

# Expansión fonética de la consulta para la recuperación de información en documentos hablados

## *Phonetic query expansion for spoken document retrieval*

Alejandro Reyes-Barragán<sup>1</sup>, Luis Villaseñor-Pineda<sup>1</sup>, Manuel Montes-y-Gómez<sup>1,2</sup>  
<sup>1</sup> Laboratorio de Tecnologías del Lenguaje, Departamento de Ciencias Computacionales,  
Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE), México.

E-mail: {alejandroreyes, villasen, mmontesg}@ccc.inaoep.mx

<sup>2</sup> Department of Computer and Information Sciences, University of Alabama at  
Birmingham (UAB), Estados Unidos.

E-mail: mmontesg@cis.uab.edu

**Resumen:** El enfoque tradicional para la búsqueda de información en grandes colecciones de documentos hablados consiste en integrar métodos de reconocimiento automático del habla (RAH) y técnicas de recuperación de información (RI) usadas para texto escrito. Una desventaja de este enfoque es su dependencia a la precisión del sistema de RAH, pues los errores generados por éste impactan fuertemente en la máquina de RI. Con el objetivo de reducir el impacto de los errores de transcripción, especialmente los de sustitución, en este trabajo se propone expandir las consultas con palabras fonéticamente similares y con ello ampliar las posibilidades de emparejar las palabras mal transcritas en los documentos. Los resultados alcanzados en dos colecciones de documentos hablados con características muy disímiles demuestran la pertinencia del método propuesto, el cual logró mejorar el MAP respecto a una expansión tradicional hasta en un 3.68%.

**Palabras clave:** Recuperación de Información, Documentos hablados, Expansión de la consulta, Codificación fonética.

**Abstract:** The traditional approach for searching information in large collections of spoken documents consists of integrating automatic speech recognition (ASR) methods and traditional text retrieval (IR) techniques. One disadvantage of this approach is its dependence to the precision of the ASR system, since transcription errors strongly affect the IR machine. With the aim of reducing the impact of these errors, especially those concerning substitutions, in this paper we propose expanding the queries by means of phonetically similar words, and by this increasing the possibility of matching incorrectly transcribed words from the documents. Results on two very different spoken-document collections show the relevance of the proposed method, which outperformed the MAP from traditional expansion techniques by up to 3.68%.

**Keywords:** Information Retrieval, Spoken Documents, Query Expansion, Phonetic Codes.

## 1 Introducción

La tarea de recuperación de información en transcripciones de audio se realizó por primera vez en el foro TREC en 1997. La tarea recibió el nombre de recuperación en documentos hablados (Spoken Document Retrieval o SDR), y se utilizó un corpus de grabaciones de noticias. Bajo ese contexto se dio por sentado que la recuperación de información en

colecciones de documentos hablados era un problema resuelto (Garofolo, Auzanne y Voorhees, 1999). Tres factores influyeron para llegar a esta prematura afirmación: (i) la tarea se realizó sobre grabaciones de noticias radiofónicas, las cuales cuentan con una estructura clara que se asemeja fuertemente a los documentos escritos; (ii) la redundancia presente en el lenguaje usado –dada la naturaleza de las grabaciones– permitía que la

efectividad de la búsqueda no fuera seriamente afectada por el error introducido por el reconocedor de habla; y (iii) se construyeron sistemas de reconocimiento de habla eficaces para la transcripción de habla como la practicada por los locutores de radio (conocida como habla planeada o leída). Fueron estas tres condiciones las que permitieron obtener resultados satisfactorios en la recuperación de información.

En la actualidad, se continúa trabajando en escenarios mucho más complejos, en los cuales aún hay mucho por hacer. Ejemplos de estos escenarios son: los registros orales de patrimonio cultural (p. ej. entrevistas, debates parlamentarios), las conversaciones multipartitas (p. ej. reuniones (Carletta, 2006), discusiones académicas), las conversaciones emitidas por medios de comunicación (p. ej. programas de entrevista y análisis), etc.

Los primeros intentos para abordar la búsqueda de información en grandes colecciones de documentos hablados consistieron en integrar los métodos de reconocimiento automático del habla (RAH) y las técnicas de recuperación de información (RI) tradicionalmente usadas para texto escrito. Este enfoque utiliza un RAH para obtener una representación textual (transcripción ortográfica) del contenido de una colección de documentos hablados y después considera la aplicación de técnicas tradicionales de RI sobre dichas transcripciones.

No obstante que este enfoque es el más utilizado y uno de los que mejores resultados reporta, depende fuertemente de la precisión del RAH. Una medida comúnmente usada para determinar el error en la transcripción obtenida por un RAH es el error a nivel de palabra (*word error rate*, WER). Esta medida compara la transcripción obtenida por el proceso automático contra una transcripción manual realizada por un especialista. La comparación cuantifica el número de palabras eliminadas, insertadas y/o substituidas por el proceso automático. Actualmente los RAH generan transcripciones con un WER entre 10% y 40%, dependiendo, entre varios aspectos, de la calidad de la grabación, del tipo de habla (habla continua vs. habla espontánea), del tamaño del diccionario de pronunciaciones, del tipo de hablante esperado (dependiente vs. independiente del locutor), del ambiente (p. ej. uno con mucho ruido como en la calle vs. una cabina de grabación), etc. Dado que las

colecciones actuales de documentos hablados son muy heterogéneas es de esperarse una situación con altas tasas de error. Bajo estas circunstancias, las técnicas tradicionales de RI se ven comprometidas y los errores introducidos por el RAH afectarán seriamente su rendimiento.

Por otro lado, cabe notar que los errores introducidos por el RAH no son precisamente al azar, dado que se trata de un proceso probabilístico, la gran mayoría de las ocasiones, se obtendrá una transcripción fonéticamente cercana al contenido real del audio. Es decir, el RAH substituirá una(s) palabra(s) por otra(s) cuya pronunciación es similar. Con esto en mente, este trabajo presenta un método para expandir la consulta con palabras fonéticamente similares, ampliando las posibilidades de emparejar las palabras de la consulta con aquellas palabras mal transcritas en los documentos. Para seleccionar qué palabras agregar a la consulta, el método descansa en la transformación de la palabra escrita a *códigos fonéticos* (CF), los cuales nos permiten representar con el mismo código aquellas palabras fonéticamente similares. De manera particular, en el presente trabajo utilizamos la codificación fonética NYSIIS (Taft, 1970) (más adelante se dan detalles respecto a este tipo de codificación fonética).

El artículo está estructurado de la siguiente manera. En la sección 2 se describen los principales enfoques utilizados en la tarea de recuperación de información en documentos hablados. La sección 3 detalla el método propuesto. La sección 4 describe los experimentos realizados, así como los resultados alcanzados. Por último, la sección 6 muestra las conclusiones y el trabajo futuro que se propone a partir de la presente investigación.

## 2 Trabajo Relacionado

La tarea de recuperación de información en documentos hablados (SDR, por sus siglas en inglés) consiste en buscar información relevante a una consulta en una colección de transcripciones automáticas de documentos hablados. La investigación en esta área ha sido principalmente abordada por dos conferencias internacionales, inicialmente por el TREC (Garofolo, et al., 1998, Garofolo, Auzanne y Voorhees, 1999) y más recientemente por el CLEF (Oard, et al., 2004, Pecina, et al., 2008).

En estos foros se han presentado diversos métodos considerando el impacto del error introducido por el RAH en el proceso de RI. Dentro de esta gran variedad de propuestas se presentan a continuación aquellas que han utilizado algún método de expansión de la consulta o de los documentos para incrementar el rendimiento del sistema de RI.

El trabajo de (Logan and Van Thong, 2002) se presenta un método donde la expansión de la consulta está orientado específicamente a resarcir el problema de las palabras desconocidas por el RAH. Esta situación se presenta cuando una palabra de la consulta no se encuentra en el vocabulario del RAH, de modo que toda ocurrencia de dicha palabra en la colección fue substituida por alguna otra palabra (o palabras). Este tipo de palabras no pueden ser correctamente transcritas pues el RAH no tiene conocimiento de su existencia. La idea de este trabajo consiste en tratar de emular los errores introducidos por el RAH y expandir la consulta con aquellas palabras presentes en el vocabulario y que fonéticamente son similares a las palabras presentes en la consulta y ausentes en el vocabulario del RAH. Este trabajo tiene un alcance limitado dado que los errores introducidos por el RAH no se limitan a palabras ausentes del vocabulario. Esta técnica depende además de disponer de acceso al vocabulario del RAH.

En (Moreau, Kim and Sikora, 2004) se propone representar los documentos usando trigramas de fonemas, y se implementa el sistema de indexado y recuperación utilizando estas unidades. En este caso las consultas (secuencias de trigramas) son expandidas a partir de la información aportada por la matriz de confusión. Así un n-grama con una alta confusión se expande con aquellos n-gramas con los que se confunde. Los resultados reportados son alentadores pero aún son preliminares, ya que estos se realizaron sobre un conjunto de sólo diez consultas, de una sola palabra; y los documentos hablados son grabaciones de lecturas (disminuyendo los errores introducidos por el RAH). Independientemente de la configuración de los experimentos, el gran inconveniente del método es que es necesario disponer de un buen número de transcripciones manuales de referencia para poder realizar el cálculo de la matriz de confusión.

Otro trabajo es el reportado por (Nishizaki and Nakagawa, 2002), cuya idea consiste en

expandir los documentos de la colección. Para ello, los autores utilizan varios RAH (desarrollados de forma independiente) con la intención de que la combinación de las diferentes transcripciones ayude a recuperarse de los errores. Este enfoque mejora ligeramente los resultados respecto a utilizar un sólo RAH. El principal problema que enfrenta este enfoque es que en caso de habla espontánea se tiene una alta tasa de error, en cuyo caso la combinación de varias transcripciones erróneas tendrá un efecto inverso provocando una mayor confusión.

En (Mamou and Ramabhadran, 2008) se propone realizar una búsqueda a nivel fonético, a través de dos índices al transformar la transcripción a nivel de palabras a fonemas y usando un segundo índice con n-gramas de fonemas reconocidos directamente por el decodificador (i. e. modificando el reconocedor para reconocer n-gramas de fonemas). A partir de estos elementos, se expande la consulta con términos (tanto palabras como n-gramas) fonéticamente similares. Para determinar los términos fonéticamente similares los autores proponen un método difuso incluyendo una valoración sobre la probabilidad de pronunciación de dicho término. Los resultados alcanzados son relevantes, sin embargo, el método es dependiente del RAH, ya que debe acceder a los procesos internos del reconocedor para obtener las unidades a indexar.

Por último, otro trabajo es el presentado por (Singhal and Pereira, 1999) donde se demuestra que la expansión de los documentos puede usarse exitosamente. Esta expansión se realiza al agregar términos que coocurren con términos presentes en un documento hablado. Para determinar la coocurrencia de los términos, se utiliza una colección de documentos escritos temáticamente relacionados a los documentos hablados. Desafortunadamente, este método tiene el gran inconveniente de que es necesario contar con un corpus prácticamente paralelo a la colección hablada.

En el caso particular del método propuesto en este artículo se desea abordar situaciones cercanas a la realidad donde no se tiene acceso a los procesos internos del RAH, ni a ningún otro recurso adicional, como el vocabulario del RAH o a transcripciones manuales de referencia. En definitiva, sólo se cuenta con las transcripciones automáticas a nivel palabra entregadas por el RAH. Bajo estas condiciones se debe determinar si dos términos son

fonéticamente similares a partir de su transcripción ortográfica.

### 3 Método propuesto

Antes de detallar el método propuesto, a continuación se describe brevemente en qué consisten los códigos fonéticos, parte fundamental de la idea presentada en este trabajo.

#### 3.1 Códigos fonéticos

El uso de los códigos fonéticos fue propuesto originalmente para representar bajo un mismo código (p. ej. una cadena alfanumérica) los nombres y/o apellidos de personas, que por razones particulares de una región, cultura o el paso del tiempo, variaron su escritura pero cuya pronunciación se conserva. Los códigos fonéticos realizan una transformación partiendo de la palabra escrita para obtener un código que aproxima la pronunciación de dicha palabra. Este tipo de métodos tienen años utilizándose en bases de datos (Holmes y McCabe 2002, Zobel and Dart 1996) precisamente para buscar nombres y apellidos de personas con pequeñas variantes ortográficas pero pronunciación similar. Por supuesto, un algoritmo de codificación fonética puede aplicarse a cualquier palabra. La Tabla 1 muestra un ejemplo de cómo con un mismo código es posible representar palabras con pronunciaciones similares.

Código fonético	Palabras representadas por el código
LANAN	lehman lemon lenient lenin linen linens

Tabla 1. Ejemplo de cómo un mismo código fonético (NYSIIS) representa distintas palabras.

En particular en este trabajo se utilizó el algoritmo de codificación fonética NYSIIS (New York State Identification Intelligence System), creado en 1970 como parte de la división de servicios de justicia criminal del estado de New York. NYSIIS. La transformación se realiza a través de reglas específicas para el inglés y realizan una generalización de la palabra escrita para aproximar su pronunciación. Es importante mencionar que no se trata de un sistema que lleva el texto a una cadena de fonemas, el

código fonético es una abstracción de la pronunciación mucho más severa (p. ej. convierte toda vocal a letra A). NYSIIS ha demostrado su efectividad al alcanzar incrementos del 2.7% sobre algoritmos tradicionales como Soundex en la tarea de búsqueda de nombres de personas en bases de datos.

El usar la codificación fonética nos permite que el método propuesto sea independiente del RAH y de cualquier otro recurso externo (p. ej. Corpus escritos o diccionarios de pronunciaciones), también con el uso de la codificación fonética es como esperamos identificar el conjunto de palabras con las cuales expandir la consulta y nos permitan reducir el impacto que tienen los errores introducidos por el RAH en la RI.

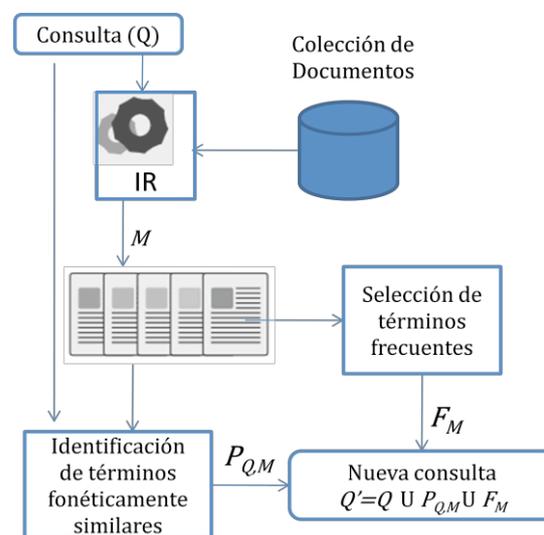


Figura 1: Arquitectura del método Propuesto

#### 3.2 Proceso de expansión

En la Figura 1 se muestra la arquitectura general del método propuesto, para lo cual hemos retomado la retroalimentación por relevancia ciega y, donde además de expandir con términos temáticamente asociados, hemos incluido términos fonéticamente similares. Los términos fonéticamente similares se identifican gracias a que comparten el mismo código fonético. El método procede de la siguiente manera:

1. Se ingresa la consulta  $Q$  a la máquina de RI, la cual nos entregara el conjunto  $M$  de documentos recuperados.

- Se identifican aquellos términos del conjunto de documentos  $M$  que comparten el mismo CF de alguno de los términos de la consulta  $Q$ , con lo que obtenemos el conjunto de términos  $P_{Q,M}$ .

$$P_{Q,M} = \bigcup_{t_i \in Q} C_M^{t_i}$$

$$C_M^{t_i} = \left\{ t_k \mid \begin{array}{l} code_{NYSIIS}(t_i) = code_{NYSIIS}(t_k) \\ \wedge t_k \in voc(M) \end{array} \right\}$$

donde  $code_{NYSIIS}(t_i)$  transforma el término  $t_i$  en el código fonético correspondiente y  $voc(M)$  es el vocabulario del conjunto de documentos  $M$ .

- Se construye el conjunto de términos  $F_M$  con aquellos términos más frecuentes en el conjunto de documentos  $M$ . (p. ej. siguiendo la técnica de retroalimentación ciega tradicional).
- Se construye la nueva consulta  $Q'$  al unir el conjunto de términos de la consulta original  $Q$ , el conjunto de términos frecuentes  $F_M$  y el conjunto de términos fonéticamente similares  $P_{Q,M}$ .

$$Q' = Q \cup P_{Q,M} \cup F_M$$

- Finalmente, se ingresa la consulta expandida  $Q'$  a la máquina de RI.

Para ejemplificar el método, la Tabla 2 muestra la consulta inicial  $Q$ , el conjunto de términos frecuentes  $F_M$ , el conjunto de términos fonéticamente similares  $P_{Q,M}$ , y la consulta extendida  $Q'$ . Como puede observarse, la consulta expandida incluye las palabras *roll* y *real* que podrían empalmarse con posibles errores del RAH al tratar de reconocer la palabra *raoul*. Otro caso es la simple inclusión de *Adolph* debido a *Adolf* error que podría haber introducido el RAH al no tener más que una variante en su diccionario de pronunciaciones.

La razón para buscar los términos fonéticamente similares únicamente en el conjunto  $M$  y no hacerlo sobre toda la colección, se debe a que la CF generaliza demasiado, de manera que expandir con todas ellas provocaría una caída en el desempeño del sistema de RI.

<i>Consulta Original Q</i>	describe personalities actions raoul wallenberg adolf Eichmann
<i>Términos frecuentes F<sub>M</sub></i>	ghetto time people jewish find
<i>Términos fonéticamente similares P<sub>Q,M</sub></i>	roll real Adolph
<i>Consulta Expandida Q'</i>	describe personalities actions raoul wallenberg adolf Eichmann ghetto time people jewish find roll real Adolph

Tabla 2: Ejemplo de la expansión de una consulta.

## 4 Configuración experimental

A continuación, describimos los corpus utilizados, las medidas de evaluación y los experimentos realizados para mostrar la utilidad de este método.

### 4.1 Descripción del corpus

Para nuestros experimentos hemos escogido dos colecciones, una perteneciente a la tarea Speech Retrieval (SR) del foro de evaluación CLEF y otra a la tarea Spoken Document Retrieval (SDR) perteneciente al TREC. En la Tabla 3 se muestran las características de estas colecciones.

La colección SDR está compuesta de transcripciones de noticias con una temática claramente definida y donde las grabaciones originales son de habla continua. Por el contrario, la colección SR comprende transcripciones de entrevistas cuyos temas son muy variados y donde las grabaciones originales son de habla espontánea. Así ambas colecciones son en extremo diferentes.

<i>Características</i>	<i>SDR (Trec)</i>	<i>SR (Clef)</i>
Número de consultas	50	63
Prom. Palabras/consulta	11.7	11.1
Prom. Doc. Relevantes	44.3	83
Prom. Palabras/doc.	79	136
Horas de Audio	502	589
Documentos	21,754	8,104
Documentos Relevantes	2,216	5,229

Tabla 3: Algunas características de las colecciones utilizadas.

## 4.2 Resultados de referencia

Para tener un punto de comparación, primero utilizamos una máquina de RI tradicional, cuyos resultados están en la Tabla 4, posteriormente mostramos los resultados al utilizar una técnica clásica de retroalimentación de relevancia ciega, Tablas 5 y 6, donde se muestran resultados de diferentes pruebas al usar los primeros 1, 5 ó 10 documentos recuperados y de ellos obtenemos los 5, 10 y 20 términos con mayor frecuencia. Se presentan resultados en las dos colecciones SR y SDR.

La medida utilizada para reportar los resultados es el MAP, que es una medida que integra la precisión con el orden de los documentos relevantes entre todos los documentos recuperados.

Asimismo, la significancia estadística de los resultados se evaluó mediante la prueba "paired t-student" considerando un intervalo de confianza  $\alpha=0.05$ . Los resultados marcados con un asterisco (\*) en las Tablas 5-8 indican que la mejora en MAP fue estadísticamente significativa con respecto a los resultados base mostrados en la Tabla 4.

Cabe resaltar que el único preprocesamiento de las colecciones consistió en remover las palabras vacías y la máquina de RI utilizada fue Lémur configurado con el modelo TFIDF.

Corpus	MAP
TREC	0.2656
CLEF	0.0620

Tabla 4: Resultados de referencia obtenidos con Lemur usando la configuración TFIDF.

Doc\Ter	5	10	20
1	0.2856	0.2738	0.2660
5	0.2928	0.2968	0.2880
10	0.3131*	0.3015	0.3086*

Tabla 5: Resultados en MAP, utilizando retroalimentación en la colección SDR (TREC).

Doc\Ter	5	10	20
1	0.0602	0.0561	0.0498
5	0.0643	0.0637	0.0634
10	0.0628	0.0644	0.0667

Tabla 6: Resultados en MAP, utilizando retroalimentación en la colección SR (CLEF).

Es interesante notar que para la colección del TREC con agregar 5 términos de un sólo documento se logra mejorar el MAP, sin embargo en la colección SR la mejora no se obtiene hasta considerar al menos 5 documentos. Esto se debe a la precisión en las primeras posiciones en donde la colección SDR tiene un mejor comportamiento.

## 5 Resultados

Los experimentos del método propuesto se llevaron a cabo siguiendo los mismos parámetros de los experimentos de referencia, añadiendo la inclusión de los términos fonéticamente similares.

En la Tabla 7 y Tabla 8 se muestran los resultados al expandir la consulta tanto con los términos temáticamente asociados como con los fonéticamente similares (+PF). Como puede observarse, en todos los casos el MAP aumenta. Se indica la mejora relativa –entre paréntesis– al comparar los resultados del método contra los mostrados en las tablas 5 y 6 respectivamente.

Doc\Ter	5+PF	10+PF	20+PF
1	0.2961 (3.68%)	0.2803 (2.37%)	0.2726 (2.48%)
5	0.3014 (2.94%)	0.3058* (3.03%)	0.2892 (0.42%)
10	0.3175* (1.41%)	0.31* (2.82%)	0.3172* (2.79%)

Tabla 7: Resultados para la colección SDR, utilizando el método propuesto. Se indica el porcentaje de mejora en paréntesis.

Doc\Ter	5+PF	10+PF	20+PF
1	0.0608 (1.00%)	0.0566 (0.89%)	0.0502 (0.80%)
5	0.0662* (2.95%)	0.0649 (1.88%)	0.0641 (1.10%)
10	0.0641 (2.07%)	0.0654 (1.55%)	0.0674 (1.05%)

Tabla 8: Resultados para la colección SR, utilizando el método propuesto. Se indica el porcentaje de mejora en paréntesis.

Un punto interesante del método es que “a pesar de que en promedio se expanden las consultas” con muy pocos términos fonéticamente similares se logra mejorar el

rendimiento del sistema. La Tabla 9 muestra estos promedios.

<i>Doc\Colección</i>	<i>SR</i>	<i>SDR</i>
<b>1</b>	0.57	0.66
<b>5</b>	1.63	1.86
<b>10</b>	2.5	3.16

Tabla 9: Promedio de términos fonéticamente similares agregados a las consultas

Como puede observarse en la colección SR se añaden menos términos que en la colección SDR. A primera vista este hecho parece contradictorio, ya que los documentos en SR son más extensos y con mayores probabilidades de encontrar términos similares. No obstante, las consultas de esta colección fueron realizadas por historiadores o analistas y cuyo nivel de abstracción dista mucho de los términos usados en la colección, hay que recordar que se trata de entrevistas.

Por último, cabe mencionar que al comparar los mejores resultados obtenidos con el método propuesto y los resultados de referencia sin retroalimentación, alcanzan una mejora relativa del 19.54% en MAP para la colección del TREC y de 8.71% en la colección del CLEF.

## 6 Conclusiones y trabajo futuro

En este artículo experimentamos con el uso de códigos fonéticos para enriquecer la expansión de una consulta en la tarea de recuperación de información de documentos hablados. Basados en la idea de retroalimentación de relevancia ciega, el método propuesto añade a la expansión tradicional términos fonéticamente similares a las palabras de la consulta. El método sólo requiere de los documentos transcritos y del algoritmo de codificación fonética para operar, sin intervenir o depender en los procesos internos del RAH o de recursos lingüísticos externos.

Los resultados preliminares alcanzados hasta ahora muestran la pertinencia del método al aplicarla a dos colecciones de documentos hablados con características muy disímiles. El método hasta ahora sólo aborda errores de sustitución, es decir, sólo aquellos errores en el que RAH substituyó una palabra por otra. Se espera que al extender esta idea a nivel de subpalabras la cobertura se incremente resolviendo otros errores introducidos por el RAH.

También como trabajo futuro se desea explorar usando otras codificaciones fonéticas como Soundex y DoubleMetaphone, ya que hasta ahora sólo se ha experimentado con NYSIIS. De igual forma, otra idea a explorar como trabajo futuro es la utilización de la codificación fonética en la expansión de los documentos y no sólo en las consultas. En este trabajo se utilizó Lemur con TFIDF, resulta de interés comprobar que tan robusto es el método utilizando otros modelos como okapi o KL-divergence, y de la misma forma, comprobar si son consistentes los resultados con otros sistemas de RI como Lucene o Terrier.

## Agradecimientos

Este trabajo fue realizado con el apoyo parcial del CONACYT/México, a través de los proyectos CB-2009-01-134186, CB-2008-106013-Y, y la beca 204467. Los autores también agradecen al comité organizador del CLEF por las colecciones de datos proporcionadas.

## Bibliografía

- Carletta, J. 2006. Announcing the AMI Meeting Corpus. The ELRA Newsletter 11(1), January-March, p. 3-5.
- Garofolo, J. S., E.M. Voorhees, C. Auzanne, V. Stanford, and B. Lund. 1998. TREC-7 spoken document retrieval track overview and results. *In Proceedings of the seventh Text Retrieval Conference (TREC-7)*. 79 - 89. NIST, Gaithersburg, Md, USA: NIST Special Publication.
- Garofolo, John S., Cedric G. P. Auzanne, y Ellen M. Voorhees. 1999. The TREC Spoken Document Retrieval Track: A Success Story. *In Proceedings of the 8th Text REtrieval Conference*. 107 - 129. NIST.
- Holmes, D., y M.C. McCabe. 2002. Improving Precision and Recall for Soundex Retrieval. *ITCC Proceedings of the International Conference on Information Technology: Coding and Computing* (IEEE Computer Society), 22 - 26.
- Logan, Beth, and JM Van Thong. 2002. Confusion-based query expansion for OOV words in spoken document retrieval. *Proceedings of ICSLP*.
- Mamou, Jonathan, and Bhuvana Ramabhadran. 2008. Phonetic Query Expansion for Spoken Document Retrieval." *Interspeech*.

- Moreau, Nicolas, Hyoung-Gook Kim, and Thomas Sikora. 2004. Phonetic Confusion Based Document Expansion for Spoken Document Retrieval. *SIGIR*.
- Nishizaki, Hiromitsu, and Seiichi Nakagawa. 2002. Japanese spoken document retrieval considering OOV keywords using LVCSR system with OOV detection processing. *Proceedings of the second international conference on Human Language Technology Research* (Morgan Kaufmann Publishers Inc). 157 - 164.
- Oard, D. W.; Soergel, D.; Doermann, D.; Huang, X.; Murray, G. C.; Wang, J.; Ramabhadran, B.; Franz, M.; Gustman, S.; Mayfield, J.; Kharevych, L.; Strassel, S. 2004. Building an information retrieval test collection for spontaneous conversational speech. *Proc. SIGIR*. 41 - 48.
- Pecina, Pavel, Petra Hoffmannová, Gareth J. F. Jones, Ying Zhang, and Douglas W. Oard. 2008. Overview of the CLEF-2007 Cross-Language Speech Retrieval Track. Edited by In Carol Peters et al. Budapest, Hungary: Springer-Verlag. 674 - 686.
- Singhal, Amit, and Fernando Pereira. 1999. Document Expansion for Speech Retrieval. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. 34-41.
- Taft, R. L. 1970. Name Search Techniques. New York State Identification and Intelligence System. Albany, New York.
- Zhang, Lei. 2009. Topic indexing of spoken documents based on optimized N-best approach. 302-205.
- Zobel, J., and P. Dart. 1996. Phonetic String Matching: Lessons from Information Retrieval. 166 – 172. *Sigir Forum, Association for Computing Machinery*.