

Búsqueda de respuestas multilingüe: Clasificación de preguntas en español basada en aprendizaje.*

Miguel Ángel García Cumbreiras
Fernando Martínez Santiago

L. Alfonso Ureña López
Arturo Montejo Raez

Grupo Sistemas Inteligentes de Acceso a la Información
Departamento de Informática
Universidad de Jaén
<http://sinai.ujaen.es>
e-mail: {magc, laurena, dofer, amontejo}@ujaen.es

Resumen: Este artículo presenta un sistema de clasificación de preguntas en español basado en aprendizaje, utilizando colecciones en inglés y diversas *máquinas de traducción* online como recursos de traducción de la pregunta original en español al inglés. En este estudio se mide de forma cuantitativa la bondad de cuatro traductores automáticos para la pareja de idiomas español-inglés, comparando los resultados obtenidos para las preguntas originales en inglés con los obtenidos de cada una de las traducciones. El sistema se ha implementado de forma modular utilizando varios métodos de aprendizaje tales como LibSVM, Bayesian Logistic Regression o PLAUM. En la tarea de clasificación de preguntas se demuestra que la pérdida de precisión debida a la traducción automática es moderada, situándose entorno a un 5 %.

Palabras clave: Clasificación de Preguntas, Sistemas de Búsqueda de Respuestas, aprendizaje automático, traductores automáticos

Abstract: This paper presents an Spanish question classification system based on machine learning, that uses English collections, different online machine translators and other NLP English resources. The original Spanish questions are translated into English. Four machine translators are evaluated in terms of precision and the results are compared with the result obtained by using original English questions. Our system has been developed into separated modules and we have tested several machine learning methods, such as LibSVM, Bayesian Logistic Regression or PLAUM. The obtained results show that these online machine translators, used for the language pair Spanish-English, and for the query translation task in a multilingual question answering system, work well. It is showed that the loss of precision because of the machine translation, in a question classification task, is reasonable, around 5 %.

Keywords: Question Classification, Question Answering Systems, machine learning, machine translation

1. Introducción

La cantidad de información digital ha experimentado un fuerte crecimiento, así como el número de usuarios finales que a través de ordenadores personales interactúan con esta información. Esto

implica el creciente interés por los sistemas de recuperación de información multilingüe (CLIR, del inglés Cross Language Information Retrieval) así como por los sistemas de búsqueda de respuestas (BR) tanto monolingües como multilingües.

Un sistema CLIR es un sistema de recuperación de información que tiene capacidad para operar sobre una colección de documentos o pasajes multilingüe, esto

* Este trabajo ha sido financiado por el Ministerio de Ciencia y Tecnología mediante el proyecto TIC2003-07158-C04-04

es, un sistema capaz de recuperar todos los documentos o pasajes relevantes que se encuentran en la colección, independientemente del idioma utilizado tanto en la consulta como en los propios documentos o pasajes (Grefenstette, 1998).

La búsqueda de respuestas se puede definir como el proceso automático que realizan los ordenadores para encontrar respuestas concretas a preguntas precisas formuladas por los usuarios. Los sistemas de BR no sólo localizan los documentos o pasajes relevantes sino que también encuentran, extraen y muestran la respuesta al usuario final, evitándole la búsqueda o la lectura de la información relevante para encontrar de forma manual la respuesta final.

La primera tarea de un sistema de búsqueda de respuestas consiste en procesar la pregunta y determinar por lo que se está preguntando. En muchos casos esto implica simplemente tomar las palabras o características adecuadas y determinar el tipo de la pregunta. Por ejemplo en la pregunta "*¿Quién fue el primer presidente en España?*" estamos buscando el nombre de una persona como tipo de la pregunta. Una vez analizada la información de las colecciones donde se busca la respuesta es fundamental conocer el tipo de la pregunta, para dar con la respuesta adecuada. Además, las conferencias actuales para la evaluación de sistemas de BR, tales como TREC QA (Voorhees, 2001) o CLEF (Peters et al., 2003), ya restringen el tamaño de la respuesta esperada a un máximo de 50 bytes, lo que implica una mayor complejidad a la hora de encontrar la respuesta adecuada en un texto relevante.

Es, por lo tanto, la clasificación de preguntas (QC, del inglés Question Classification) una tarea fundamental para los sistemas de búsqueda de respuestas.

Los sistemas QC tienen unas limitaciones (Hacioglu y Ward, 2003), entre las que cabe destacar las siguientes:

- La clasificación de preguntas en QA tradicionalmente viene realizándose mediante un juego de reglas, como por ejemplo "las preguntas que empiezan por *Who* son de tipo persona". Estas reglas se escriben manualmente, lo que implica que se tenga que revisar cada caso distinto para mejorar los

resultados.

- Las reglas son muy frágiles, ya que cuando aparecen nuevas preguntas el sistema no está preparado para determinar su tipo.
- Cada vez que utilizamos un tipo de preguntas distinto las reglas tienen que ser revisadas y en algunos casos habrá que escribirlas de nuevo.

En cuanto al idioma la mayoría de las investigaciones previas realizadas sobre clasificación de preguntas se han centrado casi exclusivamente en inglés, dado también que la mayor parte de los recursos útiles están disponibles para este idioma. Es por esta razón por la que hemos utilizado el inglés como idioma pivote.

Los sistemas de traducción automática (MT, del inglés Machine Translation) son sistemas muy apreciados en CLIR (McNamee, Mayfield, y Piatko, 2000). Por un lado estos sistemas están en constante evolución y perfeccionamiento pero por otro lado no para todos los idiomas hay disponibilidad de los mismos y además la calidad de traducción varía mucho según el par de idiomas utilizado.

Nosotros hemos desarrollado un sistema de clasificación de preguntas que intenta solventar principalmente dos puntos:

1. Por un lado se trata de un sistema de QC basado en aprendizaje, en el cual no hay ninguna regla manual definida. Hemos utilizado como sistemas de aprendizaje LibSVM, BBR y Plaum. Además, el uso de técnicas basadas en aprendizaje automático facilitan la aplicación del modelo a otros idiomas y/o tipos de preguntas.
2. Por otro lado, y dada la dificultad de encontrar recursos útiles para idiomas distintos del inglés, nuestro sistema utiliza varios traductores online para traducir la pregunta del español al inglés y a partir de ahí siempre utiliza el inglés como idioma. En este punto queremos comprobar la bondad de varios sistemas de traducción automática online en comparación con el uso de las preguntas *escritas* directamente en inglés.

2. Clasificación de preguntas.

Una importante tarea en los Sistemas de Búsqueda de Respuestas es el análisis de la pregunta, ya que proporciona la información necesaria para localizar la respuesta adecuada en el texto seleccionado de la colección.

La clasificación de preguntas es la asignación de clases semánticas a las preguntas, lo que posibilita la búsqueda posterior de forma más restringida. Los errores de algunos sistemas de búsqueda de respuestas han sido generados por un mal clasificador de preguntas. Esto ha generado un interés creciente por el desarrollo de buenos clasificadores de preguntas. Sin embargo, la mayoría de ellos son para inglés.

Muchas de los sistemas de clasificación de preguntas actuales están centrados en el uso de expresiones regulares y reglas gramaticales generadas de forma manual (Durme et al., 2003).

Sistemas más recientes han utilizado diversos métodos de aprendizaje. (Zhang y Lee, 2003) proponen un sistema de clasificación de preguntas utilizando Support Vector Machines (SVM) como el mejor método de aprendizaje, comparando los resultados obtenidos con Nearest Neighbors, Naive Bayes, Decision Tree y Sparse Network of Winnows (SNOW). Obtienen un buen resultado utilizando como conjunto de entrenamiento 21500 preguntas etiquetadas manualmente y como conjunto de prueba 1000 preguntas también etiquetadas manualmente.

(Li y Roth, 2002) proponen un sistema QC basado en la arquitectura de aprendizaje SnoW, discriminando en un primer paso entre 5 categorías generales de preguntas, y en un segundo paso en 50 subcategorías de las generales. Obtienen categorías léxicas, sintácticas y semánticas.

Otras aproximaciones han utilizado un kernel propio para SVM para obtener mejores resultados, partiendo de la clasificación ya citada de Li y Roth o utilizando una propia.

Nosotros presentamos un estudio sobre la robustez de estos modelos basados en aprendizaje frente a preguntas que han sido elaboradas en un idioma diferente al sistema de QA. Por ello, se impone el uso de máquinas de traducción para superar la barrera lingüística que se plantea entre el sistema y el usuario. En concreto se trata

de clasificar preguntas formuladas en español utilizando para ello un sistema QC basado en aprendizaje para inglés.

2.1. Descripción de la tarea.

Hemos planteado nuestro sistema de clasificación como tres módulos independientes, de forma que fácilmente se pueda sustituir un módulo por otro para comprobar la bondad tanto de los diferentes métodos de aprendizaje como de los traductores automáticos online utilizados.

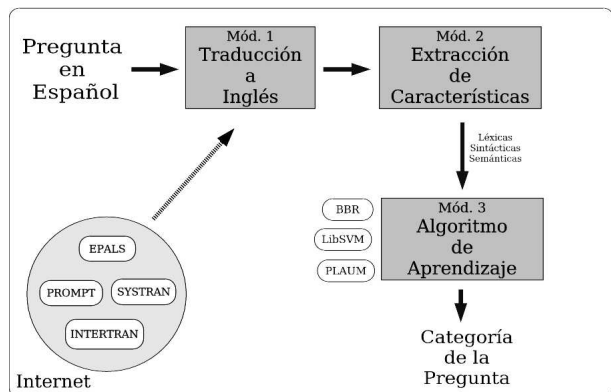


Figura 1: Módulos del sistema QC.

El primer módulo toma la pregunta y la traduce a los idiomas contemplados. En nuestro caso la entrada al sistema es en español y se traduce al inglés utilizando cuatro recursos de traducción automática online: Epals, Prompt, Systran y Wordlingo. Este módulo es fácilmente ampliable a otros idiomas siendo sólo necesario disponer de un traductor para cada par de idiomas. Uno de nuestros objetivos era probar la bondad de dichos traductores para la tarea de QC como un primer paso para un sistema de BR multilingüe.

Tras este módulo ya tenemos las preguntas en inglés. El siguiente módulo se encarga de extraer diversas características de estas preguntas, que serán utilizadas tanto para la tarea de QC como para reconocer partes importantes de la pregunta que serían utilizadas posteriormente en un sistema completo de BR. En nuestro caso hemos analizado y obtenido de cada pregunta diversos conjuntos de características de carácter léxico, sintáctico y semántico y utilizando en todo momento recursos disponibles para inglés:

- Características Léxicas:

1. Las dos primeras palabras de la pregunta.
 2. Todas las palabras de la pregunta en minúscula.
 3. Las raíces de todas las palabras (stemming).
 4. Los lemas de todas las palabras.
 5. Bigramas de la pregunta.
 6. Cada palabra junto con el orden que ocupa dentro de la pregunta.
 7. El adverbio interrogativo de la pregunta.
 8. Los lemas sólo de los nombres y verbos.
 9. La primera subcadena de la pregunta donde aparezca un verbo (primer sintagma verbal).
- Características Sintácticas:
 1. El adverbio interrogativo junto con el Part Of Speech (POS) del resto de palabras.
 2. Los Part Of Speech (POS) de todas las palabras.
 3. Las partes en las que un analizador sintáctico parcial divide la pregunta (chunking).
 4. La longitud de la pregunta.
 - Características Semánticas:
 1. El foco de la pregunta.
 2. Utilizar los POS junto con el tipo de entidad en aquellas entidades reconocidas.
 3. Si el foco de la pregunta es una entidad se utiliza el tipo de la misma si ha sido reconocido.
 4. Hiperónimos de WordNet para los nombres y Sinónimos de WordNet para los verbos.

En esta fase de preproceso de la pregunta en inglés hemos utilizado diversos recursos tal como TreeTagger (Schmid, 1994) para obtener el POS y el lema de las palabras, Lingpipe para hacer reconocimiento de entidades ¹, y el Stemmer de Porter (Porter, 1980). También hemos utilizado la base de

¹LingPipe, un software libre para Procesamiento de Lenguaje Natural, desarrollado por by Alias-I. (<http://www.alias-i.com/lingpipe/>)

datos léxicos WordNet con el fin de enriquecer ciertas palabras (hiperónimos de WordNet para los nombres y sinónimos de WordNet para los verbos).

El último módulo de nuestro sistema QC son los métodos de aprendizaje. Hemos obtenido resultados utilizando tres métodos:

- Library for Support Vector Machines o LibSVM. Se trata de una implementación de la Support Vector Machine de Vapnik (Vapnik, 1995). SVM utiliza propiedades geométricas para calcular el hiperplano que de forma óptima separa los ejemplos de entrenamiento (Stitson et al., 1996). Es un software integrado para support vector classification (C-SVM), regresión y estimación de distribuciones. Soporta multclasificación. ²
- Bayesian Logistic Regression o BBR. Se trata de una implementación de la regresión logística Bayesiana que contempla el uso de dos tipos de funciones: Gausiana y Laplace. ³
- Perceptron learning algorithm with uneven margins o Plaum. Se trata de otro clasificador lineal muy cercano a SVM (Li et al., 2002) (Robertson, Walker, y Zaragoza, 2001). Tal como SVM se basa en la idea de encontrar un margen y sus autores aseguran que funciona mejor que SVM para tareas de clasificación de texto. La clasificación de hace de forma similar que con SVM:

$$\hat{\Phi}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w} + b \quad (1)$$

Pero la forma en que \mathbf{w} y \mathbf{b} son calculados es muy simple lo que produce que sea muy fácil su implementación comparado con las necesidades de cálculo numérico que necesita un algoritmo SVM. El algoritmo se muestra en la figura 2.

3. Experimentos y Resultados

3.1. Metodología de Experimentación

Los experimentos se han realizado utilizando como colecciones de entrenamiento

²(<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)

³(<http://www.stat.rutgers.edu/~madigan/BBR/>)

Require:
 n linearly separable training samples X_t
 A learning rate $\eta \in \mathbb{R}^+$
 A maximum epochs parameter T
 Two margin parameters $\tau_{+1}, \tau_{-1} \in \mathbb{R}$

Algorithm:
 epoch $\leftarrow 0$; $i \leftarrow 1$; update $\leftarrow \text{m}$
 $\mathbf{w} \leftarrow \vec{0}$; $b \leftarrow 0$; $R \leftarrow \max_{\mathbf{x}_i \in \mathbf{X}} |\mathbf{x}_i|$
repeat
 if $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \leq \tau_{y_i}$ **then**
 $\mathbf{w} \leftarrow \mathbf{w} + \eta y_i \mathbf{x}_i$
 $b \leftarrow b + \eta y_i R^2$
 updated $\leftarrow i$
 end if
 $i \leftarrow i + 1$
 if ($i > n$) **then**
until ($i = \text{updated}$) **or** ($\text{epoch} \geq T$)
return(\mathbf{w}, b)

Figura 2: Algoritmo de aprendizaje PLAUM

y prueba los conjuntos de datos disponibles de forma pública que facilitan el USC (Hovy et al., 1999), UIUC y TREC ⁴.

Estos conjuntos de datos han sido etiquetado manualmente por el UIUC de acuerdo con las categorías generales y detalladas definidas a continuación.

ABBR: abbreviation, expansion.

DESC: definition, description, manner, reason.

ENTY: animal, body, color, creation, currency, disease/medical, event, food, instrument, language, letter, other, plant, product, religion, sport, substance, symbol, technique, term, vehicle, word.

HUM: description, group, individual, title.

LOC: city, country, mountain, other, state.

NUM: code, count, date, distance, money, order, other, percent, period, speed, temperature, size, weight.

Los datos de entrenamiento los conforman 5500 preguntas mientras que el conjunto de prueba contiene 500 preguntas etiquetadas de la conferencia TREC-10 QA.

La distribución de las 5500 preguntas del fichero de entrenamiento, de acuerdo con el adverbio interrogativo o el principio de la consulta se puede ver en el cuadro 1.

Igualmente la distribución de las categorías de las pregunta para este fichero de entrenamiento se puede observar en el cuadro 2.

Cuadro 1: Distribución de preguntas del fichero de Entrenamiento de acuerdo con el Adverbio interrogativo.

Tipo	Número
What	3242
Who	577
How	764
Where	273
When	131
Which	105
Why	103
Name	91
In	67
Define	4
Whom	4
Otras	91

Cuadro 2: Distribución de preguntas del fichero de Entrenamiento de acuerdo con la Categoría General de la pregunta.

Categoría	Número
ABBR	86
DESC	1162
ENTY	1251
HUM	1223
LOC	835
NUM	896

De igual forma la distribución de las 500 preguntas del fichero de prueba, de acuerdo con el adverbio interrogativo o el principio de la consulta se puede ver en el cuadro 3, y la distribución de las categorías se puede observar en el cuadro 4.

En un primer intento nuestro sistema intenta clasificar las preguntas en una de estas seis categorías generales, dejando la clasificación detallada para un trabajo posterior. Nótese que el principal objetivo de este estudio es evaluar de alguna forma cuantitativa la traducción automática online y la pérdida que genera en esta primera fase de un sistema de BR multilingüe.

El rendimiento del sistema de QC lo hemos medido utilizando como medidas la exactitud global (Accuracy) de forma general y la precisión para cada categoría o clase.

⁴(<http://trec.nist.gov/>)

Cuadro 3: Distribución de preguntas del fichero de Prueba de acuerdo con el Adverbio interrogativo.

Tipo	Número
What	343
Who	47
How	35
Where	26
When	26
Which	6
Why	4
Name	2
In	5
Otras	6

Cuadro 4: Distribución de preguntas del fichero de Prueba de acuerdo con la Categoría General de la pregunta.

Categoría	Número
ABBR	9
DESC	138
ENTY	94
HUM	65
LOC	81
NUM	113

$$Accuracy = \frac{\#deprediccionescorrectas}{\#depredicciones} \quad (2)$$

$$precision(c) = \frac{\#deprediccionescorrectasdelaclass}{\#deprediccionesdelaclass} \quad (3)$$

3.2. Resultados obtenidos.

Hemos realizado diversos experimentos utilizando varios métodos de aprendizaje y para cada uno de ellos hemos utilizado cinco conjuntos de entrenamiento y prueba, como se describen a continuación:

- 5500 preguntas de entrenamiento en inglés y 500 preguntas de test también en inglés. Este es el caso base con el que comparamos el resto de resultados obtenidos.
 - 5500 preguntas de entrenamiento en inglés y 500 preguntas de test traducidas a inglés mediante el recurso de MT *Epals*.
 - 5500 preguntas de entrenamiento en inglés y 500 preguntas de test traducidas a inglés mediante el recurso de MT *Prompt*.
 - 5500 preguntas de entrenamiento en inglés y 500 preguntas de test traducidas a inglés mediante el recurso de MT *Systran*.
 - 5500 preguntas de entrenamiento en inglés y 500 preguntas de test traducidas a inglés mediante el recurso de MT *Wordlingo*.
- Los recursos de MT online están disponibles en las siguientes urls:
- *EPALS*
<http://www.epals.com>
 - *PROMPT*
<http://translation2.paralink.com>
 - *SYSTRAN*
<http://www.systransoft.com>
 - *WORDLINGO*
<http://www.worldlingo.com>
- Atendiendo a las características léxicas, sintácticas y semánticas, se han realizado siete grupos de características, con la finalidad de explorar cuales resultan más útiles para la tarea:
- Características léxicas: El adverbio interrogativo de la pregunta.
 - Características léxicas: Las dos primeras palabras de la pregunta + Todas las palabras de la pregunta en minúscula + Las raíces de todas las palabras (stemming) + Los lemas de todas las palabras
 - Características léxicas: las cuatro del punto anterior + Cada palabra junto con el orden que ocupa dentro de la pregunta + El adverbio interrogativo de la pregunta + Los lemas sólo de los nombres y verbos + La primera subcadena de la pregunta donde aparezca un verbo
 - Características sintácticas: las cuatro definidas en la sección anterior.
 - Características semánticas: las tres primeras definidas en la sección anterior
 - Las siete léxicas + las cuatro sintácticas + las tres semánticas

7. Características léxicas: Las dos primeras palabras de la pregunta + El adverbio interrogativo de la pregunta + Los lemas sólo de los nombres y verbos + La primera subcadena de la pregunta donde aparezca un verbo; Características sintácticas: El adverbio interrogativo junto con el Part Of Speech (POS) del resto de palabras + La longitud de la pregunta; Características semánticas: Utilizar los POS junto con el tipo de entidad en aquellas entidades reconocidas

Podemos ver en los cuadros 5, 6, 7 y 8 una muestra de los resultados más relevantes obtenidos, combinando los tres factores mencionados: algoritmo de aprendizaje, conjunto de características y máquina de traducción.

Nótese que el tamaño del conjunto de entrenamiento y prueba utilizado no es comparable al utilizado en otros artículos (21500 preguntas de entrenamiento frente a las 5500 disponibles en el momento de realizar estos experimentos. En cualquier caso, el interés principal de este artículo no es tanto el desarrollo de un nuevo modelo de clasificación de preguntas sino el estudio de la incidencia de la traducción automática en sistemas QC basados en aprendizaje.

De los resultados obtenidos comparados obtenemos los siguientes datos:

1. Siempre el fichero original de TREC-10 en inglés ha obtenido los mejores resultados, superando en todos los métodos de aprendizaje utilizados el 50% de exactitud (accuracy) cuando se toma la primera categoría. En el caso de tomar la segunda categoría con mejor resultado cuando la diferencia del valor de ranking entre la primera y la segunda es muy baja, el resultado supera el 65% también en términos de accuracy.
2. El método de aprendizaje que mejores resultados ha dado ha sido LibSVM, teniendo también en cuenta que el tiempo de entrenamiento y de prueba no es nada elevado, apenas unos segundos por experimento.
3. Respecto al mejor caso, con las preguntas originales en inglés, las preguntas traducidas con Epals tiene un decremento medio de accuracy en

los resultados de un 4%. En el caso de las preguntas traducidas por Prompt la pérdida es similar, sobre un 4-5%. El tercer caso de prueba, las preguntas traducidas haciendo uso de Systran, la pérdida sube un poco hasta un 7% en algunos casos. Por último, las preguntas traducidas con Wordlingo llegan hasta el 9% y de media los resultados obtenidos son peores que los que nos dan el resto de traducciones.

4. Por lo tanto, la pérdida en términos de exactitud (accuracy) no es demasiado grande, no superando en ningún caso el 10%, lo que indica que para este tipo de sistemas de BR multilingües, donde lo que se necesita es una buena traducción de las preguntas, el uso de los actuales traductores automáticos no implica un empeoramiento de los resultados. Esto nos lleva a pensar si merece la pena desarrollar en un futuro cercano un sistema multilingüe donde se utilicen recursos para cada idioma o por otro lado utilizar estos traductores y utilizar como base del sistema el idioma inglés.
5. De los cuatro traductores utilizados los que mejor resultado han dado son Epals y Prompt, quedándose Systran a poca distancia y Wordlingo un poco más separado del resto. Este comportamiento ya lo conocíamos por un estudio previo de varios traductores, pero hasta ahora no lo habíamos cuantificado.
6. De los siete casos de prueba que hemos creado, como era previsible, los mejores resultados en todos los casos los obtenemos con el último caso de prueba, que combina características de los tres tipos, con una diferencia en términos de exactitud (accuracy) más que notable respecto, por ejemplo, del primer caso donde sólo se utiliza el adverbio interrogativo de la pregunta como característica relevante.

4. Conclusiones y Trabajo Futuro

Los sistemas de Búsqueda de Respuestas Multilingües están abriendo un nuevo campo de investigación, donde el uso de un Clasificador de Preguntas es imprescindible tanto para saber por lo que se está preguntando, y poder así guiar la

Cuadro 5: Conjunto de Características 1 (sólo el Adv. Interrogativo). Valores representados en términos de exactitud (accuracy).

	Inglés original	Epals	Prompt	Systran	Wordlingo
LibSVM	0,232	0,234	0,236	0,236	0,228
Plaum	0,370	0,370	0,370	0,370	0,366
BBR	0,130	0,130	0,130	0,130	0,130

Cuadro 6: Conjunto de Características 4 (sólo carac. sintácticas). Valores representados en términos de exactitud (accuracy).

	Inglés original	Epals	Prompt	Systran	Wordlingo
LibSVM	0,492	0,460	0,474	0,462	0,452
Plaum	0,368	0,364	0,368	0,364	0,362
BBR	0,328	0,340	0,318	0,300	0,292

Cuadro 7: Conjunto de Características 3 (todas las léxicas). Valores representados en términos de exactitud (accuracy).

	Inglés original	Epals	Prompt	Systran	Wordlingo
LibSVM	0,372	0,372	0,366	0,360	0,352
Plaum	0,440	0,436	0,432	0,436	0,426
BBR	0,438	0,448	0,428	0,432	0,422

Cuadro 8: Conjunto de Características 7 (combinación de características léxicas sintácticas y semánticas). Valores representados en términos de exactitud (accuracy).

	Inglés original	Epals	Prompt	Systran	Wordlingo
LibSVM	0,500	0,478	0,462	0,484	0,474
Plaum	0,392	0,358	0,384	0,368	0,366
BBR	0,256	0,190	0,238	0,210	0,204

búsqueda de la respuesta, como para obtener información útil de la pregunta analizada.

Una de las opciones para estos sistemas es utilizar los recursos más generalizados, y es el caso de los recursos para inglés. El uso de sistemas de traducción automática es entonces una tarea fundamental y era nuestro objetivo probar la bondad de los mismos en esta tarea concreta del clasificación de preguntas.

Hemos realizado una amplia experimentación tomando como base dos conjuntos de datos con preguntas etiquetadas con una característica general y otra detallada. Los distintos parámetros de la experimentación han sido el fichero de prueba utilizado (original o traducido del español), el método de aprendizaje empleado y el caso de prueba (combinaciones de las distintas características léxicas, sintácticas y semánticas extraídas de cada pregunta).

Los mejores resultados los hemos obtenidos al utilizar las preguntas originales en inglés junto con el método de aprendizaje

LibSVM (en modo de multclasificación) y sobre un caso de prueba que combina los tres tipos de características.

Analizando los resultados obtenidos hemos llegado a la conclusión que para el par de idiomas español-inglés los traductores automáticos online utilizados no suponen un empeoramiento significativo de los resultados para el clasificador de preguntas, y por extensión de un futuro sistema de búsqueda de respuestas multilingüe, en ningún caso superando un 10% de pérdida en términos de accuracy.

Como trabajo futuro tenemos pendiente la ampliación de este sistema de QC, aplicando otros métodos de aprendizaje y probando con métodos de validación cruzada, para comprobar así la bondad de las colecciones de entrenamiento y prueba utilizadas.

Otro objetivo futuro es también realizar una mejora significativa del clasificador añadiéndole reglas básicas para cada idioma contemplado y analizando los tipos de preguntas que peor resultado han dado, con

el fin de mejorar la precisión del sistema.

Bibliografía

Durme, B. Van, Y. Huang, A. Kupsc, y E. Nyberg. 2003. Towards light semantic processing for question answering. En *Proceedings of Human Language Technology conference (HLT-NAACL)*.

Grefenstette, Gregory, editor. 1998. *Cross-Language Information Retrieval*, volumen 1. Kluwer academic publishers, Boston, USA.

Hacioglu, K. y W. Ward. 2003. Question classification with support vector machines and error correcting codes. En *Proceedings of Human Language Technology conference (HLT-NAACL)*.

Hovy, E., L. Gerber, U. Hermjakob, C. Lin, y D. Ravichandran. 1999. Towards semantics-based answer pinpointing. En *Proceedings of the DARPA Human Language Technology conference (HLT)*.

Li, X. y D. Roth. 2002. Learning question classifiers. En *In COLING'02*.

Li, Y., H. Zaragoza, R. Herbrich, J. Shawe-Taylor, y J. Kandola. 2002. The perceptron algorithm with uneven margins. En *Proceedings of the International Conference of Machine Learning (ICML'2002)*.

McNamee, P., J. Mayfield, y C. Piatko. 2000. The jhu/apl haircut system at trec- 8. En *Proceedings of the Eighth Text Retrieval Conference (TREC8)*.

Peters, C., M. Braschler, J. Gonzalo, y M. Kluck. 2003. Advances in cross-language information retrieval. En *Lecture Notes in Computer Science*.

Porter, M. F. 1980. An algorithm for suffix stripping. En *Program 14*.

Robertson, S., S. Walker, y H. Zaragoza. 2001. Microsoft cambridge at trec-10: filtering and web tracks. En *Text Retrieval Conference (TREC-10)*.

Schmid, H. 1994. Probabilistic part-of-speech tagging using decision trees. En *Proceedings of International Conference on New Methods in Language Processing*.

Stitson, M. O., J. A. E. Wetson, A. Gammerman, V. Vovk, y V. Vapnik. 1996. Theory of support vector machines.

Vapnik, V. 1995. The nature of statistical learning theory.

Voorhees, E. 2001. Overview of the trec 2001 question answering track. En *Proceedings of the 10th Text Retrieval Conference (TREC-10)*.

Zhang, D. y W. Sun Lee. 2003. Question classification using support vector machines. En *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.