

Evaluación de las herramientas comerciales y métodos del estado del arte para la generación de resúmenes extractivos individuales

Griselda Areli Matias Mendoza¹, Yulia Ledeneva¹,
René Arnulfo García-Hernández¹, Grigori Sidorov²

¹ Universidad Autónoma del Estado de México
Unidad Académica Profesional Tianguistenco
Instituto Literario #100, Col. Centro, Toluca, 50000, Estado de México
griselda.matias.mendoza@gmail.com, yledeneva@yahoo.com,
rearnulfo@hotmail.com

² Laboratorio de Lenguaje Natural y Procesamiento de Texto,
Centro de Investigación en Computación, Instituto Politécnico Nacional, DF 07738,
México
sidorov@cic.ipn.mx

Resumen. Hoy en día se tiene conocimiento del avance significativo de las herramientas y métodos del estado de arte para la generación de resúmenes extractivos individuales (GREI). Para evaluar la calidad de las herramientas y los métodos de GREI es posible comparar los resúmenes generados automáticamente con los resúmenes generados por humanos. Utilizando el mismo corpus (DUC-2002) y el mismo método de evaluación independiente (ROUGE), en este trabajo se evaluaron seis herramientas comerciales y siete métodos del estado del arte que se desarrollaron para GREI. Con respecto a las herramientas comerciales de GREI, interesantemente se puede observar el avance significativo de las herramientas en línea con respecto a las herramientas instalables.

Palabras Clave: Resumen automático, Svhoong Summarizer, Pertinence Summarizer, Tool4noobs Summarizer, Open Text Summarizer.

1 Introducción

Actualmente la información en formato digital se ha ido incrementando de forma exponencial, se estima que crece a un ritmo de 57% anual [1]. El incremento acelerado de la información dificulta la obtención de documentos sobre temas específicos. Por lo que sería conveniente que los documentos recuperados de una consulta en Internet fueran relevantes y no redundantes [2]. Sin embargo, normalmente no es así, por lo que es necesario contar con herramientas de procesamiento automático de texto para generar resúmenes.

Un resumen tiene como objetivo presentar la información más importante en un texto más corto que el original; manteniendo las ideas principales [3]. Cuando se tiene como entrada un documento se puede generar un solo resumen, pero si la entrada fuera una colección de documentos se puede generar un resumen de todos ellos [4]. Según por su estrategia de condensación, los resúmenes se clasifican en extractivos y abstractivos [5].

Un resumen por extracción es aquel que está compuesto íntegramente por partes del documento original. Por el contrario, un resumen por abstracción puede contener texto que no se encuentra en el documento original. Generalmente, los humanos realizan resúmenes de tipo abstractivo, mientras que la mayoría de las herramientas comerciales y también los métodos del estado del arte generan resúmenes de forma extractiva.

Las herramientas comerciales que permiten la generación de resúmenes extractivos individuales (GREI) se pueden instalar en equipos propios o ejecutar en línea desde Internet. Algunas de las herramientas instalables son *Copernic Summarizer* [6] y Microsoft Office Word. Entre las herramientas en línea se encuentran *Svhoong Summarizer* [7], *Pertinence Summarizer* [8], *Tool4noobs Summarizer* [9] y *Open Text Summarizer* [10].

Además de las herramientas comerciales se tienen métodos propuestos de investigación sobre GREI, entre los cuales se encuentran, Secuencias Frecuentes Maximales (SFM) [4], agrupamiento con Secuencias Frecuentes Maximales [11], SFM (1best first)[12], *TextRank* [13] y por último los basados en algoritmos genéticos [5].

También se tienen las heurísticas *Baseline* [14] y *Baseline:aleatorio* [15] las cuales se utilizan para medir el avance que presentan los métodos del estado del arte. Sin embargo, poco se sabe de la calidad que tienen las herramientas comerciales en comparación con los métodos propuestos en el estado del arte de GREI. En especial, las herramientas en línea son de interés para la investigación de GREI porque su diseño debe ser eficiente tanto en espacio como en tiempo ya que pueden ser ejecutados simultáneamente por varios usuarios. En cambio las herramientas instalables tienen métodos exhaustivos que pueden disponer de todos los recursos de la computadora para su trabajo.

Por ello, el interés en este trabajo de conocer la calidad de las herramientas en línea de GREI en comparación con las herramientas instalables y con los métodos del estado del arte.

Es posible conocer cuál es la calidad de las herramientas en comparación con los métodos del estado del arte, si se utiliza la misma colección de documentos y se evalúan sus resúmenes generados en comparación con los humanos. En este trabajo se utilizó la colección de documentos estándar DUC-2002, la cual está compuesta por 567 noticias en inglés de diferentes temas y la herramienta de evaluación ROUGE, que mide el parecido del resumen automático en comparación con el generado por el humano.

El artículo está organizado de la siguiente forma. En la sección 2, se describe el trabajo relacionado a este artículo. En la sección 3 se describen las

herramientas en línea que se utilizan en este trabajo para la comparación de GREI.

En la sección 4 se describe la colección de documentos y la herramienta de evaluación que se utilizará. En la sección 5 se muestran los resultados de los experimentos realizados con las herramientas comerciales en comparación con los métodos del estado del arte. Finalmente en la sección-6 se muestran las conclusiones del trabajo realizado.

2 Trabajo Relacionado

En el área del Procesamiento del Lenguaje Natural ha habido un gran avance en la creación de nuevos métodos de GREI. Sin embargo, poco se sabe sobre la calidad de las herramientas comerciales para la comparación de GREI. Se tiene conocimiento de trabajos como el de [16] y el de [17] donde se comparan herramientas comerciales. En el trabajo de [16] se comparan los resúmenes de una colección de artículos científicos por medio de las herramientas comerciales: *Microsoft Word*, *SweSum*, *Shvoong* y *Online Brevity Document Summarizer*. La forma de evaluación que se utiliza en (Vlainic, et al, 2013), consiste en que un grupo de expertos son los que comparan los resúmenes generados por las herramientas con el resumen del artículo. Cabe mencionar que los trabajos generados por los métodos del estado del arte utilizan en su mayoría colecciones de documentos estándar. Por lo que en este trabajo solo se tomará como referencia el de García [17], donde utiliza la colección de documentos DUC-2002 y la herramienta de evaluación ROUGE.

Las herramientas comerciales que se utilizaron en el trabajo de García [17] son: *Copernic Summarizer* y *Microsoft Office Word* en sus versiones 2003¹ y 2007²². *Copernic Summarizer* es un software diseñado exclusivamente para la tarea de GREI, el cual trabaja con cuatro lenguajes (inglés, alemán, francés y español) [18]. *Microsoft Office Word* es una suite ofimática para el procesamiento y edición de texto que incluye la opción de GREI.

Entre los métodos del arte que se compararon en el trabajo de García [17] están los métodos estadísticos, los cuales son: Secuencias Frecuentes Maximales (SFMs) [4], agrupamiento con Secuencias Frecuentes Maximales [11], SFM (1best first) [12] y ponderación basada en grafos *TextRank* [13]. También se comparó con los métodos heurísticos llamados *Baseline* y *Baseline:aleatorio*; los cuales sirven de referencia para medir el avance que presentan los métodos del estado del arte. *Baseline* es una heurística que consiste en tomar las primeras n líneas u oraciones del texto para generar el

¹ Microsoft ® Office Word 2003. SP3 Parte de Microsoft Office Professional Edition 2003 Copyright © 1983-2003 Microsoft Corporation.

² Microsoft ® Office Word 2007. Parte de Microsoft Office Professional 2007 © 2006 Microsoft Corporation.

resumen [14]. *Baseline:aleatorio* es una heurística cuyo funcionamiento consiste en tomar algunas oraciones del texto al azar [15]. Por lo que cualquier método que se comporte como *Baseline:aleatorio* no tendría razón de ser.

En el gráfico 1 se pueden observar la comparación de las herramientas instalables y los métodos del estado del arte presentados en [17].

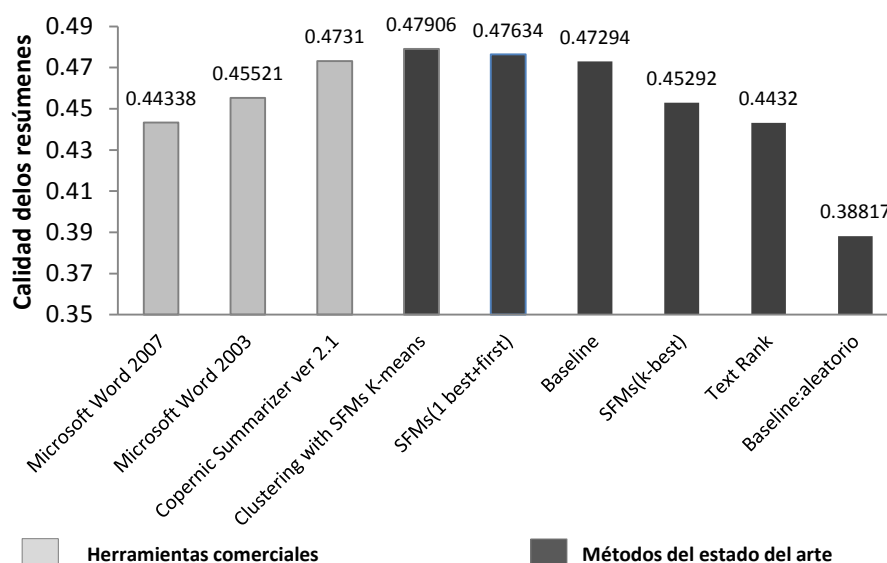


Gráfico 1. Resultados obtenidos con ROUGE 1.5.5 para la colección de resúmenes obtenidos por las herramientas comerciales y los métodos del estado del arte [17].

Tomando como base la heurística *baseline:aleatorio*, se puede observar en el gráfico 1 que tanto las herramientas comerciales instalables y los métodos del estado del arte superan esta heurística. Sin embargo, tomando como base la *Baseline*, se puede observar que la herramienta comercial instalable *Copernic Summarizer* es la única que supera ésta heurística. Los métodos del estado del arte que superan a *Baseline* son: agrupamiento con SFMs y SFMs(1 best first), lo que nos permite observar el avance significativo de los métodos propuestos en el estado de arte en comparación con las herramientas comerciales instalables.

Tomando como base la heurística *baseline:aleatorio*, se puede observar en el gráfico 1 que tanto las herramientas comerciales instalables y los métodos del estado del arte superan esta heurística. Sin embargo, tomando como base la *Baseline*, se puede observar que la herramienta comercial instalable *Copernic Summarizer* es la única que supera ésta heurística. Los métodos del estado del arte que superan a *Baseline* son: agrupamiento con SFMs y SFMs(1 best first),

4 Experimentación

Para poder realizar la comparación entre las herramientas en línea y los métodos del estado del arte de GREI se utilizará la colección *Document Understanding Conference* (DUC, 2002) [20]. DUC-2002 fue creada por *National Institute of Standards and Technology (NITS)* para el uso de los investigadores en el área de generación de resúmenes, la cual está compuesta por 567 noticias en inglés de diversas longitudes y sobre diferentes temas. Cada noticia de DUC-2002 tiene dos resúmenes de 100 palabras creados por dos expertos humanos.

Para evaluar los resúmenes generados automáticamente por las herramientas comerciales se va a utilizar ROUGE 1.5.5 [21], otra posible manera de evaluar los resúmenes es presentada en [22]. La evaluación consiste en estimar el parecido de los resúmenes generados automáticamente con los dos resúmenes realizados por los expertos humanos.

Los resúmenes generados por las herramientas comerciales instalables así como las herramientas en línea fueron generados con un mínimo de 100 palabras, por lo que se analizó cada herramienta para satisfacer la longitud mínima del resumen automático.

Con el objetivo de saber cuál es la calidad de las herramientas en línea de GREI se compararon las siguientes herramientas *Svhoong Summarizer*, *Pertinence Summarizer*, *Tool4noobs Summarizer* y *Open Text Summarizer (OTS)*.

4.1 Evaluación de las herramientas en línea en comparación con las herramientas comerciales instalables

En este trabajo se compararon cuatro herramientas en línea y dos herramientas instalables. En el gráfico 2 se puede observar que las herramientas en línea *Shvoong* y *OTS* obtuvieron los mejores resultados que las herramientas de Microsoft Office Word. Sin embargo, las otras herramientas en línea *tools4Noons* y *Pertinence* fueron las que obtuvieron los más bajos resultados. No obstante el mejor resultado lo obtuvo la herramienta instalable *Copernic Summarizer*.

Cabe mencionar que en el trabajo realizado por García [17] se hicieron experimentos con la herramienta instalable *Microsoft Office Word* en sus versiones 2003 y 2007, pero los resúmenes que generó *Microsoft Office Word* dependieron del sistema operativo. Para completar el trabajo anterior, en este trabajo se generaron resúmenes con *Microsoft Office Word* en sus versiones 2003 y 2007 con el sistema operativo *Windows 7 Ultimate*.

Los resultados obtenidos con *Microsoft Office Word* instalado en *Windows 7 Ultimate* no superaron a los resúmenes en las versiones 2003 y 2007 en el sistema operativo *Windows Vista*. Por lo que el valor que se considerará para la comparación de la herramienta instalable *Microsoft Office Word* con los

métodos del estado del arte y las herramientas en línea, será el obtenido con la versión 2003 con el sistema operativo Windows Vista.

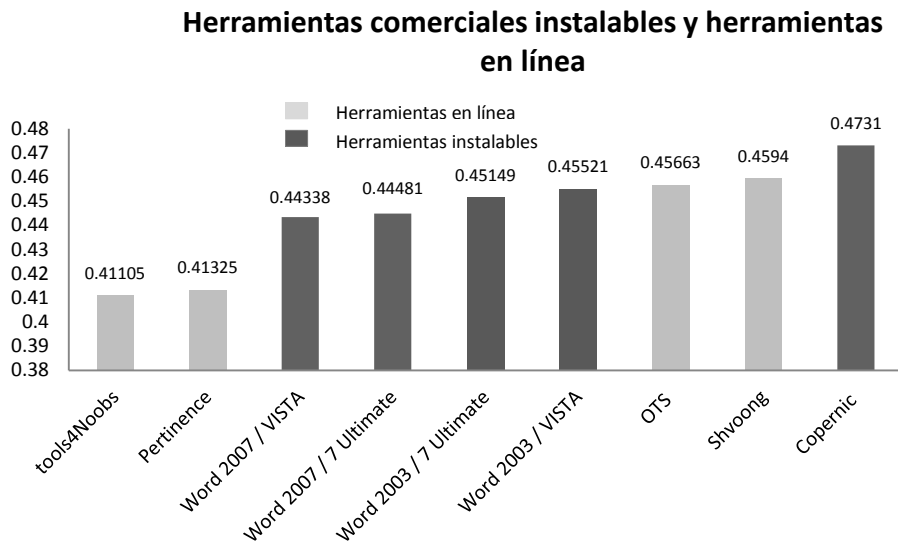


Gráfico 2. Comparación de las herramientas comerciales instalables y en línea

4.2 Evaluación de las herramientas comerciales y los métodos del estado del arte

Con el objetivo de conocer el avance que han tenido las herramientas comerciales en comparación con los métodos del estado del arte, se incluyeron los resultados anteriores junto con siete métodos del estado del arte en el gráfico 3.

En el gráfico 3 se puede observar que los resultados de las herramientas comerciales están por debajo de algunos métodos propuestos en el estado del arte. Es decir los métodos del estado de arte son de buena calidad.

Una de las heurísticas a superar por las herramientas comerciales instalables y en línea, así como por los métodos propuestos en los estados del arte es *Baseline*. Como se puede observar en el gráfico 3 solamente *Copernic Summarizer* (herramienta comercial instalable) supera esta heurística, y sólo *Clustering* con *SFM's*, *SMF's (1 best+first)* y *AG-Bigramas* métodos propuestos en el estado del arte superaron esta heurística.

Cabe mencionar que aunque las herramientas en línea no superan a la heurística de *Baseline*, algunas de ellas como *OTS* y *Shvoong* si están por encima de los resultados de la herramienta *Microsoft Office Word* y de los métodos del estado del arte *TextRank* y *SFM(K-best)*.

Evaluación de herramientas comerciales y métodos del estado del arte de GREI

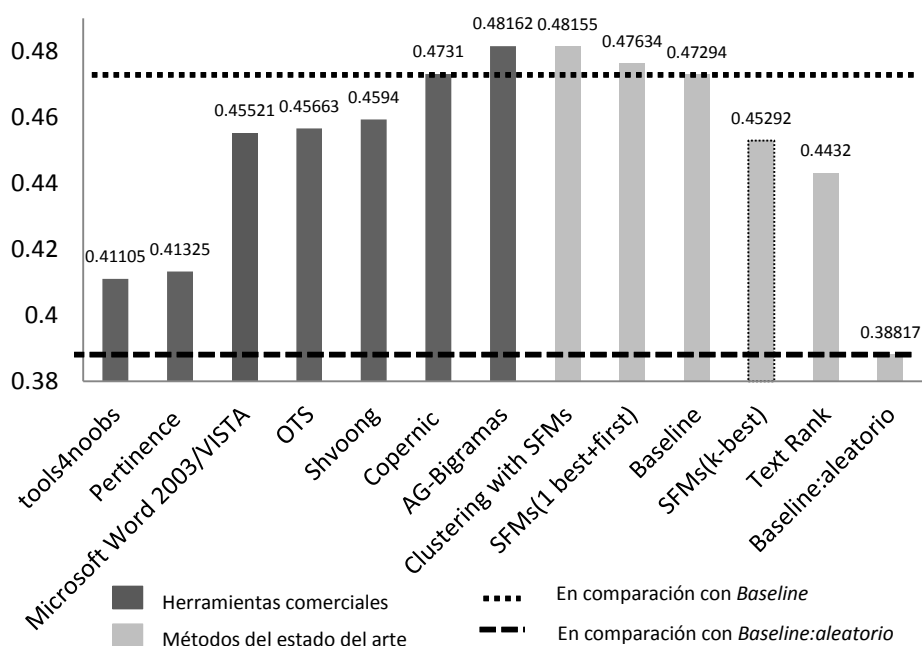


Gráfico 3. Evaluación de herramientas comerciales y métodos del estado del arte de GREI

5 Conclusiones

En este trabajo se evaluó la calidad de seis herramientas comerciales de generación de resúmenes extractivos individuales, utilizando la misma colección de noticias DUC-2002 y la misma herramienta de comparación ROUGE. La forma de trabajo anterior permitió comparar los resultados de las herramientas comerciales con siete métodos del estado del arte que utilizan la misma colección y la misma herramienta de comparación.

Por lo tanto, este trabajo contribuye con un estado del arte más amplio sobre la generación automática de resúmenes contemplando las aplicaciones y los métodos desarrollados sobre esta tarea.

En particular, se encontró que de las cuatro herramientas en línea para GREI, la mejor fue *Shvoong Summarizer*. Sin embargo, ninguna de las cuatro herramientas en línea de GREI superó a la heurística *Baseline*. Como resultado de la comparación de todas las herramientas comerciales, tanto herramientas en

línea como instalables, se observó que *Copernic Summarizer* sigue siendo la única herramienta comercial de GREI que supera a *Baseline*.

De la comparación que se muestra en el gráfico 3 se puede observar que son más los métodos del estado del arte que superan a la heurística *Baseline*. Sin embargo, aunque algunas de las herramientas comerciales no superan a *Baseline* la diferencia que existe entre estas herramientas y los métodos del estado del arte no es muy grande.

Este trabajo nos da un panorama considerando una sola colección en el lenguaje inglés, sería interesante ver si los resultados se comportarían igual considerando otros idiomas y colecciones con dominios diferentes.

Referencias

1. John F. Gantz, David Reinsel, Christopher Chute, Wolfgang Schlichting, John McArthur, Stephen Minton et al. (2007). A Forecast of Worldwide Information Growth through 2010. *The Expanding Digital Universe*, 4.
2. Laura Plaza Morales Uso de Grafos Semánticos en la Generación Automática de Resúmenes y Estudio de su Aplicación en Distintos Dominios: Biomedicina, Periodismo y Turismo. Madrid, España: Tesis presentada en la Facultad de Informática de la Universidad Complutense de Madrid, para obtención del grado de Doctor, 2010.
3. Saunmali, Ladda, Salim, Naomie y Binwahlan, Mohammed Salem, Genetic algorithm based sentence extraction for text summarization. 2011, *International Journal of Innovative Computing*, vol. 1, p. 2180-4370.
4. Yulia Ledeneva. Automatic Language-Independent Detection of Multiword Descriptions for Text Summarization. México. D.F.: Tesis presentada en el Instituto Politécnico Nacional, para obtención del grado de Doctor, 2008.
5. Griselda Areli Matias Mendoza, Generación de Resúmenes Automáticos Usando Algoritmos Genéticos, 2013. Edo de México. Tesis de licenciatura; Unidad Académica Profesional Tianguistenco de la Universidad Autónoma del Estado de México.
6. Copernic Inc. Copernic Inc. Recuperado el 28 de 10 de 2013, de la página principal de copernic. [En línea] <http://www.copernic.com/en/products/summarizer/>.
7. Svhoong. Svhoong Summarizer. Recuperado el 20 de 11 de 2013, de la página principal de Svhoong. [En línea] <http://es.shvoong.com/summarizer/>.
8. Pertinence. Pertinence Summarizer. Recuperado el 28 de 14 de 2010, de la página principal de pertinence . [En línea] http://pertinence.net/index_en.html.
9. Tools4noobs. Tools4noobs Summarizer. Recuperado el 2 de 11 de 2013, de la página principal de tools4noobs. [En línea] <http://www.tools4noobs.com/summarize/>.
10. OTS. Open Text Summarizer. Recuperado el 10 de 12 de 2013, de la página principal de OTS. [En línea] <http://libots.sourceforge.net/>.
11. Yulia Ledeneva, René García Hernández, Romyna Montiel Soto, Rafael Cruz Reyes, Alexander Gelbukh. EM Clustering Algorithm for Automatic Text

- Summarization. 2011, Springer, Lecture Notes in Artificial Intelligence, N 7094, p. 305–315.
12. Yulia Ledeneva, Alexander Gelbukh, René García Hernández. Terms Derived from Frequent Sequences for Extractive Text Summarization. Lecture Notes in Artificial Intelligence, N 4919, 2008, Springer, pp. 593–604.
 13. Rada Mihalcea, Graph-based Ranking Algorithms for Sentence Extraction. Texas; EUA: University of North Texas, 2004, Applied to Text Summarization; Department of Computer Science.
 14. Esaú Villatoro Tello. Generación automática de resúmenes de múltiples documentos. Puebla: Tesis de Maestría; Instituto Nacional de Astrofísica, Óptica y Electrónica, 2006.
 15. René Arnulfo García Hernández, Romyna Montiel, Yulia Ledeneva, Eréndira Rendón, Alexander Gelbukh, Rafael Cruz. Text Summarization by Sentence Extraction Using Unsupervised Learning. 2008, MICAI 2008; Lecture Notes in Artificial Intelligence, vol 5317, Springer, pp. 133–143.
 16. Marta Vlainic, Nieves Mikelic Preradovic. A Comparative Study of Automatic Text Summarization System Performance, Proceedings of the 7th European Computing. 2013, pp. 222–227.
 17. René Arnulfo García Hernández, Yulia Ledeneva, Griselda Areli Matias Mendoza, Ángel Hernández Domínguez, Jorge Chávez, Alexander Gelbukh, et al. Comparing Commercial Tools and State-of-the-Art Methods for Generating Text Summaries. 2009, IEEE Computer Society Press, pp. 92–96.
 18. Copernic. Copernic Summarizer. Recuperado el 10 de 01 de 2014, de Copernic Summarization Technologies White Paper. [En línea] <http://www.copernic.com/data/pdf/summarization-whitepaper-eng.pdf>.
 19. Andreas Gohr, Splitbrain.org. Recuperado el 20 de enero de 2014, de Electronic brain surgery science, 2001. [En línea] Open Text Summarizer, 2001–2013, <http://www.splitbrain.org/services/ots>.
 20. DUC. Recuperado el 01 de 02 de 2014, de Document Understanding Conference, 2002. [En línea] 2002. <http://www-nlpir.nist.gov/projects/duc/guidelines/2002.html>.
 21. Chin-yew Lin, ROUGE: A package for Automatic Evaluation of Summaries. Spain: 2004, Proceedings of Workshop on Text Summarization of ACL.
 22. Pinaki Bhaskar, Partha Pakray, Alexander Gelbukh, Sivaji Bandyopadhyay. Entailment-based Fully Automatic Technique for Evaluation of Summaries. Research in Computing Science, Vol. 65, 2013, pp. 11–23.