Procesamiento del Lenguaje Natural, Revista nº 62, marzo de 2019, pp. 45-52 recibido 22-11-2018 revisado 07-01-2019 aceptado 04-02-2019

Hacia una generación de resúmenes sin sesgo a partir de contenido generado por el usuario: Un enfoque preliminar

Towards Unbiased Automatic Summarization from User Generated Content: A Preliminary Approach

Alejandro Reyes[1], **Elena Lloret**[2]
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
Apdo. de Correos 99
E-03080, Alicante, Spain
[1]ara65@alu.ua.es, [2]elloret@dlsi.ua.es

Resumen: En este trabajo se propone un enfoque novedoso de generación automática de resúmenes capaz de sintetizar grandes cantidades de información generada por diferentes tipos de usuarios en Internet y producir un nuevo texto coherente que presente la información de forma objetiva, es decir, evitando proporcionar información parcial o sesgada, a la par que aportando múltiples perspectivas sobre el tema en cuestión. En concreto, el escenario en el que se enmarca esta investigación es el ámbito turístico, centrándonos en las opiniones sobre distintos aspectos de varios hoteles proporcionadas por 5 tipos de perfil de usuario. La evaluación realizada con usuarios demuestra que i) la calidad de los resúmenes generados es adecuada y ii) que este tipo de resúmenes ayudaría a los usuarios a tomar mejores decisiones.
Palabras clave: Generación de resúmenes, contenido generado por el usuario, resúmenes abstractivos, resúmenes multi-perspectiva, información sin sesgo

Abstract: In this paper a novel approach for automatic summarization is proposed. This approach is able to synthetize huge amounts of information generated by different types of users on the Internet and produce a new coherent text that presents the information in an objective way, i.e., avoiding biased information and giving multiple perspectives for an specific aspect/topic. This study is mainly focused on the tourism sector, especially on the opinions abour different topics existing in multiple hotels and given by 5 user types. The user evaluation conducted proves that i) the generated summaries have an appropriate quality, and ii) they would really help users to make better decisions.
Keywords: Automatic Summarization, User Generated Content, Abstractive Summarization, Multi-perspective Summarization, Unbiased information

1 Introducción

La aparición de la Web 2.0 ha supuesto un enorme aumento de la cantidad de datos a disposición de las personas. Este hecho, que a priori podría considerarse como una ventaja —a mayor información, mejor toma de decisiones—, en realidad supone una desventaja al no poder gestionar de forma eficaz y eficiente la sobrecarga de información disponible (Luo et al., 2013).

Por ejemplo, en el sector del turismo, nos encontramos con foros, redes sociales y otras plataformas que incluyen información que puede ser relevante para un usuario a la hora de decidir sobre su futuro viaje. Las opiniones, valoraciones proporcionadas por otros usuarios —"Contenido Generado por el Usuario" (CGU)— que han visitado el mismo lugar, así como otra información adicional en estos medios, suponen una gran ayuda a los usuarios a la hora de tomar decisiones. Sin embargo, resulta muy difícil sintetizar tal cantidad de información y, generalmente, el usuario acaba leyendo unas sólo unas cuantas opiniones, usualmente las primeras, pudiendo obtener así una idea sesgada o parcial, elabo-

ISSN 1135-5948. DOI 10.26342/2019-62-5

Hotel	#Coment. (03/2018)	#Coment. (11/2018)
Luxor Hotel & Casino Las Vegas	32,007	35,897
Melia Alicante	4,752	5,428

Tabla 1: Evolución en el número de comentarios de dos hoteles entre marzo y noviembre de 2018 (fuente de datos: TripAdvisor)

rada en base a un subconjunto de opiniones.

A modo ilustrativo, la Tabla 1 muestra el incremento en el número de comentarios generados por los usuarios de la página TripAdvisor[1] para dos hoteles elegidos al azar entre las fechas 5 de marzo y 5 de noviembre de 2018. Como se puede observar, el volumen de comentarios es considerablemente elevado, lo que hace inviable para una persona leer y analizar cada uno de ellos de forma manual.

Además, el hecho de encontrar opiniones diversas y contradictorias para un mismo aspecto hace que sea complicado para el usuario escoger cuáles de ellas se adecúan más a su perfil o necesidades. Esto se debe a que los comentarios que proporcionan los usuarios están basados en su experiencia, por lo que un aspecto específico puede ser malo para un usuario, mientras que para otro no. Por ejemplo, podemos ver que, dependiendo del tipo de viajero, algunos comentarios se contradicen para un mismo hotel: *"Mala ubicación, ya que está lejos del centro del Strip."*, *"Su ubicación y comunicación con otros centros de diversión [...]le hacen estratégica."*[2].

Por todo ello, la hipótesis de nuestro trabajo es que un enfoque que sea capaz de resumir objetivamente toda esta información, captando los distintos puntos de vista sobre, por ejemplo, los diferentes aspectos de un hotel específico de acuerdo al perfil del viajante, sería de gran ayuda para los usuarios que desean saber de forma fácil y sencilla cuál es la mejor opción de alojamiento para su viaje. En base a esta hipótesis, nos planteamos como objetivo principal el análisis y uso de técnicas de Procesamiento de Lenguaje Natural (PLN) para desarrollar un enfoque de generación de resúmenes abstractivo, multidocumento y multi-perspectiva, capaz de producir resúmenes que no contengan información sesgada o parcial. Los resultados obtenidos verifican la hipótesis de partida y demuestran que los resúmenes generados, además de ayudar en la toma de decisiones, son coherentes y están bien escritos.

[1] https://www.tripadvisor.es/

[2] https://www.tripadvisor.es/Hotel_Review-g45963-d111709-Reviews-Luxor_Hotel_Casino-Las_Vegas_Nevada.html

2 Estado de la cuestión

En la actualidad, los métodos y sistemas de resúmenes existentes, como por ejemplo el propuesto en (Paulus, Xiong, y Socher, 2017), se centran en la extracción y/o abstracción de la información más relevante de un texto, teniendo en cuenta varios factores, entre los que destacan la detección de la redundancia y la detección de polaridad, sobre todo cuando se generan resúmenes de opiniones o a partir de información subjetiva. Sin embargo, casi ninguno de ellos tiene en cuenta si existe información contradictoria sobre un aspecto específico ni tampoco el perfil de usuario para el que va dirigido el resumen. Por ello, en este artículo nos centraremos únicamente en aquellos enfoques que intentan ofrecer distintos puntos de vista sobre opiniones (resúmenes multiperspectiva), ya que son los más recientes relacionados con resúmenes a partir de CGU.

Lloret (2016) propone un método para la explotación de los metadatos existentes en la información procedente de CGU para enfocar la generación de resúmenes de distinta manera dependiendo de las necesidades del usuario en concreto. Para ello, realizó un estudio en el que terminó subdividiendo el proceso de generación de resúmenes en 3 fases: i) extracción de información básica; ii) identificación del tópico y de la polaridad; iii) selección de información relevante para la generación del resumen. La principal limitación de este trabajo es que solamente se queda a nivel de propuesta y no existe una implementación ni evaluación al respecto para poder determinar cuán útiles y buenos son los resúmenes generados.

Por otra parte, en el trabajo presentado por Esteban y Lloret (2017a) también se propone un sistema similar al de este trabajo, llamado TravelSum[3], que consiste en una aplicación web de la cual los usuarios pueden obtener un resumen generado automáticamente en español a partir de CGU (Esteban y Lloret, 2017b). La principal diferencia entre el enfoque propuesto en este artículo y TravelSum, es que TravelSum únicamente agrupa

[3] http://travelsum.gplsi.es/

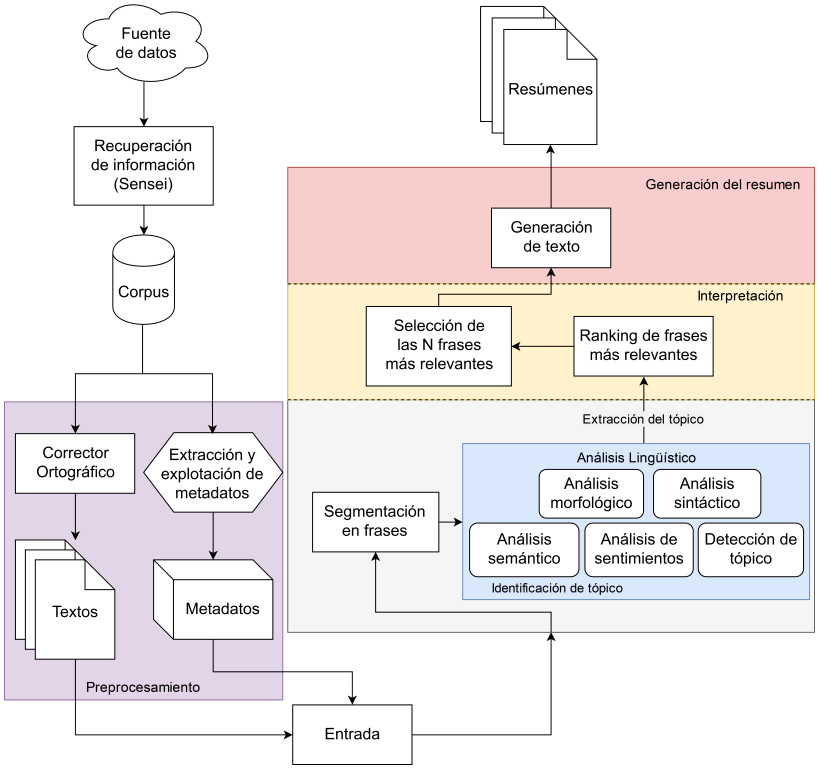

Figura 1: Arquitectura para el enfoque de resúmenes propuesto

frases por polaridad, mientras que el nuestro, además de aplicar un proceso de minería de opiniones, tiene en cuenta también la variedad y diferencias de opiniones de los usuarios para un mismo aspecto, así como el perfil de usuario que realiza un comentario (p. ej. si ha viajado solo, con amigos, etc.).

Entre las principales novedades de nuestro trabajo respecto a los comentados, destacan: i) mejora e implementación completa de un enfoque de resúmenes basada en la idea conceptual de Lloret (2016); i) la explotación e integración de los metadatos de los textos para el cálculo de la información relevante; y iii) análisis de la contradicción entre los distintos tópicos, para que los resúmenes generados no estén sesgados y/o se puedan adaptar a un perfil de usuario.

3 *Enfoque de generación de resúmenes sin sesgo*

La Figura 1 muestra la arquitectura del enfoque propuesto. Primero, se determina la fuente de información y se recopila el conjunto de documentos que servirán como entrada al enfoque de resúmenes. En nuestro caso, se decidió utilizar reseñas de hoteles, ya que las opiniones de los usuarios de una zona turística son variadas, y se recopiló, de forma automática utilizando el crawler SENSEI[4], un corpus compuesto por comentarios en inglés de 10 hoteles distintos, de una ciudad europea elegida al azar, junto con todos los metadatos disponibles (p.ej., utilidad del comentario, fecha de publicación, número de contribuciones o la valoración general del servicio). Finalmente, se obtuvo un total de 3615 comentarios, con una media de 340 comentarios por hotel.

Una vez recopilado el corpus, el siguiente paso es el preprocesado del corpus para corregir los errores ortográficos derivados del CGU. Para ello, se utiliza un diccionario en inglés[5] ligeramente modificado para la corrección de las erratas. También se extraen y guardan los metadatos del corpus.

Con esto, ya tendríamos la entrada preparada para el enfoque de generación de resúmenes, para cuyo diseño y desarrollo se tomó como base las tres fases propuestas por Hovy (2003), utilizando Python 2.7 como lenguaje de programación. A continuación, se explicará más en detalle como se ha aborda-

[4]https://gplsi.dlsi.ua.es/gplsi13/es/node/301

[5]https://introcs.cs.princeton.edu/java/data/words.utf-8.txt

do cada una de las fases y qué herramientas de PLN se han usado.

3.1 Identificación del Tópico

En esta fase se llevan a cabo distintos tipos de análisis lingüísticos para la detección e identificación de tópicos.

3.1.1 Análisis morfológico, sintáctico y semántico

Tras el preprocesado, se realiza una segmentación del texto en frases, a las cuales se les realiza un proceso de análisis morfológico y sintáctico, haciendo uso de la herramienta NLTK (Bird y Loper, 2004). Además, se realiza un análisis semántico, utilizando Wordnet (Miller, 1998), para agrupar las palabras por conceptos sinónimos (*synsets*), lo cual ayudará posteriormente a la detección de tópicos relevantes y al análisis de los sentimientos. Cabe mencionar que en esta propuesta no se realiza ningún proceso de desambiguación del sentido de las palabras, por lo que nos quedamos con el sentido más frecuente para cada palabra, puesto que se trata de un enfoque que, a pesar de su simplicidad, obtiene resultados bastante aceptables con respecto al estado de la cuestión para la tarea de desambiguación (McCarthy y Navigli, 2007).

3.1.2 Análisis de sentimientos

Analizar el sentimiento de una frase nos va a permitir conocer la opinión que un usuario ha proporcionado sobre un determinado aspecto y si esta opinión es la misma que la de otras frases o no. La polaridad de una frase se calculará sumando las polaridades individuales de las palabras que la componen, previa eliminación de las *stopwords*, gracias a la información que proporciona la herramienta SentiWordnet (Esuli y Sebastiani, 2006), que utiliza los *synsets* previamente obtenidos en el análisis semántico. Se decidió utilizar esta herramienta tras una investigación preliminar de las diferentes opciones existentes para el análisis de sentimientos, mostrados en trabajos como el de Denecke (2009).

Sin embargo, el análisis de sentimientos no nos permite saber si hablan sobre el mismo aspecto o tópico, y para ello necesitamos incluir un procesamiento adicional para la detección del tópico.

3.1.3 Detección de tópico

La detección del tópico de una frase es el proceso por el cual se obtiene de qué está hablando la frase. Generalmente, las frases escritas por usuarios para mostrar una opinión o hacer una reseña de algún producto o servicio tienden a hablar de varias cosas en la misma frase. Es por ello, que nuestro enfoque tiene en cuenta este hecho y obtiene los tópicos relevantes para, finalmente, determinar cuál es el tópico predominante para cada frase del texto/s que se quiere/n resumir.

Para realizar la detección de tópicos relevantes, primero se eliminan las partes de la frase que podrían obstaculizar el correcto análisis de los datos, tales como *stopwords* y signos de puntuación. Posteriormente, a partir de los lemas obtenidos de cada una de las palabras del texto, se conforma un diccionario del documento y se pasa a un modelo Latent Dirichlet Allocation (LDA), obtenido con la librería de Python Gensim (Řehůřek y Sojka, 2010), para obtener la lista ordenada de los tópicos más frecuentes de una frase.

A partir de la lista generada y ordenada por frecuencia de aparición, se realiza un filtrado en el que se descartan las palabras que no son sustantivos, ya que los verbos, adjetivos y adverbios no se pueden utilizar para la definición de una característica, al expresar acciones o utilizarse para describir o modificar un sustantivo o verbo. De este modo podemos seleccionar, del conjunto filtrado y ordenado por frecuencia, la palabra más comentada como el tópico predominante de la frase, almacenándolo para usarlo en los siguientes apartados.

3.2 Interpretación

En esta fase se toma la información de la etapa anterior y se interpreta para detectar la información más relevante que deberá conformar el resumen.

3.2.1 Ranking de relevancia

Para determinar la información relevante, se tuvieron en cuenta dos aspectos. Por un lado, se realizó un cálculo de la relevancia del autor del comentario en base a los metadatos del corpus utilizado, y por otro lado se aplicaron una serie de heurísticas que nos permitieron saber cuán relevante es una frase de una reseña con respecto a las otras.

La relevancia del autor la calculamos utilizando 3 metadatos principales: i) la fecha de publicación del comentario; ii) el número de contribuciones realizadas por el autor; y iii) la cantidad de comentarios de este autor que han sido considerados útiles por otros usuarios. En base a estos tres factores se es-

tablece una puntuación para cada metadato que se sumará para obtener un número positivo que representará la relevancia del autor en el corpus.

Para calcular la relevancia de una frase, se realizó previamente un análisis sobre los tópicos más relevantes para un tipo específico de usuario, así como un análisis de la contradicción existente entre varias frases que expresaban polaridades distintas al hablar sobre un tópico específico.

En concreto, para determinar la relevancia de un tópico, nos basamos en el número de apariciones de dicho tópico para un tipo de usuario concreto, de tal manera que, al final, tendremos una lista de tópicos relevantes para cada tipo de usuario individual, y una lista de tópicos relevantes en general, que aúna todos las anteriores.

Adicionalmente, se realizó un proceso de agrupación de los tópicos con el que obtuvimos una lista de palabras relacionadas utilizando el corpus de palabras vectorizadas de GloVe (Pennington, Socher, y Manning, 2014). Por ejemplo, para la palabra "bar" se detectaron palabras relacionadas como, por ejemplo, "cafe", "restaurant", etc.

Una vez obtenidas la lista de tópicos relevantes y la de tópicos agrupados, seleccionamos de esta última, aquellos tópicos que se encuentran dentro de la lista de los relevantes, de modo que podamos conocer si los tópicos de nuestras frases son relevantes o están relacionados con alguno de los relevantes.

Teniendo en cuenta la relevancia del tópico y la del autor, calculamos una puntuación de relevancia para una frase, la cual utilizaremos para ordenar las frases de mayor a menor según su relevancia para el usuario.

Además, gracias al análisis del sentimiento y la detección del tópico de las fases anteriores, podemos determinar cuándo dos frases están hablando sobre un tópico específico y diferenciar si la opinión hacia ese tópico es positiva o negativa, permitiéndonos así ajustar los resúmenes a varios perfiles de usuario y compararlos entre sí. Esto se realizó a partir de un conjunto de reglas, diseñadas de forma manual, que nos permitirán identificar opiniones contradictorias (con diferente polaridad) sobre un mismo tópico en los textos de entrada para el enfoque de resúmenes propuesto. Con estas reglas se confeccionaría un conjunto de tópicos y frases sobre los que no hay un consenso entre los usuarios, para utilizarlos a la hora de producir el resumen final.

3.2.2 Selección de frases relevantes

Una vez determinada la relevancia de cada frase, el siguiente paso consistió en seleccionar las más relevantes para incorporarlas al resumen final. Para ello, se decidió seleccionar las frases que se incluyeran en estas categorías, puesto que serían las secciones en las que se organizará el resumen a generar:

- Frase de introducción
- Opiniones positivas sobre el hotel
- Opiniones negativas sobre el hotel
- Aspectos más comentados del hotel
- Diversidad de opinión (Contradicción)
- Opiniones sobre aspectos específicos (Metadatos)

Dependiendo de la sección del resumen en la que nos encontremos la selección de frases variará. Algunas secciones utilizan la totalidad de las frases analizadas para sacar estadísticas sobre los tópicos, mientras que otras escogen entre las frases con mayor frecuencia basándose en el sentimiento que muestran, o devuelven las frases de los 3 tópicos relevantes más comentados junto con el porcentaje de aparición de estos en la totalidad del texto.

Por último, hay una parte concreta que selecciona frases contradictorias dentro del corpus para mostrar al usuario la existencia de diversidad de opiniones entre los datos recolectados y evitar así que el resumen esté sesgado.

Para todas estas selecciones se utilizan una serie de plantillas definidas para cada una de las partes del resumen, las cuales explicamos a continuación en la última fase del enfoque propuesto.

3.3 Generación del resumen

En esta última fase, utilizamos la información recolectada en la fase de Interpretación para generar un nuevo texto, basado en plantillas, que sintetice toda la información de los comentarios en un pequeño resumen abstractivo y proporcione la información de un modo claro y conciso.

Para ello, diseñamos y construimos varias plantillas, creadas a partir de reglas combinadas con lenguaje natural predefinido, y separadas por las secciones comentadas en el apartado anterior, que se montarán una tras

otra para la obtención del resumen final. En total se crearon 17 plantillas centradas en el ámbito de las opiniones de hoteles, que, combinadas, formarían un resumen adaptado a un tipo de usuario específico.

La primera parte del resumen consiste en una frase de introducción que nos presenta el tipo de usuario[6], el nombre del hotel y la opinión general del hotel con respecto a un tipo concreto de usuario. A continuación, se muestra un ejemplo, donde *travellerType*, X e Y serían valores a rellenar según los datos calculados.

(1) The *travellerType* who stayed at Y had a generally good opinion of this hotel, as the $X\,\%$ of the comments are positive.

Seguidamente se introduce el tópico más comentado entre los positivos, así como algunos ejemplos del mismo, que sustituirán a *"Y"*, *"XYZ"* y *"ZYX"*:

(2) The users seem comfortable talking about "Y" because they express positiveness on the X % of the comments related to this topic. For example, you can see that on comments like "XYZ" or "ZYX".

Del mismo modo, se muestra el tópico negativo con mayor aparición junto con frases en las que se ha detectado este tópico. Tras haber introducido, respectivamente, los tópicos positivo y negativo más comentados para el hotel y el tipo de usuario elegido, se comenta el número de frases analizadas (X) y el total de tópicos detectados (Y), así como los 3 tópicos más comentados entre todos los existentes (A, B, y C), mencionando también el porcentaje de su aparición en la totalidad de las frases analizadas (N, M y P):

(3) We can also see that, from the X reviews crawled for this type of traveler and the Y topics detected, the most commented aspects for this hotel were the "A", noticeable on the N % of the comments, "B" with a M % of appearance, and "C" with a P %.

Seguidamente introducimos la diversidad de opiniones (*"A"*, *"B"*) que existe entre frases que hablan sobre el mismo tópico para un hotel específico, mostrando algunos ejemplos (*"ABC"*, *"CBA"*) y el porcentaje de tópicos contradictorios ($X\,\%$) que hay en el corpus:

(4) This type of traveller has shown different opinions about some topics like "A" and "B". For the topic "A" they comment "ABC" and also "CBA". The same occurs for the X % of the previously commented topics on this hotel, so you cannot establish a definitive conclusion about them.

Por último, mostramos una lista de los aspectos específicos de un hotel con la puntuación media obtenida entre todos los usuarios, y que hemos extraído y calculado de los metadatos.

(5) Finally, we list below the average scores of some especific aspects of the hotel that have been ranked by the users: Service: 2/5 Cleanliness: 3/5 Value: 3/5 Sleep Quality: 3/5 Rooms: 2/5 Location: 2/5.

De esta manera juntando todos los ejemplos anteriores, tendríamos el resumen final generado por el enfoque propuesto.

4 Experimentación y evaluación

A partir del corpus recopilado y del enfoque de resúmenes propuesto, se generaron un total de 50 resúmenes en inglés, 5 resúmenes por cada perfil de usuario de cada hotel del corpus.

Con el objetivo de verificar si los resúmenes generados eran adecuados y útiles, se realizó una evaluación preliminar de forma manual, en la que participaron 10 evaluadores con altos conocimientos de inglés. Los evaluadores recibieron una encuesta[7] que debían rellenar tras haber leído los resúmenes y cuyas preguntas estaban relacionadas con sus hábitos a la hora de buscar información sobre un hotel y elegirlo, la coherencia, corrección ortográfica y gramatical y utilidad de los resúmenes generados. Las preguntas relacionadas con la evaluación directa de los resúmenes generados se respondieron siguiendo una escala de Likert de 5 puntos. Finalmente, en la última pregunta del cuestionario de evaluación quisimos conocer cómo de fácil sería descubrir si el resumen generado había sido realizado por una persona o una máquina, también conocido como Test de Turing.

[6] En nuestro enfoque trabajamos con cinco perfiles de usuarios: personas que viajan en familia, en pareja, solos, con amigos o que viajan por trabajo.

[7] https://goo.gl/forms/HyXFKb8E6DGFAWJx2

Según los resultados obtenidos en esta evaluación, el 90 % de los usuarios utiliza Internet para informarse sobre los hoteles en los que se alojarán durante su viaje, frente a un 10 % que prefiere preguntar directamente a amigos o familiares.

Los evaluadores también consideraron útil la posibilidad de disponer de un resumen de los comentarios de TripAdvisor, ya que el 70 % de ellos consideran que la gran cantidad de información existente hace imposible su completa lectura y comprensión.

La Figura 2 muestra la puntuación que los usuarios dieron a los resúmenes generados por el sistema propuesto, utilizando en las estadísticas los 50 resúmenes generados. Puntuaciones más altas reflejan que el resumen generado tiene mayor coherencia. Podemos observar que la mayoría de los resúmenes han sido puntuados con más de un 3 en coherencia, lo que indica que los evaluadores los han encontrado entendibles y con una estructura adecuada.

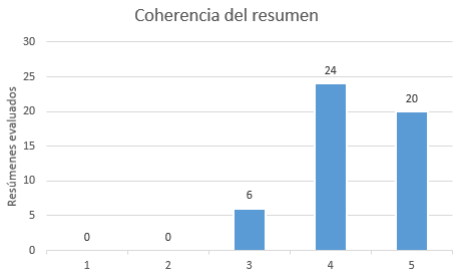

Figura 2: Grado de coherencia en los resúmenes según los evaluadores

Por otra parte, en la Figura 3 se muestra la evaluación de la falta de errores existentes en los resúmenes. A menor cantidad de errores, mayor puntuación para el resumen. Podemos observar también que la mayoría de los resúmenes han obtenido una buena puntuación con respecto a la cantidad de errores que presentan, lo que ayuda a que los usuarios los comprendan con mayor facilidad.

Sobre la utilidad de los resúmenes generados por el enfoque propuesto, los resultados de la Figura 4 indican que los evaluadores consideran los resúmenes adecuados para la toma de decisiones, lo que significa que a pesar de haber sido realizados de forma automático, serían de gran ayuda para sintetizar la información disponible y poder proporcionar una idea del hotel en cuestión.

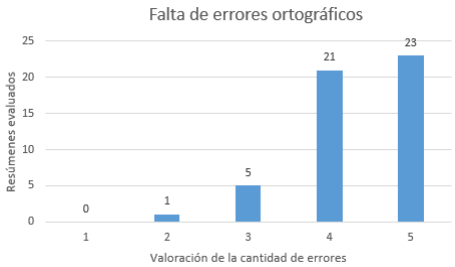

Figura 3: Ausencia de errores ortográficos según los evaluadores

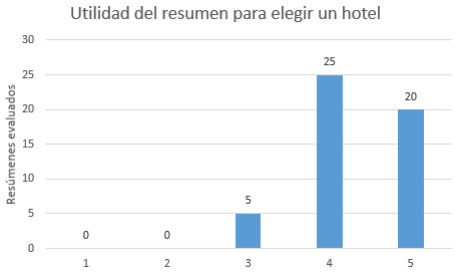

Figura 4: Grado de utilidad del resumen a la hora de elección de un hotel según los evaluadores

Sobre el Test de Turing realizado, en un 48 % de los casos, los evaluadores no consiguieron distinguir correctamente si el texto había sido escrito por un humano o no, hecho que nos indica que el enfoque de generación propuesto, en su estado actual, es bastante competitivo y tiene potencial para ser integrado en una aplicación real.

Sin embargo, al evaluar únicamente 5 resúmenes por cada hotel es difícil que los evaluadores no se den cuenta de los resúmenes han sido generados de forma automática, ya que todos siguen una estructura y un patrón similar.

Por último, remarcar que, además de la evaluación de los usuarios, también se realizó un análisis en detalle de cada uno de los resúmenes generados, para detectar las limitaciones de la propuesta y posibles áreas de mejora como, por ejemplo, la falta de detección de aspectos característicos que sean poco mencionados o la posibilidad de crear el texto sin el uso de plantillas.

5 Conclusión y trabajo Futuro

En este artículo se ha propuesto un enfoque de generación de resúmenes a partir de con-

tenido generado por usuario, con el fin de proporcionar al usuario un texto que aúne los diversos puntos de vista de las opiniones existentes y evitar, así, posibles sesgos que influyan a la hora de tomar decisiones. Concretamente, nos hemos centrado en opiniones sobre hoteles, recopilando un corpus de prueba, y aplicando técnicas de análisis lingüístico a distintos niveles junto con heurísticas para identificar información relevante y seleccionar aquella que podría ser más importante para el usuario.

Los resultados obtenidos a partir de la evaluación de los resúmenes generados han sido satisfactorios, por lo que se puede concluir que el enfoque propuesto es capaz de generar resúmenes de una calidad suficientemente buena para que los usuarios queden satisfechos. Sin embargo, el enfoque propuesto se podría mejorar integrando algunos aspectos que nos planteamos como trabajo futuro. El primer aspecto a corto plazo sería mejorar el análisis de sentimientos, teniendo en cuenta la negación, puesto que este fenómeno puede cambiar completamente la polaridad de una frase. A medio y largo plazo, nos gustaría adaptar el enfoque propuesto para el español, y probarlo y evaluarlo, en inglés y español, no sólo con reseñas de hoteles, sino también en otros ámbitos similares como opiniones sobre restaurantes, productos o servicios, de los que existen gran cantidad de información y mucha subjetividad en función de la experiencia.

Agradecimientos

Este proyecto ha sido financiado parcialmente por la Generalitat Valenciana a través del proyecto PROMETEU/2018/089.

Bibliografía

Bird, S. y E. Loper. 2004. NLTK: the natural language toolkit. En *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, página 31. Association for Computational Linguistics.

Denecke, K. 2009. Are SentiWordNet scores suited for multi-domain sentiment classification? En *Proceedings of 4th International Conference on Digital Information Management*, páginas 1–6. IEEE.

Esteban, A. y E. Lloret. 2017a. Propuesta y desarrollo de una aproximación de generación de resúmenes abstractivos multigénero. *Procesamiento del Lenguaje Natural*, 58:53–60.

Esteban, A. y E. Lloret. 2017b. TravelSum: A spanish summarization application focused on the tourism sector. *Procesamiento del Lenguaje Natural*, 59:159–162.

Esuli, A. y F. Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. En *Proceedings of the 5th International Conference on Language Resources and Evaluation*, páginas 417–422.

Hovy, E. 2003. Text summarization. En *The Oxford Handbook of Computational Linguistics 2nd edition*. Oxford University Press.

Lloret, E. 2016. Introducing the key Stages for Addressing Multi-perspective Summarization. En *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, páginas 321–326.

Luo, C., Y. Lan, C. Wang, y L. Ma. 2013. The effect of information consistency and information aggregation on eWOM readers' perception of information overload. En *Proceedings of Pacific Asia Conference on Information Systems*, página 180.

McCarthy, D. y R. Navigli. 2007. Word sense disambiguation: An overview. *Proceedings of the 4th International Workshop on Semantic Evaluations*, páginas 7–12.

Miller, G. 1998. *WordNet: An electronic lexical database*. MIT press.

Paulus, R., C. Xiong, y R. Socher. 2017. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*.

Pennington, J., R. Socher, y C. D. Manning. 2014. Glove: Global vectors for word representation. En *Proceedings of Empirical Methods in Natural Language Processing*, páginas 1532–1543.

Řehůřek, R. y P. Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. En *Proceedings of the LREC Workshop on New Challenges for NLP Frameworks*, páginas 45–50.