

Desambiguación de términos basada en IDF aplicada a recuperación de información*

Word Sense Disambiguation based on IDF applied to Information Retrieval

José M. Perea-Ortega Fernando Martínez-Santiago
Miguel Á. García-Cumbreras Arturo Montejo-Ráez

Departamento de Informática, Escuela Politécnica Superior
Universidad de Jaén, E-23071 - Jaén
{jmperea,dofer,magc,amontejo}@ujaen.es

Resumen: Por regla general, la aplicación de la desambiguación del sentido de las palabras (Word Sense Disambiguation, WSD) viene determinada exclusivamente por la confianza en el sistema desambiguador utilizado. En este trabajo se realiza un estudio en el ámbito de la Recuperación de Información (Information Retrieval, IR) sobre la incidencia que tiene en la desambiguación otros factores, tales como la confianza en la herramienta de desambiguación, el grado de polisemia o granularidad y la diferencia en la fuerza de discriminación entre el término original y el desambiguado. Por tanto, se propone un enfoque para decidir si un término debe ser desambiguado o no y aplicarlo a recuperación de información. Finalmente, se muestra que una desambiguación selectiva de términos basada en Frecuencia Inversa de Documento (Inverse Document Frequency, IDF) mejora ligeramente el rendimiento de un sistema de recuperación de información.

Palabras clave: Desambiguación de términos, recuperación de información, IDF

Abstract: The application of Word Sense Disambiguation (WSD) is usually determined exclusively by the trust in the disambiguation system used. In this paper, a study in the Information Retrieval (IR) field is carried out about the impact of others factors in WSD such as the confidence of the WSD tool, the grade of polisemy or granularity and the difference in the discrimination strength between the original term and the disambiguated one. Thus, a proposal to decide whether a word should be disambiguated or not according to Inverse Document Frequency (IDF) is presented. Finally, it is shown that a selective disambiguation based on IDF improves slightly the performance of an IR system.

Keywords: Word Sense Disambiguation, Information Retrieval, IDF

1. *Introducción*

La aplicación de la desambiguación de términos (Word Sense Disambiguation, WSD) en recuperación de información (Information Retrieval, IR) ha sido ampliamente estudiada durante las últimas décadas, pero hoy día la comunidad científica no ha llegado todavía a un acuerdo sobre cuál es la manera óptima de aplicarla. Parece razonable asumir que

un sistema IR mejorará su rendimiento si los documentos se representan mediante los sentidos de las palabras en lugar de los propios términos, sobre todo si los términos ambiguos se desambiguan correctamente. Sin embargo, estudios recientes relacionados con la aplicación de WSD a sistemas IR no muestran ninguna mejora, como se puede comprobar en la siguiente sección. Los malos resultados de los sistemas de desambiguación actuales y la granularidad demasiado fina de algunos sentidos de las palabras son el principal inconveniente.

En este trabajo se pretende analizar el comportamiento de una desambiguación se-

* Este trabajo ha sido cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER), proyecto TIN2009-13391-C04-02 (MICINN), proyecto GeOasis (P08-TIC-41999) de la Junta de Andalucía, proyecto UJA2009/12/14 (Universidad de Jaén) y por el proyecto Geocaching Urbano (RFC/IEG2010)

lectiva de términos aplicada a IR: solo algunas palabras serán desambiguadas atendiendo a su poder discriminativo (Inverse Document Frequency, IDF) con respecto al término desambiguado. Primeramente, se realizará un breve recorrido del estado del arte sobre la conveniencia de utilizar WSD en tareas IR. En segundo lugar, se explicará el marco de trabajo utilizado para llevar a cabo los experimentos y se mostrarán los resultados obtenidos, haciendo un breve análisis de los mismos. Finalmente, se expondrán las conclusiones y el trabajo futuro.

2. Trabajo relacionado

La desambiguación automática del sentido de las palabras es una tarea antigua que ha sido estudiada durante muchos años (Gale, Church, y Yarowsky, 1992) y que ha sido aplicada con éxito a ciertas tareas como categorización de texto (Ureña-López, Buenaga-Rodríguez, y Gómez, 2001) y recuperación de información (Gonzalo et al., 1998). El primer intento de utilizar un desambiguador dentro de un sistema IR consiguió una mejora del 1% en el rendimiento del sistema IR al desambiguar cinco palabras ambiguas en la colección utilizada (Weiss, 1973). En un trabajo posterior, se concluyó que el impacto de la ambigüedad de los sentidos de las palabras en IR no era dramático, aunque la desambiguación fuera probablemente beneficiosa para recuperar cuando había pocos términos en común entre el documento y la consulta (Krovetz y Croft, 1992).

Los primeros experimentos a gran escala para aplicar desambiguación en un sistema IR fueron desarrollados por Voorhees (Voorhees, 1993) y Wallis (Wallis, 1993). Voorhees construyó un desambiguador de sentidos basado en un tesoro de WordNet aplicándolo a diferentes colecciones. Desafortunadamente, los resultados obtenidos con estas colecciones desambiguadas conllevaron un decremento en el rendimiento del sistema IR. Por otro lado, Wallis utilizó un desambiguador como parte de un sistema en el que sustituía las palabras de la colección por sus definiciones obtenidas de un diccionario. Para ello utilizó sinónimos de las palabras, de tal modo que los documentos que contuvieran estos sinónimos también serían recuperados. Cuando se reemplazaba una palabra por su definición, se utilizaba el desambiguador para seleccionar el término más representativo. Los experimentos lleva-

dos a cabo sobre varias colecciones tampoco obtuvieron una mejora significativa en el rendimiento de la recuperación de información.

Un excelente estudio sobre WSD en general se puede encontrar en el libro de Agirre y Edmonds (Agirre y Edmonds, 2006). Además, otro importante trabajo sobre WSD e IR es el presentado por Sanderson (Sanderson, 2000). En este trabajo, Sanderson sugiere tres razones principales por las que la desambiguación no acaba de rendir en recuperación de información: la colocación de palabras, la distribución de frecuencias de los sentidos y la desambiguación errónea.

Con respecto a las campañas internacionales recientes que han tratado de incorporar la desambiguación como una tarea a evaluar, se pueden destacar dos principalmente: SemEval¹ y CLEF². En concreto, en el CLEF del año 2008 se introdujo una nueva tarea llamada *Robust-WSD* que se centró en estudiar y evaluar cómo contribuía la desambiguación de términos a la recuperación de información monolingüe y multilingüe (Agirre et al., 2008). Para ello, la organización proporcionó a los participantes la colección ya desambiguada. Las conclusiones generales que se obtuvieron en esta edición fueron diversas: por un lado hubo grupos que consiguieron mejorar el resultado de la recuperación utilizando la colección desambiguada pero, por otro lado, los mejores resultados en términos de precisión media se obtuvieron por sistemas que no utilizaban dicha información.

En la edición de 2009, la tarea *Robust-WSD* siguió centrándose en el mismo objetivo (Agirre et al., 2009), pero los organizadores proporcionaron las consultas y la colección ya etiquetados con sentidos de WordNet, utilizando para ello dos sistemas WSD automáticos: UBC (Agirre y Lacalle, 2007) y NUS (Chan, Ng, y Zhong, 2007). Estos sistemas proporcionaban un peso determinado al sentido establecido para cada nombre, verbo, adjetivo y adverbio que el sistema podía desambiguar. Las conclusiones obtenidas fueron similares a las del 2008. La principal conclusión en la subtarea monolingüe fue que la evidencia de utilizar WSD en IR no estaba clara, ya que solo algunos grupos consiguie-

¹Evaluation Exercises on Semantic Evaluation <http://semeval2.fbk.eu/semeval2.php>

²Cross Language Evaluation Forum <http://www.clef-campaign.org/>

```

<DOC>
  <DOCNO>GH950102-000000</DOCNO>
  <DOCID>GH950102-000000</DOCID>

  <HEADLINE>
    <TERM ID="GH950102-000000-1" LEMA="alien" POS="JJ">
      <WF>Alien</WF>
      <SYNSET SCORE="0.6" CODE="01295935-a"/>
      <SYNSET SCORE="0.4" CODE="00984080-a"/>
    </TERM>

    <TERM ID="GH950102-000000-2" LEMA="treatment" POS="NN">
      <WF>treatment</WF>
      <SYNSET SCORE="0.827904118008605" CODE="00735486-n"/>
      <SYNSET SCORE="0" CODE="03857483-n"/>
      <SYNSET SCORE="0.172095881991395" CODE="00430183-n"/>
      <SYNSET SCORE="0" CODE="05340429-n"/>
    </TERM>
  </HEADLINE>

```

Figura 1: Extracto de un documento de la colección desambiguada

ron pequeñas mejoras en términos de precisión media, mientras que los mejores resultados globales fueron conseguidos por sistemas que no utilizaban WSD. Con respecto a la subtarea multilingüe, el comportamiento de los sistemas fue muy similar. Nuestro grupo participó en la tarea *Robust-WSD* del CLEF 2008 exponiendo un primer paso para caracterizar consultas donde la desambiguación de términos podía ser útil (Martínez-Santiago, Perea-Ortega, y García-Cumbreras, 2008).

3. Marco de experimentación

Para llevar a cabo los experimentos de este trabajo se ha utilizado la colección de documentos proporcionada por la organización del CLEF para la tarea *Robust-WSD* del 2008 (Agirre et al., 2008). Esta tarea utilizaba, a su vez, distintas colecciones existentes en el CLEF pero añadiendo información sobre la desambiguación de las palabras. La colección está compuesta por 113.000 noticias publicadas en el periódico “*Los Angeles Times*” en 1994 y 56.500 noticias publicadas en el periódico “*Glasgow Herald*” en 1995. Un extracto del documento con identificador GH950102-000000 de esta colección se puede observar en la Figura 1, donde aparecen los términos *alien* y *treatment* con sus respectivos códigos de sentido (*synset code*) y el peso otorgado por el desambiguador a dicho sentido (*score*).

Con respecto a las consultas, se han utilizado todos los *topics* proporcionados en las diferentes campañas del CLEF desde el año 2001 hasta el 2006, consiguiendo un total de

160 consultas. Las consultas están compuestas por una parte “título” (T) y una parte “descripción” (D). Por otro lado, el sistema de desambiguación utilizado tanto para la colección como para las consultas ha sido el proporcionado por el grupo NUS (Chan, Ng, y Zhong, 2007), que se sirve de la base de datos léxica WordNet 1.6³ para establecer los códigos de los sentidos.

Por último, para el proceso de recuperación de información, se han generado dos índices diferentes:

- **TS index**, que para cada documento de la colección considera un *token* como la unión de la palabra (el lema) y el código del *synset* asociado que tiene el mayor *score*. Un ejemplo de *token* almacenado en este tipo de índice sería *survey00644503*. Para los experimentos llevados a cabo en este trabajo se generó además una variante de este tipo de índice que consistió en separar el lema del *synset*, con el objetivo de poder realizar consultas *mixtas*, es decir mezclando lemas y *synset* por separado.
- **S index**, que almacena únicamente el código del *synset* asociado con mayor *score*. Para el ejemplo anterior, el token almacenado en este tipo de índice sería simplemente 00644503.

4. Experimentos y resultados

Una vez presentado el marco de trabajo, en esta sección se explican los experimentos

³<http://wordnet.princeton.edu>

llevados a cabo para evaluar el impacto de la desambiguación de palabras en recuperación de información. Se proponen diferentes experimentos que se describen brevemente a continuación:

- **Aplicación tradicional de WSD en IR** (sección 4.1). Consiste en evaluar la recuperación de información utilizando las consultas y la colección ya desambiguadas.
- **Análisis de la desambiguación de cada palabra en la consulta** (sección 4.2). Este experimento ha consistido en medir el impacto de la desambiguación de una única palabra en cada consulta, con el objetivo de caracterizar los beneficios de la desambiguación a nivel de palabra. Para ello, se ha empleado el poder discriminativo de cada término en base a su frecuencia inversa de documento (IDF), como se describe a continuación. También se ha analizado el proceso inverso, es decir, se ha estudiado cómo varía el rendimiento de la recuperación cuando solo una palabra en la consulta no se ha desambiguado.
- **Análisis de una desambiguación selectiva en IR** (sección 4.3). Finalmente, se ha evaluado el comportamiento de una desambiguación selectiva en recuperación de información, según las conclusiones extraídas de los experimentos anteriores.

La medida utilizada para la evaluación de los experimentos ha sido el valor de precisión media (Mean Average Precision, MAP), cuya fórmula es la siguiente:

$$MAP = \frac{\sum_{d=1}^N P(d)}{N} \quad (1)$$

donde:

- N es el número de documentos relevantes recuperados.
- d es un documento relevante recuperado.
- $P(d)$ es la precisión del documento d recuperado.

4.1. Aplicación tradicional de WSD en IR

En este primer grupo de experimentos tratamos de aplicar WSD de una manera

Experimento	Índice	MAP
<i>Baseline</i>	lema	0,374
WSD-TS	lema + <i>synset</i>	0,368
WSD-S	solo <i>synset</i>	0,376

Tabla 1: Resultados de la aplicación tradicional de WSD en IR

tradicional, desambiguando la parte “título” y “descripción” de las consultas (TD) con el sistema NUS y utilizando la colección desambiguada con el mismo sistema NUS. Por otro lado, se evaluó el caso base (*baseline*), que consistió en lanzar las consultas sin desambiguar sobre la colección también sin desambiguar, preprocesada con *Porter stemmer* (Porter, 1980) para extraer los lemas de las palabras y eliminando las palabras vacías. Debido a que la colección desambiguada contenía multipalabras de WordNet marcadas como *tokens* únicos, se optó por marcar de igual manera las mismas multipalabras en la colección sin desambiguar. Esto fue necesario porque pretendíamos comparar el rendimiento de ambos enfoques. Ejemplos de multipalabras detectadas fueron: *look-for*, *give-in*, *military-service*, *anti-smoking*...

Los resultados de estos experimentos se muestran en la Tabla 1. Como era de esperar, la aplicación de WSD de forma indiscriminada en IR no funcionó bien. Además, concluimos que no se obtendrían mejores resultados si se utilizara un sistema de desambiguación alternativo al empleado en este trabajo porque, como ya se ha comentado, el sistema NUS es considerado uno de los mejores sistemas WSD según el estado del arte actual. Por otro lado, se puede observar una ligera mejora (2,17 %) del índice basado únicamente en *synset* con respecto al índice basado en *lema + synset*. La mejora obtenida al utilizar WSD con respecto al caso base fue únicamente del 0,5 %, por lo que se puede concluir que la aplicación tradicional de WSD en IR no consigue mejores resultados.

4.2. Análisis de la desambiguación de cada palabra en la consulta

Después de un análisis profundo del trabajo desarrollado en (Sanderson, 2000) sobre las cuestiones relacionadas con la conveniencia o no de utilizar WSD en IR, nos planteamos varias formas de aplicar la desambiguación. Por tanto, durante este experimento se estudiaron cuatro indicadores principales para me-

dir la conveniencia de aplicar WSD en IR: el impacto de la desambiguación sobre palabras aisladas, la confianza del sistema de desambiguación, el impacto de la granularidad de los sentidos y el poder discriminativo (IDF) de la palabra desambiguada. A continuación se comentan los experimentos llevados a cabo con cada uno de estos indicadores.

Impacto de la desambiguación sobre palabras aisladas. En este experimento nos planteamos la hipótesis de que, dada una consulta, algunas palabras deben ser desambiguadas y otras no. Para ello, realizamos un análisis en profundidad de la mejora obtenida para cada palabra desambiguada, intentando caracterizar qué términos realmente debían ser desambiguados. Como algunos términos estaban representados por su identificador o código de *synset* se creó un nuevo índice *mixto* uniendo palabras y sus respectivos *synset* (*TS index*). Para las consultas se tuvo en cuenta que la combinación *lema + synset* era mutuamente exclusiva: cada palabra de la consulta original fue reemplazada por un lema o un *synset*, pero no por ambos. Atendiendo a la parte sustituida, se propusieron dos conjuntos de consultas:

- *nT_1S*: en el que solo un término es reemplazado por su *synset* correspondiente, y así para cada palabra de la consulta. Por tanto, para una consulta de *m* palabras se obtienen *m* consultas.
- *nS_1T*: en el que solo un *synset* es sustituido por su término correspondiente, y así para cada palabra de la consulta. De nuevo, para una consulta formada por *m* *synsets* se obtienen *m* consultas.

En este experimento se trató de caracterizar cuándo la desambiguación mejoraba la recuperación observando las diferencias entre las consultas *nT_1S* y el caso base por un lado, y las consultas *nS_1T* y los *synset* por otro. Para ambos conjuntos se generaron un total de 1.113 consultas. La Tabla 2 muestra los resultados obtenidos al reemplazar un término por su correspondiente *synset* y viceversa. Estos resultados son totalmente compatibles con los mostrados en la Tabla 1, ya que no hay ninguna variación importante en las sustituciones. De todos modos, se puede destacar un resultado: el 30-32% de sustituciones no tuvieron ningún impacto sobre el resultado final. Esto se debió a que el 45,2%

de las palabras de las consultas solo utilizaban un sentido en la colección. Nótese que esto no conlleva necesariamente que se tratara de palabras monosémicas, sino que, de todos sus posibles sentidos, sólo uno estaba presente en la colección utilizada.

Confianza del sistema de desambiguación. La confianza en el algoritmo de desambiguación debe tenerse en cuenta. Una correcta desambiguación no asegura mejores precisiones, pero sí es cierto que desambiguaciones incorrectas aseguran una mala precisión en el sistema. La Tabla 3 muestra un resultado interesante: la confianza media del algoritmo de desambiguación cuando ésta mejora el caso base (sin desambiguar las palabras) es muy similar a la confianza del sistema cuando la desambiguación obtiene peores resultados para el caso base. En resumen, una confianza alta en el sistema de desambiguación es una condición necesaria pero no suficiente.

Impacto de la granularidad de los sentidos. La granularidad tan fina que emplea WordNet es una de las habituales críticas a esta base de datos léxica. Aunque existen trabajos en los que se han estudiado varios enfoques para aplicar agrupamiento de sentidos (*clustering*) en WordNet, las conclusiones no aportan una estrategia claramente validada por los resultados (Agirre y Lacalle, 2003). En este experimento se ha calculado el impacto de la granularidad de WordNet teniendo en cuenta el número medio de sentidos para cada grupo de sustituciones de un *synset* con su lema. Los resultados mostrados en la Tabla 4 son decepcionantes, ya que no existen diferencias entre el número medio de sentidos para cada grupo de consultas. Además, no se observa una tendencia clara, como algunos trabajos sugieren, de que al tener más sentidos se debe aplicar WSD. Por tanto, se puede concluir que el número de sentidos no es una cuestión relevante, al menos cuando se trabaja de una manera aislada.

Poder discriminativo de la palabra desambiguada. El último aspecto que investigamos fue la diferencia del poder discriminativo (IDF) entre la palabra y su correspondiente *synset*. El IDF es bastante utilizado para medir la importancia de una palabra en una determinada colección (Papineni, 2001), de manera que si un documento contiene una palabra poco usual, dicha palabra podría representar el significado de ese docu-

Conjunto de consultas	#consultas	Mejora media
Prec(nT_1S)=Prec(baseline)	360	0
Prec(nT_1S)>Prec(baseline)	403	+0,042
Prec(nT_1S)<Prec(baseline)	350	-0,046
Prec(nS_1T)=Prec(baseline)	343	0
Prec(nS_1T)>Prec(baseline)	389	+0,034
Prec(nS_1T)<Prec(baseline)	381	-0,030

Tabla 2: Experimentación a nivel de palabra, sustituyendo cada lema por su correspondiente *synset* y viceversa

Conjunto de consultas	consultas	#términos desamb.	Confianza media del sistema WSD
Prec(nT_1S) - Prec(baseline) $\geq 0,05$	71	71	0,612
Prec(nT_1S) - Prec(baseline) $\leq -0,05$	75	75	0,605
Prec(nT_1S) - Prec(baseline) $\geq 0,01$	39	39	0,675
Prec(nT_1S) - Prec(baseline) $\leq -0,01$	44	44	0,667
Baseline	160	1701	0,695

Tabla 3: Confianza media del sistema de desambiguación atendiendo a varios subconjuntos de consultas

Conjunto de consultas	Nº medio sentidos
Prec(nT_1S) - Prec(baseline) $\geq 0,05$	4,3
Prec(nT_1S) - Prec(baseline) $\leq -0,05$	4,5
Prec(nT_1S) - Prec(baseline) $\geq 0,01$	4,1
Prec(nT_1S) - Prec(baseline) $\leq -0,01$	5
Prec(nS_1T) - Prec(<i>synset</i>) $\geq 0,05$	4,5
Prec(nS_1T) - Prec(<i>synset</i>) $\leq -0,05$	5,6
Prec(nS_1T) - Prec(<i>synset</i>) $\geq 0,01$	4,1
Prec(nS_1T) - Prec(<i>synset</i>) $\leq -0,01$	5,1

Tabla 4: Número medio de sentidos por *synset* sustituido

mento mejor que aquellas palabras que son más frecuentes en la colección. La siguiente fórmula representa el cálculo del valor IDF:

$$idf_i = \lg \frac{|D|}{|d : t_i \in d|} \quad (2)$$

donde

- $|D|$ es el número total de documentos en la colección.
- $|d : t_i \in d|$ es el número de documentos donde aparece el término t_i .

De una manera similar al número de sentidos, investigamos también el impacto del IDF. (Krovetz y Croft, 1992) mostraron que los sentidos de una palabra no aparecen todos por igual y que a menudo un único sentido de una palabra se utiliza de forma más frecuente que los otros. La hipótesis que planteamos en este experimento fue que los sentidos que fueran muy poco usuales no deberían ser desam-

biguados. La Tabla 5 muestra los resultados obtenidos para este experimento.

Una primera cuestión a analizar con respecto a estos resultados es que el IDF medio de los *synsets* es más alto que el IDF medio de los lemas. Dado que al desambiguar se le asigna un token a cada sentido, esto indica que es más frecuente la polisemia que la sinonimia. Este resultado es consistente con el mostrado en el trabajo de Krovetz y Croft. Por lo tanto, un primer resultado de la desambiguación es que las palabras (lemas) con un valor bajo de IDF aumentan el IDF desambiguando dicha palabra. Esto no es necesariamente un problema porque es usual que palabras comunes sean altamente polisémicas, pero facilita que tales palabras vean incrementado su IDF artificialmente. Este resultado sugiere que cuando la diferencia entre el IDF de la palabra y el IDF del *synset* es muy alta, no debería aplicarse la desambiguación.

Conjunto de consultas	IDF medio del lema	IDF medio del <i>synset</i>	Diferencia
Prec(nT_1S) - Prec(baseline) $\geq 0,05$	3,4	4,3	0,9
Prec(nT_1S) - Prec(baseline) $\leq -0,05$	3,2	4,4	1,2
Prec(nT_1S) - Prec(baseline) $\geq 0,1$	3,3	4,2	0,9
Prec(nT_1S) - Prec(baseline) $\leq -0,1$	3,4	4,7	1,3
Prec(nS_1T) - Prec(synsets) $\geq 0,05$	3,1	4,4	1,3
Prec(nS_1T) - Prec(synsets) $\leq -0,05$	3,1	4,3	1,2
Prec(nS_1T) - Prec(synsets) $\geq 0,1$	3,0	4,2	1,2
Prec(nS_1T) - Prec(synsets) $\leq -0,1$	3,4	4,3	1,1

Tabla 5: Diferencia media entre el IDF de un lema y su correspondiente *synset* y viceversa

Experimento	WSD selectiva	% de palabras desambiguadas	MAP
<i>Baseline</i>	no	0 %	0,374
WSD-S	no	100 %	0,376
WSD-S selectiva 1	Sí, Desambiguados aquellos términos con confianza $> 0,826$	44,2 %	0,379
WSD-S selectiva 2	Sí, Solo términos con diferencia IDF $< 1,3$	80,6 %	0,370
WSD-S selectiva 3	Sí, Solo términos con confianza $> 0,826$ y diferencia IDF $< 1,3$	38,3 %	0,40

Tabla 6: Aplicación selectiva de la desambiguación para el conjunto completo de consultas (160 consultas)

4.3. Análisis de una desambiguación selectiva en IR

Para aplicar las conclusiones obtenidas en los experimentos anteriores, se propuso utilizar una desambiguación selectiva de los términos para cada consulta original. Este algoritmo tuvo en cuenta dos factores: la confianza en el algoritmo de desambiguación y la diferencia entre el IDF del *synset* y el IDF del lema. Los resultados obtenidos se muestran en la Tabla 6. Como se puede comprobar, es necesario aplicar ambos factores para conseguir una mejora del 7 % con respecto al caso base. Esta diferencia en la mejora sugiere que el enfoque propuesto de aplicar una desambiguación selectiva es útil para algunos casos, aquellos que se pueden caracterizar teniendo en cuenta la confianza del algoritmo de desambiguación y el IDF.

5. Conclusiones y trabajo futuro

En este trabajo se ha desarrollado un estudio exhaustivo sobre el impacto de la desambiguación de palabras en recuperación de información. La idea más habitual sobre IR y WSD es que la desambiguación es únicamente útil cuando el algoritmo WSD es casi per-

fecto. Sin embargo, los resultados obtenidos con los experimentos realizados muestran que una perfecta desambiguación de los términos de nuestra colección y consultas no obtiene una mejora en los resultados.

Otro factor a tener en cuenta es la granularidad del algoritmo WSD. El número de sentidos de un término es una buena medida de la granularidad de ese término. Para estudiar el impacto de la granularidad se generaron consultas mixtas, es decir, consultas no desambiguadas donde solo un término estaba desambiguado. Después, se comparó el rendimiento de la consulta original y la consulta con los términos sustituidos. El mismo estudio se realizó para las consultas completamente desambiguadas sustituyendo un *synset* por su palabra original. La conclusión fue que el número de sentidos no discrimina las peores y las mejores sustituciones.

Por último, se estudió el impacto del IDF, obteniendo como conclusión que si el IDF del *synset* es mucho más alto que el IDF del término, es muy probable que la desambiguación de dicho término decremente el rendimiento de la consulta. Además, comprobamos un algoritmo sencillo para llevar a cabo una desambiguación selectiva, de manera que

solo se desambiguaban aquellas palabras con una confianza muy alta del algoritmo WSD y un valor similar de IDF con el *synset*. Este último experimento consiguió mejorar un 7% el rendimiento obtenido con el caso base sin aplicar desambiguación.

Como trabajo futuro, sería interesante analizar el impacto de la polisemia y de la granularidad de una manera más detallada, construyendo *clusters* jerárquicos de sentidos. Los enfoques habituales basados en *clustering* de sentidos no funcionan bien debido a que los *clusters* son conjuntos cerrados y todos los elementos pertenecen al conjunto de la misma forma. Por tanto, se podría generar un conjunto difuso de sentidos para cada palabra atendiendo a factores que se han investigado en este trabajo: confianza en el algoritmo WSD, polisemia e IDF. Finalmente, también se podría aplicar nuestro enfoque de desambiguación selectiva a otras tareas de Procesamiento del Lenguaje Natural como los sistemas de búsqueda de respuestas y el reconocimiento de la implicación textual.

Bibliografía

- Agirre, E. y P.G. Edmonds. 2006. *Word Sense Disambiguation: Algorithms and Applications*. Springer.
- Agirre, E. y O.L. Lacalle. 2003. Clustering wordnet word senses. En *RANLP*, volumen 260 de *Current Issues in Linguistic Theory (CILT)*, páginas 121–130.
- Agirre, E. y O.L. Lacalle. 2007. Ubc-alm: combining k-nn with svd for wsd. En *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, páginas 342–345, Morristown, NJ, USA. Association for Computational Linguistics.
- Agirre, E., G.M. Di Nunzio, N. Ferro, T. Mandl, y C. Peters. 2008. Clef 2008: Ad hoc track overview. En *CLEF*, volumen 5706 de *Lecture Notes in Computer Science*, páginas 15–37. Springer.
- Agirre, E., G.M. Di Nunzio, T. Mandl, y A. Otegi. 2009. Clef 2009 ad hoc track overview: Robust-wsd task. En *CLEF*, volumen 6241 de *Lecture Notes in Computer Science*, páginas 36–49. Springer.
- Chan, Y.S., H. Tou Ng, y Z. Zhong. 2007. Nus-pt: Exploiting parallel texts for word sense disambiguation in the english all-words tasks. En *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval 2007)*, páginas 253–256.
- Gale, W., K.W. Church, y D. Yarowsky. 1992. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. En *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, páginas 249–256. ACL.
- Gonzalo, J., F. Verdejo, I. Chugur, y J. Cigarran. 1998. Indexing with WordNet synsets can improve Text Retrieval. En *Proceedings of the COLING/ACL '98 Workshop on Usage of WordNet for NLP*, páginas 38–44.
- Krovetz, R. y W.B. Croft. 1992. Lexical ambiguity and information retrieval. *ACM Trans. Inf. Syst.*, 10(2):115–141.
- Martínez-Santiago, F., J.M. Perea-Ortega, y M.A. García-Cumbreras. 2008. Evaluating word sense disambiguation tools for information retrieval task. En *CLEF*, volumen 5706 de *Lecture Notes in Computer Science*, páginas 113–117. Springer.
- Papineni, K. 2001. Why inverse document frequency? En *NAACL '01: Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, páginas 1–8. ACL.
- Porter, M. F. 1980. An algorithm for suffix stripping. *Program*, 14(3):130–137.
- Sanderson, M. 2000. Retrieving with good sense. *Information Retrieval*, 2(1):45–65.
- Ureña-López, L.A., M. Buenaga-Rodríguez, y J.M. Gómez. 2001. Integrating Linguistic Resources in TC through WSD. *Computers and the Humanities*, 35(2):215–230.
- Voorhees, E.M. 1993. Using wordnet to disambiguate word sense for text retrieval. En *Proceedings of ACM SIGIR Conference*, páginas 171–180.
- Wallis, P. 1993. Information retrieval based on paraphrase. En *Proceedings of PACLING Conference*.
- Weiss, S.F. 1973. Learning to disambiguate. *Information Storage and Retrieval*, 9(1):33–41.