien existen algunos documentos que si tienen la asociación objeto-atributo esperada, no tienen mayor relevancia que otros documentos con la asociación contraria.

En el buscador **BING** los resultados aparentan ser mejores pues si se hacen las dos consultas alterando las relaciones objeto-atributo se obtienen documentos distintos y en uno de los casos aparecen algunos documentos relevantes primeros que otros aunque tengan la asociación contraria a la esperada, pero en muchos otros documentos está correcta la asociación.

En **Duck Duck Go**, si bien para un tipo de asociación objetos-atributos – *black cat White dog* - los resultados parecen ser correctos, si se cambia el orden los resultados comienzan a tener bastantes errores.

Por último en **Hakia** se detectan muchos inconvenientes, se recuperan documentos donde los atributos no están asociados a ningún objeto de los introducidos en la consulta.

3.2.2.2. Palabras polisémicas

En los buscadores tradicionales no existen mecanismos para darle a elegir al usuario en qué contexto se encuentran las palabras buscadas o por lo menos a que categoría pertenece alguna de las palabras con posibles ambigüedades.

En **Duck Duck Go** lo que se obtiene es: cuando se ingresa una palabra polisémica, se detallan los distintos significados de la misma, pero los documentos recuperados no dependen del significado, es decir, no se puede seleccionar el significado a buscar.

Por lo tanto se puede concluir que según la situación actual y las pruebas realizadas, ningún buscador de los probados, cumple con el requisito de resolver las búsquedas por palabras polisémicas.

3.2.2.3. Expansión por sinónimos

En este tipo de pruebas se espera que al ingresar una consulta, se obtengan documentos que contengan palabras que son sinónimos de las palabras buscadas, además de documentos con las palabras exactas.

Por ejemplo en el caso de la búsqueda “**enormous dog**”, se podrían obtener documentos que digan “**huge dog**”, “**big dog**” entre otros.

Google maneja un operador especial para sinónimos el cual cumple con las expectativas, ya que como se puede visualizar en las pruebas realizadas se encuentran documentos que hablan sobre **biggest dog**, **giant dog**. De todas formas al hacer la consulta sin ese operador, se puede notar que en los documentos recuperados se obtienen sinónimos de lo que se está buscando por lo que sigue funcionando para este punto.

Otro buscador que devuelve documentos con sinónimos es **Hakia** aunque en menor medida. Esto se puede deber al conjunto de documentos sobre los que se busca o por cómo está resuelto el problema, pero esto es algo que desconocemos.

El resto de los buscadores simplemente devuelven lo que se busca como palabra clave, sin expandir la consulta por sinónimos.

3.2.2.4. Semántica temporal

En los buscadores tradicionales se demuestra que, como hacen búsqueda por palabras claves, salvo que las fechas indicadas en las consultas estén contenidas en los documentos, no se recuperan documentos que contengan fechas similares, rangos y menos la asociación de eventos a las fechas.

En cambio en **Hakia** y **Duck Duck Go** existe alguna resolución de esta funcionalidad ya que por ejemplo con la búsqueda que se detalla en el anexo, **Maradona goals in the decade of 80**, algunos documentos recuperados, hablan sobre los goles de Maradona y mencionan la década del ochenta, pero parece más ser una resolución por palabras y que justo coinciden los años introducidos a una resolución de semántica temporal.

3.2.2.5. Operadores sobre predicados (negación)

No existen en los buscadores operadores para los predicados para denotar por ejemplo la negación. Si se introducen consultas utilizando la palabra NOT no se obtienen buenos resultados, lo que se puede hacer es excluir documentos que contengan la palabra a la que está aplicada el operador NOT, pero no ordenar documentos.

Este requerimiento no es resuelto entonces por ningún buscador.

4. Solución propuesta

El objetivo es diseñar un sistema de recuperación de información donde dada una consulta que indica la información que se debe recuperar, retorne un conjunto de documentos relacionados ordenados en el orden de mayor a menor relevancia según los criterios definidos para la búsqueda.

A partir de este objetivo se plantea la necesidad de describir la arquitectura del mismo, así como los procesos y módulos claves para poder cumplir con los criterios de búsqueda definidos anteriormente.

4.1. Diseño de la solución

4.1.1. Arquitectura del sistema

Los sistemas de recuperación de información (IR) tienen como problemas principales primero la indexación correcta de la biblioteca de documentos y segundo la búsqueda de los mejores resultados de una consulta sobre los documentos indexados.

A continuación se presenta un diseño general del sistema propuesto con el alcance completo para incluir desde la consulta del usuario hasta la visualización de los documentos recuperados.

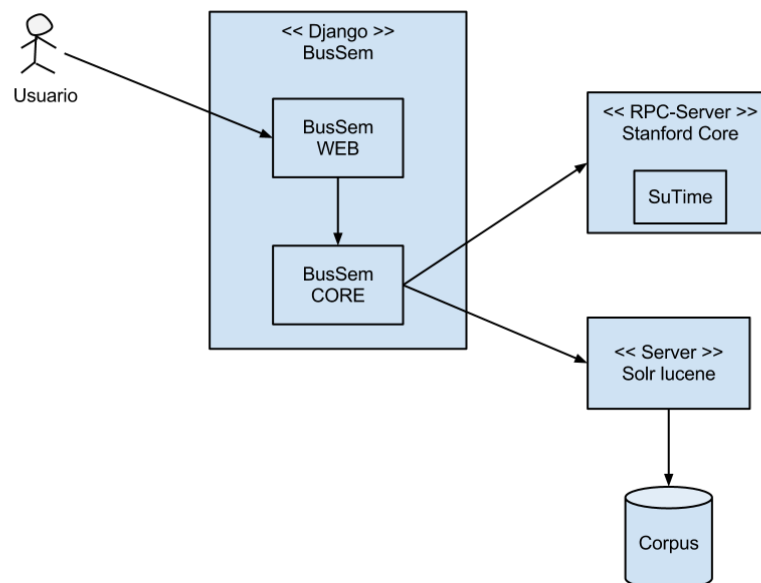


Figura 1 – Diseño general del sistema

Este diseño define los siguientes módulos que permiten la ejecución de una consulta:

- interfaz de usuario web donde el usuario ingresa su búsqueda
- el núcleo del sistema se comunica con un servidor de indexación que obtiene a partir de una búsqueda los documentos relevantes del conjunto de documentos total

- el núcleo también se comunica con un analizador sintáctico para resolver los problemas de semántica y ordenar los documentos según los criterios definidos
- La misma interfaz de usuario web que permite visualizar los documentos recuperados ordenados

A continuación se detalla el diseño de cada parte fundamental del sistema implementado.

4.1.2. SemQL - Lenguaje de consulta semántico

En un buscador el usuario ingresa una consulta en lenguaje natural, lenguaje que llamamos **UserQL**. Con el objetivo de aparear las búsquedas creadas por un usuario común en lenguaje **UserQL** con la información extraída de los documentos que fueron indexados, se define un lenguaje propio **SemQL**, que tiene distintas anotaciones para especificar las búsquedas deseadas.

Este lenguaje tiene como objetivo acortar las distancias entre el lenguaje **UserQL** y los documentos indexados del sistema de recuperación de información.

4.1.2.1. Especificación del lenguaje SemQL

El lenguaje se define mediante una serie de términos y variables que representan distintos conceptos dentro de un documento o una búsqueda particular.

Un **término** puede ser cualquier palabra lematizada, que aparezca en un documento, el cual va a representar un objeto o un atributo dentro del **SemQL** y que puede a su vez estar negado, o puede ser una expresión temporal para representar rangos de tiempo o fechas.

Los términos siempre se asocian a variables que permiten relacionar los diferentes términos entre sí. En el caso de las expresiones temporales, éstas se van a asociar además a una fecha o dos dependiendo de la expresión, y opcionalmente a un modo que permite introducir más detalle sobre el rango. Los valores de las fechas se basan en el formato *TimeX3* [29].

A los términos léxicos se les denomina **atributos**, a los negados **negaciones** y el resto son las **expresiones temporales**. Estas son una lista de palabras claves como:

- *Between*
- *Before*
- *After*
- *On*
- *Since*
- *Until*
- *Till*
- *At*
- *From*
- *To*

Todos los términos se separan con un \wedge (and), que es el único operador de concatenación permitido de términos, pudiendo existir tantos términos como se quiera, de cualquier tipo pero como mínimo un término como atributo.

Una **variable** puede tomar distintos valores como x, y, z, o cualquier otro valor. Representa un sintagma, grupo nominal o una oración, es decir que consta de un núcleo y constituyentes que se relacionan a este para integrarse en la oración, o dicho de otra forma, es una secuencia de palabras que funcionan como unidad dentro de la estructura jerárquica de una oración. Un cambio de variable dentro del **SemQL** representaría otro sintagma o grupo nominal. Por lo tanto los términos que contengan la misma variable consecutivamente van a estar asociados entre sí respecto al árbol sintáctico.

Los términos asociados a una misma variable dentro del **SemQL** están ordenados de izquierda a derecha según su relevancia dentro del sintagma o grupo nominal. El término más importante es el primero comenzando desde la izquierda, o dicho de otra manera, a más a la izquierda, más arriba en el árbol sintáctico.

A modo de ejemplo en un caso de una dependencia entre un sustantivo y su modificador, el sustantivo es el gobernador, el más importante por lo que se encuentra al inicio del **SemQL** y posteriormente el modificador.

4.1.2.2. Notación (BNF)

```

<semQL> ::= <termino variable> ['^' <sentencia>] [<semQL>]
<sentencia> ::= ['!'] <termino variable> | <expresion temporal>
<termino variable> ::= <termino> '(' <variable> ')'
<expresion temporal> ::= <expresion unaria> | <expresion binaria>
<expresion unaria> ::= <expr1> '(' <variable> ',' <expr value mod> ')'
<expresion binaria> ::= <expr2> '(' <variable> ',' <expr value mod> ',' <expr value mod> ')'
<expr value mod> ::= 'value' '=' ''' <fecha> ''' ['mod' '=' ''' <modo> ''']
<termino> ::= cualquier palabra en su forma de lema
<variable> ::= 'x' | 'y' | 'z' | ... {es una variable por cada SemQL que arranca}
<expr1> ::= 'before' | 'from' | 'after' | 'since' | 'to' | 'on' | 'until' | 'till' | 'at'
<expr2> ::= 'between'
<fecha> ::= yyyy | yyyy/MM | yyyy/MM/dd | yyyy/MM/ddTHH:mm:ss | yyXX | yyX *
<modo> ::= 'early' | 'mid' | 'late'
    
```

* El formato fecha tiene ciertas características especiales por lo que se especifica a continuación el significado de cada término posible:

Termino	Significado
yyyy	Año exacto
yyyy/MM	Año y mes exacto
yyyy/MM/dd	Día exacto
yyyy/MM/ddTHH:mm:ss	Día y hora exacta
yyXX	Siglo
yyX	Decada

Ejemplo

A partir de la oración “**the red and yellow flag was not raised after January of 1991 by an American soldier**”, si se utilizara un analizador sintáctico podríamos ver a partir del análisis de las dependencias distinta información para realizar la traducción a **SemQL**.

Por ejemplo, ver que existen dos sintagmas o grupos nominales. El primero contiene a **flag** como núcleo, es decir como término más relevante, **red** y **yellow** como atributos, **raised** como negación y la expresión temporal **after**. El segundo contiene a **soldier** como objeto principal y **american** como atributo, por lo que se arma el **SemQL** con dos variables, de la siguiente forma:

flag(x) ^ red(x) ^ yellow(x) ^ !rise(x) ^ after(x, value="1991/01") ^ soldier(y) ^ american(y)

Notar que parte de la información se puede perder, por ejemplo en este caso la **flag** no fue izada por el **soldier**, pero la recuperación y utilización de esta información adicional queda por fuera del alcance del proyecto.

4.1.3. Análisis de dependencias

Uno de los principales problemas a resolver es la interpretación semántica de los enunciados buscados y la comparación con la búsqueda proporcionada por el usuario.

El análisis de dependencias [30] de una oración tiene como resultado un grafo de dependencias donde existen pares de palabras relacionadas entre sí, una palabra es el “gobernador” sobre otra palabra “dependiente”.

A partir del análisis de dependencias realizado por herramientas externas se realiza un procesamiento para convertir esas dependencias en un formato único que será el formato de entrada para el módulo de ranking, el formato es el lenguaje **SemQL** definido anteriormente. Si bien la representación de estas dependencias depende de la herramienta utilizada para realizar el análisis, la traducción sería de la misma forma. Es decir, recorrer los distintos sub-grafos existentes e ir generando una secuencia de palabras que están asociadas a una misma variable por cada sub-grafo.

Una de las mayores ventajas de transformar los enunciados a **SemQL**, que es el mismo lenguaje de consulta, es la posibilidad de desacoplar lo que es el algoritmo de ordenamiento que se mencionará a continuación de cómo se forman los enunciados **SemQL** permitiendo una mayor facilidad para extender el sistema a otros idiomas o utilizar otras herramientas para el análisis de dependencias.

Para obtener la sentencia SemQL se analiza la oración con el módulo provisto por Stanford [31] para el análisis de dependencias y de expresiones temporales, realizando un *mapping* entre cada tipo de dependencia y su sentencia asociada en SemQL, estudiando solo los tipos de dependencias que son de interés para cumplir con los requerimientos del proyecto.

A partir de la colección de dependencias de **Stanford** se distinguen algunas dependencias de interés para el proyecto y son agrupadas en cinco categorías que permiten facilitar la traducción posterior.

A continuación se presentan las clases de dependencias definidas y ejemplos a partir del análisis de dependencias que se obtiene con **Stanford** de la frase “**the black soldier that died on 1914 at first world war was not guilty**”.

El cuadro de dependencias obtenido es:

```
det ( soldier-3 , the-1 )
amod ( soldier-3 , black-2 )
nsubj ( guilty-14 , soldier-3 )
nsubj ( died-5 , that-4 )
rcmod ( soldier-3 , died-5 )
prep_on ( died-5 , 1914-7 )
amod ( war-11 , first-9 )
nn ( war-11 , world-10 )
prep_at ( died-5 , war-11 )
cop ( guilty-14 , was-12 )
neg ( guilty-14 , not-13 )
```

Atributos

Es la clase de dependencias que indican la modificación de sintagmas nominales, siendo la primera clase evaluada para realizar la traducción al lenguaje SemQL.

Un ejemplo puede ser un modificador de un adjetivo, cuya dependencia es **AMOD** según **Stanford**. La frase anterior por ejemplo cuenta con la dependencia **amod(soldier, black)** que indica entonces que al sintagma nominal **men** se le asocia la palabra **black**. Además cuenta con la dependencia **amod(war, first)** y **nn(war, world)** que comparten el mismo gobernador, por lo que son dos modificadores asociados a la misma palabra.

El SemQL generado hasta el momento por esta dependencia es **soldier(x) ^ black(x) ^ war(y) ^ first(y) ^ world(y)**

Expresiones verbales

Es la clase de dependencias que indican la relación entre un verbo y un sintagma nominal, o se utiliza esta clase también para otros tipos de dependencias en las que el gobernador (según la traducción a SemQL) sería un objeto secundario mientras que el dependiente sería el atributo inicial.

Del ejemplo anterior se detectan por ejemplo las dependencias **nsubj(guilty, soldier)**, **nsubj(died, that)**. Estas dos dependencias generarían un **SemQL** de la forma **soldier(x) ^ guilt(x) ^ that(y) ^ died(y)**, pero como el objeto **soldier** ya es parte del **SEMQL** anterior, se adjunta el nuevo atributo a la misma variable, obteniendo el **SemQ**:

soldier(x) ^ black(x) ^ guilt(x) ^ war(y) ^ first(y) ^ world(y) ^ that(z) ^ die(z).

Expresiones temporales

Las dependencias que marcan expresiones temporales son generalmente preposiciones como **prep_around**, **prep_till** entre otras que contiene además de una palabra una marca de tiempo. Las expresiones temporales son almacenadas por separado de los atributos y modificadores obtenidos de las partes anteriores, pues implican un tratamiento diferente debido a los términos que tienen valores con marcas de tiempo.

Una vez que se está armando el **SemQL** a partir de los atributos, se verifican si los atributos están relacionados con alguna expresión temporal y a partir de ahí se anexa el **SemQL** correspondiente.

En la frase anterior se detecta la dependencia **prep_on(died, 1914)** que relaciona el evento del verbo con la fecha. En la traducción propuesta luego del análisis de dependencias, se obtiene el **SemQL** **die(x) ^ on(x, value="1914")** .

Agregando al **SemQL** anterior entonces tenemos un **SemQL** de la forma

soldier(x) ^ black(x) ^ guilt(x) ^ war(y) ^ first(y) ^ world(y) ^ that(z) ^ die(z) ^ on(z, value="1914").

Negaciones

Las negaciones son las dependencias más simples porque indican que una palabra está relacionada con la palabra **not** o negaciones similares. Estas negaciones son almacenadas en forma separada a las colecciones de atributos y expresiones temporales encontradas pues se utilizan simplemente para negar o no la expresión cuando se traduce a **SemQL**.

En el ejemplo se encuentra la dependencia **neg(guilty, not)** la que se traduce a un **SemQL** de la forma **!guilt(x)**, por lo que el **SemQL** de la frase sería **soldier(x) ^ black(x) ^ !guilt(x) ^ war(y) ^ first(y) ^ world(y) ^ that(z) ^ die(z) ^ on(z, value="1914")**.

Otras dependencias

Por último se encuentran otras dependencias que son adicionales que tienen menos pesos que las primeras dos categorías para generar un **SemQL** pero de todas formas influyen y pueden ser sub-grafos interesantes para las búsquedas.

En el ejemplo de la frase, una dependencia de este estilo es la **rcmod(soldier, died)** donde relaciona dos objetos que ya están contemplados en otras dependencias.

Si se traduce esto al lenguaje **SemQL** se puede ver que como **soldier** es el dependiente, queda el verbo **die** aplicado al mismo.

La traducción final de la frase al **SemQL** a partir del análisis de dependencias es entonces

soldier(x) ^ black(x) ^ !guilt(x) ^ die(x) ^ war(y) ^ first(y) ^ world(y) ^ that(z) ^ die(z) ^ on(z, value="1914").

4.1.3.1. *A continuación se presentan las dependencias capturadas y cómo cada una es interpretada según las cinco clasificaciones que se mencionaron.* **Dependencias capturadas**

A continuación se presentan los tipos de dependencia básicos que serán interpretados y traducidos al lenguaje **SemQL**, teniendo en cuenta la clasificación mencionada anteriormente. Existen otros tipos de dependencias que podrían ser incluidos o no para mejorar los resultados, analizando cuál es la clasificación correcta.

Atributos

Adjetivos

Amplían o precisan la significación del sustantivo o sintagma nominal. En **Stanford Parser** se denota de la siguiente forma **AMOD(NP, JJ)** y la traducción a **SemQL** es de la forma **NP(x) ^ ADJ(x)**.

Ejemplo: “*The first black american soldier died on 1914*” tiene las dependencias **AMOD(soldier, first)** y **AMOD(soldier, black)**.

Modificador Sustantivo

Un modificador sustantivo es cualquier nombre que sirve para modificar un sustantivo, explica un aspecto del sustantivo al que determina. La notación es **NN(NN1, NN2)** cuya traducción es **NN1(x) ^ NN2(x)**.

Ejemplo: “*The first black american soldier died on 1914*” tiene la dependencia **NN(soldier, american)**.

Expresiones verbales

Complemento directo

El complemento directo de una frase verbal es un sintagma nominal que es el objeto del verbo, es aquello que resulta afectado o modificado por el verbo. En las dependencias de **Stanford** se denota como **DOBJ(VP, NP)** y su traducción al **SemQL** es **NP(x) ^ VP(x)**.

Ejemplo: “*He ate a very tasty apple*” tiene la dependencia **DOBJ(eat, apple)**

Sujeto nominal

Un sujeto nominal es un sintagma nominal que es el sujeto sintáctico de una cláusula. Si bien está dentro de la categoría expresiones verbales, no siempre el gobernador de esta dependencia es un verbo. La salida del **Stanford Parser** es de la forma **NSUBJ(VP, NP)** o **NSUBJ(JJ, NP)** por ejemplo y su traducción es **NP(x) ^ VP(x)** o **NP(x) ^ JJ(x)** respectivamente.

Ejemplo: “*The first black american soldier died on 1914*” tiene la dependencia **NSUBJ(died, soldier)**.

Expresiones temporales

Todas las dependencias que contienen expresiones temporales son preposiciones. Estas dependencias indican el evento y el momento temporal y se dividen en cuatro tipos según el tipo de marca temporal que corresponda, en **SemQL** se traducen así:

- **BEFORE(x, TimeX3)** para las dependencias **PREP_BEFORE**, **PREP_TILL**, **PREP_UNTIL**
- **ON(x, TimeX3)** para las dependencias **PREP_AROUND**, **PREP_ON**, **PREP_PREP_AT**
- **AFTER(x, TimeX3)** para las dependencias **PREP_AFTER**, **PREP_SINCE**
- **BETWEEN(x, TimeX3, TimeX3)** para las dependencias **PREP_BETWEEN**

Negaciones

El modificador de negación interpretado es de la forma **NEG(WORD, NOT)** donde **WORD** puede ser cualquier palabra. Esta dependencia marca que la palabra **WORD** está negada en la oración original. En **SemQL** se traduce como **!WORD(x)**.

Otras dependencias

Adverbios

Un sintagma adverbial es un adverbio o frase adverbial que agrega información sobre la palabra que acompaña. En **Stanford** se denotan como **ADVMOD(WORD, ADVP)** donde la traducción a **SemQL** sería **WORD(x) ^ ADVP(x)**.

Ejemplo: *“He ate a very tasty apple”* tiene la dependencia **ADVMOD(tasty, very)**

Preposiciones

Todas las preposiciones que no tienen una fecha en formato **TimeX3** en su relación, son interpretadas en esta clase de dependencias.

Ejemplo: *“He sat in a chair”* tiene la dependencia **PREP_IN(sat, chair)** donde ninguna de esas palabras son expresiones temporales, por lo que la representación en **SemQL** sería **sat(x) ^ chair(x)**.

4.1.4. Motor de búsqueda

Un motor de búsqueda tiene diferentes problemas a resolver que se pueden dividir en las siguientes tareas [32]:

- *Crawling* – búsqueda de documentos que deben ser indexados
- Indexación – almacenamiento de documentos para permitir recuperación a partir de consultas
- Interpretación de consulta – mecanismo para comprender la consulta generada por el usuario
- Búsqueda – recuperación de documentos según la consulta deseada
- Ranking – ordenamiento de los documentos recuperados según una métrica de relevancia
- Presentación - despliegue de documentos ordenados.

En la implementación para la búsqueda semántica no se implementará un sistema de *crawling* ya que se parte de un set de documentos definido para el estudio. De igual manera no se implementará la tarea de interpretación ya que no entra en el alcance del proyecto, es decir se partirá de una consulta **SemQL** para realizar las búsquedas y el pasaje de **UserQL** a **SemQL** se deja para trabajo a futuro.

De todas formas, la traducción **UserQL** a **SemQL** implica el análisis sintáctico de la consulta **UserQL** y a partir de las dependencias generar el **SemQL** de búsqueda, por lo que podría ser resuelta bajo ciertas restricciones en la consulta del usuario.

A continuación se presenta cómo se resuelven el resto de las tareas más importantes de un motor de búsqueda.

4.1.4.1. Indexación

Los documentos son indexados en una base de datos donde el índice esta generado por las palabras claves que aparecen en cada documento. Como se introdujo anteriormente, se utiliza al igual que **Google** la técnica de índice invertido, teniendo como ventaja que para cada palabra, se tiene la referencia a cada documento en la que aparece.

Para indexar primero entonces se deben realizar ciertos pasos para que los documentos no contengan información redundante o innecesaria y las palabras claves sean fácilmente identificables.

Dentro de estos pasos que debe resolver el módulo de indexación se destacan [33]:

- Análisis léxico para determinar si los textos son legibles, si las puntuaciones están correctas y si no existen caracteres extraños que compliquen las búsquedas.
- Eliminación de *stopwords* que son aquellas que brindan poca información para la recuperación pues son palabras frecuentes en todo tipo de documentos.
- Extracción de palabras claves, obteniendo los lemas de cada palabra y generando los índices por dichos lemas.

4.1.4.2. Búsqueda

En cuanto a la búsqueda para recuperar documentos, será utilizada mediante las palabras claves de la consulta como realizan los buscadores estudiados anteriormente, obteniendo el conjunto de documentos relevantes en primera instancia.

En la *Figura 2 – Proceso de búsqueda de un documento* se puede visualizar el proceso por el que pasa una consulta para recuperar los documentos, el cual ejemplifica el proceso de búsqueda y los módulos que se deberán implementar.

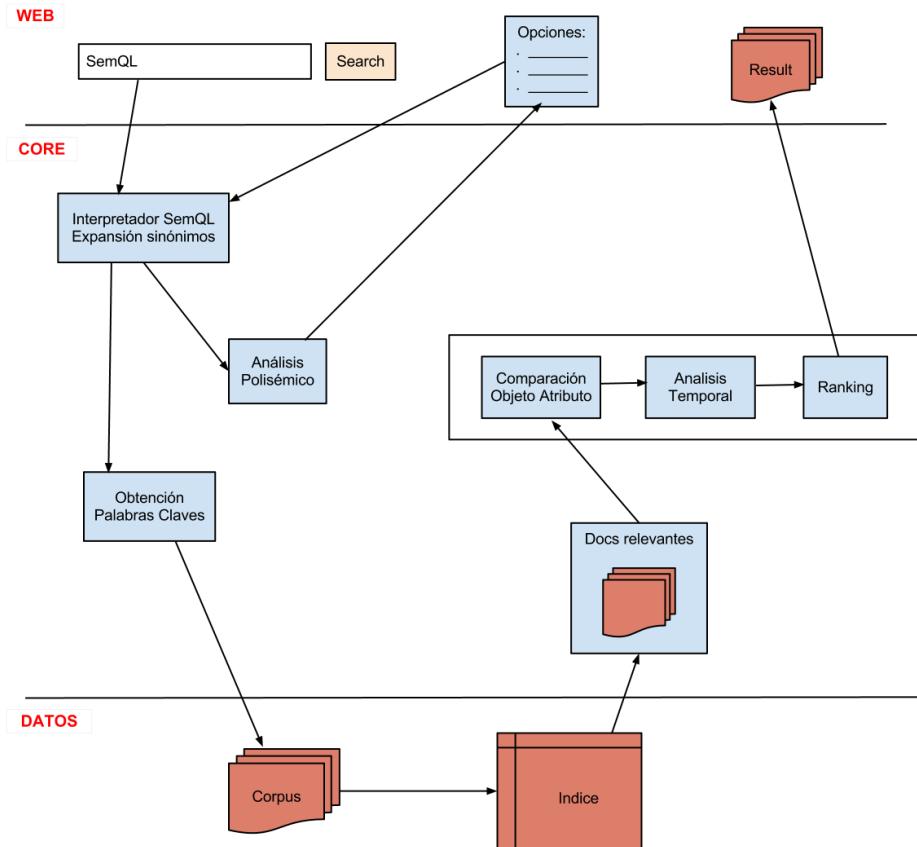


Figura 2 – Proceso de búsqueda de un documento

Posteriormente se ejecuta la etapa de ranking que se detalla a continuación.

4.1.4.3. Ranking

Como se puede ver en el ANEXO II - Modelos de recuperación de información, los modelos actuales se basan más que nada en palabras claves, la frecuencia de las ocurrencias de esas palabras, la distancia entre las mismas, probabilidades de relevancia, o en el caso de la web, los links entrantes y salientes de una página.

En nuestro caso, con un previo análisis de dependencias realizado, nos enfocamos más en una búsqueda semántica, adoptando un sistema de puntajes que le otorga mayor relevancia a las ocurrencias de los

objeto-atributo y análisis temporal, y que en un segundo plano ordena según la frecuencia de las palabras en caso de documentos similares.

El ranking semántico **SemQL** consiste en puntuar una colección de documentos relevantes de los que se conoce semánticamente su contenido, comparando con la consulta **SemQL** deseada.

Mediante un sistema de puntaje establecido, se puede determinar el orden con el cual los documentos serán devueltos al usuario. Con el objetivo de establecer el sistema de puntaje se definen los siguientes criterios y posteriormente se especifica la función de ranking.

Criterios

Un **documento relevante** es aquel documento que pertenece al corpus y contiene al menos un término como palabra clave.

Un **objeto** es el primer término asociado a una variable cualquiera, mientras que los términos consecutivos asociados a la misma variable son **atributos** de dicho objeto. El **objeto** en realidad puede ser cualquier término, es decir un verbo, adjetivo, nombre o cualquier palabra y esto va a depender de donde ocurra la primera coincidencia.

Se define una **jerarquía de relevancia** según las coincidencias

- La coincidencia del objeto, que sería el primer término que coincide, es el que dará mayor puntaje.
- La coincidencia de los atributos tienen el mismo puntaje independientemente de la posición
- Las coincidencias de expresiones temporales tienen el mismo puntaje que los atributos
- Las coincidencias de un atributo cuya semántica esté negada en el documento relevante, restará puntos salvo que el atributo en la consulta también esté negado y en ese caso sumará puntos.
- La coincidencia de los sinónimos de los atributos u objetos deben tener un menor puntaje que la ocurrencia del atributo u objeto mismo.
- El puntaje final estará dado en la suma de los puntos positivos, menos la resta de los negativos, fraccionando según la posición del objeto.
- En caso de empate en el puntaje de documentos, se definirá por la frecuencia que aparecen en el documento, en primer lugar los atributos y en segundo lugar las ocurrencias de negaciones.
- No se toman en cuenta distancias entre palabras ni repeticiones de palabras adicionales al orden que brinda la búsqueda por palabras claves

Puntajes y función de ranking

Tomando en cuenta los criterios anteriores, se define un sistema de puntaje el cual intenta reflejar la relevancia real que tendría un documento con respecto a lo que se quiere buscar semánticamente.

Un documento entre la colección de documentos que tenemos que ordenar, adoptará un puntaje del 0 al 1.

Se propone un sistema de puntajes predeterminados el cual puede ser modificado mediante configuración, donde los puntajes iniciales detallados a continuación son con los que se realizaron las pruebas del sistema.

A continuación se presentan los puntajes y las observaciones correspondientes, pensando en una consulta de una sola variable, como podría ser un sintagma nominal.

Tipo	Puntaje	Observaciones
Objeto	0.4	Puntúa por la coincidencia del primer término de la consulta. En caso de que el documento no contenga este término, el documento para esta consulta tendrá un puntaje de cero puntos.
Atributo	0.3 / #AtributosConsulta	Puntúa por coincidencia de cada atributo, en caso de que no existan se ignora el atributo de la consulta.
Expresión Temporal	0.2 / #ExprTempConsulta	Puntúa por coincidencia de cada expresión temporal, en caso de que no existan, se ignora la expresión.
Negación	0.1 / #Negaciones	Puntúa por coincidencia de la negación, es decir, el documento tiene el mismo atributo negado.
Negación-Atributo	-0.35	Resta puntos cuando la consulta contiene una negación que aparece como atributo en el documento, es decir se busca por la negativa y se encuentra un positivo.
Atributo negado	-0.25	Resta puntos cuando la consulta contiene un atributo y en el documento aparece negado, es decir se busca por la positiva y se encuentra negativa.
% Sinónimo	0.9	Puntúa menos que una coincidencia exacta de un objeto o atributo, se multiplica el puntaje de ese objeto o atributo por este porcentaje.
Diferencia por nivel	0.2	Indica cual es el ajuste a realizar por la aparición del objeto en un nivel menos de jerarquía (más a la derecha del resultado).

EJEMPLOS

Para demostrar manualmente como sería el cálculo del puntaje de un documento, se toman dos enunciados de un documento y se le aplica el análisis de dependencias de **Stanford** para obtener el **SemQL** para cada uno. Por otro lado, se generan consultas **SemQL** para realizar la evaluación y obtener el puntaje correspondiente.

Enunciado A

He returned to the studio in the 1980s and continued to record and perform into his late seventies.

Análisis de dependencias	nsubj(returned-2, He-1) nsubj(continued-10, He-1) xsubj(record-12, He-1) root(ROOT-0, returned-2) det(studio-5, the-4) prep_to(returned-2, studio-5) det(1980s-8, the-7) prep_in(returned-2, 1980s-8) conj_and(returned-2, continued-10) aux(record-12, to-11) xcomp(continued-10, record-12) xcomp(continued-10, perform-14) conj_and(record-12, perform-14) poss(seventies-18, his-16) amod(seventies-18, late-17) prep_into(record-12, seventies-18)
SemQL	he(x1) ^ return(x1) ^ studio(x1) ^ IN(x1, <TIMEX3 tid="t1" value="198X" type="DATE">the 1980s</TIMEX3>) ^ continue(x1) ^ record(x1) ^ seventy(x2) ^ late(x2)

Enunciado B

Born in Mississippi in the Depression, he escaped the misery of rural poverty and segregation by moving north to Detroit.

Análisis de dependencias	partmod(escaped-9, Born-1) prep_in(Born-1, Mississippi-3) det(Depression-6, the-5) prep_in(Mississippi-3, Depression-6) nsubj(escaped-9, he-8) root(ROOT-0, escaped-9) det(misery-11, the-10) dobj(escaped-9, misery-11) amod(poverty-14, rural-13) prep_of(misery-11, poverty-14) prep_of(misery-11, segregation-16) conj_and(poverty-14, segregation-16) prepc_by(escaped-9, moving-18) advmod(moving-18, north-19) prep_to(north-19, Detroit-21)
SemQL	misery(x1) ^ poverty(x1) ^ segregation(x1) ^ escape(x1) ^ poverty(x2) ^ rural(x2) ^ he(x3) ^ escape(x3) ^ bear(x4) ^ mississippi(x4) ^ depression(x4) ^ detroit(x4) ^ move(x4)

Consulta para estudio 1

Supongamos que un usuario quiere consultar sobre "estudios del año 1982 que continúan sin ser grandes sobre la gran miseria".

$studio(x) \wedge continue(x) \wedge !big(x) \wedge on(x, value="1982") \wedge misery(y) \wedge great(y)$

A partir de los enunciados descritos anteriormente y la consulta a puntuar, el procedimiento para la puntuación es el siguiente.

<p>Se consideran los términos relacionados con la variable x y se comparan con SemQL de enunciado A</p> <ul style="list-style-type: none"> + 0.4 por coincidencia del objeto 'studio' + 0.3 por coincidencia del atributo 'continue' + 0.2 por coincidencia de expresión temporal <p>El total es 0.9, no se obtuvo puntaje por el término negado pues no existe coincidencia</p>
<p>Se consideran los términos relacionados con la variable y se comparan con SemQL de enunciado B</p> <ul style="list-style-type: none"> + 0.4 por coincidencia del objeto 'misery' + 0.2 ya que la consulta no tiene expresión temporal + 0.1 ya que la consulta no tiene términos negativos <p>El total es 0.7, no se obtuvo puntaje por atributos</p>
<p>De ambos puntajes se realiza el promedio y se obtiene el puntaje del documento para la consulta realizada que en este caso es 0.8</p>

Consulta para estudio 2

Supongamos ahora que se quiere consultar sobre "grandes estudios que continúan en el año 1982 sobre la pobreza y la miseria".

$great(x) \wedge studio(x) \wedge continue(x) \wedge on(x, value="1982") \wedge misery(y) \wedge poorness(y)$

A partir de los enunciados descritos anteriormente y la consulta a puntuar, el procedimiento para la puntuación es el siguiente.

Se consideran los términos relacionados con la variable x y se comparan con SemQL de enunciado A

+ 0.4 por coincidencia del objeto 'studio'

+ 0.15 por coincidencia del atributo 'continue' y tener dos atributos asociados

+ 0.2 por coincidencia de expresión temporal

+ 0.1 ya que la consulta no tiene términos negativos

El total es 0.85 pero como el primer término de la consulta no coincide, se multiplica por el coeficiente de nivel, llegando a un puntaje de $0.85 \cdot 0.8 = 0.68$

Se consideran los términos relacionados con la variable y se comparan con SemQL de enunciado B

+ 0.4 por coincidencia del objeto 'misery'

+ 0.27 por coincidencia de un sinónimo de un atributo

+ 0.2 ya que la consulta no tiene expresión temporal

+ 0.1 ya que la consulta no tiene términos negativos

El total es 0.97.

De ambos puntajes se obtiene un puntaje total de 0.825

4.2. Implementación

La implementación de la solución requiere de la integración de diversas herramientas disponibles para el análisis de dependencias, de expresiones temporales y para resolver otros problemas como la búsqueda de palabras claves y la visualización de resultados, así como el desarrollo de ciertos módulos o clases que permiten generar un sistema de recuperación de información con todas las características planteadas.

El sistema está implementado en **Python** debido a que es un lenguaje que tiene librerías muy útiles para trabajar con el procesamiento del lenguaje natural y se tienen conocimientos debido a cursos y proyectos realizados anteriormente. Además su integración con todas las herramientas de Procesamiento de Lenguaje Natural es sencilla, como por ejemplo a partir del uso de la plataforma NLTK [34] que integra diversos módulos que pueden ser utilizados para el proyecto, como por ejemplo el módulo de **wordnet**.

A continuación se detallan los componentes externos utilizados y los módulos internos que componen el núcleo del sistema que interactúan con los anteriores.

4.2.1. Módulos externos

En el *ANEXO III - Herramientas evaluadas* se detallan las distintas herramientas que se evaluaron, de las cuales para la construcción del sistema, algunas de ellas han sido utilizadas con diversos objetivos.

4.2.1.1. Indexación y recuperación de documentos

Ante las dos herramientas disponibles, **Lucene** y **Lucene+SOLR** se opta por la utilización de **Lucene+SOLR** ya que los siguientes puntos son determinantes:

- **Lucene+SOLR** es simple de integrar con **Python**, disminuyendo los tiempos de integración de los módulos
- **Lucene+SOLR** contiene facilidades para la búsqueda y subrayado que podrían ser útiles que **Lucene** no tiene
- **Lucene** en las pruebas de instalación contenía errores para sistemas operativos **Windows** y **Ubuntu** de 64 bits

Para la comunicación con **Lucene+SOLR** se genera una interfaz que permite indexar y recuperar documentos comunicándose por **JSON**.

4.2.1.2. Interfaz de usuario

La herramienta **Django** brinda todas las funcionalidades que se necesitan para la implementación de una simple interfaz de usuario que permite al usuario ingresar una consulta y visualizar los documentos.

Se genera una interfaz dentro del núcleo del proyecto que le brinda servicios al proyecto web que contiene la interfaz gráfica.

4.2.1.3. Analizador de dependencias

En primera instancia se evalúa la herramienta de **Stanford** como analizador y debido a que brinda todas las características que son requeridas para poder realizar un correcto análisis de dependencias en inglés, se decide continuar utilizando esta herramienta que es potente, estable y fácil.

La comunicación con **Stanford** a diferencia de las otras herramientas utilizadas no es tan sencilla ya que requiere que esté levantado un servidor **RPC** ofreciendo el servicio para el análisis.

Para la implementación entonces se creó un paquete **Stanford** que permite levantar el servidor dejando un puerto abierto para aceptar pedidos.

Por otra parte se implementa una interfaz que se comunica con el servidor y hace el *request* para cada oración que se desea analizar, devolviendo la respuesta a quien utiliza la interfaz.

Luego de obtenida la respuesta del parser de **Stanford**, comienza el análisis particular para obtener las dependencias y poder generar entonces las sentencias **SemQL** asociadas a cada oración. Esta implementación se encuentra dentro del módulo de análisis de dependencias.

4.2.2. Módulos internos

A continuación se presenta una breve descripción de cómo se implementa cada modulo del sistema de recuperación de información para lograr la solución deseada.

4.2.2.1. *Indexador de documentos*

Es el módulo utilizado para generar el índice de documentos y de esta forma tener un conjunto de documentos indexados en **SOLR**. Este conjunto de documentos surge a partir del análisis de uno varios conjuntos de documentos que se denominan corpus.

La indexación consiste en analizar cada corpus brindado recorriendo cada oración de un documento para obtener las palabras claves a partir de su análisis de **Stanford** y almacenarlos utilizando la interfaz con **SOLR** de **python**. Esta indexación además de almacenar en la base de datos de **SOLR** cada documento, genera un archivo con la salida de **Stanford** con el análisis de dependencias completo con el objetivo de minimizar los tiempos de búsquedas posteriores, debido a que los documentos ya fueron analizados en primera instancia.

Es un módulo que se puede procesar varias veces para indexar distintos corpus o para generar un índice completo de todo el corpora.

Una característica del indexador es que cada vez que se ejecute, elimina y crea todo el índice nuevamente ya que dadas la características del proyecto no era necesario la indexación de nuevos documentos. De todas formas es posible reutilizar el módulo para realizar una indexación de un documento, pero queda por fuera del alcance del proyecto.

A continuación se presenta un pseudo-código que ayuda a comprender el funcionamiento.

```

Borrar Indice
FOR corpus in ColeccionCorpus {
    palabrasClaves = {}
    FOR documento IN ColeccionDocumentos {
        FOR oracion IN ColeccionOraciones {
            ColeccionTerminos = términos SemQL del analizador de dependencias
            FOR termino IN ColeccionTerminos {
                IF palabra(termino) NOT IN palabrasClaves {
                    IF palabra(termino) NOT IN stopWords {
                        palabrasClaves += {palabra(termino)}
                    }
                }
            }
        }
        solrDocumento = objeto basado en documento + palabrasClaves
        Indexa articulo anotado con dependencias en disco
        Indexa solrDocument
    }
}
    
```

4.2.2.2. *Recuperador de documentos relevantes*

Este módulo es el encargado de, a partir de una consulta de palabras claves, poder buscar en el índice de documentos total, los documentos que contienen dichas palabras relevantes.

El objetivo es brindar una colección de documentos relevantes para luego ser analizados correctamente por el analizador de dependencias.

La implementación de la recuperación de documentos está fuertemente relacionada con la forma en que los documentos son indexados. De esta forma, como se ha descrito anteriormente, los documentos son indexados con diversos valores (id, texto, keywords) por lo que para la búsqueda con el propósito de optimizar más los tiempos, se utilizará la búsqueda por el *field* de **SOLR keywords**.

Una particularidad de esta búsqueda es que debe devolver todos los documentos que contienen las distintas palabras claves solicitadas y solamente los que tienen esas palabras claves, independientemente del orden en el que aparecen ya que esta afinación de resultados es responsabilidad del analizador de dependencias y de módulo de ranking.

A continuación se presenta un pseudocódigo sobre esta funcionalidad suponiendo que se recibe una consulta “**consultaWeb**” desde la interfaz de **Django** que se menciona más adelante en el texto.


```

ColeccionPalabrasClaves = {conjunto de palabras claves de la consultaWeb}
consultaSolr = ""
FOR palabraClave in ColeccionPalabrasClaves {
    consultaSolr = 'keywords:' + palabraClave
    IF busquedaSinonimos {
        sinonimos = {sinonimos en WordNet de palabraClave}
        FOR sinonimo in sinonimos {
            FOR lema in lema(sinonimos)
                IF lema NOT IN ColeccionPalabrasClaves {
                    consultaSolr += 'OR keywords:' + lema
                }
            }
        }
    }
}
// Se obtienen documentos a partir de la consulta generada usando la interfaz con solr
ColeccionDocumentosSolr = solrInterface.recuperarDocuentos(consultaSolr)
FOR DocumentoSolr IN ColeccionDocumentosSolr {
    FOR oracion IN DocumentoSolr {
        IF ((ColeccionPalabrasClaves INT ColeccionPalabrasOracion) >= MinRequerido) {
            semQL = sentencia SemQL obtenida del analizador de dependencias
            Se almacenan palabras, dependencias y SemQL al documento
        }
    }
    Documento = [palabras, dependencias, SemQL]
    ColeccionDocumentos += {Documento}
}
Ordenar ColeccionDocumentos con SemQLRanking

```

4.2.2.3. *Analizador de dependencias*

Este módulo es el responsable de interpretar semánticamente oraciones, obteniendo para cada una, la sentencia SemQL que permite luego al módulo de ranking semántico tener los insumos necesarios para comparar si los documentos son relevantes para la sentencia **SemQL** de búsqueda.

El objetivo es recibir una oración y devolver la consulta **SemQL** correspondiente, además de proveer otras funcionalidades útiles para todo el proyecto.

Como se mencionó anteriormente, si bien existen diversos analizadores de dependencias, el analizador utilizado es el **Stanford Parser**.

A pesar de utilizar este analizador de dependencias, se decide modelar el sistema para que la sustitución de un tipo de analizador por otro sea lo más sencillo posible. De esta forma es posible que a futuro se pueda evolucionar el sistema de recuperación para funcionar con más idiomas donde cada uno puede tener un analizador que interpreta mejor sus dependencias.

El analizador simplemente recibe una oración y luego analizando la salida del **Stanford Parser**, contemplando las decisiones de diseño que se detallan en la sección **4.1.3 Analizador de Dependencias**, se devuelve el lenguaje semántico **SemQL** asociado a cada oración para luego poder ser utilizada por los algoritmos de ordenamiento.

4.2.2.4. Ranking Semántico SemQL

En la sección **4.1.5.1 Ranking Semántico**, se ha visto el diseño del sistema de puntajes para comparar un **SemQL** de consulta con otro **SemQL** del documento, pero hasta el momento se ha realizado tomando un **SemQL** de una sola variable. A partir de ahora se detalla cómo se realiza el cálculo del puntaje total.

Se define clase **SemQL** como el conjunto formado por los términos (atributos, negaciones, expresiones temporales) consecutivos con la misma variable, es decir que un cambio de variable genera una nueva clase **SemQL**, por lo tanto una consulta **SemQL** se transforma en una lista de clases **SemQL**.

Los documentos se transforman mediante el analizador en una lista de sentencias **SemQL**, y por el mismo motivo que vimos recién, cada una de estas se transforman en varias clases **SemQL**.

Estas clases son las que se comparan mediante el sistema propuesto hasta ahora, pero hay que comparar todos los **SemQL** de consulta contra todos los **SemQL** generados a partir del documento para obtener el puntaje total del mismo.

Una consulta tiene una lista de clases **SemQL**, para cada una de ellas se obtiene un puntaje al compararla con todas las clases **SemQL** del documento. El puntaje estará determinado por el máximo puntaje obtenido en dichas comparaciones.

Por lo tanto al hacer esto con cada una de las clases **SemQL** de consulta, se obtendrán distintos puntajes para cada una de ellas.

El puntaje total para ese documento será el de sumar cada uno de esos puntajes obtenidos de las clases **SemQL** de consulta y dividirlo por la cantidad total de clases **SemQL** de consulta.

Así se obtendrá el puntaje total para el documento en cuestión lo cual reflejará la ubicación final que dispondrá dicho documento en la lista final de documentos.

Acá se puede ver un pseudocódigo de lo descripto anteriormente:

```

listaSemQLcons = <semQL_1, semQL_2, ..., semQL_n>
listaSemQLdoc = <semQLdoc_N, semQLdoc_n, ..., semQLdoc_n>
puntosTotal = 0
for (semQLcons in listaSemQLcons):
    puntos = 0
    for (semQLdoc in listaSemQLdoc)
        // obtiene el puntaje de comparar sel semQLcons con el semQLdoc
        // se aplica sistema de puntajes explicado anteriormente
        pts = puntaje(semQLcons, semQLdoc);
        // me quedo con el maximo
        if (pts > puntos)
            puntos = pts
    end for
    // sumo los puntos obtenidos por el semQLcons al total
    puntosTotal = puntosTotal + puntos
end for
// divido finalmente el puntaje total por la cantidad de clases semQL de consulta
puntosTotal = puntosTotal / len(listaSemQLcons)

```

Por último, en caso de que dos o más documentos obtengan el mismo puntaje, se mostrará antes en el ranking el de mayor frecuencia de las ocurrencias del objeto en el documento, si persiste el empate se definirá por la frecuencia de los atributos y por último ocurrencias de negaciones en el documento. En caso de que continúe la paridad, el orden de los documentos será indeterminado ya que se supone que tienen la misma relevancia.

4.2.2.5. *Visualizador de documentos*

Desde la web se realiza la consulta donde el usuario la ingresa con el formato **SemQL** correspondiente.

La web se comunica con el core donde se expone el servicio principal que es el módulo para recuperar documentos y obtiene como resultado una lista de documentos ordenada según el ranking de **SemQL**. En la web entonces se visualizan los documentos recuperados del corpus ordenados según su relevancia de mayor a menor.

Los documentos tienen un conjunto de propiedades (título, texto, puntaje, entre otros) y serán visualizados en el orden en que aparecen en la colección.

A partir de modificaciones en el diseño gráfico se podría fácilmente incluir más propiedades a los documentos como links, porciones de texto importantes, autores o lo que fuera necesario según las necesidades que se requieran. Para la visualización de las pruebas simplemente se visualiza el título y la frase de mayor puntuación.

5. Validación de la solución

Con el objetivo de determinar si el sistema implementado mejora respecto a algún buscador se deben generar consultas con diversas características sobre un conjunto de documentos de prueba y poder verificar que los resultados obtenidos son mejores que los que se podrían obtener en otros buscadores.

En primer lugar se estudió cómo tener un conjunto de pruebas definido en el que se pudieran realizar las pruebas y verificar las mejoras de una manera tangible y luego de definido el conjunto de documentos para las pruebas se ejecutaron las mismas.

A continuación se presenta el estudio realizado para definir el conjunto de documentos que llamamos corpus y posteriormente se presentan los resultados obtenidos de las pruebas.

5.1. Corpus para validación de datos

Una de las bases para la implementación de sistemas de recuperación de información es contar con un conjunto de documentos confiable, que permita ejecutar las pruebas necesarias y demostrar los puntos fuertes y débiles de un trabajo.

Es por eso que se estudiaron diversos corpora, se evaluaron las ventajas y desventajas de cada uno y se tomó la decisión de generar un corpus personalizado que cumpla las características que el proyecto requiere y que a su vez sea confiable.

A continuación se presenta cada corpus estudiado y por último el corpus definido particularmente para este proyecto.

5.1.1. Corpus Time

Este corpus es un conjunto de cuatrocientos artículos extraídos de la revista **Time** y un conjunto de ochenta consultas sobre dichos artículos donde se marca la relación entre cada consulta y sus artículos correspondientes.

El estudio de este corpus consistió de las siguientes etapas:

- Análisis de consultas
- Análisis de artículos

Análisis de consultas

En primera instancia se estudiaron las diversas consultas que se brindan en el corpus, para evaluar si son consultas que contemplan los distintos problemas a resolver en este proyecto.

La particularidad de estas consultas es que son creadas por humanos con el objetivo de encontrar documentos que respondan los problemas planteados en las consultas.

Un ejemplo se da con la consulta “**PERSONS INVOLVED IN THE VIET NAM COUP**” que claramente no intenta recuperar documentos que contengan esas palabras, sino documentos que puedan devolver insumos como para responder esa pregunta. Un ejemplo similar sería la consulta “**WHAT COUNTRIES HAVE NEWLY JOINED THE UNITED NATIONS**”.

Los artículos del corpus no necesariamente tienen que tener esas palabras dentro de su texto, sino simplemente permitir a una persona luego de leer el artículo, contestar la interrogante.

Análisis de artículos por palabras claves obtenidas de consultas

A partir del análisis anterior, donde se ve que las consultas no son útiles tal y como se presentan en el corpus, se plantea la posibilidad de hacer búsquedas por palabras claves dentro de los artículos, a partir de palabras claves obtenidas de las consultas.

Esto implica indexar los artículos guardando las palabras claves de cada uno, analizar las consultas para identificar palabras que puedan ser útiles y luego realizar las búsquedas para evaluar los resultados.

La estrategia para indexar los artículos y luego poder hacer consultas se basa en utilizar **SOLR** con el objetivo de ya tener implementado los mecanismos de búsquedas por palabras claves.

Se generó un programa con **PERL** con el siguiente pseudocódigo:

```

Para cada artículo A dentro de corpus
    K = {} -- palabras claves del artículo
Para cada oración O dentro de A
    L = Lemas obtenidos de analizar O con Stanford Parser
    K += L - {stopwords}
Se genera documento Solr con texto de A y palabras claves K
    
```

De esta forma se creó el índice en **SOLR** y luego se comienza a crear las consultas con palabras claves utilizando las consultas brindadas del corpus.

A modo de ejemplo se analiza la consulta “**WHAT IS THE "HOT LINE" PROPOSAL**” donde se detecta que por ejemplo, los artículos que permitirían resolver esta consulta son artículos que deberían tener incluidas al menos las palabras **HOT**, **LINE** y **PROPOSAL** o algún sinónimo.

Se escribieron todas las consultas de esta forma, analizando una por una y se obtuvieron 83 consultas por palabras claves.

Luego se programó en **PERL** otra rutina, que recorre el conjunto de consultas por palabra clave, genera la consulta en **SOLR** y almacena en un archivo el resultado de los artículos que se pudieron obtener en comparación con los artículos que se deberían haber obtenido. Esta búsqueda de artículos se realizó en diversas etapas evaluando distintos factores: utilizando solo lemas, luego sinónimos, operadores booleanos de AND primero y de OR posteriormente.

Los resultados obtenidos se pueden visualizar en las planillas que se anexan con la documentación.

De los mismos en primera instancia se detectó un desfase en los índices de los artículos, dado que por ejemplo con la consulta “**HOT or LINE or PROPOSAL**” se obtenía un artículo, tal como se mencionaba en la información del corpus, pero la diferencia era que el artículo obtenido, no era el mismo que el que se brindaba como dato.

Por lo tanto se procede a re-indexar los artículos para corregir esos desfasajes y lograr mejorar la cantidad de hits, es decir, para varias consultas encontrar varios documentos que contenían las palabras claves, y algunos de estos artículos correspondieran con el dato provisto en la especificación del corpus.

A partir de estos cambios, se obtienen al menos treinta y cinco documentos cuyas búsquedas por palabras claves recuperan tantos o más documentos que los especificados en la definición del corpus, por lo que probablemente se podría utilizar como corpus final.

Luego se continuaron realizando pruebas pero se detectó que el hecho de que el corpus estuviera todo escrito en mayúsculas hacía que el **Stanford Parser** no se comportara de la mejor forma pues no se distinguen los sintagmas nominales de otros grupos sintácticos.

5.1.1.1. Evaluación

Luego de todas las pruebas realizadas se llega a la conclusión que el corpus tiene algunas desventajas claves que no permitirían generar buenos resultados o reutilización para futuros trabajos:

- Artículos todos con letra mayúscula
- Artículos con puntuaciones incorrectas
- Desfasaje en los índices de los artículos
- Poca cantidad de artículos
- Pocas marcas temporales dentro de los artículos
- Consultas de alto nivel

A raíz de todas estas desventajas se decide la no utilización de este corpus y se plantea analizar el corpus Cranfield.

5.1.2. Corpus Cranfield

El corpus de **Cranfield** es evaluado de la misma forma que el corpus **Time** y rápidamente se llega a la conclusión de que la situación es similar, principalmente por la complejidad de los artículos y las consultas relacionadas.

Una consulta del estilo “**is it possible to find an analytical, similar solution of the strong blast wave problem in the newtonian approximation**” donde los artículos a obtener esperaban poder responder esa pregunta, claramente no es una consulta que se pueda traducir a un lenguaje semántico sencillo.

Analizando el resultado esperado se visualiza que la respuesta no incluye palabras de la consulta, sino que se incluyen resultados que tienen alguna relación con el tema de estudio, relación imposible de detectar con una búsqueda por términos como se pretende en el proyecto.

Además de esta desventaja al igual que el corpus Time tiene el problema de que los artículos son todos escritos con minúscula, los textos son muy complejos y las puntuaciones no parecen ser del todo adecuadas.

Por lo tanto, se decide que lo mejor para poder demostrar si lo implementado en este proyecto puede contribuir a mejorar las búsquedas en determinados buscadores a partir de la semántica, es crear un corpus personalizado.

5.1.3. Corpus BusSem-2012

Se definen ciertos requisitos básicos que deben cumplir los artículos para ser incorporados como documentos dentro del corpus personalizado:

- Misma fuente de origen
- Textos claros y correctos en inglés
- Textos con marcas temporales
- Textos que permitan la búsqueda por términos
- Estáticos - no cambian con el tiempo

Analizando estas características, una de las mejores fuentes de origen de los documentos son los artículos de diarios ya que generalmente tienen buenas redacciones, son claros, incluyen marcas temporales y pueden ser utilizados para diversas consultas.

Se define que se creará un corpus a partir de artículos obtenidos por el sitio web del diario británico **The Times** [35] al que se puede ingresar a través del sitio web y utilizando el buscador, generar diversos conjuntos de artículos.

Cada conjunto de artículos es el resultado de una consulta, contemplando los primeros veinticinco artículos que se obtienen del buscador, solo tomando en cuenta los contenidos de **The Times**, excluyendo los contenidos de **The Sunday Times**.

Se implementa un sencillo sistema de recopilación de datos que permite facilitar el relacionamiento de diversas consultas con cada artículo encontrado, ya que consultas similares generalmente devuelven artículos similares y es de especial interés mantener esta relación. La información obtenida en este sistema es información que se especifica como información del corpus, que permite comprender más el alcance del mismo.

Se generan consultas con ciertos criterios, para cada consulta se hace la búsqueda en **The Times**, se almacenan los primeros veinticinco artículos y se van asociando a cada consulta. Un artículo puede pertenecer a muchas consultas y cada consulta tiene varios artículos asociados.

A continuación se presenta el detalle de las consultas generadas:

Consulta	Cantidad Noticias
high cost and low profits	25
low cost and high profits	25
high profits	25
low profits	25
low cost	25
high cost	25
afghan soldier american civilian	25
afghan civilian american soldier	25
afghan civilian	25
afghan soldier	25
american soldier	23
american civilian	25
low blood pressure high cholesterol	23
high blood pressure low cholesterol	24
low cholesterol	24
high cholesterol	24
low pressure	25
high pressure	24
pressure cholesterol	25

La idea de las consultas es generar artículos que contuvieran las mismas palabras para poder visualizar luego con un mejor algoritmo de búsqueda que pueden existir documentos más cercanos a los que el buscador del diario *The Times* brinda, simplemente mejorando la consulta por términos agregando la notación semántica.

Una vez obtenidos todos los artículos - se obtienen 351 - se procede a generar el corpus para que pueda ser interpretado correctamente por la herramienta de Búsqueda Semántica. Para esto se requiere analizar algunos textos y mejorar algunos temas de puntuación donde por ejemplo aparecen puntos que no definen oraciones sino son parte de sintagmas nominales. De todas formas se procesan e indexaron correctamente, permitiendo a partir de este corpus, poder realizar pruebas sobre lo implementado.

5.2. Resultados y evidencias

Las pruebas consisten en realizar las mismas o similares búsquedas que se realizaron para crear el corpus de prueba, comparando los documentos obtenidos por este sistema de recuperación con lo obtenido utilizando el buscador del sitio *The Times*.

5.2.1. Resultados por funcionalidad

En la evaluación de un sistema de recuperación de información se utilizan dos medidas que sirven para determinar si un sistema tiene un mejor comportamiento que otro sobre un corpus. Estas medidas son la **precisión** y el **recall** [36].

Usando la *Tabla 1 - Tabla de contingencia*, se puede definir que la **precisión** es la división entre los **verdaderos positivos** (documentos relevantes recuperados) sobre la suma de los **positivos** (documentos recuperados relevantes y no relevantes); mientras que el **recall** es la división entre los **verdaderos positivos** y la suma de estos con los **falsos negativos** (documentos relevantes no recuperados).

	RELEVANTES	NO RELEVANTES
RECUPERADOS	VERDADEROS POSITIVOS	FALSOS POSITIVOS
NO RECUPERADOS	FALSOS NEGATIVOS	VERDADEROS NEGATIVOS

Tabla 1 - Tabla de contingencia

A partir de la poca cantidad de documentos que se encuentran en el corpus que utilizamos de prueba y la cantidad de consultas realizadas sobre el mismo, no se toman como criterios de evaluación estas medidas ya que no serían objetivas.

La estrategia para evaluar el sistema implementado es realizar pruebas sobre cada requerimiento por separado, analizando los resultados y proyectando como sería el comportamiento del sistema sobre cada medida.

5.2.1.1. Objetos y atributos

Se pretende probar si la asociación de distintos términos a una variable en una consulta **SemQL** permite que los documentos que son obtenidos en primer lugar, con mayor puntaje tengan esta asociación.

Si se realiza la prueba con la consulta **american(x) ^ civilian(x) ^ afghan(y) ^ soldier(y)** se recuperan varios documentos donde los primeros dos tienen puntaje ideal. Los documentos que se obtienen son:

- *Sixteen Americans dead as helicopters crash in Afghan raids*, que contiene la frase "**as well as 14 Afghan soldiers and one American civilian**"
- *Americans killed in shooting on Afghan base*, que contiene la frase "**Two American civilians and two Afghans were killed in a shooting on an Afghanistan military base after an Afghan soldier opened fire yesterday**"

En cambio, si a la misma consulta se le cambia la asociación de los términos con las variables, es decir, la consulta es **american(x) ^ soldier(x) ^ afghan(y) ^ civilian(y)**. Con esta consulta, muchos documentos recuperados tienen puntaje ideal, pero existen otros que tienen menor puntaje. Particularmente los dos documentos mencionados anteriormente se encuentran con puntaje de 0.7 ya que no se brindan puntos por la asociación de atributos.

La asociación de objetos y atributos implica que los documentos que son recuperados a partir de las palabras claves de la búsqueda, son ordenados según esa asociación, por lo que se puede suponer que la precisión podría mejorar, debido a que los primeros documentos recuperados van a ser más relevantes. Es decir, si se recuperan diez documentos ordenados según la semántica, es probable que los documentos menos relevantes no sean recuperados en ese primer conjunto.

5.2.1.2. Expresiones temporales

En este tipo de pruebas se desea verificar si ingresando consultas con expresiones temporales se puede mejorar la recuperación de documentos, ordenando los documentos que tengan información sobre las expresiones temporales y que los rangos de tiempo y objetos relacionados coincidan.

A partir de la consulta **soldier(x) ^ between(x,value="2000",value="2013")** se recuperan varios documentos, pero tan solo dos con puntaje ideal. Esto es porque esos documentos tienen expresiones temporales dentro del rango consultado y relacionadas con el término asociado. Los documentos obtenidos son:

- *Nato investigating 'rogue' attack by Afghan soldier*, que contiene la frase **"In October, an Afghan soldier fired a rocket-propelled grenade"**
- *Insider attacks wreck trust and add risk to Afghanistan withdrawal*, que contiene **"die at the hands of an Afghan policeman or soldier this year, compared with just one in 2011, three the previous year and five in 2009"**

Se analizan los documentos encontrados y se visualiza que en el primer documento se asocia a **soldier** con la fecha **october/2013** mientras que en el segundo documento con el año **2009**. Generando nuevas consultas se puede ver que si los rangos son modificados, se obtiene un documento o el otro, dependiendo de las fechas.

A modo de ejemplo:

- **soldier(x) ^ between(x,value="2000",value="2010")** – devuelve el segundo documento en primer lugar con puntaje ideal
- **soldier(x) ^ on(x, value="2013")** – devuelve el primer documento como primero
- **soldier(x) ^ before(x, value="2009")** – devuelve todos los documentos con igual puntaje 0.8 pues ninguno cumple con la expresión temporal.

Las consultas con expresiones temporales podrían mejorar la precisión en las búsquedas ya que los documentos serían ordenados según la coincidencia temporal y eso haría que los documentos relevantes sean recuperados en primera instancia.

Por otra parte, dado que las expresiones temporales incluyen rangos o marcas temporales de épocas, el recall podría mejorar, dado que por ejemplo, si en una consulta se busca por un rango de diez años se recuperan documentos que tengan cualquier año como marca temporal, mientras que si no se pudiera incluir esa semántica temporal, se recuperarían solo los documentos que tienen el año exacto.

5.2.1.3. Sinónimos

La inclusión de búsquedas por sinónimos permite que los documentos recuperados puedan ser obtenidos y ordenados según el puntaje a pesar de no tener los términos exactos introducidos en la consulta **SemQL**. De esta forma un documento que no tiene ningún término exacto pero contiene sinónimos de cada uno de ellos y el significado semántico de las oraciones es similar al buscado por la consulta, puede ser más relevante que un documento que tiene algún término exacto.

Por ejemplo teniendo la consulta **deal(x) ^ good(x)** se recuperan los siguientes dos documentos en primer lugar, con un punto cada uno:

- *Pop a pill if you're not ill? Dr Google says no*, que contiene la frase "**who would like to live a good deal longer**"
- *Insurers plan their own legal DIY to hammer down cost*, que contiene la frase "**structure that convinces me I'm getting a good deal**"

A continuación en el ranking se encuentran diversos documentos con puntaje 0.96 que se deben a la aparición en los documentos de sinónimos de **deal** asociados al atributo **good**. A continuación aparecen otro grupo de documentos con puntaje 0.93 en donde no aparece ningún término exacto, pero si aparecen los sinónimos relacionados entre ellos.

Si se modifica la consulta **SemQL** para buscar por **manage(x) ^ well(x)**, los documentos recuperados son los mismos que con la consulta anterior pero en distinto orden, ya que se puntúa las apariciones de **deal** y **good** como sinónimos por ejemplo.

En cuanto a las mejoras del sistema de recuperación esta funcionalidad claramente mejora el recall, pues se recuperan documentos que en caso de no contemplar el uso de sinónimos y utilizar solo el criterio de palabras claves no se recuperarían, pero no por eso no son relevantes.

5.2.1.4. Negaciones

Se debe verificar que el hecho de incluir negaciones en las consultas **SemQL** mejora los resultados obtenidos, ya sea visualizando los documentos en primer lugar que se corresponden con la negación, o brindando menos prioridad a los documentos que no tienen el atributo en cuestión negado.

A partir de la consulta **banker(x) ^ anything(x) ^ !responsible(x)** se obtiene un documento relevante con puntaje ideal ya que existe tanto el objeto-atributo inicial como la negación asociada al mismo objeto. El documento obtenido es:

- *Spanish woes weigh heavily on markets*, que contiene la frase "**Politicians and bankers here are never responsible for anything**"

Por otro lado, con una consulta como $\text{banker}(x) \wedge \text{!big}(x)$ el documento anterior aparece en menor posición en la lista, pues la negación no se manifiesta en el documento.

Continuando con el mismo ejemplo, si utilizamos la consulta $\text{banker}(x) \wedge \text{anything}(x) \wedge \text{responsible}(x)$, que es muy similar a la primera solo que el atributo **responsible** no está negado, en la lista de documentos recuperados el documento anterior aparece con mucho menor puntaje, pues resta puntos (0.25) por tener un atributo negado en el documento. Sin embargo, si a los dos atributos le agregamos la negación es decir, la consulta sería de la forma $\text{banker}(x) \wedge \text{!anything}(x) \wedge \text{!responsible}(x)$, en este se suman puntos por coincidir la negación en el atributo **responsible** y se restan puntos por tener un atributo que se busca como negado, en el documento sin negación.

La negación al igual que la asociación de objetos-atributos mejora la precisión de un sistema de recuperación, pues los documentos que son recuperados son más relevantes, se obtienen documentos entre los primeros documentos que tienen más relevancia que los devueltos por un sistema de información sin esta funcionalidad.

5.2.2. Comparación de resultados con *The Times*

Se realizan pruebas comparativas entre lo que sería una búsqueda en *TheTimes* y una búsqueda utilizando el sistema de recuperación implementado que llamamos **BusSem**. Para simplificar debido a que *TheTimes* no maneja operadores con marcas de tiempo y ese estilo, se realizan comparaciones con búsquedas del estilo objetos-atributos, ya que las pruebas más específicas fueron realizadas anteriormente en la sección 5.2.1 *Resultados por Funcionalidad*.

5.2.2.1. Consulta: American Civilian Afghan Soldier

En la Tabla 2 - Búsqueda comparativa - American Civilian Afghan Soldier se aprecian algunos puntos favorables y otros no tanto de utilizar el buscador semántico. La traducción a SemQL sería **american(x) ^ civilian(x) ^ afghan(y) ^ soldier(y)**

Titulo Documento	Pos. BusSem	Pos. Times	Puntaje
Sixteen Americans dead as helicopters crash in Afghan	1	18	1.0
Americans killed in shooting on Afghan base	2		1.0
Two US soldiers killed as Afghan teacher opens fire	3	12	0.85
Focus on how soldier was able to leave his base	4	6	0.85
Soldiers shot on Helmand guard duty were just days from	5	5	0.85
US military death toll in Afghanistan reaches 2,000	6	1	0.85
Survivors of Afghanistan massacre to give evidence	7		0.85
US troops shot dead after Ramadan meal invitation	8		0.85
Timeline: Attacks by rogue forces	9	24	0.78
Plan for Afghan exit as protests grow	10	22	0.78
Soldier could face death for Afghan 'killing spree'	11		0.78
Afghan retreat	12		0.78
Video shows US 'massacre' soldier hid weapon under shaw	13	7	0.70
US 'massacre' soldier flown out of Afghanistan	14	9	0.70
US forces fear reprisals over pictures of desecrated dead	15		0.70
Taliban vow revenge after US soldier shoots Afghan	16	13	0.70
US soldier on 3am rampage walked in and shot families	17	21	0.70
America insists it is on the path to success in Afghanistan	18		0.70
Staff Sergeant Robert Bales, accused of killing 16 Afghans	19	4	0.70
2,000th US soldier killed in Afghan green on blue attack	20	2	0.70
Real cost to the West could be a collapse of confidence	21	23	0.70
Crocker to step down as Ambassador to Afghanistan	22		0.70
US soldier Robert Bales accused of killing Afghans said	23	8	0.70
US Army drops charges against soldier in 'kill for sport'	24		0.63
You can't guard completely against these horrors	25		0.63
US soldier kills 16 Afghans in dawn shooting spree	26	17	0.63
Soldier accused of murder rampage is named	27		0.63
Soldier accused of gunning down 16 Afghan civilians goes	28	3	0.63
Green-on-blue murders must not deflect us	29		0.63
Suicide bus attack kills at least 13 US troops in Kabul	30	19	0.63
Insider attacks wreck trust and add risk to Afghanistan	31	20	0.63
Handover is put at risk by attacks from within	32	10	0.63
Mosque blast at Afghan festival kills 36	33		0.63
Allies look to speed up Afghan exit next year	34		0.63
Wife of Afghan 'murder rampage' soldier speaks out	35		0.63
Book of the week: Hiroshima by Paul Ham	36		0.63
No Cut And Run	37	15	0.63
We haven't lost in Afghanistan, says General Sir Richard	38		0.63
Motorbike doctor	39		0.63
Haqqani bombs Kabul in revenge for terror designation	40		0.63

Tabla 2 - Búsqueda comparativa - American Civilian Afghan Soldier

En primer lugar se aprecia que ninguno de los dos primeros documentos devueltos por **BusSem**, que tienen puntaje ideal, tenían buena relevancia en las búsquedas en **The Times**. El primero está en la posición dieciocho, mientras que el segundo documento ni siquiera era devuelto en la consulta en **The Times**.

Existen muchos documentos que son recuperados por **BusSem** y que no están dentro de los primeros veinticinco documentos devueltos por **The Times**, mientras que también existen algunos documentos recuperados por **The Times** y no son recuperados dentro de los primeros cuarenta de **BusSem**.

Analizando algunos de los casos podemos ver que por ejemplo el artículo *“American killed in shooting on Afghan base”* tiene puntaje ideal según **BusSem** y no es recuperado por **The Times**. Este artículo está en el corpus debido a que se obtuvo en el buscador de **The Times** con la consulta *“american civilian”* pero se puede apreciar en el primer enunciado del documento la frase *“Two American civilians and two Afghans were killed in a shooting on an Afghanistan military base after an Afghan soldier opened fire yesterday.”* que tiene las dos relaciones objeto-atributo deseada.

Hay algunos documentos que existen en el corpus al haber sido recuperados por la consulta en **The Times** pero que no están dentro de los primeros cuarenta documentos recuperados por **BusSem**. Un ejemplo puede ser el documento titulado *“The State of the Taleban: extracts from the Nato report”*. Si se analiza el contenido se encuentra la frase *“Afghan civilians frequently prefer Taleban governance over GIROA, usually as a result of government corruption, ethnic bias and lack of connection with local religious and tribal leaders”*. El puntaje al analizar este enunciado es de 0.63 al igual que muchos documentos que fueron recuperados por **BusSem**. Si se amplía la cantidad de documentos recuperados por palabras claves de **SOLR** (en vez de cuarenta a cien por ejemplo), este documento es recuperado en la posición **37** por lo que nos hace pensar, que se debe quitar la restricción de cuarenta documentos a recuperar por **SOLR**, porque pueden existir algunos documentos que no aparezcan pues para **SOLR** no son relevantes, pero sí contienen la información semántica que se busca.

Otro punto que resulta interesante ver de la tabla comparativa es que existen algunos documentos con igual puntaje de **BusSem** y que en **The Times** tenían una mayor relevancia sobre otros que tienen el mismo puntaje de **BusSem**. Esto se produce por la aleatoriedad por la que se ordenan los documentos con mismo puntaje **BusSem**, que podría mejorarse con el objetivo de mantener el orden según la cantidad de ocurrencias de las palabras o algún otro criterio de ordenamiento.

5.2.2.2. Consulta: High cost and low profit

A partir de lo comentado en las pruebas de la consulta anterior, para realizar la prueba para esta consulta se amplía la cantidad de documentos recuperados por **SOLR** a doscientos, con el objetivo de poder evaluar todos los documentos y que el ordenamiento a partir del ranking semántico no deje afuera a documentos que tienen pocas apariciones de palabras claves, pero que con una sola aparición pueden ser más relevantes que otros documentos. De todas formas se analizan los primeros veinticinco documentos recuperados en la *Tabla 3 - Búsqueda comparativa - High cost and low profit* Donde la traducción a **SemQL** sería $\text{cost}(x) \wedge \text{high}(x) \wedge \text{profit}(y) \wedge \text{low}(y)$

Titulo Documento	Pos. BusSem	Pos. Times	Puntaje
Greek bailout agreed ahead of key meeting	1	15	0.96
German court could delay euro ruling	2	24	0.85
Ryanair looks down on rivals as it plots a course for Tel Aviv	3	1	0.85
Adapt or die, Indian challenger warns consulting's Big Five	4		0.85
Shares surge on Draghi's plan to save the euro	5	18	0.83
Seize the power to go green and save	6		0.83
Child benefit cuts put mortgages at risk	7		0.83
Advertising decline forces down profit at The New York Times	8		0.83
Sir George Jefferson	9		0.83
Have your say...	10		0.83
Leveson plan for complaints triggers compensation fears	11		0.83
German parliament approves Greek bailout	12	13	0.83
The year of the cock-up	13		0.83
Contracts fill in the margins	14	11	0.83
Tate & Lyle has nothing to sweeten City	15		0.83
Milford Haven counts cost of gas shortage	16	2	0.83
Plain packaging 'will lead to big rise in illicit tobacco'	17		0.83
EasyJet doubles dividend as profits soar	18	23	0.83
What to say about: Big-ticket theatre	19		0.83
The nightmare of basement conversions	20		0.815
Spanish unemployment reaches 25 per cent	21		0.795
Time to accentuate the positive	22		0.795
We must reduce the corporation tax burden	23		0.78
High energy costs 'are constraining growth'	24		0.78
Childcare revolution urged by TUC leader	25		0.78

Tabla 3 - Búsqueda comparativa - High cost and low profit

La situación es similar a la consulta anterior, existen documentos que no fueron recuperados en los primeros veinticinco de **The Times** pero son muy relevantes pues contienen al menos una parte completa de la consulta **SemQL** buscada.

Una de las mayores diferencias que se ven es que en esta consulta es que la intersección entre los primeros veinticinco resultados de cada buscador es un conjunto de documentos muy chico, de apenas siete documentos. Esto se debe a que las palabras buscadas son muy frecuentes en los textos y al buscar en todo el conjunto de documentos y no solo en el conjunto que fue devuelto por **The Times** para esta consulta, existen más documentos que contienen al menos algunas de los dos objetos y quizás algún atributo.

Si se verifican estos documentos se puede llegar a la conclusión de que son relevantes para la consulta **SemQL** ingresada, por lo que en realidad en el ranking de **The Times** son menos relevantes debido a que no se interpreta semánticamente los textos.

5.2.2.3. Consulta: Low blood pressure high colesterol

En este caso los resultados fueron muy similares a los obtenidos anteriormente, se comprueban las mismas particularidades. El detalle de los resultados se encuentra en la *Tabla 4 - Búsqueda comparativa - Low blood pressure high colesterol*, donde la traducción a **SemQL** sería **pressure(x) ^ blood(x) ^ low(x) ^ colesterol(y) ^ high(y)**

Titulo Documento	Pos. BusSem	Pos. Times	Puntaje
Dr Thomas Stuttaford	1		0.925
The DASH diet: the weight-loss plan approved by doctors	2	1	0.925
Over 35? Know your blood pressure	3	3	0.925
Food v drugs	4		0.925
Don't gamble with dementia	5		0.925
Blood test predictors for heart disease and strokes	6		0.925
Can superfruit sea buckthorn lower cholesterol?	7		0.925
Middle-aged people could die 15 years earlier because of lifestyle, study says	8		0.925
Our modern diet is giving us 'diabesity'	9	21	0.925
Dr Copperfield: inside the mind of a GP	10		0.925
Why a bit of flab is good for your health	11	11	0.925
Fried food not all bad, says study	12		0.925
A little knowledge can be a good thing	13		0.85
Alternate awards of 2012: Dopers, chokers, no-hopers, the good, the bad and Joey Barton	14		0.83
Conditions: High cholesterol	15	8	0.78
An egg a day is good for you, says British Nutrition Foundation	16		0.78
Screen babies to ensure early detection of high cholesterol, say doctors	17		0.78
Heart risk ten-year-olds to be tested for cholesterol	18		0.78
New diet rules - eat red meat, don't eat grapes	19	16	0.78
High cholesterol link to prostate cancer	20		0.78
Britons ignore advice on healthy diet	21		0.78
Blood pressure	22		0.775
Mid-life diet: what you should eat after 40	23	10	0.775
For a healthy life, aim for a total cholesterol level of less than five	24		0.775
Obese children face heart risk	25	2	0.775

Tabla 4 - Búsqueda comparativa - Low blood pressure high colesterol

6. Conclusiones y trabajo a futuro

6.1. Conclusiones

El proyecto consistió en una etapa de estudio sobre la existencia de buscadores que puedan resolver las búsquedas semánticas de manera eficiente y una etapa de diseño de una solución que tenía como objetivo mejorar los resultados al realizar consultas de este estilo.

En la primera etapa se estudiaron varios buscadores, desde los tradicionales a buscadores creados por grupos de investigación de lenguaje natural para relevar qué tan maduros estaban en la resolución de estos problemas.

El estudio nos indicó que ninguno de ellos resuelve de forma eficiente estos puntos en la actualidad. El único que resuelve uno de los temas planteados con satisfacción es *Google* con el tratamiento de sinónimos, pero el resto de los buscadores que se pudieron analizar resolvían algún punto pero en determinados contextos y no de manera genérica.

Además durante el estudio del estado de arte en esta temática, descubrimos que existen grupo dedicados a estudiarla y diseñar mejores soluciones que serán los posibles impulsores de mejoras futuras en los buscadores tradicionales. Un ejemplo es el caso de *Powerset* que fue adquirido por *Bing* para introducir a su motor de recuperación los conceptos semánticos de *Powerset* y es posible que *Google* en algún momento comience a resolver estos problemas ya sea adquiriendo soluciones de estos grupos o desarrollando las propias.

En la segunda etapa se estudió una posible solución al problema planteado, intentando diseñar un sistema de recuperación de información que a partir de una búsqueda con conceptos semánticos pueda retornar documentos de forma ordenada según la coincidencia semántica.

Durante el análisis de la solución se investigaron diversas herramientas para poder interpretar la semántica de las oraciones y de esta forma poder utilizar esa información para ordenar los documentos de manera eficiente. La herramienta de mayor utilidad para obtener este tipo de información fue el *parser* de **Stanford** que a partir de una oración devolvía los lemas que aparecen en la oración y su árbol de dependencias.

Con el objetivo de independizar la solución de qué sistema y lenguaje se utilice para obtener el árbol de dependencias, se resolvió realizar una traducción del árbol de dependencias a un lenguaje definido para este tipo de consultas que llamamos **SemQL (Semantic Query Language)**. De esta forma, si la búsqueda se traduce a un **SemQL**, luego analizando los contenidos del corpus obteniendo los arboles de dependencias y su traducción al **SemQL**.

Una vez recuperados los documentos según las palabras claves se ordenan los mismos según los criterios semánticos. Este ordenamiento surge del diseño de un algoritmo que contempla los requisitos planteados y define la relevancia de cada documento, comparando las sentencias **SemQL** de búsqueda y de los contenidos.

Luego de implementada esta solución se procedió a generar los casos de prueba deseados con contenidos obtenidos del diario electrónico **The Times** y se realizaron diversas pruebas para demostrar que la solución planteada puede mejorar notoriamente al introducir el análisis de dependencias y un correcto algoritmo de ordenamiento.

Estas pruebas fueron auspiciosas y nos hace creer que incorporando este tipo de análisis y algoritmos a buscadores tradicionales, se podría mejorar notoriamente la experiencia de búsqueda de los usuarios.

6.2. Trabajo a futuro

A continuación se presentan posibles mejoras a introducir al sistema de recuperación diseñado que no fueron incluidas dentro del alcance del proyecto.

Una de las mejoras necesarias para llevar este sistema de recuperación a que sea utilizado por los usuarios de un buscador web por ejemplo, es implementar la traducción del lenguaje natural al lenguaje **SemQL**. Con la solución implementada la consulta debe ser ingresada en lenguaje **SemQL** lo que puede llegar a ser dificultoso para el usuario común. Si se analiza la consulta del usuario con un *parser* y se obtienen las dependencias sería sencillo traducir la misma a un lenguaje **SemQL** que sea transparente para el usuario.

Una de las desventajas del sistema implementado es que el conjunto de documentos esta acotado y no se puede modificar en tiempo real. Se podría mejorando los módulos de crawling e indexación para que generen nuevos contenidos de forma simple y en modo online. Por otra parte, los contenidos con información semántica están almacenados en disco y quizás la mejor solución sea incorporarlos al índice de **SOLR**.

Por otro lado, existe la posibilidad de ampliar los idiomas con los que se recupera información de manera sencilla. Los cambios a realizar para tener la solución para otro idioma son tener un analizador de dependencias similar al **Stanford Parser**, para lograr una traducción al lenguaje **SemQL** correcta según el idioma, y el diccionario de sinónimos que está basado en el idioma inglés. Los algoritmos de recuperación y ordenamiento no se ven afectados por el idioma a utilizar.

Por último, el manejo de palabras polisémicas si bien esta planteado como requisito inicial del proyecto, luego de analizar las herramientas disponibles para resolver este problema y de no disponer con corpus correctamente anotado, fue dejado fuera del alcance. Para implementar esta solución se debería extender el lenguaje **SemQL** para que acepte una marca sobre el significado de un lema polisémico, y tener, para documento indexado, el significado correcto de cada palabra polisémica. De esta forma la mejora sería previo al rdenamiento de los documentos recuperados, ya que dentro de los documentos a ordenar, no existirían documentos con palabras con un significado distinto al que el usuario busca.

7. Referencias

- [01] "What is Freebase?", Freebase, http://wiki.freebase.com/wiki/What_is_Freebase%3F, último acceso 05/06/2012.
- [02] Antonio Ortiz, "Freebase y la web semántica", Error 500, <http://www.error500.net/freebase-web-semantica>, último acceso 05/06/2012.
- [03] Scott Prevost, "Powerset & Bing", Bing, http://www.bing.com/community/site_blogs/b/powerset/archive/2009/06/03/b.aspx, último acceso 12/08/2012.
- [04] Barney Pell, "Powerset: Natural Language and the Semantic web", Videolectures, http://videolectures.net/iswc07_pell_nlpsw, último acceso 12/08/2012.
- [05] Scott Prevost, "Powerset hybrid approach to semantic search", Channel 9, <http://channel9.msdn.com/blogs/jonudell/scott-prevost-explains-powersets-hybrid-approach-to-semantic-search>, último acceso 12/08/2012.
- [06] "About DuckDuckGo", DuckDuckGo, <https://duckduckgo.com/about>, último acceso 18/04/2013
- [07] "DuckDuckGo Sources", DuckDuckGo, <http://help.duckduckgo.com/customer/portal/articles/216399-sources>, último acceso 22/07/2012.
- [08] Fernando Arias Márquez, "Kosmix, o la búsqueda bien organizada", Acaprensa, <http://www.aceprensa.com/articles/kosmix-o-la-busqueda-bien-organizada>, último acceso 08/06/2012.
- [09] "Kosmix, buscador semántico", Gnos, <http://red.gnos.com/comunidad/nextweb/recurso/Kosmix-The-web-organized-for-you-Buscador-semant/f8e7409c-3f8f-4295-9246-c28f2a6caff5>, último acceso 07/06/2012.
- [10] "Kosmix, mucho más que un buscador", DesarrolloWeb, <http://www.desarrolloweb.com/actualidad/kosmix-mucho-mas-buscador-1653.html>, último acceso 07/06/2012.
- [11] Arun Radhakrishnan, "Kosmix: search start-up looking for Meaning in search queries", Search engine Journal, <http://www.searchenginejournal.com/kosmixcom-search-start-up-looking-for-meaning-in-search-queries/5216>, último acceso 07/06/2012.
- [12] "About Hakia", Hakia, <http://company.hakia.com/about.html>, último acceso 18/04/2013. [13] "Query Detection and Extraction - QDEX", Hakia, <http://company.hakia.com/qdex.html>, último acceso 01/08/2012.
- [14] "SemanticRank Algorithm", Hakia, <http://company.hakia.com/semanticrank.html>, último acceso 01/08/2012.
- [15] "Swoogle Manual", Swoogle, http://swoogle.umbc.edu/index.php?option=com_swoogle_manual, último acceso 21/07/2012.

- [16] "Swoogle Publications", UMBC eBiquity, <http://ebiquity.umbc.edu/project/html/id/53/Swoogle?pub=on#pub>, último acceso 22/07/2012.
- [17] "How Google Works", GoogleGuide, http://www.googleguide.com/google_works.html, último acceso 22/07/2012.
- [18] "How Google Searches the Entire Web in Half a Second", Mashable, <http://mashable.com/2012/04/23/how-google-search-works>, último acceso 23/07/2012.
- [19] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, "An Introduction to Information Retrieval", 2009, cap. 1, secciones 1 y 2.
- [20] Sergey Brin, Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", 1998, <http://infolab.stanford.edu/~backrub/google.html>.
- [21] Amit Singhal, "Introducing the Knowledge Graph: things, not strings", Knowledge Graph - <http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>, último acceso 20/08/2012.
- [22] Peet Morris, "Wolfram Alpha: How does it work?", Computer Weekly, <http://www.computerweekly.com/opinion/Opinion-Wolfram-Alpha-How-does-it-work>, último acceso 02/08/2012.
- [23] "About Ask", Ask, <http://www.ask.com/about>, último acceso 15/08/2012.
- [24] "Siri", Apple, <http://www.apple.com/ios/siri>, último acceso 15/08/2012.
- [25] David Flanagan, "MQL Reference Guide", Freebase, <http://mql.freebaseapps.com/index.html>, último acceso 02/08/2012.
- [26] "CQL, Contextual Query Language", Library of Congress, <http://www.loc.gov/standards/sru/specs/cql.html>, último acceso 01/09/2012.
- [27] "Search Operators", GoogleGuide, http://www.googleguide.com/advanced_operators.html, último acceso 23/08/2012.
- [28] "Bing Query Language", Microsoft, <http://msdn.microsoft.com/en-us/library/ff795667.aspx>, último acceso 01/09/2012.
- [29] Grupo de trabajo de TimeML, "Guidelines for Temporal Expression Annotation for English for TempEval", 2009, <http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/timex3guidelines-072009.pdf>.
- [30] Daniel Jurafsky, James H. Martin, "Speech and Language Processing (2nd Ed)", 2009, cap. 12, sección 7.

- [31] Marie-Catherine de Marneffe, Christopher D. Manning, "*Stanford typed dependencies manual*", 2008, http://nlp.stanford.edu/software/dependencies_manual.pdf.
- [32] Anwar A. Alhenshiri, "*Web Information Retrieval and Search Engines Techniques*", 2007, <http://www.alsatil.edu.ly/alsatili/issues/issue32007/e-n32007-4.pdf>.
- [33] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, "*An Introduction to Information Retrieval*", 2009, cap. 1, sección 2.
- [34] "*Natural Lenguaje Toolkit*", NLTK, <http://nltk.org>, último acceso 10/06/2012.
- [35] "*The Times Of London*", The Times, <http://www.thetimes.co.uk>, último acceso 01/09/2012.
- [36] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, "*An Introduction to Information Retrieval*", 2009, cap. 8, sección 3.
- [37] "*Introduction to Information Retrieval*", Stanford NLP, <http://nlp.stanford.edu/IR-book/newslides.html>, último acceso 10/10/2012.
- [38] Joydip Datta, "Ranking in Information Retrieval", 2010, <http://www.cse.iitb.ac.in/internal/techreports/reports/TR-CSE-2010-31.pdf>.
- [39] "*Apache Lucene Core*", Apache, <http://lucene.apache.org/core>, último acceso 01/07/2012.
- [40] "*Apache Solr*", Apache, <http://lucene.apache.org/solr>, último acceso 02/07/2012.
- [41] "*SolrPy*", Google, <https://code.google.com/p/solrpy>, último acceso 02/07/2012.
- [42] "*The Django Framework*", Django, <https://www.djangoproject.com>, último acceso 20/07/2012.
- [43] "*The Tarsqi Toolkit*", TimeML, <http://www.timeml.org/site/tarsqi/toolkit>, último acceso 29/06/2012.
- [44] "*Stanford Temporal Tagger: SUTime*", Stanford NLP, <http://nlp.stanford.edu/software/sutime.shtml>, último acceso 29/06/2012.
- [45] "*Stanford CoreNLP*", Stanford NLP, <http://nlp.stanford.edu/software/corenlp.shtml>, último acceso 10/07/2012.

ANEXO I – Consultas Capturadas

A continuación se presentan distintas pruebas realizadas con los buscadores estudiados.

Solo se visualizan los resultados de una consulta representativa para cada caso de estudio y para cada buscador, aunque las pruebas se realizaron para todas las consultas detalladas en el informe.

Algunos documentos recuperados están marcados con ROJO que serían los resultados no deseados en la consulta y en VERDE los resultados que sí se esperaban encontrar.

Objeto - Atributo

Google

The image shows two side-by-side Google search result pages. The left page is for the query 'black dog white cat' and the right page is for 'black cat white dog'. Both pages show search results with various links, ratings, and images. Several results are circled in red, indicating they are unwanted results. The red circles are around the following titles: 'White Dog Black Cat Cafe | 201 S Broadway, GB, WI 920 445 0440', 'Menu | White Dog Black Cat Cafe', 'White Dog Black Cat Cafe - Downtown - Green Bay | Urbanspoon', 'White Dog Black Cat Cafe - Green Bay, WI - Yelp', 'White Dog Black Cat Cafe - Green Bay, WI - Yelp', 'The Happy Hour Tour | White Dog, Black Cat', and 'Black Cat, White Cat (1998) - IMDb'. The right page also has a red circle around the title 'White Dog Black Cat Cafe - Green Bay, WI - Yelp'.

Bing

bing

2.690.000 RESULTADOS Filtrar por idioma Filtrar por región

[White Cat Black Dog | Ocean to Mountain...](#) Traducir esta página
whitecatblackdog.com

[Ocean to Mountain Tops ... I broke down to desire. But I'm on the taller side and am prone to falling.](#)


Black Cat White Dog Traducir esta página
blackcatwhitedognews.com

Black Cat White Dog is a popular Vancouver based newspaper designed for Pet Lovers!

[Black Dog, White Cat - Wrestling Match...](#) Traducir esta página
www.metacafe.com/watch/249039/black_dog_white_cat_wrestling_match

Our dog Bobo and our cat Zaki - are they friends or not? Watch Video about People & Stories.Dogs.Cats by Metacafe.com

[Imágenes de black dog white cat](#)
bing.com/images



Black Cat, White Dog - YouTube Traducir esta página
www.youtube.com/watch?v=88hev72ZY4

De Victoria Masina - 2 min - 62 vistas - Agregado 19/08/2012

My cat Bagheera watches with fascination our neighbor's dog Yoshi showing off his dashing skills in the hallway.

[Photobucket | black dog white cat...](#) Traducir esta página
photobucket.com/images/Black+Dog+White+Cat

View black dog white cat Pictures, black dog white cat Images, black dog white cat Photos on Photobucket. Share them with your friends on MySpace or upload your own!

Black Cat White Dog Traducir esta página
blackcatwhitedognews.com

Michi & Prisca's joint blog - little snippets from our daily lives in Singapore

[February | 2013 | White Cat Black Dog](#) Traducir esta página
whitecatblackdog.com/2013/02

10 posts published by ebelz during February 2013 ... I am not leaving the house today. A fantasy harbored for weeks. Save for procuring eggplant parmesan ingredients ...

[Black Dog, White Cat | Pink Moss](#) Traducir esta página
pinkmoss.wordpress.com/2013/03/13/black-dog-white-cat

13/03/2013 - My beautiful white cat named Kia and frisky black dog Lotti love to pose for photos.

[Amazon.com: White Cat, Black Dog: Dan...](#) Traducir esta página
www.amazon.com/White-Cat-Black-Dog/dp/B008KDUJ0U

Dan Matheny - MP3 - 0 opiniones
Original Release Date: June 1, 2012 ; Format - Music: MP3; Compatible with MP3

BUSQUEDA S RELACIONADAS

Black Cat Cafe
LPS Black and White Cat
Black and White Cat Photos
Cat Names with White Paws
Black and White Cat Drawings
Cafe Green Bay WI
The Bar Green Bay East
Whitedogblackcat

bing

2.340.000 RESULTADOS Filtrar por idioma Filtrar por región

[Black Cat White Dog](#) Traducir esta página
blackcatwhitedognews.com

Black Cat White Dog is a popular Vancouver based newspaper designed for Pet Lovers!

[Black Cat White Dog](#) Traducir esta página
splack.blogspot.com


Michi & Prisca's joint blog - little snippets from our daily lives in Singapore

[Black Cat, White Dog - YouTube](#) Traducir esta página
www.youtube.com/watch?v=88hev72ZY4

De Victoria Masina - 2 min - 62 vistas - Agregado 19/08/2012

My cat Bagheera watches with fascination our neighbor's dog Yoshi showing off his dashing skills in the hallway.

[Imágenes de black cat white dog](#)
bing.com/images



[Black Cat & White Dog - YouTube](#) Traducir esta página
www.youtube.com/watch?v=U7Alqs0yQ4Y

De HOODOBOS808 - 1 min - 4 vistas - Agregado 11/03/2013

Sign in with your YouTube Account (YouTube, Google+, Gmail, Orkut, Picasa, or Chrome) to add HOODOBOS808's video to your playlist.

[White Cat Black Dog | Ocean to Mountain...](#) Traducir esta página
whitecatblackdog.com

Ocean to Mountain Tops ... I broke down to desire. But I'm on the taller side and am prone to falling.

[Black Cat](#)
www.blackcatwhitedognews.com/editions/BCWD_summar2011.pdf - archivo de PDF

Black CatWhite Dog eel Summer Issue 2011 Adopt-A-Pet Pets Need Homes Cat Licensing Greyhounds Veterinarian Tips

[Black Cat White Dog News | Facebook](#) Traducir esta página
www.facebook.com/BlackCatWhiteDog

Black Cat White Dog News, North Vancouver, BC. 258 likes · 164 talking about this.

[The Happy Hour Tour: White Dog, Black Cat](#) Traducir esta página
thehappyhourtour.blogspot.com/2012/03/white-dog-black-cat.html

30/03/2012 - Drink Deals: 3 Domestic Beers for \$4; 2 Single Shot Rail Mixers for \$5; Buy One Shot, get 2nd half-off

[Menu | White Dog Black Cat Cafe](#) Traducir esta página
whitedogblackcatqb.com/food

201 S Broadway, GB, WI 920 445.0440 ... SOUPS-N-SALADS *Chili \$ 5.50. Loaded with beef & boasting the freshest garden peppers, beans, garlic & chilies accompanied by ...

Duck Duck GO

The image shows two side-by-side browser windows from Duck Duck GO. The left window has the search query 'black dog white cat' and the right window has 'black cat white dog'. Both windows display a list of search results. Several terms in the search results are circled in red: 'White Dog Black Cat Cafe', 'White Cat Black Dog', 'White Dog Roadhouse Black Cat Cafe', 'Black and White Cat Names', 'Beautiful black and white cats photos', 'Black Cat White Cat', and 'Black Cat White Cat (1998)'. The results include links to various websites such as whitecatblackdog.com, diningfollo.com, greenbaynightlife.com, catnamesmeow.com, petfinder.com, rottentomatoes.com, wikipedia.org, whitecatwhiteblackcatgcb.com, blackcatwhiteblackcatgcb.com, blackcatwhiteblackcatgcb.com, personalizationmall.com, amazon.com, and splanck.blogspot.com.

Hakia

hakia

Surface Semantics

Web
[News](#)
[Blog](#)

Deep Semantics

[PubMed](#)
[Credible](#)
[Galleries](#)

Regular Search

[Image](#)
[Twitter](#)

[White Cat Black Dog](#)
Ocean to Mountain Tops ... ebeliz. Born and raised in NYC, I spend all of my time with a **dog, cat,** garden, and several biology books.
<http://whitecatblackdog.com/2013/04/14/509/>

[White Cat Black Dog | Ocean to Mountain Tops](#)
Ocean to Mountain Tops ... I broke down to desire. But I'm on the taller side and am prone to falling.
<http://whitecatblackdog.com/>

[Cats & Dogs - Black and White Photography Posters and Prints ...](#)
[Cats & Dogs - Black and White Photography Posters and Prints - Find Cats & Dogs, Black and White Photography Posters and Prints at Art.com](#) or select a print or poster from Art.com's Galleries
<http://www.art.com/gallery/id--b240521-c23951/cats-dogs-black-and-white-photography-prints.htm>

[Photos/Pictures of Dogs \(and a cat\) \(Black & White\)](#)
Photos/Pictures of **Dogs (and a cat) (Black & White)** (Cuba)
<http://www.danheller.com/cuba-dogs-bw.html>

[Black and White Pictures of Dogs - Boost Inspiration](#)
40 Funny Pictures of Dogs and Cats; **Black and White Pictures: 25 Stunning Examples; 20 Amazing Examples of Black and White Photography, Random Pictures for Your Inspiration – N03; Surreal Pictures of Animals, Written by Waheed Akhtar.**
<http://www.boostinspiration.com/photography/black-and-white-pictures-of-dogs/>

[Black cat Sock Illustration Images. 5755 black cat ...](#)
White dog and black cat. Foto search Clip Art RF Royalty Free. Black cat. Foto search Clipart RF Royalty Free. Relax. Black cat silhouette for your design. Foto search Clipart RF Royalty Free. I love cats! Black cats silhouette in heart shape for your design.
<http://www.fotosearch.com/illustration/black-cat.html>

[Black cat - Wikipedia, the free encyclopedia](#)
Kutsushita Nyanko (Japanese for "SocksCat") is a traveling black cat that wears white socks. The character was created by San-X. ... **Black dog bias; Bombay (cat) Familiar spirit; Gobbolino, the Witch's Cat; Halloween; Vilification of black cats; White cat; Witch**
http://en.wikipedia.org/wiki/Black_cat

[White Dog Roadhouse / Black Cat Cafe Green Bay WI](#)

[White Cat Black Dog | Ocean to Mountain Tops](#)
White Cat Black Dog | Ocean to Mountain Tops
Ocean to Mountain Tops ... I broke down to desire. But I'm on the taller side and am prone to falling.
<http://whitecatblackdog.com/>

[Come on Back to Me | White Cat Black Dog](#)
Come on Back to Me | White Cat Black Dog
Ocean to Mountain Tops ... I'm looking at you Springtime sunshine! Yesterday was my birthday and it was disappointingly cold and gray.
<http://whitecatblackdog.com/2013/03/18/come-on-back-to-me-2/>

[Black Cat White Dog](#)
Black Cat White Dog is a popular Vancouver based newspaper designed for Pet Lovers!
<http://blackcatwhitedognews.com/>

[black VS white! Evil black cat makes white dog go crazy ...](#)
Take your side between **black and white** as an **evil black cat** launches its sharpest claws on a **white hapless white dog.**
<http://www.youtube.com/watch?v=eQOq3Ht8TJQ>

[White Dog Roadhouse / Black Cat Cafe Green Bay WI](#)
White Dog Roadhouse, Black Cat Cafe in Green Bay, Wisconsin. map, features, description and more on greenbaynightlife.com. write your own review.
<http://www.greenbaynightlife.com/White-Dog-Roadhouse-Black-Cat-Cafe.html>

[White Dog Roadhouse / Black Cat Cafe - Green Bay, WI](#)
Photos, menus, events, details, and more for **White Dog Roadhouse / Black Cat Cafe** in Green Bay, WI on Dining Follo.
http://diningfollo.com/Green_Bay/768/

[Cats & Dogs - Black and White Photography Posters and Prints ...](#)
[Cats & Dogs - Black and White Photography Posters and Prints - Find Cats & Dogs, Black and White Photography Posters and Prints at Art.com](#) or select a print or poster from Art.com's Galleries
<http://www.art.com/gallery/id--b240521-c23951/cats-dogs-black-and-white-photography-prints.htm>

[Photos/Pictures of Dogs \(and a cat\) \(Black & White\)](#)
Photos/Pictures of **Dogs (and a cat) (Black & White)** (Cuba)
<http://www.danheller.com/cuba-dogs-bw.html>

Sinónimos

Google


Google

[Web](#) [Imágenes](#) [Videos](#) [Más](#) [Herramientas de búsqueda](#)

Cerca de 17.500.000 resultados (0,20 segundos)

[Top Giant Dog Breeds - Best of the World's Largest Dogs](#)
[dogs.about.com/.../dogbre...](#) - Traducir esta página
De Jenna Stregowski - en 188 círculos de Google+
Learn all about the largest dogs in the world. Find out about **giant dog** breeds and learn about the pros and cons of **big dogs**.

[Imágenes de Enormous dog](#) - Notificar imágenes



[The Competitive World of Absurdly Enormous Dogs \(PHOTOS ...](#)
[www.manolith.com/...](#) - [Estados Unidos](#) - Traducir esta página
24/12/2012 - The stakes, and the **dogs**, just keep getting higher in the **dog-eat-dog** world of preposterously tall Great Danes.

[Boy And His Enormous Dog Hit the Big Time | Stuff.co.nz](#)
[www.stuff.co.nz/.../Just-a-boy-and-hi...](#) - Traducir esta página
12/04/2013 - Could it get any cuter? A boy and his **huge pet** dog have become world renowned after mum started taking daily photos.

['Giant George: Life with the World's Biggest Dog' written by Dave ...](#)
[www.dailymail.co.uk/.../Giant-Georg...](#) - Traducir esta página
10/04/2012 - 'Giant' George the world's largest **dog** is used to **enormous** amounts of attention and is all set for more with the release of his new book.

[Giant dog breed - Wikipedia, the free encyclopedia](#)
[en.wikipedia.org/.../Giant_dog_breed](#) - Traducir esta página
A **giant dog** breed is any of many dog breeds whose height and weight put them at the largest end of all dog breeds. There is no universal height or weight at ...

[World's largest dog? Truth! & Fiction! - Truth or Fiction](#)
[www.truthorfiction.com/.../hercules.h...](#) - Traducir esta página
12/04/2007 - For some reason the story began circulating widely on the Internet in March, 2007, and was accompanied by a picture of an **enormous dog** ...

[Enormous Dog Delighted to Have Military Dad Home - Mamiverse](#)
[www.mamiverse.com/giant-dog-delig...](#) - Traducir esta página
23/09/2012 - VIDEO: Few things are as heartwarming as the sheer delight of a dog welcoming his dad home from a tour of duty.

Bing



Enormous dog



1.020.000 RESULTADOS Filtrar por idioma ▼ Filtrar por región ▼

[Enormous Dog Reaching for ball Great Dane...](#) Traducir esta página

www.collegehumor.com/upick/6817237/enormous-dog

Dog's special Christmas "present" For Christmas break my friend and I drove from California to Colorado in the span of 2 days along with my dog Charlie.

[Fun with enormous dog - YouTube](#) Traducir esta página

www.youtube.com/watch?v=qjd5q8f9CXQ

De sigridsf · 1 min · 180 vistas · Agregado 05/11/2010

The boy and Enormous Dog ... Sign in with your YouTube Account (YouTube, Google+, Gmail, Orkut, Picasa, or Chrome) to add sigridsf's video to your playlist.

[Click here to see an enormous dog penis....](#) Traducir esta página

www.collegehumor.com/picture/38860

A funny picture from CollegeHumor. View "Click here to see an enormous dog penis." and more funny pictures on CollegeHumor

[Imágenes de Enormous dog](#)

bing.com/images



[Boy And His Enormous Dog Hit the Big Time...](#) Traducir esta página

www.stuff.co.nz/life-style/cutestuff/8542459

Have you got a giant pet? Share your stories, photos and videos. Could it get any cuter? A boy and his huge pet dog have become world renowned after mum started ...

[GoFun!.com - Picture - Enormous dog](#) Traducir esta página

www.gofuni.com/pictures/v/Enormous_dog/193

Enormous dog ... © G-FUN.com 2008.

[Enormous Dogs - Funny Animals, Funny Cats...](#) Traducir esta página

www.damnfunnypictures.com/funny-animals/308/enormous-dogs

- Funny Animal Pictures Crazy Cats, Funny Dogs brings daily updates of Funny Crazy Animals.

[Lindsay and summer the enormous dog...](#) Traducir esta página

www.youtube.com/watch?v=aS3hEOxcvDs

De garrett willard · 0 min · 133 vistas · Agregado 01/07/2010

This video was uploaded from an Android phone. ... Sign in with your YouTube Account (YouTube, Google+, Gmail, Orkut, Picasa, or Chrome) to add garrett ...

[Enormous dog house](#) Traducir esta página

memphis.craigslist.org/grd/3704244205.html

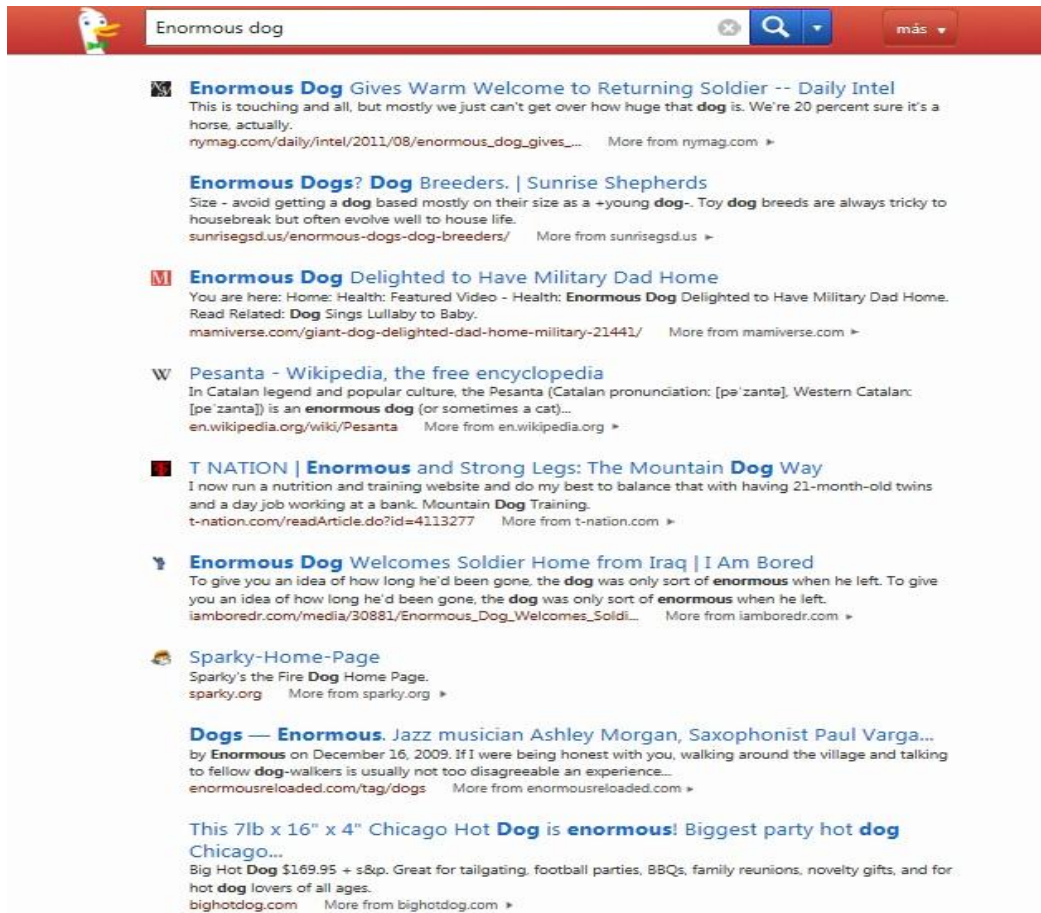
02/04/2013 · Enormous dog house 8ft long by 5ft wide by 6 ft tall at The point shingled roof. Dog house is 4 years old paid 600. Willing to let go for \$250 or best offer.

[Probono Enormous Dog Biscuits - Dandy Dogs](#) Traducir esta página

[www.dandydogs.co.za/pPROBONO%20ENORMOU\\$%20DOG%20BISCUITS/Pr...](http://www.dandydogs.co.za/pPROBONO%20ENORMOU$%20DOG%20BISCUITS/Pr...)

Enormous Dog Biscuits which are individually wrapped to maintain freshness and

Duck Duck Go



The screenshot shows the Duck Duck Go search interface with the search term "Enormous dog". The search bar is at the top, and the results are listed below. Each result includes a small icon, a title, a brief description, and a link to the source website.

Enormous Dog Gives Warm Welcome to Returning Soldier -- Daily Intel
 This is touching and all, but mostly we just can't get over how huge that **dog** is. We're 20 percent sure it's a horse, actually.
nymag.com/daily/intel/2011/08/enormous_dog_gives... More from nymag.com ▶

Enormous Dogs? Dog Breeders. | Sunrise Shepherds
 Size - avoid getting a **dog** based mostly on their size as a +young **dog**-. Toy **dog** breeds are always tricky to housebreak but often evolve well to house life.
sunrisegsd.us/enormous-dogs-dog-breeders/ More from sunrisegsd.us ▶

Enormous Dog Delighted to Have Military Dad Home
 You are here: Home: Health: Featured Video - Health: **Enormous Dog** Delighted to Have Military Dad Home.
 Read Related: **Dog** Sings Lullaby to Baby.
mamiverse.com/giant-dog-delighted-dad-home-military-21441/ More from mamiverse.com ▶

Pesanta - Wikipedia, the free encyclopedia
 In Catalan legend and popular culture, the Pesanta (Catalan pronunciation: [pe'zanta], Western Catalan: [pe'zanta]) is an **enormous dog** (or sometimes a cat)...
en.wikipedia.org/wiki/Pesanta More from en.wikipedia.org ▶

T NATION | Enormous and Strong Legs: The Mountain Dog Way
 I now run a nutrition and training website and do my best to balance that with having 21-month-old twins and a day job working at a bank. Mountain **Dog** Training.
t-nation.com/readArticle.do?id=4113277 More from t-nation.com ▶

Enormous Dog Welcomes Soldier Home from Iraq | I Am Bored
 To give you an idea of how long he'd been gone, the **dog** was only sort of **enormous** when he left. To give you an idea of how long he'd been gone, the **dog** was only sort of **enormous** when he left.
iamboredr.com/media/30881/Enormous_Dog_Welcomes_Soldi... More from iamboredr.com ▶

Sparky-Home-Page
 Sparky's the Fire **Dog** Home Page.
sparky.org More from sparky.org ▶

Dogs — Enormous. Jazz musician Ashley Morgan, Saxophonist Paul Varga...
 by **Enormous** on December 16, 2009. If I were being honest with you, walking around the village and talking to fellow **dog**-walkers is usually not too disagreeable an experience...
enormousreloaded.com/tag/dogs More from enormousreloaded.com ▶

This 7lb x 16" x 4" Chicago Hot Dog is enormous! Biggest party hot dog Chicago...
 Big Hot **Dog** \$169.95 + s&p. Great for tailgating, football parties, BBQs, family reunions, novelty gifts, and for hot **dog** lovers of all ages.
bighotdog.com More from bighotdog.com ▶

Hakia



Enormous dog



Surface Semantics

[Click here to see an enormous dog penis. - CollegeHumor Picture](#)

A funny picture from CollegeHumor. View "Click here to **see an enormous dog penis.**" and more funny pictures on CollegeHumor
<http://www.collegehumor.com/picture/38600>

[Web](#)
[News](#)
[Blog](#)

Deep Semantics

[Piccsy :: Enormous Dog](#)

Beautiful, inspirational and creative images from Piccsy. Thousands of Piccs from all our streams, for you to browse, enjoy and share with a friend.
<http://piccsy.com/2011/08/enormous-dog/>

[Pubmed](#)
[Credible](#)
[Galleries](#)

Regular Search

[Giant dog](#) | [very big dogs](#) | [tall dog](#) | [advantages](#) | [disadvantages](#)

A **giant dog**, a really tall **dog** is **enormous**, but usually these **big dogs**, are very gentle and make excellent family pets.
<http://www.dogbiz.com/giant-dog-breeds.html>

[Image](#)
[Twitter](#)

[World's largest dog? Truth! & Fiction! - TruthOrFiction.com ...](#)

..., 2007, and was accompanied by a picture of an **enormous dog** being walked near a horse. That picture is ... a site about large dogs. Updated 4/12/07. A real example of the eRumor as it has appeared on the Internet: Now this ...
<http://www.truthorfiction.com/rumors/h/hercules.htm>

[Hercules: The World's Biggest Dog Ever According to Guinness ...](#)

enormous; hercules the **dog** more by Nicole. View Profile. MJ and MC sittin' in a tree. SHE laid THAT? Magic Eyes. Recommended; Comments Picture "Possibly the worlds largest frosted flake!! Almost 3 in. diameter.
<http://www.collegehumor.com/picture/929133>

['Giant George: Life with the World's Biggest Dog' written by ...](#)

'Giant' George the world's largest **dog** is used to **enormous** amounts of attention and is all set for more with the release of his new book. Written by his owner, Dave Nasser, 'Giant George: Life with the World's Biggest Dog' charts the gigantic allowances that the 230 pound Great Dane ...
<http://www.dailymail.co.uk/news/article-2127960/Giant-George-Life-Worlds-Biggest-Dog-written-Dave-Nasser-7ft-Great-Danes-owner.html>

Expresiones temporales

Google

The image shows a Google search interface. The search bar contains the text "Maradona goals in the decade of 80". Below the search bar, there are tabs for "Web", "Imágenes", "Más", and "Herramientas de búsqueda". The search results are displayed below, showing several links with brief descriptions. The results include Wikipedia, Biography, YouTube, Football Pantheon, Bleacher Report, In Bed With Maradona, Guardian, and The Score.

Google Maradona goals in the decade of 80

Web Imágenes Más Herramientas de búsqueda

Cerca de 12.400.000 resultados (0,23 segundos)

[Diego Maradona - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/.../Diego_Maradona - Traducir esta página
 During his two injury-hit seasons at Barcelona, **Maradona** scored 38 goals in 58 as the most important personality in Sports in the last decade in his country. [80]: Over the years following his retirement his health seriously deteriorated.

[Diego Maradona Facts](#)
biography.yourdictionary.com/diego-... - Traducir esta página
 Diego **Maradona** facts: Diego **Maradona** (born 1961) is an icon in Argentina, much like Evita Peron. ... The second goal, however, guaranteed him a place in soccer history. and selection as athlete of the decade by the British, whom he quite possibly had cheated out ... An 80-piece tango orchestra entertained the crowd.

[Newcastle United Best Goals - EPL 1st decade - YouTube](#)
www.youtube.com/watch?v...list... - Traducir esta página
 This clip is of Newcastle United few best goals compiled together to make a ... 58 **Maradona** vs Ronaldinho vs Zidane by Alberto Maldonado; 59 **maradona** vs zidane ... 79 The 10 Legacy by TheThreeGods; 80 Arsenal Fight Manchester United ...

[The 100 greatest players of all time - Football Pantheon](#)
footballpantheon.com/.../the-100-gre... - Traducir esta página
 20/09/2011 – Who is will finish on top, Diego **Maradona** or Pele? Pele was the best player in the world for a whole decade. whether pele would've scored even 200 goals if he played in the 80s... whereas **maradona** constantly played ...


[Picking the Greatest Soccer Player of Every Decade Since 1920 ...](#)
bleacherreport.com/.../1423543-picki... - Traducir esta página
 28/11/2012 – Here are the greatest players, one from each decade, from the 1920s to now. Just like scoring over 1000 goals, winning 3 World Cups, being elected ... ever and **Maradona** won by a solid clip but Fifa came the award together. a forward involved as much in the build-up play (80-80passes per game).

[JOURNAL Archive — IBWM - In Bed With Maradona](#)
inbedwithmaradona.com... - Reino Unido - Traducir esta página
 hace 4 días – That goal in 1977 sealed victory for Liverpool over St Eienne and would ... Starved of senior football for over a decade the people of Derry got behind ... 'derby' game and that was against Sligo Rovers, a mere 80 miles away.

[Classic YouTube | Diego Maradona's warm-up wonders, Pietro ...](#)
www.guardian.co.uk/.../classic-youtu... - Traducir esta página
 28/03/2013 – Diego **Maradona's** warm-up wonders, Pietro Mennea and 80s monster trucks ... Redbridge: a striker in Georgia misses an open goal from less than a yard out in ... on the sport in that decade; and some classic mud-bog-racing.

[21 reasons why the 80s were the best sporting decade - The Score](#)
www.thescore.ie/21-reasons-80s-best-... - Traducir esta página

BING



Maradona goals in the decade of 80 🔍


739.000 RESULTADOS Filtrar por idioma ▼ Filtrar por región ▼

[La década del 80: Personajes, Diego Armando Maradona](#)
nostalgia.porlos80s.blogspot.com/2006/12/personajes-diego-armando... 13/12/2006 · En tierra azteca, apareció en todo su esplendor, Diego Armando Maradona ... Mas películas y series de la **década del 80**.


[Who's the greatest? Messi or Maradona...](#) Traducir esta página
www.goal.com/.../03/21/3842131/whos-the-greatest-messi-or-maradona · The previous occasion on which we were shown a **Maradona goal** from '88 replayed in ... wandering onto the Camp Nou field as a squat, scruffy teenager around a **decade** ago.

[Who's the greatest? Messi v Maradona...](#) Traducir esta página
www.goal.com/.../01/01/3838090/whos-the-greatest-messi-or-maradona · Draw £ 32.50: Boca Juniors £ 28.80 ... The previous occasion on which we were shown a **Maradona goal** ... Nou field as a squat, scruffy teenager around a **decade** ...


Vídeos de Maradona goals in the decade of 80
[bing.com/videos](#)




Top 5 Best Football Players
YouTube



Cristiano Ronaldo 2012 | Skills & Goals
YouTube



Mario Alberto Kempes, mejor jugador de la década de los 80
YouTube



real madrid 2009-2010 skills and tactics
YouTube

[Diego Maradona - Wikipedia, the free encyclopedia](#) Traducir esta página
en.wikipedia.org/wiki/Diego_Maradona · Early years · Club career · International career · Playing style
Maradona's second goal, just four minutes after the hotly disputed hand-goal, was later voted ... in Argentina, as the most important personality in Sports in the last **decade** ...

[Maradona's mother – Dona Tota has died at...](#) Traducir esta página
maradonagoal.com/2011/11/21/maradona-s-mother---dona-tota-has... · **Maradona Goals – Pele Maradona | Maradona hand of god** ... Legendary footballer Diego Armando Maradona's mother, Dona Tota has died because ...

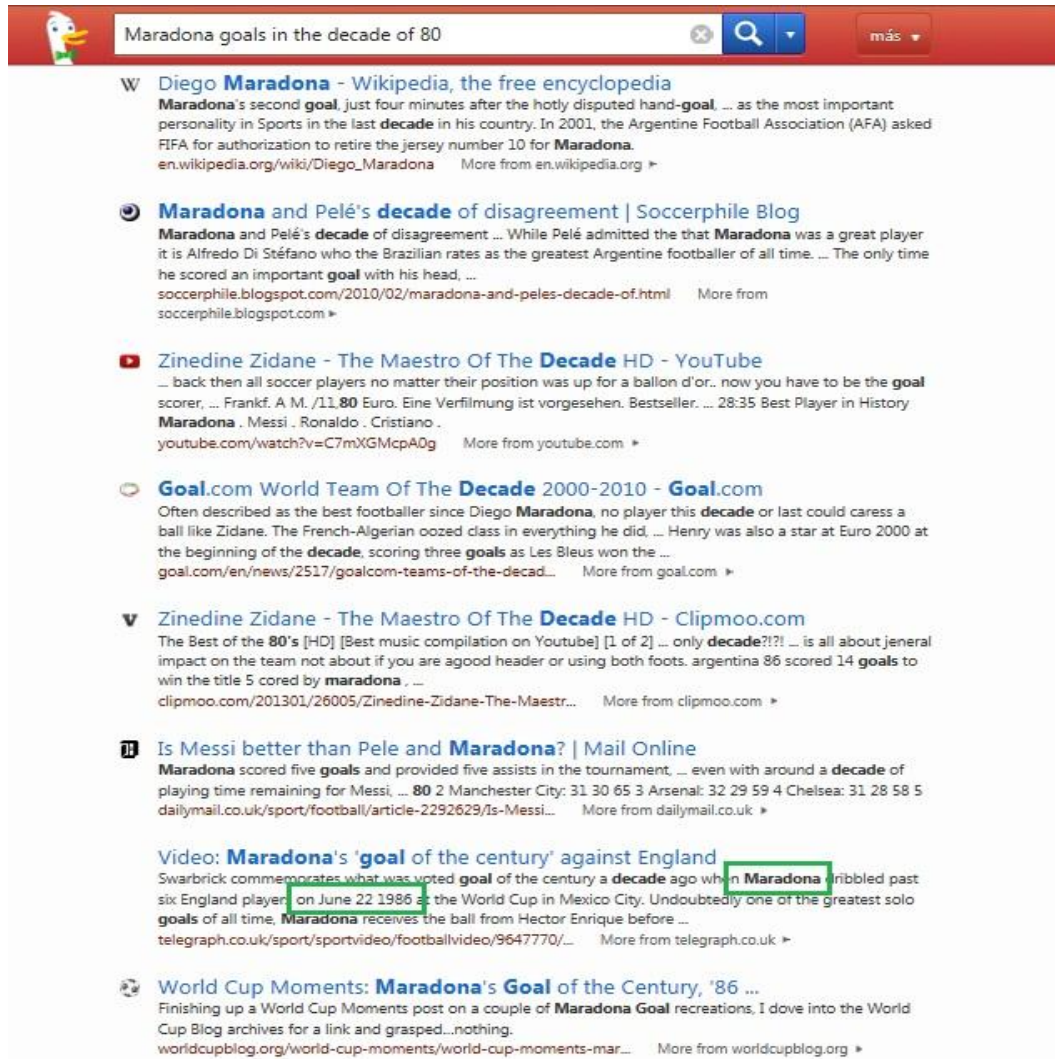
[1980s - Wikipedia, the free encyclopedia](#) Traducir esta página
en.wikipedia.org/wiki/1980s · Politics and wars · Disasters · Assaults · Technology · Economics · Society
The **1980s decade** was the period of time between January 1 ... The drought caused \$60 billion in damage (between \$80 ... first new American make since the Edsel), with the **goal** ...

[Maradona Goal, Argentina Vs. England....](#) Traducir esta página
www.youtube.com/watch?v=AsPUYL_qen0 · De argentina54 · 1 min · 130 vistas · Agregado 27/07/2010
10:59 Zinedine Zidane - The Maestro Of The **Decade** HD by MoodySZ 7,601,700 views ... 0:49 **Maradona Goal V England 1986**. Goal impossible by Rouf Bin Omer 3,228 views

BÚSQUEDAS RELACIONADAS

- [Diego Maradona History](#)
- [1982 Maradona Video](#)
- [Maradona Pictures](#)
- [Maradona Hand of God](#)
- [Maradona Video](#)
- [Maradona Goals](#)
- [Video Best Goals Maradona](#)
- [Messi](#)


Duck Duck GO



The image shows a screenshot of the Duck Duck Go search engine interface. The search bar at the top contains the text "Maradona goals in the decade of 80". Below the search bar, there are several search results. Each result includes a small icon, a title, a brief description, and a link to the source. The results are as follows:

- Wikipedia**: **Diego Maradona** - Wikipedia, the free encyclopedia. **Maradona's** second **goal**, just four minutes after the hotly disputed hand-**goal**. ... as the most important personality in Sports in the last **decade** in his country. In 2001, the Argentine Football Association (AFA) asked FIFA for authorization to retire the jersey number 10 for **Maradona**. en.wikipedia.org/wiki/Diego_Maradona More from en.wikipedia.org ▶
- Soccerphile Blog**: **Maradona and Pelé's decade** of disagreement | Soccerphile Blog. **Maradona** and Pelé's **decade** of disagreement ... While Pelé admitted the that **Maradona** was a great player it is Alfredo Di Stéfano who the Brazilian rates as the greatest Argentine footballer of all time. ... The only time he scored an important **goal** with his head, ... soccerphile.blogspot.com/2010/02/maradona-and-peles-decade-of.html More from soccerphile.blogspot.com ▶
- YouTube**: **Zinedine Zidane - The Maestro Of The Decade HD** - YouTube. ... back then all soccer players no matter their position was up for a ballon d'or.. now you have to be the **goal** scorer, ... Frankf. A M. /11.80 Euro. Eine Verfilmung ist vorgesehen. Bestseller. ... 28:35 Best Player in History **Maradona** . Messi . Ronaldo . Cristiano . youtube.com/watch?v=C7mXGMcpA0g More from youtube.com ▶
- Goal.com**: **Goal.com World Team Of The Decade 2000-2010** - Goal.com. Often described as the best footballer since Diego **Maradona**, no player this **decade** or last could caress a ball like Zidane. The French-Algerian oozed class in everything he did, ... Henry was also a star at Euro 2000 at the beginning of the **decade**, scoring three **goals** as Les Bleus won the ... goal.com/en/news/2517/goalcom-teams-of-the-decad... More from goal.com ▶
- Clipmoo.com**: **Zinedine Zidane - The Maestro Of The Decade HD** - Clipmoo.com. The Best of the 80's [HD] [Best music compilation on Youtube] [1 of 2] ... only **decade**?!?! ... is all about jeneral impact on the team not about if you are agood header or using both foots. argentina 86 scored 14 **goals** to win the title 5 cored by **maradona** clipmoo.com/201301/26005/Zinedine-Zidane-The-Maestr... More from clipmoo.com ▶
- Mail Online**: **Is Messi better than Pele and Maradona?** | Mail Online. **Maradona** scored five **goals** and provided five assists in the tournament. ... even with around a **decade** of playing time remaining for Messi, ... 80 2 Manchester City: 31 30 65 3 Arsenal: 32 29 59 4 Chelsea: 31 28 58 5 dailymail.co.uk/sport/football/article-2292629/Is-Messi... More from dailymail.co.uk ▶
- Telegraph.co.uk**: Video: **Maradona's 'goal of the century'** against England. Swarbrick commemorates what was voted **goal** of the century a **decade** ago when **Maradona** dribbled past six England players on June 22 1986 at the World Cup in Mexico City. Undoubtedly one of the greatest solo **goals** of all time, **Maradona** receives the ball from Hector Enrique before ... [telegraph.co.uk/sport/sportvideo/footballvideo/9647770/...](http://telegraph.co.uk/sport/sportvideo/footballvideo/9647770/) More from telegraph.co.uk ▶
- Worldcupblog.org**: **World Cup Moments: Maradona's Goal of the Century, '86** ... Finishing up a World Cup Moments post on a couple of **Maradona Goal** recreations, I dove into the World Cup Blog archives for a link and grasped...nothing. worldcupblog.org/world-cup-moments/world-cup-moments-mar... More from worldcupblog.org ▶

Hakia



Maradona goals in the decade of 80 Q

Surface Semantics

[Web](#)
[News](#)
[Blog](#)

Deep Semantics

[Pubmed](#)
[Credible](#)
[Galleries](#)

Regular Search

[Image](#)
[Twitter](#)

[The Most Influential Decades in World Football | Bleacher Report](#)
 ... Diego Maradona provided both one of the greatest and most controversial goals that the world had ever seen. Those two goals symbolize the dichotomy of Diego Maradona, who was easily the most influential player of the decade. As we approached the late '80s, ...
<http://bleacherreport.com/articles/1571932-the-most-influential-decades-in-world-football>

[Maradona's New Career?! | Maradona Goals - Pele Maradona](#)
 Owners of the Blackburn Rovers are speaking about the possibility of Diego Maradona succeeding Sam Allardyce. Chairwoman of Rovers owners Venky's Anuradha. same day loans. Maradona Goals - Pele Maradona | Maradona hand of god. Home; About; Mail; RSS;
<http://maradonagoal.com/2010/12/15/maradona%e2%80%99s-new-career/>

[World Cup Moments: Maradona's Goal of the Century, '86 ...](#)
 Finishing up a World Cup Moments post on a couple of Maradona Goal recreations, I dove into the World Cup Blog archives for a link and grasped...nothing.
<http://www.worldcupblog.org/world-cup-moments/world-cup-moments-maradonas-goal-of-the-century-86.html>

[1980's: Decade All-Star Lineup | Page 5 | Soccer Forum ...](#)
 After hours of thinking and researching, this is my ultimate '80s All Decade squad. Can't think of a better team than that really. What do you guys think? JGott, Nov 12, 2012 #101 ... this also includes 11 penalty goals whereas Maradona and Platini each scored three PK-goals for their NT.
<http://www.bigsoccer.com/community/threads/1980s-decade-all-star-lineup.1864996/page-5>

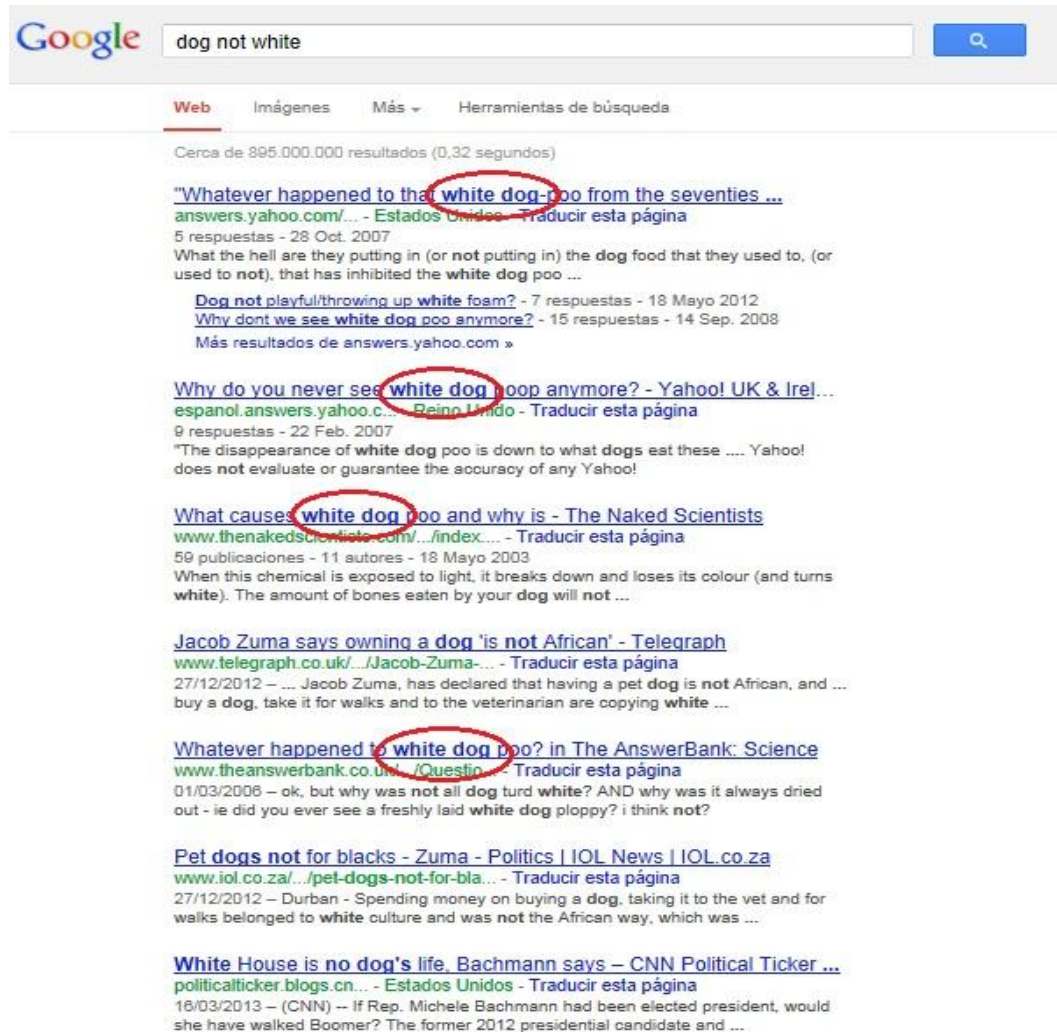
[Hand of God goal](#)
 Decades later Maradona did admit that the ball came off his hand, but maintains that it was an accident, ... Later in the same match, Maradona scored another goal, considered by many as one of the best goals in World Cup history, in which he eluded five English outfield players and Shilton.
http://www.fact-index.com/h/ha/hand_of_god_goal.html

[Who is better: Maradona or Messi? - Football Forums at ...](#)
 Pele's period was at least 10 to 20 years earlier than Maradona's. Maradona's era which in the 80s is really the tough one. ... Maradona- 342 goals over 21 seasons, ... I find it interesting that people are already comparing the two when messi has another decade to wow us.
<http://www.footytube.com/forums/the-stands-general-footy-banter/who-is-better-maradona-or-messi-11114/>

[Diego Maradona: Greatest Soccer \(Football\) Players of All Time](#)
 It is often said that he reached his peak only in the late 80s but that is not true. He was one if not the top guy throughout the entire decade, ... Diego Maradona may score a goal with his hand but that's how smart he's he always wants to win and never wants to lose like some other.

Negaciones

Google



Google

Web Imágenes Más ▾ Herramientas de búsqueda

Cerca de 895.000.000 resultados (0,32 segundos)

["Whatever happened to that white dog poo from the seventies ... answers.yahoo.com/... - Estados Unidos - Traducir esta página](#)
5 respuestas - 28 Oct. 2007
What the hell are they putting in (or not putting in) the dog food that they used to, (or used to not), that has inhibited the white dog poo ...
[Dog not playful/throwing up white foam?](#) - 7 respuestas - 18 Mayo 2012
[Why dont we see white dog poo anymore?](#) - 15 respuestas - 14 Sep. 2008
Más resultados de answers.yahoo.com »

[Why do you never see white dog poop anymore? - Yahoo! UK & Ire... espanol.answers.yahoo.c... Reino Unido - Traducir esta página](#)
9 respuestas - 22 Feb. 2007
"The disappearance of white dog poo is down to what dogs eat these Yahoo! does not evaluate or guarantee the accuracy of any Yahoo!

[What causes white dog poo and why is - The Naked Scientists www.thenakedscientists.com/.../index... - Traducir esta página](#)
59 publicaciones - 11 autores - 18 Mayo 2003
When this chemical is exposed to light, it breaks down and loses its colour (and turns white). The amount of bones eaten by your dog will not ...



[Jacob Zuma says owning a dog 'is not African' - Telegraph www.telegraph.co.uk/.../Jacob-Zuma-... - Traducir esta página](#)
27/12/2012 - ... Jacob Zuma, has declared that having a pet dog is not African, and ... buy a dog, take it for walks and to the veterinarian are copying white ...

[Whatever happened to white dog poo? in The AnswerBank: Science www.theanswerbank.co.uk/.../Questio... - Traducir esta página](#)
01/03/2006 - ok, but why was not all dog turd white? AND why was it always dried out - ie did you ever see a freshly laid white dog ploppy? i think not?

[Pet dogs not for blacks - Zuma - Politics | IOL News | IOL.co.za www.iol.co.za/.../pet-dogs-not-for-bla... - Traducir esta página](#)
27/12/2012 - Durban - Spending money on buying a dog, taking it to the vet and for walks belonged to white culture and was not the African way, which was ...

[White House is no dog's life, Bachmann says - CNN Political Ticker ... politicalticker.blogs.cn... - Estados Unidos - Traducir esta página](#)
16/03/2013 - (CNN) -- If Rep. Michele Bachmann had been elected president, would she have walked Boomer? The former 2012 presidential candidate and ...

BING


5.820.000 RESULTADOS Filtrar por idioma ▼ Filtrar por región ▼

[White Dog](#) [Wikipedia, the free...](#) Traducir esta página
en.wikipedia.org/wiki/White_Dog
Plot · Themes · Production · Distribution · Reception
In the original Romain Gary novel, this was **not** the story that was told--the **dog** started to attack **white** people because a black man embittered by **white** racism ...

[Restaurants in Wayne, PA: Main Line...](#) Traducir esta página
www.whitedog.com
Welcome to **White Dog** Café in Philadelphia. **White Dog** Café is happy to welcome guests to ... It's about **not** only helping our local economy grow, but also preserving the ...

[My dog is not all white but...? - Yahoo!](#) Traducir esta página
answers.yahoo.com/question/index?qid=20080909004636AAjuYKE
Resuelto · 5 respuestas totales
09/09/2008 · Best Answer: The whitening shampoo will brighten the **white** but will not change the color of the light brown areas ... haha she will be okay ... yeah shell ...

[Imágenes de dog not white](#)
bing.com/images



[Dog not eating, dogs poop is also...](#) Traducir esta página
answers.yahoo.com/question/index?qid=20090726211821AAISR0Y
Resuelto · 9 respuestas totales
27/07/2009 · Best Answer: b_bard9 has lead you in the right direction, there is either a bile duct, liver, pancreas problem, all would make the **dog** stop eating because ...

[Dog Breeds, Breed Information and Picture](#) Traducir esta página
www.nextdaypets.com/directory/breeds
Find **dog** breed information and **dog** breed pictures for ... BREED GROUP: Not AKC Recognized; WEIGHT: Male: 120; Female ... COLOR(S) light gray through to black, with **white** shading ...

[White House "Holiday" Card Spotlights Dog...](#) Traducir esta página
radio.foxnews.com/.../top-stories/white-house...dog-not-christmas.html
10/12/2012 · The 2012 **White House** "Holiday" card spotlights the Obama's family Portuguese water **dog** — instead of Christmas.

[Why is dog poo not white anymore |...](#) Traducir esta página
www.answerbag.com/q_view/663043
Why is **dog** poo not **white** anymore The dung beetles eat it before it has a chance to turn **white**.

[Samuel Fullers "White Dog": Natural is...](#) Traducir esta página
mubi.com/notebook/posts/samuel-fullers-white-dog-natural-is-not-in-it

BÚSQUEDAS RELACIONADAS
Training Dogs Not to Bark
Training Dogs Not to Jump
Reasons for Dogs Not Eating
How Much Are Dogs
I AM a Good Dog
2013 Westminster Dog Show
Man Beats Dog in Elevator
Dogs Not Labradors

Duck Duck GO

The image shows a Duck Duck GO search interface with the query 'dog not white'. The search bar is at the top, and the results are listed below. Several search results have red circles highlighting the words 'White Dog' and 'White Dogs'.

W White Dog - Wikipedia, the free encyclopedia
White Dog is a 2002 American drama film directed by Samuel Fuller using a screenplay written by Fuller and Curtis Hanson loosely based on Romain Gary's 1970 novel of the same title.
en.wikipedia.org/wiki/White_Dog More from en.wikipedia.org >

White Coat Color in Dogs - Webpages - Individual Webpages on ...
White Coat Color in Dogs. Albinism; **White** Coat with Pigmented Eyelids **"White"** Wolf; This webpage is part of a series on **Dog** Coat Color Genetics and was last updated on June 17, 2007 by Sheila Schmutz.
homepage.usask.ca/~schmutz/white.html More from homepage.usask.ca >

Dog Breeds, Breed Information and Picture
 BREED GROUP: Working; WEIGHT: Male: 85; Female: 75 lbs; HEIGHT: Male: 25; Female: 23 inches; COLOR(S) light gray through to black, with **white** shading and a **white** mask or cap; also shades of sable or red with shading, or all **white**
nextdaypets.com/directory/breeds/ More from nextdaypets.com >

Amazon.com: Service Dog WORKING - DO NOT PET Red on White 2.5 ...
 This Sew-on Patch is printed on lightweight polyester fabric using Dye-Sublimation technology, so the graphic won't crack or fade like heat transfers. The graphics look and ...
amazon.com/Service-Dog-WORKING-White-Sew/dp/B007X04T6G More from amazon.com >

Samuel Fuller's "White Dog" Natural is not in it, on ...
 An allowance for irrationality **sees** the difference between a shrewd and a calculating eye. This arrangement (of facts, of phenomena, of the empirical) is what
mubi.com/notebook/posts/samuel-fullers-white-dog... More from mubi.com >

How to teach a dog not to bite | Black and White Dogs
 Biting is a normal behavior of dogs. Biting is a **dog's** way of interacting with other dogs, their way of keeping puppies in line and their way of establishing dominance over dogs of weaker personality. Puppies bite everything their paws and mouth can reach as this is their way of discovering ...
blackandwhitedogs.com/rearing/training-dogs/how-to-teach-a-do... More from blackandwhitedogs.com >

Dog Breeds - Dog.com
Dog supplies from **dog.com** includes a huge variety of **dog** supplies & products at wholesale discounted prices. **Dog.com** satisfies your **dog** supplies & **dog** information needs. **Dog.com** Horse.com Garden.com Bike.com Bird.com Fish.com. ... Bull Terrier (**white**) Bulldog Bullmastiff ...
dog.com/Breeds/Default.aspx More from dog.com >

White Dog (1982) - IMDb
Reprogramming a dog who kills Blacks is the ultimate challenge for an unorthodox African-American trainer. When a young Hollywood actress finds the injured stray, she nurses it back to health, **not** knowing it's a **"White Dog"** trained by a racist to attack only Blacks. Julie's appalled when the ...
imdb.com/title/tt0084899/ More from imdb.com >

[PDF] Why Not White - Sterling MS Miniature Schnauzers
 Why **Not White**? by Marcia Feld First, please understand that any an AKC registration means only that the parents of the **dog** were also registered with AKC and the validity of that registration depends on the

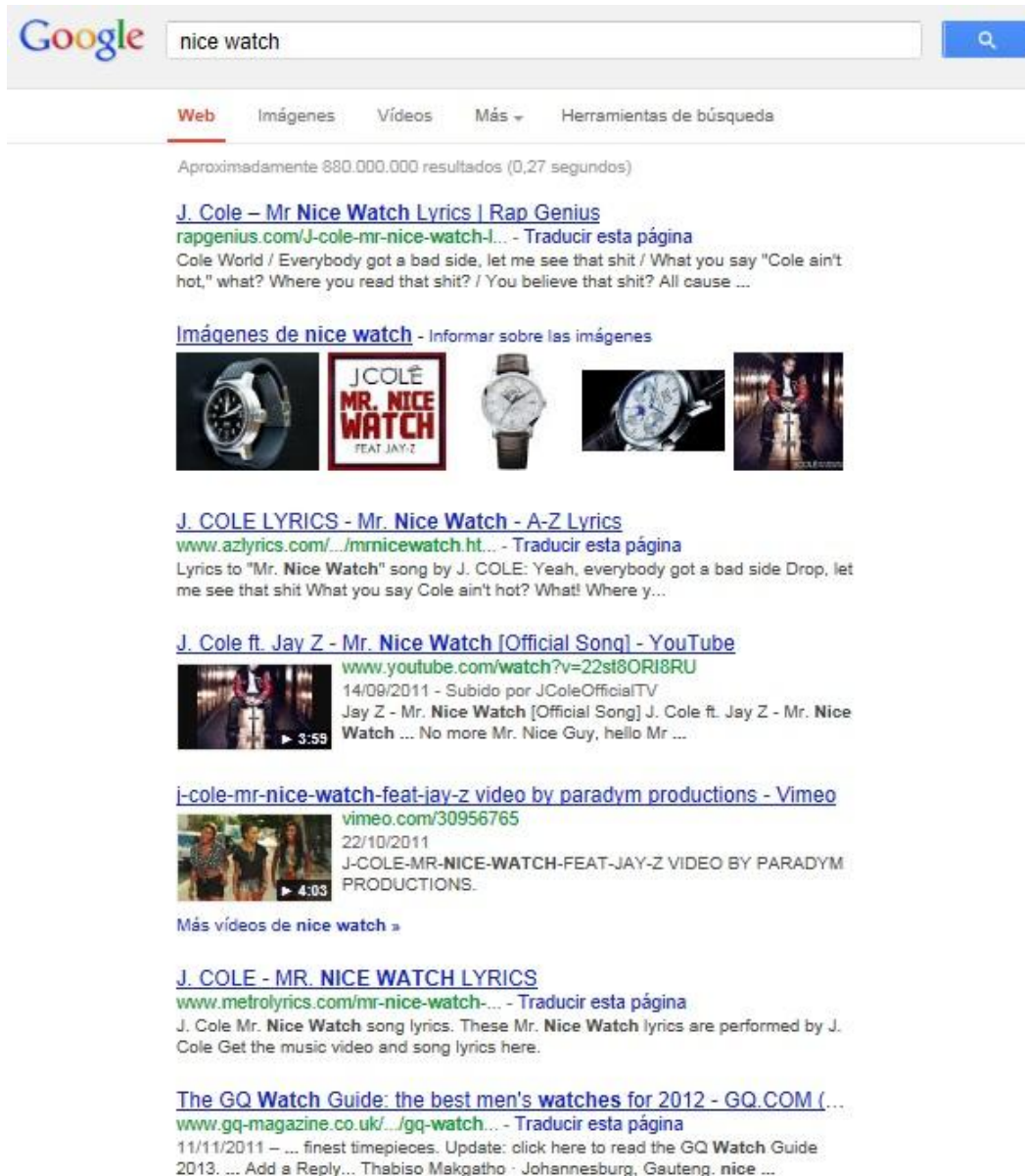
Hakia

The screenshot shows the Hakia search engine interface. At the top left is the 'hakia' logo. To its right is a search bar containing the text 'dog not white' and a magnifying glass icon. Below the search bar, there are several categories of search results:

- Surface Semantics:**
 - [White Dogs](#) (circled in red): **White dogs** have white hopes, white dreams and white problems. <http://whitedogs.tumblr.com/>
 - [Web News Blog](#)
- Deep Semantics:**
 - [What is WHITE DOG?](#) (circled in red): What is **WHITE DOG?** M What will tell you the definition or meaning of What is WHITE DOG <http://mrwhatis.com/white-dog.html>
 - [Common Dog Health Problems: Dog Throwing Up White Foam](#): Welcome to **Common Dog Health Problems: Dog Throwing Up White Foam** <http://www.commondoghealthproblems.com/dog-throwing-up-white-foam.php>
- Regular Search:**
 - [Image Twitter](#)
 - [My dog is throwing up white foam. - Questions & Answers...](#): **My dog** is throwing up white foam. She had a piece of burger last night at a BBQ. I'm worried something is wrong. <http://www.vetinfo.com/vets/answers/my-dog-throwing-up-white-foam>
 - [my dog is not pregnant but has a white discharge today and ...](#): **Question - my dog** is not pregnant but has a white discharge today and Find the answer to this and other Dog questions on JustAnswer. <http://www.justanswer.com/pet-dog/Oy65h-dog-not-pregnant-white-discharge-today.html>
 - [Dog Vomiting White Foam | Dog Throwing Up, Vomits White Foam ...](#): A **dog vomiting white** foam however must be taken more seriously if that is not the only symptom, and there are other signs like restlessness, heavy panting, excessive salivating, appearing confused, attempting to vomit <http://www.pethealthandcare.com/articles/247/dog-diseases/dog-vomiting-white-foam.html>
 - [Why is My Dog Coughing Up White Foam?](#): **Dogs coughing up white** foam can be attributed to a variety of causes. Assess the many possible reasons why dogs cough white foam and how it can be treated. <http://www.whyguides.com/why-is-my-dog-coughing-up-white-foam.html>
 - [Why is dog poo not white anymore | Answerbag](#): Why is **dog poo not white** anymore The dung beetles eat it before it has a chance to turn white. http://www.answerbag.com/q_view/663043
 - [notcoming.com | White Dog](#) (circled in red): Even though **White Dog** tells the story of a white-furred German Shepherd taught to attack black skin on sight, Sam Fuller did not intend his film to be a controversial one.

Polisemia

Google




Google nice watch

Web Imágenes Vídeos Más ▾ Herramientas de búsqueda

Aproximadamente 880.000.000 resultados (0,27 segundos)

[J. Cole – Mr Nice Watch Lyrics | Rap Genius](#)
[rapgenius.com/J-cole-mr-nice-watch-l... - Traducir esta página](#)
Cole World / Everybody got a bad side, let me see that shit / What you say "Cole ain't hot," what? Where you read that shit? / You believe that shit? All cause ...

[Imágenes de nice watch](#) - Informar sobre las imágenes



[J. COLE LYRICS - Mr. Nice Watch - A-Z Lyrics](#)
[www.azlyrics.com/.../mnicewatch.ht... - Traducir esta página](#)
Lyrics to "Mr. Nice Watch" song by J. COLE: Yeah, everybody got a bad side Drop, let me see that shit What you say Cole ain't hot? What! Where y...

[J. Cole ft. Jay Z - Mr. Nice Watch \[Official Song\] - YouTube](#)
[www.youtube.com/watch?v=22st8ORi8RU](#)
14/09/2011 - Subido por JColeOfficialTV
Jay Z - Mr. Nice Watch [Official Song] J. Cole ft. Jay Z - Mr. Nice Watch ... No more Mr. Nice Guy, hello Mr ...


[j-cole-mr-nice-watch-feat-jay-z video by paradym productions - Vimeo](#)
[vimeo.com/30956765](#)
22/10/2011
J-COLE-MR-NICE-WATCH-FEAT-JAY-Z VIDEO BY PARADYM PRODUCTIONS.

Más vídeos de nice watch »

[J. COLE - MR. NICE WATCH LYRICS](#)
[www.metrolyrics.com/mr-nice-watch-... - Traducir esta página](#)
J. Cole Mr. Nice Watch song lyrics. These Mr. Nice Watch lyrics are performed by J. Cole Get the music video and song lyrics here.

[The GQ Watch Guide: the best men's watches for 2012 - GQ.COM \(...\)](#)
[www.gq-magazine.co.uk/.../gq-watch... - Traducir esta página](#)
11/11/2011 – ... finest timepieces. Update: click here to read the GQ Watch Guide 2013. ... Add a Reply... Thabiso Makgatho · Johannesburg, Gauteng. nice ...

BING

bing Beta nice watch 


74.100.000 RESULTADOS Filtrar por idioma ▼ Filtrar por región ▼

[Nice Watches from Christopher Ward...](#) Traducir esta página
www.nicewatches.co.uk
Nice Watches. Nice Watches is dedicated to helping you find nice mens and womens luxury watches and quality timepieces at the best prices. Whether you fancy a Casio ...

[Amazon.com: Products tagged with nice...](#) Traducir esta página
www.amazon.com/tag/nice%20watch
 Products tagged with nice watch

[How to Choose a Nice Watch: 8 Steps...](#) Traducir esta página
www.wikihow.com/Choose-a-Nice-Watch
 How to Choose a Nice Watch. There are many different watches in the market. Find one that fits your style. Decide what kind of watch you want. Analog? Digital? Both?

Videos de nice watch
bing.com/videos



J. Cole ft. Jay Z - Mr. Nice Watch [YouTube] J. Cole Mr. Nice Watch (feat. Ja... [YouTube] J. Cole - "Mr. Nice Watch" Ft Jay-Z [YouTube] Pat Cruz - Mr. Nice Watch [YouTube]

[Nice Watch - YouTube](#) Traducir esta página
www.youtube.com/watch?v=5USc83s8jG0
 De tonedeffication · 4 min · 52 vistas · Agregado 03/03/2012
 3:48 Rolling by tonedeffication 63 views 2:49 Kidd 2 deff Spiteful Chant by tonedeffication 44 views 3:32 Faded-Tyga ft. Lil Wayne/High School of The ...

[J. COLE LYRICS Mr. Nice Watch ^ Z...](#) Traducir esta página
www.azlyrics.com/lyrics/jcole/mrnicewatch.html
 Lyrics to "Mr. Nice Watch" song by J. COLE: Yeah, everybody got a bad side Drop, let me see that shit: What you say Cole ain't hot? What! Where y...

[nice watch - YouTube](#) Traducir esta página
www.youtube.com/watch?v=AEjPkS9Lvyl
 De toronto talac · 6 min · 2 vistas · Agregado 04/07/2012
 Sign in with your YouTube Account (YouTube, Google+, Gmail, Orkut, Plasa, or Chrome) to add toronto talac's video to your playlist

[Jim A \(nice_watch\) on Twitter](#) Traducir esta página
twitter.com/nice_watch
 1.151 seguidores · 3.290 tweets · Sigue a 635 usuarios
 The latest from Jim_A (@nice_watch). schizophrenic Good kid | IDGAFinder | basketball | rugby | Nairobi half-life

BÚSQUEDAS RELACIONADAS
[Nice Watch Jay Z](#)
[Nice Watch Brands](#)
[Nice Watch for Men](#)
[How to Choose a Watch](#)
[Nice Watches for Women](#)
[Nices! Watch!](#)
[Mr. Nice Watch](#)
[Mr. Nice Watch Lyrics](#)

Duck Duck GO

The screenshot shows the Duck Duck GO search engine interface. At the top, there is a search bar containing the text "nice watch". To the left of the search bar is the Duck Duck GO logo (a duck). To the right is a search icon and a "más" button. Below the search bar, there are several search results, each with a small icon on the left, a title, a brief description, and a URL.

- J. Cole ft. Jay Z - Mr. Nice Watch [Official Song] - YouTube**
 J. Cole ft. Jay Z - Mr. **Nice Watch** [Official Song] J. Cole ft. Jay Z - Mr. **Nice Watch** [Official Song] J. Cole ft. Jay Z - Mr. **Nice Watch** [Official Song] J. Cole ft. Jay Z - Mr. ...
youtube.com/watch?v=22st8ORIBRU More from youtube.com
- How to Choose a Nice Watch: 8 Steps - wikiHow**
 How to Choose a **Nice Watch**. There are many different watches in the market. Find one that fits your style. Decide what kind of **watch** you want. Analog? Digital? Both?
wikihow.com/Choose-a-Nice-Watch More from wikihow.com
- Nice watch - TheFind**
Nice watch - Find the largest selection of **nice watch** on sale. Shop by price, color, locally and more. Get the best sales, coupons, and deals at TheFind.
thefind.com/jewelry/info-nice-watch More from thefind.com
- Nice Italy Watches. - WatchWear.com - Men's Watches, Fashion ...**
NICE ITALY WATCHES NICE was created in Italy by Mr. Andrea Palama, a renowned Italian designer of fashion **watch** and accessories design. Started in 1991 in Italy, our company has been expanding its worldwide operation capacity, strengthening our design team and production effectiveness.
watchwear.com/nice-italy.htm More from watchwear.com
- Mr. Nice Watch (ft. Jay-Z) - Dirty by J. Cole on SoundCloud ...**
 Listen to Mr. **Nice Watch** (ft. Jay-Z) - Dirty by J. Cole: "Cole World: The Sideline Story" in stores September 27th. | Explore the largest community of artists, bands, podcasters and creators of music & audio.
soundcloud.com/j-cole/mr-nice-watch-ft-jay-z-dirty More from soundcloud.com
- very nice watch | eBay - Electronics, Cars, Fashion ...**
 Visit eBay for great deals on a huge selection very **nice watch**. Shop eBay!
rover.ebay.com/rover/1/711-53200-19255-0/1?icp_ff3=9&... More from rover.ebay.com
- That Is A Nice Watch - Custom T-Shirts, Unique Gifts, Posters ...**
 Welcome! The growing demand for quality timepieces via the internet has exploded over the years. Finding a stylish designer **watch** online can be done easier than ever before.
cafepress.com/thatisanicewatch More from cafepress.com
- Nice Mens Watches - Esquire Watches for Men - Esquire**
 Low-key luxury **nice** men's **watches** were the name of the **watch** game in the height of the recession, and this diversity of options might still help your purchase now that the recession's officially over
esquire.com/style/how-to-buy-a-watch/nice-mens-watc... More from esquire.com
- Nice Watches from Christopher Ward, Citizen, Nautica and ...**
Nice Watches. Nice Watches is dedicated to helping you find **nice** mens and womens luxury **watches** and quality timepieces at the best prices. Whether you fancy a Casio or a Citizen, a Seiko or a Sekonda or a classy timepiece from Christopher Ward, Rotary or Tag Heuer, we'll bring you everything you ...
nicewatches.co.uk More from nicewatches.co.uk
- Watch Mr. Nice online - Watch Movies Online, Full Movies ...**
Watch Mr Nice online for free. Download Mr **Nice** movie at movie2k.to
www.movie2k.to/mr-nice-watch-2005.html

Hakia



nice watch



Surface Semantics

- [Web](#)
- [News](#)
- [Blog](#)

Deep Semantics

- [Pubmed](#)
- [Credible](#)
- [Galleries](#)

Regular Search

- [Image](#)
- [Twitter](#)

[UGGH...NICE WATCH](#)

2011 was a slow year for Ugggh...Nice Watch, but a great year for music. THE TOP 25 ALBUMS OF 2011: 1. Drake Take Care 2. The Weeknd House of Balloons 3. Frank Ocean nostalgia, ULTRA. 4. Jay-Z & Kanye West Watch The Throne 5. Toro Y Moi Underneath The Pine 6. <http://uggghnicewatch.com/>

[Nice Watches for Men](#)

There are many different nice watches for men. These can be simple, or quite elaborate, with a number of bells and whistles. They can be analog or digital, as well as many hybrids. Because of this, the choices can be a little overwhelming. There are great nice watches that are cheap and others ...

<http://www.nicewatchesformen.net/>

[Mr. Nice Watch" lyrics by J. COLE - Urban Lyrics - R&B Rap ...](#)

Completed J. COLE ... [feat. Jay-Z] Send "Mr. Nice Watch" Ringtone to your Cell [Verse 1:] Yeah, everybody got a bad side

<http://www.urbanlyrics.com/lyrics/jcole/mrnicewatch.html>

[J. Cole – Mr. Nice Watch f. Jay-Z \[Dirty\] | 2dopeboyz](#)

419 Responses to "J. Cole – Mr. Nice Watch f. Jay-Z [Dirty]" first bitches! casper wordsmith said this on September 14th, 2011 at 4:35 pm

<http://www.2dopeboyz.com/2011/09/14/j-cole-mr-nice-watch-f-jay-z/>

[J. Cole Welcomes Jay-Z Verse on 'Mr. Nice Watch' -- Listen ...](#)

Because life is short, J. Cole advises we gather our Hublots and Rollies while we may. On the just-released 'Mr. Nice Watch,' the sixth track to surface from his hotly anticipated debut, 'Cole World: The Sideline Story,' the North Carolina rapper and producer uses the fleeting nature of human

<http://www.theboombox.com/2011/09/14/j-cole-mr-nice-watch/>

[J. Cole - Mr. Nice Watch Lyrics - Lyrics On Demand - Song ...](#)

J. Cole Lyrics. Mr. Nice Watch Lyrics [Verse 1:] Yeah, everybody got a bad side Drop, let me see that shit What you say Cole ain't hot? What! Where you read that shit?

<http://www.lyricsondemand.com/j/jcoleyrics/mrnicewatchlyrics.html>

[J. Cole & Jay-Z "Mr. Nice Watch" \[NEW MUSIC\] | The Urban ...](#)

This week J. Cole previewed his album, and this collabo with Jay-Z "Mr. Nice Watch" left the room buzzing! Take a listen to this new fire from J. Cole, and cop his debut album "Cole World: The Sideline Story" when it comes out on September 27th

<http://theurbandaily.com/1532935/j-cole-jay-z-mr-nice-watch-new-music/>

[Mr. Nice Watch \(ft. Jay-Z\) - Dirty by J. Cole on SoundCloud ...](#)

Listen to Mr. Nice Watch (ft. Jay-Z) - Dirty by J. Cole: "Cole World: The Sideline Story" in stores September 27th. | Explore the largest community of artists, bands, podcasters and creators of

ANEXO II - Modelos de recuperación de información.

Se presentan a continuación algunos de los principales modelos existentes de recuperación de información [37][38].

Modelo booleano

La consulta es una expresión del algebra booleana, como por ejemplo $q = K_a \text{ and } (K_b \text{ or } (\text{not } K_c))$.

Esta consulta recupera los documentos que contienen K_a y K_b o no K_c , o sea que si por ejemplo no contiene K_a , no se incluirá en la respuesta.

Es un método muy eficiente y sencillo de implementar pero no representa resultados parciales según la relevancia, es decir que resultados parciales no serán recuperados y además los resultados tampoco serán entregados en orden.

Modelo de espacio vectorial

Según este modelo cada expresión del lenguaje natural puede representarse como un vector de pesos de términos (entendiendo término como la unidad mínima de información de una palabra). En el caso de la recuperación de información se representan los documentos y la consulta:

- documento = (peso_de_término_1, peso_de_término_2, ..., peso_de_término_n)
- consulta = (peso_de_término_1, peso_de_término_2, ..., peso_de_término_n)

Para determinar la similitud que existe entre un documento y una consulta se calcula la distancia que existe entre los vectores que los representan, a menor distancia, mayor similitud.

La distancia puede calcularse usando la diferencia de ángulos, para esto se aplica el Teorema del Coseno. Así un valor de coseno de cero significa que la búsqueda y el documento son ortogonales el uno al otro, y eso significa que no hay coincidencia.

Estos modelos solucionan el problema de los resultados parciales pero no capturan la semántica de la consulta o del documento ya que los pesos de los términos pueden ser independientes entre sí.

Modelos probabilísticos

Los modelos probabilísticos son capaces de calcular el grado de similitud existente entre cada documento de la colección y la consulta, consiguiendo ordenar los documentos de la colección en orden descendente de probabilidad de relevancia en relación a la consulta. El cálculo de probabilidad se deriva del teorema de Bayes.

Se toman algunas asunciones previas para luego poder refinar los resultados a medida que se realizan las búsquedas. Una de estas es asumir el conjunto inicial de documentos relevantes para una consulta dada, otra sería que la probabilidad de que un término se encuentre en los documentos del conjunto relevante es la misma para todos los términos.

Existen varios modelos probabilísticos, entre ellos podemos encontrar el "**Modelo binario independiente**", el cual considera solo la presencia o la ausencia de un término y no la frecuencia en la cual aparece en el documento. Otro sería el "**Okapi BM25**" es una función de ranking la cual nos permite ordenar por relevancia los documentos basado más que nada en las palabras claves buscadas y en la frecuencia en que aparecen en el documento.

Estos modelos generalmente hacen demasiadas asunciones las cuales pueden afectar la calidad y fiabilidad de los resultados obtenidos, por lo tanto no son usualmente implementados en procesos de recuperación de información de grandes escalas.

PageRank (Google)

Es una serie de algoritmos utilizado por el buscador Google para asignar de forma numérica la relevancia de las páginas web o documentos devueltos en una consulta.

Estos algoritmos se basan fuertemente en los enlaces de una página, tantos los entrantes como los salientes.

ANEXO III – Herramientas evaluadas

Con el objetivo de reutilizar herramientas existentes que resuelvan los problemas comunes de los sistemas de recuperación de información, se estudiaron diversas herramientas de las cuales se optó por utilizar algunas y otras fueron descartadas.

A continuación se detallan las herramientas evaluadas haciendo una reseña de las pruebas que se realizaron sobre cada una, las ventajas y desventajas que tenían y su utilización final en el proyecto.

Indexación y recuperación de documentos

Lucene

Lucene [39] es una herramienta que permite indexar textos y luego realizar búsquedas sobre el índice creado.

La herramienta es un módulo fácil de instalar en **Tomcat** que está implementada en **JAVA** y para integrarlo con **Python** se puede utilizar un modulo llamado **pyLucene** que permite implementar con este lenguaje la comunicación con el servidor que contiene los archivos indexados y devolver el resultado de una consulta.

En **Python** la implementación consiste simplemente en importar el módulo de **pyLucene**, recorrer los textos a indexar e ir creando documentos con ciertos atributos como el identificador, texto y autor, para luego poder buscar por esos valores. Para la recuperación de documentos el desarrollo es muy similar.

Una particularidad de **Lucene** integrado con **Python** con el módulo **pyLucene** es que por un lado no funciona correctamente en ambientes **Windows** y por otro lado que se le deben configurar ciertas propiedades como el directorio donde se guardarán los documentos y otros valores que no son necesarios con herramientas más avanzadas como **SOLR**.

SOLR

SOLR [40] es una herramienta basada en **Lucene** que potencia sus funcionalidades brindando más características como por ejemplo el hecho de resaltar la aparición de palabras buscadas en un documento, que si bien tiene estas características no son en principio de gran importancia para nuestro proyecto, puede llegar a ser útiles para posibles trabajos a futuro.

SOLR al igual que **Lucene** se instala sobre **Tomcat** como un aplicativo web, capacitado para recibir documentos a indexar y consultas para devolver documentos relevantes.

La integración de **SOLR** con **Python** se realizó utilizando un módulo llamado **SolrPy** [41] que permite comunicar una aplicación de **Python** con el servidor donde están corriendo los servicios de **SOLR** utilizando **JSON**.

Por lo tanto la utilización de **SOLR** consiste simplemente en tener un servidor **Tomcat** con la aplicación recibiendo solicitudes en un puerto determinado y comunicar la aplicación creada con **Python** por servicios **JSON** tanto para indexar documentos como para recuperar.

Interfaz de usuario

Django

Es un *framework* de desarrollo web *open source* escrito en **Python** que utiliza el paradigma Modelo-Vista-Controlador (MVC) aunque con una leve diferencia ya que lo llama Modelo-Plantilla-Vista [42].

El modelo en ambos casos se trata del mismo concepto, en nuestro caso no manejamos un modelo de datos. Las vistas serían el controlador del MVC, en este caso donde se definen las urls, funciones, etc, sería la lógica de la parte web y la parte que interactúa con el Core. Por último tenemos las plantillas que serían las vistas del MVC, en este caso los .html donde se presentan los datos al usuario.

Otras características de **Django** aparte de hacer más rápido el desarrollo web y ser **MVC**, es que es bastante fácil de usar, está pensado para ser eficiente, tiene muy bajo acoplamiento y es muy modular y muy potente.

Python es usado en todas las partes del *framework*, incluso en configuraciones, archivos, y en los modelos de datos.

Django se comunica con el servidor web utilizando alguna de varias formas como ser `mod_python`, `mod_wsgi`, etc.

Expresiones temporales

Tarsqi Toolkit

Esta herramienta se compone de varios componentes, donde los más importantes para el proyecto son **GuTime** como reconocedor de expresiones temporales y **Evita** como reconocedor de eventos, más otros componentes que permiten relacionar los eventos y las marcas temporales para determinar a partir de un documento qué ocurre y en qué momento [43].

Esta posibilidad de crear enlaces entre los eventos y su ocurrencia en el tiempo permitiría para el proyecto evaluar las expresiones temporales de una consulta fácilmente, pero en las pruebas realizadas para su utilización se detectaron algunos problemas, como frases sencillas que no eran identificadas y por consiguiente no es adecuado para nuestros requisitos.

Por ejemplo, para la frase “**Twenty schools were built between 1920 and 1930**”, al ejecutar **Tarsqi** con todos sus módulos se identifican dos fechas de tiempo los años **1920** y **1930**, más una ocurrencia del evento **built**. La relación entre el evento **built** y las fechas detectadas no es anotada. Sin embargo con la frase “**Twenty schools were built on 1920**” **Tarsqi** detecta el año **1920** y la ocurrencia del evento **built** enlazados por la aparición de la preposición **on**.

Además de estas particularidades en el *parsing* que no contemplaban las distintas frases deseadas, la integración entre la herramienta y nuestra aplicación no es del todo sencilla ya que el módulo de **Tarsqi** está desarrollado para trabajar con archivos de entrada y salida en directorios relativos a donde se encuentra el módulo como también utilizar herramientas de anotación también relativas al módulo de **Tarsqi**.

SuTime

Es una herramienta provista por el grupo de PLN de la universidad de **Stanford**, que tiene similares características que el componente **GuTime** utilizado por *Tarsqi*, que tiene como objetivo detectar expresiones temporales en una frase, independientemente de los eventos que ocurren [44].

Por las pruebas realizadas sobre **SuTime** si bien reconoce expresiones temporales deseadas, la imposibilidad de relacionar las marcas de tiempo con la ocurrencia de los eventos hace que la herramienta por sí sola, no sea útil y es muy similar en cuanto a su funcionalidad a **GuTime**.

Luego de investigar la integración del módulo **SuTime** con el analizador de dependencias de **Stanford**, se comprobó que se puede vincular de una forma sencilla las ocurrencias de los eventos con las marcas de tiempos.

Por ejemplo si se usa **SuTime** para la frase “**Twenty schools were built between 1920 and 1930**” se detectan dos marcas de tiempo, los años **1920** y **1930** pero ninguna relación con la ocurrencia del evento **built**.

Por otro lado en la integración con el parser, mediante el **Stanford CorePLN** [45] con la misma frase se obtiene una expresión temporal de fechas como **1920/1930** y en el análisis de dependencias se refleja que la preposición **between** relaciona el evento **built** con la fecha **1920/1930**. Similares anotaciones se logran con palabras **after, on, in**.

Analizador de Dependencias

Stanford Parser y COREPLN

El único analizador de dependencias estudiado para el idioma inglés fue el **Stanford Parser** ya que con las primeras pruebas se pudo comprobar que junto con algunas otras herramientas de análisis sintáctico, tenían todo lo que se necesitaba para el proyecto.

El **Stanford Parser** es un programa que analiza la gramática de los enunciados, identificando qué palabras son los sujetos, predicados, verbos mediante la implementación de un modelo probabilístico y a partir de la aplicación se obtienen las dependencias que surgen del análisis.

Se realizaron pruebas sobre el parser y si bien identificaba palabras como se pretendía, existía alguna serie de problemas a resolver como el etiquetado, el reconocimiento de entidades con nombre, el reconocimiento de expresiones temporales que indicaban que se debía complementar en *parser* con algo más.

De esta forma surge el estudio del **Stanford COREPLN** que es un conjunto de herramientas de análisis del lenguaje natural que incluyen todas las características que se requieren para el proyecto. En la página del proveedor se encuentra un analizador online que al colocarle una frase devuelve el análisis resultante, donde se pudo comprobar que la información obtenida era necesaria y suficiente para resolver los problemas de la búsqueda semántica.