Procesamiento del Lenguaje Natural, Revista nº 51, septiembre de 2013, pp 33-40 recibido 30-04-2013 revisado 18-06-2013 aceptado 21-06-2013

# Información Lingüística en Recuperación de Imágenes Multilingüe

## *Linguistic Information in Multilingual Image Retrieval*

**David Hernández-Aranda**
NLP&IR group at UNED
daherar@lsi.uned.es

**Víctor Fresno Fernández**
NLP&IR group at UNED
vfresno@lsi.uned.es

**Resumen**: En este trabajo se evalúan diferentes modelos de indexación, así como la aplicación de técnicas de Procesamiento del Lenguaje Natural al texto asociado a las imágenes en un problema de Recuperación de Imágenes Multilingüe. Los resultados muestran que traducir el texto asociado y usar Entidades Nombradas, junto con sus categorías, permite mejorar el proceso de recuperación, mientras que la mejora obtenida con el uso de sintagmas nominales no compensa el coste computacional que conlleva.
**Palabras clave**: Recuperación de Imágenes, TBIR, Recuperación de Información Multilingüe, Procesamiento de Lenguaje Natural, Entidades Nombradas, Sintagmas Nominales.

**Abstract:** In this paper we evaluate several indexing models and the application of different Natural Language Processing techniques to textual descriptions associated with images in an Image Retrieval task. The results show that the use of a translation-based model and to take into account the Named Entities with their categories, improves the retrieval process. However, the improvement obtained with the use of noun phrases is not worth the computational cost involved.
**Keywords:** Image Retrieval, TBIR, Multilingual Information Retrieval, Natural Language Processing, Named Entities, Noun Phrases.

## *1 Introducción*

Las principales aproximaciones utilizadas en la Recuperación de Imágenes Multilingüe se centran en el análisis del texto asociado a las imágenes (*Text-Based Image Retrieval, TBIR*) o de las características visuales de las mismas (*Content-Based Image Retrieval, CBIR*), y actualmente se están haciendo grandes esfuerzos en combinar ambos enfoques.

Las aproximaciones TBIR suelen aplican técnicas de IR ad-hoc sobre textos planos. En pocos casos se trata como un problema de IR estructurada, considerando de diferente modo distintos tipos de información textual. Si bien pueden encontrarse aproximaciones a la IR Multilingüe documental, hasta donde se ha revisado la literatura no se ha encontrado ningún estudio detallado sobre cómo tratar el texto asociado a las imágenes. En este trabajo se estudia qué y cómo realizar la indexación de esta información asociada, y sobre qué campos y cómo realizar el proceso de recuperación.

Los textos asociados a imágenes suelen ser textos cortos y de carácter descriptivo. De este modo, la presencia de entidades nombradas (*Named Entities*, NEs) cobra un papel destacado a la hora de describir el contenido de una imagen. Por otro lado, la aplicación de otras técnicas lingüísticas también puede ayudar a obtener buenos resultados. Entre estas técnicas se van a estudiar: el reconocimiento de NEs, junto con sus categorías, y el uso de sintagmas nominales. Aunque estos últimos necesitan de un mayor contexto para su detección, ya que en ocasiones se requiere de una fase de análisis sintáctico, se considera que pueden representar una unidad interesante de información a la hora de describir el contenido de una imagen.

Este trabajo se centra, por tanto, en tratar de explotar al máximo esa "pequeña cantidad de contenido multilingüe" de la que se dispone en las colecciones de imágenes anotadas y, para ello, se estudia la aplicación de técnicas de IR y Procesamiento de Lenguaje Natural con la idea de que cuanto más se mejoren las aproximaciones TBIR, mejores resultados se obtendrán después combinándolas con CBIR.

ISSN 1135-5948 

## *2 Estado del Arte*

En los sistemas TBIR, las imágenes se recuperan a partir de las anotaciones de texto (o metadatos) asociados con las imágenes (Zhang, Jiang y Zhang, 2009). Estas anotaciones pueden ser el texto que rodea a la imagen, el nombre de archivo, el hipervínculo o cualquier otro texto asociado con la imagen (Su et al., 2009). Los motores de búsqueda de imágenes de Google y Yahoo son sistemas que utilizan este enfoque.

Los metadatos de las imágenes se pueden dividir en dos partes (Styrman, 2005). Una se refiere a información sobre quién ha creado la imagen, las herramientas utilizadas en su creación, el estilo artístico, etc. La otra describe las propiedades implícitas que pueden ser entendidas por la percepción de la imagen en sí.

Por tanto, el primer paso es contar con anotaciones y, para ello, se suelen aplicar dos enfoques. En el primero se crean anotaciones manualmente; el segundo consiste en anotar automáticamente las imágenes a partir de un conjunto predefinido de categorías (Chatzilari et al., 2011). Para ello, utilizando aprendizaje automático se extraen relaciones entre las características de la imagen y palabras en imágenes anotadas (Zhang, Chai y Jin, 2005).

Los métodos automáticos no suelen ser fiables, ya que la anotación automática tiene funcionalidad limitada, y sólo unos pocos objetos pueden detectarse con seguridad a partir de imágenes generales (Uchihashi y Kanade, 2005). Por otro lado, las anotaciones manuales suelen ser costosas e incompletas, debido a la subjetividad, ambigüedad e imprecisión provocada por los contenidos semánticos de las imágenes. Este problema recibe el nombre de *semantic gap*, la diferencia que se produce entre dos (o más) descripciones de un objeto en diferentes representaciones (Nguyen, 2010). Esta diferencia suele venir dada por fenómenos del lenguaje natural tales como la sinonimia o la polisemia (Saenko y Darrell, 2008). Además, dependiendo del contexto, dos usuarios pueden anotar de manera diferente una misma imagen, p.ej., para una imagen de un coche, un usuario puede anotar "coche" y otro usuario puede anotar "Seat León de color rojo". Esto puede dar lugar a incoherencias e imprecisiones en la anotación (Pavlidis, 2008).

Otro hecho que hace que las anotaciones sean incompletas es que suelen ser cortas, por lo que es difícil representar completamente el contenido de la imagen. Las anotaciones no están siempre disponibles o no describen características visuales. Por otro lado, los conocimientos previos o influencias culturales pueden dar lugar a diferentes interpretaciones, por lo que también esto influye directamente en la calidad de las anotaciones. Además, el enfoque TBIR no es ajeno al multilingüismo, ya que las anotaciones pueden encontrarse en diferentes idiomas; por lo tanto, el éxito de la recuperación va a estar ligado a si un usuario conoce o no el idioma y no tanto por el contenido de la imagen en sí mismo.

Una vez extraídas las anotaciones asociadas a las imágenes, un sistema TBIR indexa esta información convirtiéndose en un problema de IR textual, por lo que solo se recuperarán imágenes que hayan sido anotadas con alguno de los términos usados en la consulta (Pérez-Iglesias, 2008). Sin embargo, estas técnicas no han sido muy aplicadas a los textos asociados a las imágenes considerando sus características. De hecho, tomando como referencia los sistemas TBIR presentados en el foro de evaluación ImageCLEF, desde 2003 a 2012, se puede observar que no se ha ido más allá de la detección de NEs. Por ejemplo, en el trabajo presentado por la UNED (Granados et al., 2011) se generan dos índices, uno solo con metadatos y otro solo con NEs, de manera que luego fusionan las dos listas de resultados obtenidas en la recuperación.

## *3 Experimentación*

En esta sección se describen los experimentos así como el marco de evaluación en el que se llevan a cabo.

### 3.1 Colección de evaluación

La colección de evaluación utilizada en este trabajo es la empleada en la edición de ImageCLEF 2010 "Wikipedia Retrieval" (Popescu, Tsikrika y Kludas, 2010) para la recuperación de imágenes.

Esta colección está formada por 237.434 imágenes extraídas de la Wikipedia y sus correspondientes anotaciones cortas textuales realizadas por usuarios. Las anotaciones son multilingües (EN, DE y FR), y su distribución es la siguiente: 10% de imágenes con anotaciones en los 3 idiomas, 24% en 2 idiomas (11% EN+DE, 9% EN+FR, 4% FR+DE), 62% en un solo idioma (30% EN, 20% DE, 12% FR) y un 4% de imágenes en otros idiomas.

Las anotaciones de cada imagen se encuentran en los elementos <description>, <comment> y <caption>, identificados para cada idioma con el atributo "xml:lang" del elemento <text> del xml asociado. Además, existe un elemento común a todos los idiomas (<comment>) que contiene anotaciones entremezcladas en los tres idiomas, y sin formato fijo, lo que dificulta su procesado.

El método de evaluación se basa en juicios de relevancia. Se proporcionan 70 consultas en las tres lenguas, aportándose la traducción exacta de los términos de la consulta.

## 3.2 Modelos propuestos

Se evalúan tres modelos de indexación de información multilingüe. El primer modelo está formado por tres índices independientes que contendrán la información de cada idioma por separado. Un segundo modelo expande los índices anteriores traduciendo la información de las tres lenguas entre sí. El tercer modelo consiste en la creación de un único índice con la información conjunta en los tres idiomas.

Todos los índices estarán compuestos por los siguientes campos:

- **id**. Identificador del documento contenido en el atributo *id* del elemento <image>.
- **content**. Texto formado por la concatenación de los textos contenidos en los elementos <name>, <description>, <comment>, <caption> y el elemento común <comment> para cada idioma.
- **nes**. NEs detectadas en el campo *content*.
- **nes_person**. NEs de tipo PERSON detectadas en el campo *content*.
- **nes_organization**. NEs de tipo ORGANIZATION detectadas en *content*.
- **nes_location**. NEs de tipo LOCATION detectadas en el campo *content*.
- **nes_misc**. NEs de tipo MISC detectadas en el campo *content*.
- **sintagmas**. Sintagmas nominales detectados en el campo *content*.

La detección de NEs y sus categorías, tanto para el inglés como el alemán, se ha realizado con la herramienta *Stanford NER*[1], y para el francés con la herramienta *Stilus NER*[2]. Las categorías de las NEs consideradas son PERSON, ORGANIZATION, LOCATION y MISC, y para la detección de los sintagmas nominales, tanto en inglés como en francés, se ha utilizado la herramienta *Stilus Core*[3]; para el alemán, la herramienta *ParZu*[4].

En los tres modelos, el preprocesamiento del texto se ha llevado a cabo por los analizadores de *SnowBall*[5] implementados para cada idioma en *Lucene*[6], consistentes en la transformación del texto en minúsculas, eliminación de caracteres especiales, signos de puntuación y stemming. Sin embargo, la lista de stopwords ha sido sustituida por listas más completas proporcionadas por la Universidad de Neuchatel[7] (UniNE). Las consultas tendrán el mismo tratamiento que el de los documentos y para el caso de consultas multi-término, el operador utilizado para la búsqueda es OR.

### 3.2.1 Índices independientes

Este primer modelo consiste en crear tres índices independientes por idioma, que denominaremos 'EN', 'FR', 'DE' para el inglés, francés y alemán respectivamente. De esta manera, cada índice almacena únicamente la información extraída en su idioma. Se proponen dos modos de recuperación. Por un lado, se realizan búsquedas monolingües, consistentes en realizar una consulta por idioma, denominada 'en', 'fr' y 'de' para el inglés, francés y alemán respectivamente. Así, con una consulta en inglés se accederá sólo al índice en inglés y así sucesivamente para el resto de idiomas, obteniéndose tres listas de rankings. Estas consultas se denotan por las duplas EN-en, FR-fr y DE-de, siendo el primer término el idioma del índice y el segundo el de la consulta. Se propone una fusión de listas de documentos mediante el algoritmo de fusión *raw scoring* (Kwok, Grunfeld y Lewis, 1995).

### 3.2.2 Índices independientes expandidos

En este modelo se crean tres índices independientes por idioma, pero ahora la información extraída se expande con la información traducida de otros idiomas en el caso de no existir en el idioma original. Por ejemplo, si en el campo de un documento no existe información para el inglés, pero sí la hay en alemán, traduciríamos el texto en alemán al

[1] http://nlp.stanford.edu/software/CRF-NER.shtml
[2] http://www.daedalus.es/productos/stilus/stilus-ner/
[3] http://daedalus.es/productos/stilus/stilus-core/
[4] http://github.com/rsennrich/parzu
[5] http://snowball.tartarus.org/
[6] http://lucene.apache.org/java/docs/index.html
[7] http://members.unine.ch/jacques.savoy/clef/

inglés. Si no hubiera tampoco contenido en alemán, traduciríamos la del francés si la hubiera. Este proceso crea tres índices expandidos (ENexp, FRexp y DEexp). Teniendo en cuenta que en la colección la gran mayoría de la información está en inglés, seguida del alemán y francés, se establece este orden de precedencia para la búsqueda de información para traducir, y una vez encontrada en un idioma, no se seguiría buscando. La traducción se ha realizado por medio del API proporcionado por *Google Translate*[8]. Al igual que en el modelo anterior, se realizará la recuperación mediante búsquedas monolingües representadas por las duplas ENexp-en, FRexp-fr y DEexp-de y habrá que fusionar las listas de rankings obtenidas.

#### 3.2.3 Índice único

Este modelo consiste en la creación de un único índice, que llamamos ALL, con toda la información extraída de los tres idiomas. Dada la dificultad para identificar el idioma de los textos en los elementos <name> y <comment>, donde se mezclan contenidos en diferentes lenguas, se asume que los contenidos están en inglés. La recuperación se realiza mediante búsquedas monolingües al índice conjunto, representadas por las duplas ALL-en, ALL-fr y ALL-de, y otra búsqueda compuesta por la concatenación de la consulta en los tres idiomas, representada por la dupla ALL-all, donde all es la unión de las consultas en 'en', 'fr' y 'de'. Con este modelo se obtendrá, por tanto, una única lista de ranking.

### 3.3 Funciones de ranking

Se considera el empleo de la función de ranking BM25, y para poder estudiar el impacto de la información lingüística contenida en los diferentes campos de los índices, se utiliza su extensión a documentos estructurados, BM25F (Robertson y Walker, 1994).

Como valores de *b* y k se establecen valores estándar, mientras que los valores de empuje para los diferentes campos en el caso de BM25F se fijan tras probar diferentes combinaciones, algo común cuando se usan estas funciones tan parametrizables (Hernández-Aranda, 2013).

[8] http://code.google.com/intl/es-ES/apis/language/translate/overview.html

### 3.4 Medidas de evaluación

Las medidas de evaluación que se van a emplear esta experimentación son: MAP (*Mean Average Precision*), que calcula una media de la precisión hallada a distintos niveles de cobertura; y P@k (*Precision at k*) que indica la precisión obtenida en el conjunto de las primeras k imágenes recuperados.

## *4 Análisis de Resultados*

En esta sección se presentan y analizan los resultados de la experimentación llevada a cabo. Este análisis se estructura en torno a los modelos de indexación considerados y, para cada uno de ellos, los resultados se encuentran agrupados en los siguientes bloques:

- **Baseline**. La consulta se realiza sólo sobre el campo *content* y al ser un único campo, la función de ranking utilizada es BM25.
- **+ NEs**. La consulta se realiza también sobre el campo en el que se almacenan las NEs y se emplea la función de ranking BM25F.
- **+ Categorías**. La consulta se realiza sobre los campos *content* y los correspondientes a las diferentes categorías de las NEs.

A continuación, y para cada modelo de indexación, se presentan los resultados obtenidos cuando se añaden, a los experimentos anteriores, los sintagmas nominales detectados. En estos experimentos los resultados se agrupan en los siguientes bloques:

- **+ Sintagmas**. En este caso, además de lanzar la consulta sobre el campo *content* (*Baseline*), se hace también sobre el campo correspondiente a los sintagmas nominales.
- **+ Sintagmas + NEs**. Además de considerar los campos empleados en la aproximación anterior (*+Sintagmas*), la consulta se realiza ahora también sobre el campo de las NEs.
- **+ Sintagmas + Categorías**. Además de considerar los campos usados en *+Sintagmas*, la consulta se lanza también sobre los campos correspondientes a las diferentes categorías de las NEs.

Para cada uno de los experimentos se aplica el test de significancia estadística pareado de Wilcoxon, de modo que nos permita comprobar si las diferencias entre los valores obtenidos con un método u otro son estadísticamente significativas. El test se realiza tanto para valores de MAP ("p (MAP)" en las tablas), como de P@10 ("p (P@10)" en las tablas).

Antes de presentar los resultados, y para contextualizarlos de manera adecuada, se resumen los resultados obtenidos en el foro ImageCLEF 2010 y se comparan con los obtenidos en este trabajo. En ImageCLEF 2010 se presentaron 13 grupos de investigación que enviaron 113 experimentos, de los cuales 48 suponían un enfoque TBIR. Los mejores resultados obtenidos por cada grupo se correspondieron con experimentos realizados sólo en inglés o considerando todos los idiomas conjuntamente, tanto en las descripciones como en las consultas. Este hecho indica que la información textual disponible en inglés es más rica que la disponible para el resto de idiomas.

El mejor resultado para los enfoques TBIR, en el que nuestro trabajo se enmarca, lo obtuvo el grupo *XRCE* de Xerox (Clinchant et al., 2010), representando el texto por medio de modelos del lenguaje y con un modelo de información basado en la ley de potencias, combinándolo con una aproximación basada en retroalimentación. Los valores de MAP y P@10 que obtuvieron fueron 0.2361 y 0.4871 respectivamente, mientras que la media obtenida por los experimentos presentados en ImageCLEF 2010, calculada eliminando el 10% de los peores resultados, tiene como valor de MAP 0,1602 con una desviación estándar de 0,0505. Del mismo modo, en el caso de la P@10 el valor medio de los sistemas fue 0,4095, con una desviación estándar de 0,0713.

A continuación se muestran los resultados obtenidos con nuestros experimentos para cada uno de los tres modelos propuestos.

**Índices independientes.** La Tabla 1 muestra los resultados obtenidos para este modelo. En primer lugar, y como se concluyó en ImageCLEF 2010, se observa que para las búsquedas monolingües los mejores resultados se obtienen para el inglés. Una posible razón de este comportamiento es el hecho de que la información en este idioma es más extensa.

Observando estos resultados se puede concluir, de forma general, que el hecho de considerar NEs y sus categorías mejoran el baseline, excepto para el caso del alemán en el que puede no haber muchas NEs en su texto, obteniéndose mejores resultados para las NEs en conjunto en lugar de ponderar de diferente manera a cada una de las categorías.

En cuanto a los resultados de la P@10, en ningún caso se obtienen mejoras significativas, lo que quiere decir que las posibles diferencias que hubiera al considerar o no las NEs y sus categorías no modifican el conjunto de los 10 documentos más relevantes, aunque en algún caso el orden dentro de estos 10 primeros documentos pudiera ser diferente. Esto es lo que explica que exista algún caso donde se encuentren mejoras en MAP y no de P@10.

| **Baseline** | **Map** | **P@10** | | |
|---|---|---|---|---|
| BM25 (EN-en) | 0.2244 | 0.5071 | | |
| BM25 (FR-fr) | 0.1029 | 0.3500 | | |
| BM25 (DE-de) | 0.0959 | 0.3186 | | |
| BM25 (con fusión) | 0.2263 | 0.5114 | | |
| **+ NEs** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (EN-en) | **0.2314** | 0.5014 | 0.0394 | 0.4400 |
| BM25F (FR-fr) | 0.1052 | 0.3457 | 0.0003 | 0.3735 |
| BM25F (DE-de) | 0.1002 | 0.3186 | 0.4614 | 0.4761 |
| BM25F (con fusión) | **0.2359** | 0.5071 | 0.0007 | 0.4353 |
| **+ Categorías** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (EN-en) | 0.2307 | 0.5000 | 0.0072 | 0.4259 |
| BM25F (FR-fr) | 0.1052 | 0.3600 | 0,0000 | 0.2758 |
| BM25F (DE-de) | 0.0985 | 0.3186 | 0,5954 | 0.6082 |
| BM25F (con fusión) | 0.2328 | 0.5086 | 0,0002 | 0.6026 |
| **+ Sintagmas** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (EN-en) | 0.2264 | 0.4971 | 0.1259 | 0.6639 |
| BM25F (FR-fr) | 0.1081 | 0.3643 | 0.0004 | 0.2970 |
| BM25F (DE-de) | 0.0918 | 0.2814 | 0.0051 | 0.0046 |
| BM25F (con fusión) | 0.2257 | 0.4786 | 0.9714 | 0.0183 |
| **+ Sintagmas + NEs** | **Map** | **P@10** | **p(Map)** | **P(P@10)** |
| BM25F (EN-en) | 0.2313 | 0.5000 | 0.0327 | 0.3994 |
| BM25F (FR-fr) | 0.1074 | 0.3557 | 0.0001 | 0.8717 |
| BM25F (DE-de) | 0.0954 | 0.3029 | 0.3950 | 0.3568 |
| BM25F (con fusión) | 0.2324 | 0.5000 | 0.0112 | 0.4242 |
| **+ Sintagmas + Categorías** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (EN-en) | 0.2306 | 0.5000 | 0.0073 | 0.4844 |
| BM25F (FR-fr) | 0.1053 | 0.3600 | 0.0000 | 0.2155 |
| BM25F (DE-de) | 0.0949 | 0.3143 | 0.8569 | 0.1670 |
| BM25F (con fusión) | 0.2298 | 0.4971 | 0.0010 | 0.0686 |

Tabla 1 - Resultados sobre el modelo de *Índices independientes*.

Un aspecto a resaltar es el hecho de que sí se encuentra mejora en el valor de MAP en el caso de la fusión de listas. Esta mejora puede deberse a que a la hora de fusionar se están introduciendo en la lista de ranking del inglés (por ser el idioma con el que se obtiene mejor resultado), documentos del resto de idiomas que en otro caso no se hubieran recuperado por no estar escritos en inglés. De hecho, el mejor resultado con este modelo, con un valor de MAP de 0.2359, se obtiene cuando se realiza la fusión de listas con la función BM25F y teniendo en cuenta las NEs. De nuevo no se observa mejora en los valores de P@10, lo que indica que la fusión está reordenando los 1000 primeros resultados devueltos por el sistema, pero sin variar el número de documentos relevantes en las primeras 10 posiciones.

Cuando se tienen en cuenta los sintagmas no se encuentran diferencias significativas, salvo en el único caso del francés. Sin embargo,

cuando se combinan los sintagmas con las NEs y categorías se consigue superar siempre el baseline excepto para el alemán. Incluso con la fusión de listas se consiguen mejoras significativas. Sin embargo, estos resultados no superan los resultados obtenidos cuando se tienen en cuenta solamente las NEs y categorías. Por tanto, se puede concluir que esta mejora se debe al uso de las NEs y sus categorías, y no al uso de los sintagmas.

En resumen, para este primer modelo el mejor resultado monolingüe ha sido obtenido por el inglés teniendo en cuenta las NEs, con un valor de MAP de 0.2314. Sin embargo, el mejor resultado total se ha obtenido con la fusión de listas de documentos en los tres idiomas teniendo en cuenta las NEs, con un valor de MAP de 0.2359. Por tanto, se observa un buen comportamiento con el uso de las NEs y sus categorías, mientas que los sintagmas no parecen ser una buena opción. Sin embargo, a la vista de los valores de P@10, la mejora solo se da en la ordenación de los resultados, pero no se consigue recuperar más imágenes relevantes en los 10 primeros resultados.

**Índices independientes expandidos.** La Tabla 2 muestra los resultados obtenidos para este modelo. Al igual que en el modelo anterior, los mejores resultados se obtienen para el inglés. Sin embargo, se observa una importante mejora en el caso de francés y alemán gracias a la expansión de información que llevan consigo las traducciones. Esto se debe a que las traducciones permiten recuperar documentos escritos en una o dos lenguas que de otra manera sólo se hubiera podido recuperar con la consulta en su propia lengua.

Se observa que los mejores resultados se obtienen, tanto con búsquedas monolingües como con fusión de resultados teniendo en cuenta a las NEs (a excepción del alemán) y Categorías. Además, con las categorías, para la búsqueda monolingüe en inglés se obtiene el mejor resultado en el mismo experimento con NEs. De hecho, con 0.2413 es el mejor resultado en búsquedas monolingües. Por tanto, dar un peso específico a las NEs y, sobretodo, dar más importancia a las NE de tipo Persona y Organización que a las de tipo Location y Miscelánea, parece dar buenos resultados.

El mejor resultado de este modelo, con un valor de MAP de 0.2434, se vuelve a obtener cuando se realiza la fusión de listas teniendo en cuenta las NEs. De nuevo, con respecto a los valores de P@10, en ningún caso se obtienen mejoras significativas, lo que quiere decir que no se modifica el conjunto de los 10 documentos más relevantes.

| **Baseline** | **Map** | **P@10** | | |
|---|---|---|---|---|
| BM25 (ENexp-en) | 0.2323 | 0.4871 | | |
| BM25 (FRexp-fr) | 0.1766 | 0.4129 | | |
| BM25 (DEexp-de) | 0.1432 | 0.3557 | | |
| BM25 (con fusión) | 0.2302 | 0.5043 | | |
| **+ NEs** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ENexp-en) | 0.2408 | 0.5057 | 0.0125 | 0.2309 |
| BM25F (FRexp-fr) | 0.1858 | 0.4157 | 0.0130 | 0.4966 |
| BM25F (DEexp-de) | 0.1498 | 0.3614 | 0.1286 | 0.4352 |
| BM25F (con fusión) | **0.2434** | 0.5114 | 0.0007 | 0.6823 |
| **+ Categorías** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ENexp-en) | 0.2413 | 0.4971 | 0.0001 | 0.3310 |
| BM25F (FRexp-fr) | 0.1855 | 0.4200 | 0.0000 | 0.1779 |
| BM25F (DEexp-de) | 0.1490 | 0.3657 | 0.0006 | 0.9022 |
| BM25F (con fusión) | 0.2404 | 0.5057 | 0.0000 | 0.8367 |
| **+ Sintagmas** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ENexp-en) | 0.2362 | 0.4714 | 0.4084 | 0,3093 |
| BM25F (FRexp-fr) | 0.1839 | 0.4200 | 0.0011 | 0.4928 |
| BM25F (DEexp-de) | 0.1436 | 0.3443 | 0.5791 | 0.1297 |
| BM25F (con fusión) | 0.2383 | 0.5029 | 0.0172 | 0.8833 |
| **+ Sintagmas + NEs** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ENexp-en) | 0.2409 | 0.4986 | 0.0165 | 0.5743 |
| BM25F (FRexp-fr) | 0.1860 | 0.4229 | 0.0045 | 0.2195 |
| BM25F (DEexp-de) | 0.1494 | 0.3629 | 0.1282 | 0.6854 |
| BM25F (con fusión) | 0.2427 | 0.5086 | 0.0023 | 0.7186 |
| **+ Sintagmas + Categorías** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ENexp-en) | **0.2414** | 0.4914 | 0.0001 | 0.8353 |
| BM25F (FRexp-fr) | 0.1858 | 0.4214 | 0.0000 | 0.1516 |
| BM25F (DEexp-de) | 0.1485 | 0.3657 | 0.0024 | 1.0000 |
| BM25F (con fusión) | 0.2405 | 0.5086 | 0.0000 | 1.0000 |

Tabla 2 - Resultados sobre el modelo de *Índices independientes expandidos*.

Con este modelo los resultados superan a los mejores obtenidos en ImageCLEF 2010 (Xrce) en términos de MAP y P@10, tanto con consulta monolingüe en inglés como con fusión de listas considerando las NEs y sus categorías.

En este modelo, cuando se tienen en cuenta los sintagmas, tanto solos como junto a las NEs y Categorías, en la mayoría de los casos se consiguen superar los resultados del baseline, tanto en las búsquedas monolingües como en la fusión de listas. Esta mejora puede deberse, con respecto al modelo anterior, a que con las traducciones se ha ampliado el contexto de las descripciones de los documentos y, por tanto, se detectan un mayor número de sintagmas. Con respecto al modelo de índices independientes, en el que solo en un caso con sintagmas se conseguía superar el baseline, parece que con la traducción se ha superado la falta de información que supuestamente hacía que la utilización de sintagmas no fuera útil.

Para las búsquedas monolingües, teniendo en cuenta Sintagmas y Categorías, se consigue superar al mejor resultado previo. Sin embargo

esta mejora es mínima, siendo el valor de MAP de 0.2414 frente al 0.2413 que se consigue con el mismo experimento pero sin tener en cuenta los sintagmas. Para esta mejora tan mínima no merece la pena el coste computacional que supone la detección de sintagmas nominales. En cuanto a la fusión de listas se obtiene el mejor resultado considerando NEs, pero sin superar el mejor resultado obtenido dentro de este modelo.

En resumen, el mejor resultado monolingüe de este modelo se ha obtenido con inglés teniendo en cuenta los sintagmas y categorías. Por otro lado, el mejor resultado global se ha obtenido con la fusión de listas de documentos en los tres idiomas teniendo en cuenta las NEs, con un valor de MAP de 0.2434.

**Índice Único.** La Tabla 3 muestra los resultados obtenidos para este modelo. El comportamiento observado en los modelos anteriores se mantiene para el inglés. Para el inglés y alemán en el baseline, los resultados superan al baseline del modelo de Índices Independientes, lo que indica que se están recuperando más documentos relevantes al estar toda la información almacenada conjuntamente, por lo que es posible que existan términos en la consulta y en las descripciones de las imágenes que compartan su misma forma canónica o su raíz en inglés y alemán (puede ser que se trate de cognados entre ambas lenguas que compartan raíz). Esto no ocurre para el francés.

Se observa que los resultados que presentan mejoras significativas se obtienen considerando las NEs (a excepción del alemán) y sus Categorías, tanto con búsquedas monolingües como con fusión de resultados. Además con las categorías, y para el inglés, se obtienen mejoras al mismo experimento realizado con NEs. Por tanto, dar un peso específico a las NEs, y sobretodo, dar más importancia a las entidades de tipo Person y Organization que a las de tipo Location y Misc, parece dar buenos resultados.

Es importante destacar que este modelo mejora cuando la consulta se realiza con la concatenación de las *queries* en los tres idiomas (consulta *all*) y se tienen en cuentas las NEs y sus categorías. Se está logrando aumentar la cobertura (al recuperar documentos en todos los idiomas) sin penalizar valores de precisión. De hecho, los valores de P@10 mejoran con este tipo de consultas, de modo que se están consiguiendo recuperar más documentos relevantes en los 10 primeros resultados, al intercalar documentos relevantes de los tres idiomas en esas 10 primeras posiciones.

| **Baseline** | **Map** | **P@10** | | |
|---|---|---|---|---|
| BM25 (ALL-en) | 0.2285 | 0.5000 | | |
| BM25 (ALL-fr) | 0.1019 | 0.3043 | | |
| BM25 (ALL-de) | 0.1043 | 0.3000 | | |
| BM25 (ALL-all) | 0.2300 | 0.5014 | | |
| **+ NEs** | **Map** | **P@10** | **P(Map)** | **p(P@10)** |
| BM25F (ALL-en) | 0.2332 | 0.5057 | 0.0016 | 0.7485 |
| BM25F (ALL-fr) | 0.1060 | 0.3100 | 0.0000 | 0.3791 |
| BM25F (ALL-de) | 0.1025 | 0.2943 | 0.5749 | 1.0000 |
| BM25F (ALL-all) | 0.2401 | 0.5243 | 0.0298 | 0.0132 |
| **+ Categorías** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ALL-en) | **0.2336** | 0.5057 | 0.0002 | 0.7112 |
| BM25F (ALL-fr) | 0.1046 | 0.3014 | 0.0000 | 0.7750 |
| BM25F (ALL-de) | 0.1022 | 0.2943 | 0.2785 | 0.9364 |
| BM25F (ALL-all) | **0.2403** | 0.5107 | 0.0103 | 0.0191 |
| **+ Sintagmas** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ALL-en) | 0.2285 | 0.4686 | 0.9829 | 0.0098 |
| BM25F (ALL-fr) | 0.1019 | 0.3143 | 0.0935 | 0.1659 |
| BM25F (ALL-de) | 0.1022 | 0.2900 | 0.9492 | 0.6174 |
| BM25F (ALL-all) | 0.2300 | 0.5086 | 0.9117 | 0.4113 |
| **+ Sintagmas +NEs** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ALL-en) | 0.2321 | 0.4900 | 0.0597 | 0.2046 |
| BM25F (ALL-fr) | 0.1034 | 0.3057 | 0.0310 | 0.4101 |
| BM25F (ALL-de) | 0.1024 | 0.2857 | 0.8130 | 0.2386 |
| BM25F (ALL-all) | 0.2400 | 0.5071 | 0.1617 | 0.6306 |
| **+ Sintagmas + Categorías** | **Map** | **P@10** | **p(Map)** | **p(P@10)** |
| BM25F (ALL-en) | 0.2333 | 0.5100 | 0.0030 | 0.3919 |
| BM25F (ALL-fr) | 0.1034 | 0.3014 | 0.0385 | 0.7346 |
| BM25F (ALL-de) | 0.1014 | 0.2914 | 0.5725 | 0.7495 |
| BM25F (ALL-all) | 0.2402 | 0.5086 | 0.8416 | 0.5900 |

Tabla 3 - Resultados sobre el modelo de *Índice Único*

Se puede observar también que, tanto para las búsquedas monolingües como cuando se concatenan las consultas, al tener en cuenta los sintagmas no se consigue superar el baseline, ni cuando consideramos las NEs, ni las categorías. Sin embargo, cuando se tienen en cuenta los sintagmas y las categorías para el inglés, se consigue superar el baseline, aunque no se supera el resultado obtenido cuando se tienen en cuenta solamente las categorías. Por tanto, la mejora vuelve a deberse al uso de las categorías y no de los sintagmas.

En resumen, el mejor resultado monolingüe se ha obtenido con el inglés teniendo en cuenta las categorías. Sin embargo, el mejor resultado se ha obtenido cuando la recuperación se realiza con la concatenación de las consultas en los tres idiomas y teniendo en cuenta las categorías, con un valor de MAP de 0.2403.

## *5 Conclusiones y Futuros trabajos*

En este artículo se ha presentado un análisis y evaluación del uso de diferentes modelos de indexación/recuperación dentro de una tarea de recuperación multilingüe de imágenes basada

en el texto asociado a las imágenes. Se ha evaluado el uso de las NEs, sus categorías y los sintagmas nominales detectados.

Los resultados obtenidos muestran que, a pesar del elevado coste computacional, el mejor enfoque es el de traducir los documentos a todas las lenguas consideradas. Sin embargo, si no es posible traducir toda la colección, la creación de un índice único y la concatenación de consultas en diferentes idiomas ha resultado ser la mejor opción. Además, la fusión de listas de resultados es mejor solución que realizar búsquedas monolingües. El hecho de considerar las entidades nombradas y sus categorías ha resultado ser buena opción, pero sin embargo, el uso de sintagmas nominales quedaría descartado, ya que se obtiene poco beneficio para el coste computacional que requiere.

En cuanto a los futuros trabajos, sería interesante la aplicación de los enfoques presentados en otro tipo de colecciones multimedia; por ejemplo de vídeos o audios de los que se tuvieran transcripciones del habla o subtítulos. Por último, se estudiará el efecto de diferentes algoritmos de fusión de listas.

## *6 Agradecimientos*

Este trabajo no sería posible sin la participación de los autores en los proyectos de investigación MA2VICMR (S2009/TIC-1542), HOLOPEDIA (TIN2010-21128-C02) y PROYECTO UNED (2012V/PUNED/0004).

## *Bibliografía*

Chatzilari, .E, S. Nikolopoulos, Papadopoulos, C. Zigkolis y Y. Kompatsiaris. 2011. Semi-Supervised object recognition using flickr images. In *9th International Workshop on Content-Based Multimedia Indexing*, Madrid, Spain.

Clinchant, S., G. Csurka, J. Ah-Pine, G. Jacquet, F. Perronnin, J. Sánchez y K. Minoukadeh. 2010. XRCE's Participation in Wikipedia Retrieval, Medical Image Modality Classification and Ad-hoc Retrieval Tasks of ImageCLEF 2010. *CLEF*.

Granados, R., J. Benavent, X. Benavent, E. de Ves y A. García-Serrano. 2011. Multimodal information approaches for the Wikipedia collection at Image-CLEF 2011. In *CLEF 2011 Working Notes*.

Hernández-Aranda, D. 2013. Información Textual Multilingüe en Recuperación de Imágenes. *Tesis de Fin de Master en Lenguajes y Sistemas Informáticos*. UNED.

Kwok, K. L., L. Grunfeld y D. D. Lewis. 1995. TREC-3 Ad-Hoc, Routing Retrieval and Thresholding Experiments using PIRCS. In *Procs. of TREC3*, 47-56. NIST.

Nguyen, C. T. 2010. Bridging semantic gaps in information retrieval: Context-based approaches. ACM VLDB, 10.

Pavlidis, T. 2008. Limitations of cbir. In *ICPR*.

Pérez Iglesias, J. 2008. Función de ranking para documentos estructurados, basada en lógica borrosa. *DEA*. UNED.

Popescu, A., T. Tsikrika y J. Kludas. 2010. Overview of the Wikipedia Retrieval Task at ImageCLEF 2010. In *CLEF* (Notebook Papers/LABs/Workshops).

Robertson, S. E. y S. Walker. 1994. Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. In *Proc. of the SIGIR '94*, W. Bruce Croft and C. J. van Rijsbergen (Eds.). Springer-Verlag. NY, USA, 232-241.

Saenko, K. y T. Darrell. 2008. Unsupervised Learning of Visual Sense Models for Polysemous Word. In *Proc. of the 22nd Annual Conference on Neural Information Processing Systems*. Vancouver, Canada, pp.1393-1400.

Styrman, A. 2005. Ontology-based image annotation and retrieval. Doctoral dissertation, Master thesis.

Su, J., B. Wang, H. Yeh y V. S. Tseng. 2009. Ontology-Based Semantic Web Image Retrieval by Utilizing Textual and Visual Annotations. In *Web Intelligence/IAT Workshops*, pp: 425-428.

Uchihashi, S. y T. Kanade. 2005. Content-Free Image Retrieval Based On Relations Exploited From User Feedbacks. IEEE.

Zhang, Ch., J. Y. Chai y R. Jin. 2005. User Term Feedback in Interactive Text-based Image Retrieval. SIGIR'05, August 15–19, Salvador, Brazil.

Zhang, H., M. Jiang y X. Zhang. 2009. Exploring image context for semantic understanding and retrieval. In *International Conference on Computational Intelligence and Software Engineering*, pp. 1 – 4.