ISSN 1870-4069

Resúmenes de múltiples documentos guiados por consulta empleando representaciones distribucionales

Leticia Luna-Tlatelpa[1], Esaú Villatoro-Tello[2], Gabriela Ramírez-de-la-Rosa[2], Carlos J. Rivero-Moreno[2]

[1] Maestría en Diseño, Información y Comunicación (MADIC),
División de Ciencias de la Comunicación y Diseño,
Universidad Autónoma Metropolitana (UAM) Unidad Cuajimalpa, México

[2] Departamento de Tecnologías de la Información,
Universidad Autónoma Metropolitana (UAM) Unidad Cuajimalpa, México

letyludigital@gmail.com, {evillatoro,gramirez,crivero}@correo.cua.uam.mx

Resumen. Actualmente, la generación automática de resúmenes de múltiples documentos guiados por consulta ha cobrado mucha importancia dentro de la comunidad del Procesamiento de Lenguaje Natural, principalmente debido a la necesidad de información cada vez más específica por parte de usuarios especializados. En este contexto, el presente trabajo evalúa la pertinencia de emplear técnicas de representación distribucional contra técnicas tradicionales al momento de construir un resumen, el cual debe satisfacer una necesidad particular de información, *i.e.,* la consulta. Nuestra hipótesis plantea que la representación distribucional incide de manera positiva en la selección de oraciones relevantes a la consulta, ya que esta representación se aproxima mejor al contenido semántico de las palabras. Los resultados obtenidos durante la fase experimental son alentadores, pues muestran que las representaciones distribucionales pueden ser aplicadas de manera eficiente en el problema planteado.

Palabras clave: Representación de textos, representación distribucional, aproximación semántica, agrupamiento de textos, resumen automático de múltiples documentos.

Query-based Multi-document Summarization through Distributional Representations

Abstract. Recently, research in query-based multi-document summarization has gain the attention of the natural language processing community. This is mainly due to the growing need of specialized users to obtain very specific information. In this context, this paper describes a method for automatically generate a summary from multiple documents considering the user's profile. Particularly, we evaluate the pertinence

of using a distributional representation against traditional document representation techniques. Our main hypothesis suggests that through the information captured by a distributional representation, is possible to build better summaries in response to a short query. Our experimental results are encouraging and indicate that distributional representations can be efficiently applied to the posed problem.

Keywords: Document representation, distributional term representation, semantic information, text clustering, multi-document summarization.

1. Introducción

Es un hecho que la cantidad de información existente ha superado la capacidad, en tiempo y almacenamiento, que tenemos los humanos para poder procesarla; por lo que, ante tal situación, ha sido necesaria la creación de herramientas que puedan aligerar el peso de la información, y una de ellas es la generación automática de resúmenes.

Un resumen es la síntesis concisa y coherente de la información más importante contenida en uno o más documentos, por lo que un sistema generador de resúmenes tiene como objetivo presentar al usuario las ideas principales de los documentos de referencia en un texto pequeño [9,3]. Existen varias categorías para definir a este tipo de sistemas, por ejemplo, sistemas que realizan el resumen de forma general, *i.e.,* simplemente tratan de inferir qué información podría ser relevante para pertenecer al resumen. Por otro lado, sistemas más complejos toman en cuenta el perfil del usuario para la construcción del resumen, usualmente el perfil va indicado por medio de una consulta específica, a estos últimos se les denomina sistemas de generación de resúmenes guiados por consulta. En general, los sistemas de generación de resúmenes transforman los documentos de entrada en oraciones y después, extraen aquellas más importantes. El problema fundamental que enfrentan es cómo calcular la relevancia de las oraciones. Así entonces, el sistema requiere de una representación de las oraciones que permita a los distintos algoritmos identificar aquellas oraciones más relevantes, ya sea por su contenido semántico o porque satisfacen de manera adecuada una necesidad de información [7].

Dentro de este trabajo nos enfocaremos específicamente en el problema de generación de resúmenes de múltiples documentos guiados por consulta. Esto quiere decir que el usuario conoce a priori el conjunto de documentos que versan sobre un tópico de su interés. Así entonces, el usuario desea construir un resumen de estos documentos a partir de una consulta que él define. Idealmente, el resumen generado, debería contener toda la información relevante compartida entre los documentos de la colección (una sola vez), y además toda la información única y relevante a la consulta proporcionada por el usuario.

Es sabido en el área de Procesamiento de Lenguaje Natural (PLN), que el cálculo de la relevancia depende de manera importante de la representación que

se tenga de los documentos. Tradicionalmente, la representación de bolsa de palabras (BoW[3]) ha sido ampliamente utilizada debido a que es relativamente fácil de implementar y ha mostrado ser eficiente en distintas tareas de PLN [4,13,10]. Este modelo, no obstante, tiene algunas desventajas: no distingue la polisemia ni la sinonimia, se pierde el orden de las palabras y se omite información semántica. Con el fin de resolver algunos de estos problemas, surge la representación denominada "bolsa de conceptos" (BoC por sus siglas en inglés), específicamente, nos concentraremos en la representación distribucional TCOR[4] [7]. En este tipo de representación, un término se representa en función de otros términos que co-ocurren con él en el documento. De esta manera se tiene un acercamiento al significado de un término en función de otros términos que tengan una distribución o patrones de uso similar dentro del documento.

El objetivo principal de este trabajo es evaluar si la representación distribucional TCOR mejora la generación automática de resúmenes de múltiples documentos guiados por una consulta. La hipótesis es que la representación TCOR incide positivamente en la selección de las oraciones más importantes ya que ésta representación considera, hasta cierto punto, la semántica contenida en los términos de los documentos y la consulta. Para evaluar la pertinencia de nuestra hipótesis se realizaron experimentos con una colección estándar proporcionada por los organizadores del DUC (Document Understandig Conference) del 2005. Los resultados obtenidos son alentadores y muestran que el uso de representaciones distribucionales permite obtener resúmenes relevantes a una necesidad de información específica.

El resto de este documento se encuentra organizado de la siguiente manera: en la sección 2 se describen algunos trabajos relacionados a la utilización de diferentes tipos de representación y de agrupamiento. En la sección 3 se describe en detalle el método propuesto. En la sección 4 se describen la metodología experimental y la sección 5 muestra los resultados obtenidos. Finalmente en la sección 6 se plantean las conclusiones y se discuten posibles líneas a futuro.

2. Trabajo relacionado

Para la generación automática de resúmenes, existen dos enfoques principales: el método extractivo y el abstractivo. El primero, consiste en extraer (literalmente) la información que se considere importante de los documentos a resumir, y construir con éstas el resumen final. El segundo enfoque, consiste en generar nuevas oraciones a partir de la información identificada como importante; este tipo de técnicas poseen la capacidad de general lenguaje natural [9,3]. Dentro de este trabajo nos enfocaremos en resúmenes extractivos de múltiples documentos.

El proceso de generar un resumen de múltiples documentos consiste en la creación de un resumen simple de un conjunto de documentos relacionados

[3] *Bag-of-Words* por sus siglas en Inglés.
[4] *Term co-occurrence.*

temáticamente; además, dicho resumen debe satisfacer la necesidad de información del usuario, es decir, responder a la consulta proporcionada. Existen tres grandes problemas que surgen en este escenario: (i) reconocer y resolver redundancias, (ii) identificar diferencias importantes entre los documentos, y (iii) asegurar la relevancia del resumen final, tomando en cuenta que diferentes porciones de texto podrían ser relevantes a la consulta proporcionada. En general, para enfrentar estos problemas se ha seguido dos grandes pasos [6,3]: la fase de pre-procesamiento y la fase de procesamiento de la información.

En la etapa de pre-procesamiento de información se consideran acciones como el segmentado de la información (dividir documentos en oraciones), eliminación de palabras vacías, identificación de la raíz léxica de las palabras, etc. Por otro lado, en la etapa de procesamiento de la información se decide, por un lado, los atributos con los que se representará la información, y por otro, la técnica de identificación de la información más relevante. Respecto a las técnicas para la identificación de la información más relevantes es conveniente mencionar que existen una gran variedad de estrategias, por ejemplo, métodos estadísticos, métodos basados en tópicos, técnicas basadas en aprendizaje automático, basados en grafos, y basados en discurso. Para tener un referente más amplio de las distintas estrategias refiérase a [3].

Dentro de este trabajo nos enfocaremos únicamente en los atributos, *i.e.*, la representación empleada. Muchos trabajos existentes recurren a la tradicional *Bolsa-de-Términos* debido a su simplicidad [9,3], donde los *términos* pueden ser palabras simples o secuencias de palabras [4,10,2,5] los cuales son ponderados de acuerdo a su valor TF-IDF. La desventaja principal de este tipo de representación es que carece de información semántica, y para que la relevancia de las oraciones con respecto a la consulta sea determinada, requiere de que los términos de la consulta existan de manera idéntica en los documentos de la colección, de otra forma la relevancia no puede ser determinada o es prácticamente nula. Recientemente, el uso de representaciones más semánticas ha permitido enfrentar estas limitantes, por ejemplo en [12,11] se propone el uso de LSA[5] para la generación de resúmenes. Estos trabajos han mostrado la pertinencia de éste tipo de formas de representación, sin embargo tienen la principal desventaja de que requieren de un corpus de entrenamiento para poder generar la representación.

Contrario a los trabajos descritos anteriormente, dentro de este artículo proponemos utilizar una forma de representación distribucional para la generación de resúmenes de múltiples documentos a partir de consultas. A diferencia de trabajos previos, este tipo de representación captura la semántica de los términos sin la necesidad de un corpus de entrenamiento, permitiendo esto la independencia de dominio y de lenguaje.

3. Método propuesto

En la figura 1 se muestra el esquema general de nuestro método para la generación de resúmenes de múltiples documentos guiado por consulta. Básí-

[5] Latent Semantic Analysis

camente, dada una colección de documentos relacionados temáticamente y una consulta proporcionada por el usuario, nuestro método aplica algunas operaciones de pre-procesamiento previo al proceso de generación de resúmenes. Una vez dentro de la etapa de procesamiento de la información, la parte fundamental de nuestro método recae en la construcción de la representación, la cual sirve posteriormente tanto para agrupar como para extraer información relevante a la consulta, misma que es empleada para la construcción del resumen final. A continuación describimos detalladamente cada uno de los módulos involucrados en nuestro método propuesto.

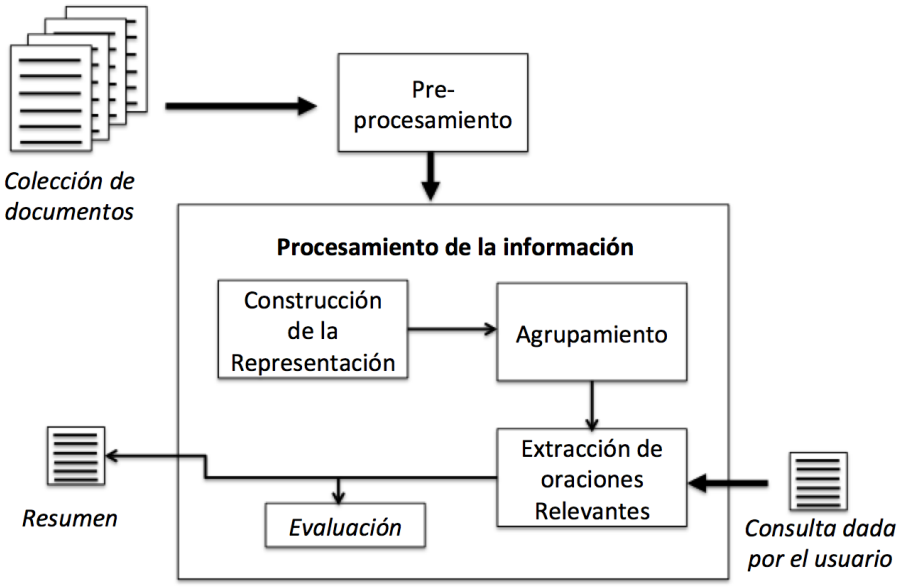

Figura 1. Arquitectura general del método propuesto.

3.1. Pre-procesamiento

En la etapa del pre-procesamiento, se eliminaron todas las palabras que no contribuyen a la semántica de los textos de entrada, como son: etiquetas html y xml, signos de puntuación y palabras vacías. Todos los textos son convertidas a minúsculas y se realizó truncamiento de palabras[6], el cual es un proceso heurístico para aproximar las palabras a su raíz léxica. En esta etapa se hace el segmentado de los documentos en oraciones, para lo cual se empleó el programa proporcionado por los organizadores de la conferencia DUC.[7]

[6] Para este proceso se utilizó el método de *Porter*
[7] `http://duc.nist.gov/`

3.2. Construcción de la representación

El primer paso obligado es el *indexado* de las oraciones (S), actividad que denota hacer el mapeo de una oración s_i en una forma compacta de su contenido. La representación más comúnmente utilizada para representar textos es un vector con términos ponderados como entradas, concepto tomado del modelo de espacio vectorial usado en recuperación de información. Esto es, cada texto s_i es representado como el vector $\overrightarrow{s}_i = \langle w_{ki}, \ldots, w_{|\tau|i} \rangle$, donde τ es el *diccionario*, *i.e.*, el conjunto de términos que ocurren al menos una vez en algún elemento de S, mientras que w_{ki} representa la importancia del término t_k dentro del contenido del documento s_i. Este método de representación, también conocido como bolsa de palabras (BoW), propone varios esquemas para definir w_{ki}, en particular, para nuestro método base se utilizó el esquema de pesado TF-IDF.

Nuestra propuesta como alternativa a las limitantes de la representación BoW es utilizar una forma de representación distribucional llamada TCOR. La representación de co-ocurrencia de términos (TCOR) se basa en las estadísticas de co-ocurrencia de los mismos términos. La idea intuitiva es que la semántica del termino t_j puede ser revelada por medio de los términos con los que co-ocurre dentro de la colección de documentos. Aquí, cada término $t_j \in \tau$ es representado por un vector de pesos $\overrightarrow{w}_j = \langle w_{1,j}, \ldots, w_{|\tau|,j} \rangle$, donde $0 \leq w_{k,j} \leq 1$ representa la contribución del término t_k a la descripción semántica de t_j:

$$w_{k,j} = tf(t_k, t_j) \cdot log \frac{|\tau|}{\tau_k}, \tag{1}$$

donde τ_k es el número de diferentes términos en el diccionario τ que co-ocurren con t_j en almenos un documento y

$$tf(t_k, t_j) = \begin{cases} 1 + \log(\#(t_k, t_j)) & si(\#(t_k, t_j) > 0), \\ 0 & \text{en otro caso}, \end{cases} \tag{2}$$

donde $\#(t_k, t_j)$ denota el número de documentos en los cuales el término t_j co-ocurre con el término t_k. La intuición detrás de este esquema de pesado es que entre más veces t_k y t_j co-ocurran más importante será t_k para describir la semántica de t_j; mientras más términos co-ocurran con t_k menos importante será éste en la definición de la semántica de t_j.

Una vez que se tiene el vector $\overrightarrow{w}_j$ de pesos de cada término, la forma de representar cada oración s_i se obtiene por medio de:

$$\overrightarrow{s}_i^{tcor} = \sum_{t_j \in s_i} \alpha_{t_j} \cdot \overrightarrow{w}_{t_j}, \tag{3}$$

donde α_j es un escalar que pondera la contribución del término $t_j \in s_i$ dentro de la representación de la oración, normalmente el valor TF-IDF del término t_j. De esta forma, la representación de una oración está dada por la suma de los vectores contextuales de sus términos.

Observe que bajo la representación TCOR cada oración s_i es representada por $\overrightarrow{s}_i^{tcor} \in \mathbb{R}^{|\tau|}$, un vector de la misma dimensionalidad que el vocabulario. Los

valores de $\overrightarrow{s}_i^{tcor}$ indican el grado de asociación entre los términos del vocabulario y los términos que ocurren dentro de s_i.

3.3. Agrupamiento

El proceso de agrupamiento tiene como principal objetivo permitir al sistema de generación de resúmenes dividir la colección inicial en sus diferentes subtemas. Idealmente, esto permitirá identificar similitudes (redundancias) entre documentos y además detectar la información única (relevante) dentro de cada uno de ellos.

Para los experimentos realizados en este trabajo se utilizó el algoritmo de agrupamiento estrella. Este es un algoritmo que ha mostrado evidencia positiva sobre la verdad de la hipótesis del grupo[8] [1]. Este es un algoritmo que induce de manera natural el número de grupos y la estructura de los temas dentro del espacio de textos. Es un algoritmo de tipo particional, y que se basa en Teoría de grafos. La salida de este algoritmo son grupos de documentos en forma de "estrella", donde se garantiza que el elemento central de cada una de las estrellas es el más representativo del grupo. Esto se logra a través de un parámetro de similitud σ que es pasado al algoritmo, el cual permite incorporar elementos al grupo cuando estos son semejantes al centro por un factor mayor o igual a σ.

Un paso fundamental del algoritmo estrella es el cálculo de un grafo umbralizado G_σ [1]. Para esto, es necesario aplicar técnicas de medición de similitud entre documentos por medio de las cuales es posible definir el umbral σ. Para nuestros experimentos se utilizó la medida cosenoidal. La idea básica de ésta es medir el ángulo entre el vector de dos oraciones cualesquiera s_i y de s_j, para hacerlo calculamos:

$$SC(s_i, s_j) = \frac{\sum_{k=1}^{|\tau|} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{|\tau|} (w_{jk})^2 \sum_{k=1}^{|\tau|} .(w_{ik})^2}} \tag{4}$$

Note que dependiendo de la representación que estemos utilizando, BoW o TCOR, w_{ik} puede tomar diferentes significados. Por un lado, puede referir al valor TF-IDF del termino k bajo una representación BoW; mientras que bajo una representación TCOR representa el vector contextual del término k calculado por medio de la expresión 1.

Finalmente, para hacer la selección del mejor umbral σ, se hace uso de la información estadística de la matriz de similitud que emplea el algoritmo estrella. Para esto, se calcula la media ($\overline{x}$) y la desviación estándar (δ) de los datos. Así entonces, se definen dos posibles valores para σ, un valor ALTO ($\overline{x} + \delta$) y un valor BAJO ($\overline{x} - \delta$).

[8] El uso del agrupamiento dentro del área de Recuperación de Información surge debido a una hipótesis *(hipótesis del grupo)*, que dice que los documentos relevantes a una petición tienden a ser más cercanos entre ellos que aquellos que no son relevantes a una petición en particular.

3.4. Extracción de oraciones relevantes

Una vez que se ha hecho el agrupamiento de la información contenida en los documentos de la colección, el siguiente paso es la extracción de la información más relevante para construir el resumen final. Para la realización de este proceso se definieron dos métodos, los cuales describimos a continuación.

- CENTRO_ESTRELLA (**CE**). Bajo este esquema, se aprovecha la estructura de salida generada por el algoritmo de agrupamiento estrella. Dado que este algoritmo garantiza que el centro de cada estrella es el elemento más representativo, se ordenan los centros de las estrellas de acuerdo a su similitud con la consulta. Una vez hecho esto, se toman los centros más similares para la construcción del resumen final hasta que se alcanza el tamaño de resumen requerido por el usuario.
- SATÉLITE_CERCANO (**SC**). En esta configuración no se toman como elementos más representativos a los centros de las estrellas. La hipótesis aquí es que, a pesar de que el elemento más representativo de cada grupo es el centro de la estrella, este agrupamiento no se hizo contemplando información de la consulta y por ende, podría haber un elemento más relevante a la consulta entre los satélites de la estrella. Así entonces, para cada estrella formada en la etapa de agrupamiento se identifica al elemento (satélite o centro) más similar con la consulta y éste es incorporado a una lista ordenada. Finalmente, se van tomando las oraciones más similares de esta lista hasta construir un resumen del tamaño requerido.

4. Configuración experimental

En esta sección se describen el conjunto de datos con el que se realizaron los experimentos. Agregado a esto se describen los métodos base con los que se compara la propuesta hecha en este trabajo, se explican las métricas de evaluación, y se definen los experimentos realizados para comprobar las hipótesis planteadas en este artículo.

4.1. Conjunto de datos

Los datos con los que se trabajó corresponden a los proporcionados por DUC 2005[9], los cuales consisten en noticias de *Los Angeles Times* y *Financial Times of London*, divididos en 50 tópicos (colecciones de documentos) independientes entre sí. Cada tópico tiene una consulta asociada, la cual consiste de un título, una pequeña narrativa y un valor de granularidad, el cual indica la especificidad con que se sugiere realizar el resumen. Para nuestros experimentos no se consideró en valor de granularidad.

[9] `http://www-nlpir.nist.gov/projects/duc/data/2005_data.html`

4.2. Métodos base

A continuación se describen los tres métodos base que se definieron para la tarea. Es conveniente mencionar que en las especificaciones del DUC 2005, el tamaño del resumen final no debería de ser superior a 250 palabras .

- Método Base 1 (**MB1**). Este método corresponde al método propuesto por los organizadores del DUC 2005. Consiste en extraer las primeras 250 palabras del documento más reciente de la colección. Aunque simple, este método base es considerado un método fuerte y difícil de superar.
- Método Base 2 (**MB2**). Este método consiste en segmentar todos los documentos de cada tópico en oraciones. Posteriormente, las oraciones son ordenadas de acuerdo a su grado de similitud con la consulta, y para la construcción del resumen se toman las oraciones más similares hasta completar 250 palabras. Para esto se empleó una representación tipo BoW con pesado TF-IDF y empleando como medida de similitud el coseno.
- Método Base 3 (**MB3**). Muy similar al MB2, con la diferencia de que en lugar de emplear palabras simples como atributos de la BoW, se empleó la combinación de uni-gramas, bi-gramas y tri-gramas de palabras como elementos del vocabulario τ. El objetivo fue evaluar si por medio de incorporar información contextual a través del uso de n-gramas se podía obtener mejores resultados.

4.3. Experimentos

Con el objetivo de comprobar la pertinencia del método propuesto descrito en la sección 3, se definieron dos grandes conjuntos de experimentos: *i)* aplicando la metodología propuesta en la figura 1 empleando una representación tradicional tipo BoW considerando palabras simples como atributos del vector de la representación; y *ii)* aplicando la metodología propuesta en la figura 1 empleando una representación TCOR.

Recuerde que el método descrito en la sección 3 tiene algunos parámetros como lo son el valor del umbral de similitud (ALTO/BAJO) y la forma en que se extraen las oraciones para conformar el resumen final (CE/SC). Así entonces, con el objetivo de cubrir todos estos aspectos, se realizaron un total de ocho experimentos. La nomenclatura para nombrar a los experimentos es como sigue: **REP-STAR-UMB-EXT**. Donde REP puede ser BOW o TCOR en referencia al tipo de representación empleado, STAR indica que se utilizó el método descrito en la figura 1, UMB indica si el experimento consideró un umbral ALTO o BAJO en su configuración, y finalmente EXT indica el método empleado (CE o SC) para la extracción y construcción final del resumen.

4.4. Evaluación

Para la evaluación de los resultados se empleó ROUGE, un sistema automático para la evaluación de resúmenes propuesto por Lin y Hovy [8]. Este

sistema está basado en el método propuesto para la evaluación de traducciones automáticas BLEU, *i.e.*, en la co-ocurrencia de n-gramas de palabras. Lin y Hovy demuestran en [8] como este tipo de métricas pueden ser aplicados para evaluar la calidad de los resúmenes generados automáticamente.

ROUGE incluye diferentes métricas para evaluar ésta correlación, en particular nos enfocaremos en ROUGE-N, la cual fue utilizada para reportar nuestros resultados. ROUGE-N es un método de evaluación basado en el recuerdo entre un resumen "candidato" (resumen generado automáticamente) y un resumen "referencia" (generado por un experto humano). ROUGE-N es calculado por medio de:

$$ROUGE{-}N = \frac{\sum_{S_i \in \{ResumenReferencia\}} \sum_{gram_n \in S_i} Count_{match}(gram_n)}{\sum_{S_i \in \{ResumenReferencia\}} \sum_{gram_n \in S_i} Count(gram_n)}, \quad (5)$$

donde S_i se refiere a la oración i dentro del resumen de referencia, n es la longitud del n-grama, $gram_n$ y $Count_{match}(gram_n)$ es el máximo número de n-gramas que co-ocurren en el resumen candidato y el conjunto de resúmenes de referencia. La idea intuitiva detrás de esta medida de evaluación radica en comparar el resumen realizado por un sistema automático contra uno o varios resúmenes generados por un humano. Entre mayor recuerdo (elementos similares) haya entre el generado por el sistema y el generado por el humano, mejor se considera su desempeño. Para reportar nuestros resultados obtenidos se utilizó tanto ROUGE-1 como ROUGE-2.

5. Resultados

Las tablas 1 y 2 muestran los resultados obtenidos por los diferentes experimentos definidos en la sección 4. Las tablas muestran los valores de *Recuerdo*, *Precisión* y *F-score* obtenidos al aplicar ROUGE-N a los resúmenes generados de forma automática. Para evaluar nuestros experimentos se consideraron todos resúmenes de referencia proporcionados por los organizadores del DUC 2005 (*i.e.,* resúmenes generados por humanos, en promedio 4 por tópico).

Una primer observación de los experimentos realizados es que los métodos base propuestos (MB2 y MB3) superan de manera importante al método base propuesto por los organizadores del DUC (F-score de 0.33 contra 0.29 en ROUGE-1, tabla 1; F-score de 0.06 contra un 0.04 en ROUGE-2, tabla 2). Esto indica, hasta cierto punto, que es posible construir resúmenes relevantes a la consulta por medio de simplemente recuperar las oraciones más similares a la consulta. No obstante, el método propuesto resulta ser mejor, en particular la configuración que emplea una representación TCOR con un umbral de similitud alto y un esquema de extracción que no depende en los centros de las estrellas formadas por el algoritmo de agrupamiento, *i.e.*, TCOR-STAR-ALTO-SC. La ventaja de ésta técnica con respecto a los métodos base es que garantiza una mayor heterogeneidad en la información contenida en el resumen final gracias a la etapa de agrupamiento de información, etapa que ayuda en la eliminación de redundancias e identificación de información relevante.

Tabla 1. Resumen de los resultados experimentales empleando como métrica de evaluación ROUGE-1. Los datos reportados representan el desempeño promedio obtenido a través de los 50 tópicos; entre paréntesis se muestra el valor de la desviación estándar.

Nombre del experimento	**Recuerdo**	**Precisión**	**F-score**
Método Base 1 (MB1)	0.3014 (0.05)	0.2921 (0.05)	0.2966 (0.05)
Método Base 2 (MB2)	0.3396 (0.06)	0.3292 (0.05)	0.3342 (0.06)
Método Base 3 (MB3)	0.3374 (0.06)	0.3271 (0.06)	0.3320 (0.06)
BOW-STAR-ALTO-CE	0.3054 (0.08)	0.3252 (0.05)	0.3115 (0.06)
BOW-STAR-ALTO-SC	0.3365 (0.06)	0.3261 (0.05)	0.3311 (0.06)
BOW-STAR-BAJO-CE	0.1718 (0.05)	**0.3674** (0.06)	0.2278 (0.05)
BOW-STAR-BAJO-SC	0.3396 (0.06)	0.3289 (0.05)	0.3341 (0.05)
TCOR-STAR-ALTO-CE	0.3021 (0.04)	0.2925 (0.04)	0.2971 (0.04)
TCOR-STAR-ALTO-SC	**0.3599** (0.06)	0.3476 (0.06)	**0.3535** (0.06)
TCOR-STAR-BAJO-CE	0.1787 (0.07)	0.2643 (0.05)	0.2025 (0.06)
TCOR-STAR-BAJO-SC	0.3099 (0.09)	0.3613 (0.06)	0.3245 (0.07)

Note que el desempeño de la configuraciones que emplean un esquema de extracción tipo SC son en general mejores que su contra parte, es decir, superan al esquema de extracción basado 100 % en los centros de las estrellas. Este comportamiento respalda nuestra intuición respecto a la relevancia del centro de las estrellas. Como se mencionó en la sección 3.4, a pesar de que el algoritmo de agrupamiento garantiza que el centro de la estrella es el elemento más representativo, este elemento no es necesariamente el más relevante a la consulta, y la razón es simple, el algoritmo de agrupamiento no considera a la consulta en el proceso de agrupamiento.

Tabla 2. Resumen de los resultados experimentales empleando como métrica de evaluación ROUGE-2. Los datos reportados representan el desempeño promedio obtenido a través de los 50 tópicos; entre paréntesis se muestra el valor de la desviación estándar.

Nombre del experimento	**Precisión**	**Recuerdo**	**F-score**
Método Base 1 (MB1)	0.0466 (0.02)	0.0445 (0.02)	0.0455 (0.02)
Método Base 2 (MB2)	0.0668 (0.03)	0.0642 (0.03)	0.0655 (0.03)
Método Base 3 (MB3)	0.0667 (0.03)	0.0641 (0.03)	0.0654 (0.03)
BOW-STAR-ALTO-CE	0.0484 (0.03)	0.0499 (0.02)	0.0487 (0.02)
BOW-STAR-ALTO-SC	0.0666 (0.03)	0.0639 (0.03)	0.0652 (0.03)
BOW-STAR-BAJO-CE	0.0236 (0.01)	0.0503 (0.02)	0.0314 (0.01)
BOW-STAR-BAJO-SC	0.0682 (0.03)	0.0652 (0.03)	0.0666 (0.03)
TCOR-STAR-ALTO-CE	0.0416 (0.02)	0.0400 (0.02)	0.0408 (0.02)
TCOR-STAR-ALTO-SC	**0.0755** (0.04)	**0.0723** (0.03)	**0.0739** (0.04)
TCOR-STAR-BAJO-CE	0.0160 (0.01)	0.0241 (0.01)	0.0182 (0.01)
TCOR-STAR-BAJO-SC	0.0598 (0.03)	0.0687 (0.03)	0.0623 (0.03)

6. Conclusiones

En este trabajo hemos propuesto un método para la generación de resúmenes de múltiples documentos guiados por consulta, el cual emplea una técnica de representación distribucional para hacer la caracterización de la información. Contrario a las técnicas tradicionales de representación de textos, las representación utilizada permite obtener una aproximación semántica de los términos contenidos en una colección de documentos y, en consecuencia, es posible satisfacer de manera más efectiva las necesidades de información del usuario al momento de construir un resumen.

Los experimentos realizados sobre un conjunto estándar de evaluación demuestran que la metodología propuesta es pertinente para la tarea en mano. Específicamente se mostró que la representación TCOR permite obtener resultados que superan a los métodos base al mismo tiempo que a técnicas de representación tradicionales.

A pesar de los buenos resultados obtenidos, hace falta mucho trabajo por delante. Por ejemplo, nos interesa incorporar en la fase de agrupamiento información de la consulta, de manera que los grupos formados puedan ser más adecuados a la necesidad de información establecida por el usuario. Además, nos interesa realizar experimentos empleando otro tipo de técnicas de agrupamiento de forma que nos sea posible determinar la pertinencia del algoritmo seleccionado para los experimentos reportados en este trabajo.

Agradecimientos. El trabajo del primer autor fue parcialmente financiado por el CONACyT a través de la beca 615483. También se agradece el apoyo otorgado a través de la Coordinación de la Maestría en Diseño, Información y Comunicación (MADIC) de la UAM Cuajimalpa, así como al Departamento de Tecnologías de la Información de la UAM Cuajimalpa.

Referencias

1. Aslam, J., Pelekhov, K., Rus, D.: A practical clustering algorithm for static and dynamic information organization. In: Proceedings of the 1999 Symposium on Discrete Algorithms. pp. 208–217 (1999)
2. Baralis, E., Cagliero, L., Jabeen, S., Fiori, A.: Multi-document summarization exploiting frequent itemsets. In: Proceedings of the 27th Annual ACM Symposium on Applied Computing. pp. 782–786. ACM (2012)
3. Gambhir, M., Gupta, V.: Recent automatic text summarization techniques: a survey. Artificial Intelligence Review 47(1), 1–66 (2017), http://dx.doi.org/10.1007/s10462-016-9475-9
4. García-Hernández, R.A., Ledeneva, Y.: Word sequence models for single text summarization. In: 2009 Second International Conferences on Advances in Computer-Human Interactions. pp. 44–48 (Feb 2009)
5. Glavaš, G., Šnajder, J.: Event graphs for information retrieval and multi-document summarization. Expert Systems with Applications 41(15), 6904 – 6916 (2014), http://www.sciencedirect.com/science/article/pii/S0957417414001985

6. Gupta, V., Lehal, G.S.: A survey of text summarization extractive techniques. Journal of emerging technologies in web intelligence 2(3), 258–268 (2010)
7. Lavelli, A., Sebastiani, F., Zanoli, R.: Distributional term representations: an experimental comparison. In: Proceedings of the thirteenth ACM international conference on Information and knowledge management. pp. 615–624. ACM (2004)
8. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Proceedings of Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL. Barcelona, Spain (2004)
9. Torres-Moreno, J.M.: Automatic text summarization. John Wiley & Sons (2014)
10. Villatoro-Tello, E., Villaseñor-Pineda, L., Montes-y Gómez, M.: Using word sequences for text summarization. In: International Conference on Text, Speech and Dialogue. pp. 293–300. Springer (2006)
11. Yao, J.g., Wan, X., Xiao, J.: Compressive document summarization via sparse optimization. In: IJCAI. pp. 1376–1382 (2015)
12. Yeh, J.Y., Ke, H.R., Yang, W.P., Meng, I.H.: Text summarization using a trainable summarizer and latent semantic analysis. Information Processing & Management 41(1), 75 – 95 (2005), http://www.sciencedirect.com/science/article/pii/S0306457304000329, an Asian Digital Libraries Perspective
13. Zhang, Y., Zincir-Heywood, N., Milios, E.: Narrative text classification for automatic key phrase extraction in web document corpora. In: Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management. pp. 51–58. WIDM '05, ACM, New York, NY, USA (2005), http://doi.acm.org/10.1145/1097047.1097059