

Evaluación de resúmenes automáticos con y sin resúmenes de referencia para el idioma inglés

Selene Arai Vilchis Sepúlveda, Yulia Ledeneva

Universidad Autónoma del Estado de México,
Estado de México
selene.vilchis96@hotmail.com, yledeneva@yahoo.com

Resumen. Durante casi 60 años se han realizado diversas investigaciones sobre diferentes métodos y herramientas comerciales para generación automática de resúmenes de texto, en consecuencia, también se ha indagado en los diferentes métodos de evaluación de resúmenes de texto automáticos. Entre los más utilizados en el estado de arte encontramos los evaluadores ROUGE, mismos que hacen la comparación de un resumen generado de manera automática contra uno generado por un humano. Por otro lado, encontramos evaluadores más robustos como Jensen-Shannon y ROUGE-C que evalúan el resumen generado de manera automática contra el documento original (debido a la ausencia de resúmenes generados por humanos). En este artículo, se evaluaron los resúmenes generados automáticamente con diferentes evaluadores, donde se observó que cada evaluador toma una parcialidad diferente.

Palabras clave: evaluación de resúmenes automáticos, resumen de referencia, medidas, Jensen Shannon, ROUGE-C, ROUGE-1, ROUGE-2.

Evaluation of Automatic Summaries with and without Reference Summaries for the English Language

Abstract. During almost 60 years, several investigations have been carried out on different methods and commercial tools for the automatic generation of text summaries, consequently, it also has been researched in the different methods of evaluation of automatic text summaries. Among the most used in the state-of-the-art, we find ROUGE evaluators, which make the comparison of a summary generated automatically against one generated by a human. On the other hand, we find more robust evaluators like Jensen-Shannon and ROUGE-C that evaluate the automatically generated summary against the original document (due to the absence of summaries generated by humans). In this paper, the automatically generated abstracts were evaluated with different evaluators, where it was observed that each evaluator takes a different partiality.

Keywords: evaluation of automatic summaries, reference summary, measures, Jensen Shannon, ROUGE-C, ROUGE-1, ROUGE-2.

1. Introducción

El ser humano ha buscado a lo largo del tiempo transmitir el conocimiento que ha adquirido en su paso por la tierra, desde las pinturas rupestres que fueron creadas hace miles de años, con el fin de transmitir los conocimientos adquiridos a sus descendientes, posterior a ello la información documentada [13] ha ido evolucionando en diferentes partes del mundo como las tablillas de arcilla en Mesopotamia, el papiro en Egipto, el pergamino en Grecia y Roma, la invención del papel en China, hasta la época actual en donde encontramos grandes volúmenes de información digitalizada en todo el mundo [1]. Con la revolución tecnológica, la ayuda de internet y de grandes albergadores de documentos, el ser humano actualmente cuenta con una gran cantidad de información. La Fundación Nacional de la Ciencia, muestra los indicadores sobre el crecimiento de la información publicada de artículos científicos y técnicos, misma en la que se percibe un crecimiento exponencial en un periodo de 2003 a 2016, con un crecimiento acelerado a nivel mundial, en 2003 se reportaron 1.12 millones de artículos publicados y para 2016 se registraron 2.37 millones de artículos publicados [5].

El crecimiento acelerado de información ha dificultado la búsqueda de temas específicos, por lo que las ciencias computacionales han intervenido a través del Procesamiento del Lenguaje Natural (PLN), al que se le conoce como una disciplina en la inteligencia artificial que se encarga de la investigación de mecanismos computacionales para la comunicación entre humanos y computadoras mediante el uso de Lenguajes Naturales (LN) [10,20,21].

El PLN ha creado técnicas, herramientas y sistemas para el manejo de la información tales como la Generación Automática de Resúmenes de Texto (GART) [8,33], los cuales se describen como la extracción de ideas principales del contenido de un texto original [28,34], donde el contenido es una cuestión tanto de información como de expresión, la extracción de ideas principales es una cuestión de lo que es esencial y lo que es relevante [15]. Los métodos de GART se clasifican, de acuerdo con el tipo de resumen de salida, resúmenes extractivos o abstractivos [6,30]. Los resúmenes abstractivos, son textos que describen el contenido y el sentido de un documento original, esto consiste en entender el contexto de un documento y después reescribirlo sin perder el sentido del texto utilizando nuevos conceptos en un número menor de palabras, sin perder el contexto del documento [28,31]. Mientras que los resúmenes extractivos consisten en, reducir el contenido textual de un documento o un grupo de documentos mediante la selección de un conjunto de frases u oraciones del texto original [14,31,32]. Generalmente, las personas realizan resúmenes de tipo abstractivo, mientras que la mayoría de las herramientas comerciales y los métodos del estado del arte generan resúmenes de forma extractiva. Actualmente existen herramientas gratuitas disponibles en internet para la generación automática de resúmenes en diferentes idiomas. Algunas de las herramientas instalables son Copernic summarizer [7] y Microsoft Office Word [23]. Algunas de las herramientas en línea son Open Text Summarizer [22], Tool4noobs Summarizer [27], Text Compactor [26], t-Conspectus [25], Summarizer [4] y SweSum [24], entre otras.

Para conocer la calidad de los resúmenes generados por las herramientas comerciales, en trabajos anteriores [11,12,29] se han realizado las evaluaciones mediante el evaluador ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [16], propuesto por Lin y Hovy, Lin & Och [17,18,19]. Este sistema calcula la calidad

del resumen generado de manera automática mediante la comparación con resúmenes generados por un humano como menciona Ledeneva [28].

Para la utilización de estos evaluadores son importantes las referencias humanas, es decir, los seres humanos deben crear resúmenes de alta calidad. Por otro lado, existen métodos de evaluación de resúmenes como Jensen Shannon y ROUGE-C. Estos métodos de evaluación automáticos no requieren la creación de resúmenes modelo humano, se evalúan tomando un conjunto de resúmenes generados de manera automática y se comparan con los textos originales [2,3], ya que se tiene la idea que los textos originales contienen la información completa.

En este artículo se realiza la evaluación de los resúmenes generados de manera automática por las herramientas comerciales con los evaluadores que utilizan un resumen modelo humano y los que utilizan el texto original.

El resto del trabajo se organiza de la siguiente manera: En la sección 2, se describe el trabajo de estado del arte que ha realizado la evaluación de diferentes herramientas comerciales. En la sección 3, se describen los diferentes evaluadores para los resúmenes automáticos, por una parte, se describen los evaluadores con el uso de resúmenes de referencia y por otra parte se mencionan los evaluadores sin resumen de referencia. En la sección 4, se describe la colección de documentos, la experimentación y los resultados obtenidos. La sección 5, se describen las conclusiones del trabajo realizado.

2. Trabajo relacionado

Durante los últimos años se han realizado diversas investigaciones sobre la Generación Automática de Resúmenes de Texto (GART) de las herramientas comerciales y de métodos del estado del arte. Sin embargo, la mayoría de estos estudios realizados se evalúan a través de una comparación entre un resumen generado de manera automática contra con un resumen generado por un humano. El problema de investigación consiste en cómo realizar la evaluación de resúmenes automáticos generados por herramientas comerciales sin resúmenes modelo. En esta sección, se describe el trabajo realizado por Matías [12], en el que realiza la evaluación de herramientas comerciales y métodos del estado del arte con la colección de datos DUC02. Esta evaluación se realizó con el sistema evaluador ROUGE.

2.1. Herramientas comerciales

En el trabajo de Matías [12] se emplearon diversas herramientas comerciales algunas en línea como:

- Shvoong: Es una herramienta que permite generar resúmenes automáticos en 21 idiomas diferentes (checo, neerlandés, danés, inglés, finlandés, francés, alemán, griego, hebreo, húngaro, indonesio, italiano, malayo, noruego, polaco, portugués, rumano, ruso, español, sueco y turco). A diferencia de otras herramientas Shvoong no devuelve el resumen, sino subraya el texto que considera más importante del documento original.
- Pertinence Summarizer: Desarrollada por la empresa francesa Pertinence Mining. Pertinence es una herramienta en línea que permite generar resúmenes en 12 idiomas

(alemán, inglés, árabe, chino, coreano, español, francés, italiano, japonés, portugués, ruso y neerlandés) de los documentos de texto en formatos diversos (html, pdf, doc, rtf y txt).

- Tools4Noobs: Es una herramienta en línea que permite generar resúmenes desde 1 al 100 % del texto original. Para la generación de un resumen Tools4Noobs tiene 3 etapas: extracción de las oraciones, identificación de las palabras claves del texto contando la relevancia de cada palabra e identificación de las oraciones de acuerdo con las palabras claves identificadas.
- Open Text Summariser: Es una herramienta de código abierto para resumir textos, que puede ser descargada de Internet de forma gratuita. Sin embargo, también puede encontrarse la interfaz de ésta en línea. OTS genera resúmenes automáticos en diferentes porcentajes, genera resúmenes en 37 idiomas.

2.2. Métodos del estado del arte

Además del uso de herramientas comerciales Matías [12] realizó la comparación con los siguientes métodos del estado del arte:

- TextRank: Es un método que consiste en gráficos ponderados, en que se construye el gráfico para representar el texto, por lo que los nodos son palabras interconectadas por vértices con relaciones significativas. Para la tarea de extraer oraciones, el objetivo es calificar oraciones completas y ordenarlas de mayor a menor importancia. Por lo tanto, los vértices se agregan al gráfico para cada oración en el texto.
- Baseline: Esta heurística garantiza que la información más importante de un documento se encuentra en las primeras secciones para las que se toman las primeras n oraciones del documento para formar el resumen. Se ha demostrado que esta simple heurística genera muy buenos resúmenes en el campo de los documentos de noticias.
- SFMs: Una secuencia que se repite en el texto es frecuente y es maximal si dicha secuencia no es subsecuencia de otra secuencia frecuente.
- Baseline-Aleatorio: Esta heurística obtiene los mejores resúmenes para determinar los resúmenes cuando solo se toma un conjunto de oraciones al azar. La idea es determinar los resultados que se pueden lograr.

3. Evaluadores de resúmenes automáticos

Los evaluadores de resúmenes automáticos han sido de gran ayuda para los investigadores de GART. A través de estos sistemas se califica la calidad de los resúmenes generados de manera automática. Generalmente los investigadores utilizan los evaluadores ROUGE, porque éstos realizan la comparación entre un resumen automático y un resumen modelo generado por un humano. Sin embargo, no todas las colecciones de datos cuentan con un resumen modelo (o también le llaman estándar de oro), lo que dificulta la evaluación. Es por ello que es un problema de investigación importante que se ha indagado en los evaluadores que no ocupan un resumen modelo como ROUGE-C y Jensen Shannon. Estas evaluaciones se realizan al comparar el

resumen automático contra el texto original, por lo que se tiene la idea que los textos originales cuentan con las ideas completas.

3.1. ROUGE

ROUGE es un sistema para calcular la calidad de los resúmenes generados de forma automática mediante la comparación de los resúmenes automáticos con resúmenes creados por humanos, fue propuesto por (Lin & Hovy, 2003; Lin & Och, 2004a,b) [17, 18, 19]. Este sistema trabaja mediante la recuperación de n-gramas entre un resumen candidato y un resumen ideal creado por un humano y se calcula de la siguiente manera:

$$ROUGE - N = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} \sum_{\epsilon \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} \sum_{\epsilon \in S} Count(gram_n)}, \quad (1)$$

donde n es la longitud del n-grama y $Count_{match}(gram_n)$ el número máximo de n-gramas que co-ocurren en el resumen candidato y el conjunto de resúmenes de referencia.

3.2. ROUGE-C

ROUGE-C es un método de evaluación automático para la evaluación de resúmenes de texto automáticos, fue propuesto por (He et al., 2008) [9]. Este método es una variación al sistema de evaluación de resúmenes de ROUGE-1.5.5, en el que se plantea la sustitución de los resúmenes de oro en el texto fuente y además se plantea el intercambio entre documentos, es decir, el documento fuente es colocado como resumen de evaluación y el resumen generado es colocado como un resumen de evaluación, mientras que el resumen regentado por el sistema es colocado como un resumen de referencia como se muestra en la figura 1. La evaluación de resúmenes con ROUGE-C está basado bajo la siguiente expresión:

$$ROUGE - C - N = \frac{\sum_{S \in \{PeerSummary\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{SourceDocument\}} \sum_{gram_n \in S} Count(gram_n)}, \quad (2)$$

donde n determina el uso del n-grama a evaluar a través de su longitud, $Count_{match}(gram_n)$ es el máximo número de co-ocurrencia de n-gramas entre el resumen generado por el sistema y el documento fuente.

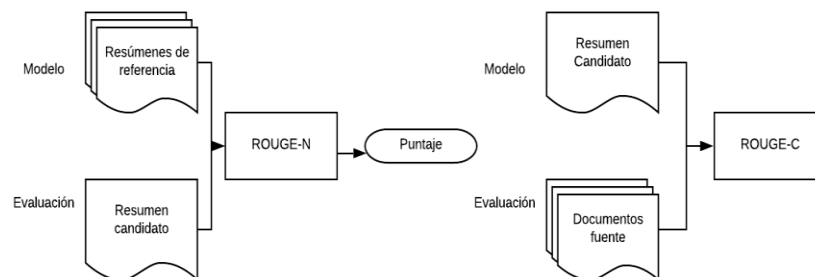


Fig. 1. Evaluación de ROUGE (Lin, 2004) y ROUGE-C (He et al., 2008) [9].

3.3. Jensen Shannon

Jensen Shannon divergence es un método de evaluación para la generación automática de resúmenes de texto, fue propuesto por (Louis & Nenkova, 2008), (Louis & Nenkova, 2009) [2, 3]. Este método evalúa un resumen generado de manera automática con el documento original, incorpora la idea de la distancia entre dos distribuciones no puede ser muy diferente del promedio de las distancias de su distribución media. La evaluación está basada bajo la siguiente expresión.

$$J(P||Q) = \frac{1}{2} [D(P||A) + D(Q||A)], \quad (3)$$

donde $A = \frac{P+Q}{2}$ es la distribución media de P y Q . La distancia JS es simétrica y siempre está definida.

4. Experimentación y resultados

Para la primera evaluación de los resúmenes generados automáticamente por las herramientas comerciales, se van a utilizar los evaluadores ROUGE-1 y ROUGE-2. La evaluación consiste en estimar el parecido de los resúmenes generados automáticamente con los resúmenes realizados por expertos humanos. Los resúmenes generados por las herramientas en línea e instalables fueron generados con un mínimo de 100 palabras, por lo que anterior a ello, se analizaron las herramientas para realizar dicha tarea.

Por otro lado, realizamos la segunda evaluación de los resúmenes automáticos generados por las herramientas comerciales, con el uso de los evaluadores ROUGE-C y Jensen Shannon. Estas evaluaciones constan de una comparación entre los resúmenes automáticos y los textos originales, ya que los autores de dichas evaluaciones plantean la idea de decir si realmente los humanos generan buenos resúmenes.

5. Corpus

DUC (Document Understanding Conference) es una serie de talleres creados por NIST (National Institute of Standards and Technology) para el desarrollo de la tarea de GART. Las colecciones de documentos elaboradas por DUC son las más usadas por investigadores en la tarea de GART de multi-documentos y para un solo documento, destacando las colecciones de datos DUC01 y DUC02.

Para este artículo se utilizó la colección de datos DUC01, la cual consta de 309 documentos en idioma inglés, agrupados en 30 colecciones, cada colección contiene en promedio de 10 a 11 documentos de noticias periodísticas, que abordan temas sobre desastres naturales, información biográfica de un individuo, temas políticos, etc. Esta colección de datos también contiene los resúmenes de oro generados por humanos.

6. Herramientas comerciales

En el artículo se utilizaron las herramientas comerciales de mayor auge en las investigaciones realizadas. La herramienta instalable en sus diferentes versiones es:

- Microsoft Office Word: Es una herramienta que se encuentra incluida en Microsoft Office Word en las versiones 2003 y 2007, permite generar resúmenes de 10 o 20 oraciones, 100 o 500 palabras. También cuenta con la opción de realizar resúmenes con porcentajes: 10%, 25%, 50% y 75% de palabras del documento original.

Las herramientas en línea que se consideran en este artículo son:

- Open Text Summarizer (OTS): Es una herramienta de código abierto analiza automáticamente textos. OTS genera resúmenes automáticos en diferentes porcentajes de 5% a un 80%. Se pueden generar resúmenes en 37 idiomas.
- Summarizer: es una herramienta que permite generar resúmenes automáticos. Está disponible como componente de Intellexer API y como aplicación de escritorio. Genera resúmenes en diferentes porcentajes de 1% a un 99%. Recibe un documento de origen, extrae texto sin formato, proporciona el procesamiento sintáctico y semántico, extrae la información para la generación de resumen del documento, finalmente, asigna un valor determinado por oraciones del texto original. Este valor define la importancia de la oración en lo que respecta a la idea del texto, tiene como opción crear diferentes tipos de resúmenes.
- Summarizing: Es una herramienta en línea para la generación automática de resúmenes de manera rápida y eficaz. La herramienta tiene las opciones de 100, 150, 200 y 300 palabras para obtener un resumen.
- SweSum: Es un generador de resúmenes multilingüe, propuesto para sueco e inglés. Utiliza múltiples aspectos para valorar las oraciones como: la posición o valor numérico de las oraciones.
- T-Conspectus: Es una aplicación web para resumir textos de un solo documento en inglés, alemán y ruso. Permite realizar resúmenes por porcentajes (va desde el valor 5% hasta el valor 70 %). El proceso se realiza en tres etapas: preprocesamiento, puntuación de resúmenes y generación del resumen.
- Text Compactor: Es una herramienta en línea para la generación automática de resúmenes creada para procesar cantidades grandes de información (Edyburn, 10), está basada en la herramienta Open Text Summarizer. Esta herramienta genera resúmenes de 1% a 100 % del texto.

7. Resultados

Como primer aspecto se realizó la generación de los resúmenes automáticos por las herramientas comerciales. Después de haber obtenido los resúmenes automáticos, se realizó la primera evaluación con el evaluador ROUGE-1. Los resultados obtenidos mostraron que la herramienta que obtuvo los resultados con mayor similitud contra uno realizado por un humano fue la herramienta SweSum, ya que obtuvo un puntaje de acuerdo con el F-measure de 0.43459. Esto significa que el resumen generado por la herramienta coincide en un 43% con un resumen generado por un humano experto.

En la segunda evaluación de los resúmenes automáticos que se realizó con el evaluador ROUGE-2, los resultados obtenidos demostraron que SweSum sigue siendo la herramienta en línea que genera resúmenes automáticos con mayor similitud a los generados por un humano, sin embargo, el puntaje obtenido con este evaluador

disminuyó en un 0.18203 de puntos, dejando a la herramienta con un puntaje final de 0.18203.

Tabla 1. Evaluación de diferentes herramientas comerciales usando ROUGE-1.

Herramienta	Recuerdo	Precisión	F-measure
SweSum	0.43626	0.43313	0.43459
OTS	0.42577	0.42386	0.42458
Text-Compactor	0.42243	0.41954	0.42084
T-Conspectus	0.41595	0.41070	0.41321
Summarizing	0.41050	0.40726	0.40874
Summarizer	0.40947	0.40646	0.40786
Word Office 2007	0.39597	0.39730	0.39651
Word Office 2003	0.39469	0.39618	0.39533

En la figura 3 se observa que la herramienta con mayor puntaje fue SweSum.

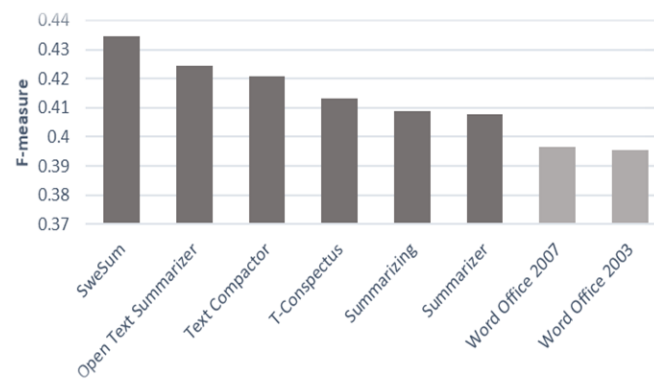


Fig. 2. Comparación del puntaje obtenido de las herramientas comerciales por el evaluador ROUGE-1.

Para las herramientas instalables Office Word 2007 sigue siendo mejor que Office Word 2003 con un puntaje de 0.14050, pero sigue teniendo la posición 8 en el ranking de las herramientas probadas en este artículo.

Tabla 2. Evaluación de diferentes herramientas comerciales usando ROUGE-2.

Herramienta	Recuerdo	Precisión	F-Measure
SweSum	0.18264	0.18153	0.18203
OTS	0.17242	0.17129	0.17177
Text-Compactor	0.16888	0.16759	0.16819
T-Conspectus	0.15718	0.15506	0.15607
Summarizing	0.14658	0.14545	0.14597
Summarizer	0.14526	0.14404	0.14462
Word Office 2007	0.14030	0.14086	0.14054
Word Office 2003	0.13915	0.13971	0.13940

En la tercera evaluación realizada se compara el resumen generado automáticamente contra el texto original y según los datos obtenidos Text Compactor es la mejor herramienta en línea para la GART, con este evaluador la herramienta Word 2007 ahora se posiciona en el lugar número 3.

Tabla 3. Evaluación de diferentes herramientas comerciales usando Jensen Shannon.

Herramienta	SMT	WSMT
Text-Compactor	0.73630	0.71277
Summarizer	0.72611	0.69972
Word Office 2007	0.72378	0.69679
OTS	0.72226	0.69674
SweSum	0.71815	0.69063
Word Office 2003	0.71165	0.68180
Conspectus	0.70925	0.68131
Summarizing	0.70185	0.67105

Las evaluaciones que se realizaron con el evaluador ROUGE-C demuestran que Text-Compactor es la herramienta con mayor similitud al texto original mientras que SweSum para esta evaluación, resulta ser una herramienta de poca competencia.

En la figura 3 se observa que la herramienta SweSum se posicionó en el quinto lugar, mientras que en las evaluaciones pasadas con resumen de referencia era que la se posicionaba en primer lugar.

Tabla 4. Evaluación de diferentes herramientas comerciales usando ROUGE-C.

Herramienta	C-1	C-2	C-3	C-L	C-SU4
Text-Compactor	0.25654	0.25029	0.24575	0.25630	0.24526
Summarizer	0.22946	0.20385	0.17852	0.22939	0.17537
Word 2007	0.22571	0.21244	0.19972	0.22549	0.19714
OTS	0.22570	0.21356	0.20340	0.22491	0.20176
SweSum	0.21080	0.19991	0.18922	0.21055	0.18730
Conspectus	0.20913	0.18442	0.15951	0.20764	0.15549
Word 2003	0.20353	0.19149	0.17966	0.20335	0.17690
Summarizing	0.19141	0.17795	0.16629	0.18711	0.16427

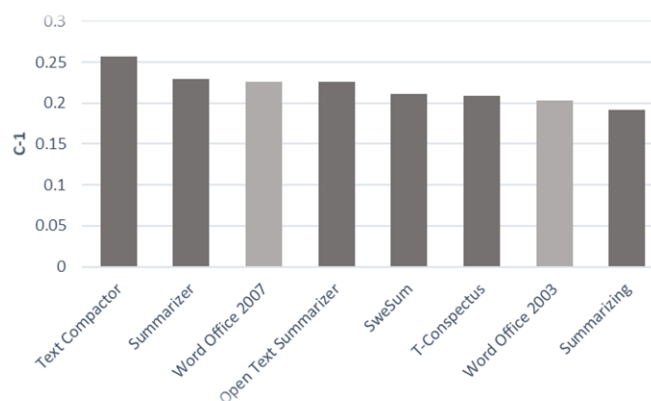


Fig. 3. Comparación de puntaje obtenido por el evaluador ROUGE-C sin resumen de referencia.

8. Conclusiones

En las evaluaciones realizadas con resúmenes de referencia y sin resúmenes de referencia, pudimos observar que la diferencia de puntajes entre evaluadores varía significativamente. Esto es debido a que los evaluadores evalúan aspectos diferentes, mientras los evaluadores con resumen de referencia evalúan la similitud que hay entre un resumen automático y uno generado por un humano. Los evaluadores sin resumen de referencia evalúan la divergencia que existe entre el texto original y el resumen automático. Por lo que se puede decir que el evaluador realiza una evaluación parcial, toma aspectos diferentes a evaluar.

De acuerdo con la evaluación realizada para los evaluadores con resumen de referencia y sin resumen de referencia, podemos apreciar que, mientras el evaluador ROUGE-C arroja puntajes mayores a 0.7, los evaluadores como ROUGE-2 y Jensen Shannon se encuentran por debajo del 0.3 y el evaluador ROUGE-1 se encuentra en un rango de 0.4 y 0.5.

De acuerdo con los puntajes obtenidos por los diferentes evaluadores, se concluye diciendo que las mejores herramientas para la generación automática de resúmenes son las herramientas: Text-Compactor y Open Text Summarizer. Esto es debido a que en 4 experimentos, estas herramientas lograron posicionarse dentro de las 4 mejores lugares, por lo que se determina que estas herramientas cumplen con el mayor número de características evaluadas por los diferentes evaluadores, en comparación con otras herramientas que se posicionaron en las primeras evaluaciones en los últimos lugares y para las últimas evaluaciones se posicionaron dentro de los primeros lugares.

Referencias

1. Sánchez-Arteche, A., Ortiz-López, H.A, Pérez-Rodríguez, N.C., Jiménez-Mercado. I.E., Cue-Vidal, J.C.: Cronología de la escritura y el libro, primera edición, SEP, Fondo editorial del Estado de México (2012)
2. Louis, A., Nenkova, A.: Automatic Summary Evaluation without Human Models. In: Notebook Papers and Results, Text Analysis Conference (TAC-2008), Gaithersburg, Maryland, USA (2008)
3. Louis, A., Nenkova, A.: Automatically Evaluating Content Selection in Summarization without Human Models. Proceedings of the Conference on Empirical Methods in Natural Language Processing: vol. 1, pp. 306–314 (2009)
4. La página de inicio de Article Summarizer, <https://www.summarizing.biz/best-summarizing-strategies/article-summarizer-online/>, ultimo acceso 2019/02/20
5. La página de inicio de Banco Mundial de Datos. <https://datos.bancomundial.org/indicador/ip.jrn.artc.sc>, ultimo acceso 2019/02/30
6. Ledeneva, Y., García-Hernández, R., Gelbukh, A.: Graph Ranking on Maximal Frequent Sequences for Single Extractive Text Summarization. LNCS Springer-Verlag, ISSN 0302-9743, vol. 8404, pp. 466–480 (2014)
7. La página de inicio de Copernic, www.copernic.com/en/products/summarizer, ultimo acceso 2019/02/20
8. Vázquez, E., García Hernández, R.A., Ledeneva, Y.: Learning Relevant Models using Symbolic Regression for Automatic Text Summarization. Computación y Sistemas 23(1), 127–141 (2018)

9. Loret, E., Plaza, L., Aker, A.: The Challenging Task of Summary Evaluation: An overview, Language Resources and Evaluation, vol. 52, pp. 101–148 (2018).
10. Martín Mateos, F.J., Ruiz Reina, J.L.: Procesamiento del Lenguaje natural. Departamento de la Computación e Inteligencia Artificial, Universidad de Sevilla (2013)
11. Matias Mendoza, G.A.: Generación automática de resúmenes independientes del lenguaje. Tesis de Maestría en Ciencias de la Computación, Universidad Autónoma del Estado de México (2016)
12. Matias Mendoza, G.A., Ledeneva, Y., García Hernández, R.A., Sidorov, G.: Evaluación de las Herramientas comerciales y métodos del estado del arte para la generación de resúmenes extractivos individuales. Research in Computing Science, vol. 70, pp. 265–274 (2013)
13. Hoffmann, E.: Defining Information: An Analysis of the information content of documents. Information Processing and Management, vol. 18, pp. 133–139 (1982)
14. Rojas Simón, J., Ledeneva, Y., García, R.: Calculating the Upper Bounds for Multi-Document Summarization using Genetic Algorithms. Computación y Sistemas 22(1), 11–26 (2018)
15. Rojas Simon, J., Ledeneva, Y., García, R.: Calculating the Significance of Automatic Extractive Text Summarization using a Genetic Algorithm. Journal of Intelligent & Fuzzy Systems. Applications in Engineering and Technology 35(1), 293–304 (2018)
16. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Text summarization branches out: Proceedings of the ACL-04 workshop Vol. 8 (2004)
17. Lin, C.Y., Hovy, E.: Automatic evaluation of summaries using n-gram co-occurrence statistics. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1, pp. 71–78 (2003)
18. Lin, C.Y., Och, F.J.: Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, p. 605 (2004)
19. Lin, C.Y., Och, F.J.: Orange: a method for evaluating automatic evaluation metrics for machine translation. In: Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, p. 501 (2004)
20. Cortez Vásquez, A., Vega Huerta, H., Pariona Quispe, J.: Procesamiento del Lenguaje Natural. Ingeniería de Sistemas e Informática vol. 6. Facultad de Ingeniería de Sistemas e Informática, julio-diciembre (2009)
21. Alonso Pardo, M.A., Alonso Ramos, M., Gómez Rodríguez, C., Gaña Gil, J., Villares Ferro, J.: La enseñanza del Procesamiento del Lenguaje Natural en la Facultad de Informática y Filología, Investigaciones sobre docencia universitaria y nuevas metodologías, Chapter: 1, pp.13–18 (2012)
22. La página de inicio de Open Text Summarizer <https://open-text-summarizer.soft112.com/>, último acceso 2019/02/20
23. La página de inicio de Microsoft Office Word <https://products.office.com/es-mx/word>, ultimo acceso 2019/02/20
24. La página de inicio de SweSum <http://swesum.nada.kth.se/index-eng.html>, ultimo acceso 2019/02/20
25. La página de inicio de T-CONSPECTUS <https://tconspectus.pythonanywhere.com/>, ultimo acceso 2019/02/20
26. La página de inicio de Text Compactor <http://textcompactor.com/>, ultimo acceso 2019/02/20
27. La página de inicio de Tools4noobs Summarizer, <https://www.tools4noobs.com/summarize/>, ultimo acceso 2019/02/20
28. Ledeneva, Y., García Hernández: Generación Automática de resúmenes. Retos, propuestas y experimentos. Automatic Generation of text Summaries Challenges Proposals and Experiments. Universidad Autónoma del Estado de México (Ed.)/Cigome (2017)

29. Rojas Sánchez, J.M.: Evaluación de herramientas comerciales y métodos del estado del arte para la generación de resúmenes en idioma ruso. Tesis para obtener el título de ingeniero en software. Universidad Autónoma del Estado de México (2016)
30. Hahn, U., Mani, I.: The challenges of automatic summarization. *Computer* 33, 29–36 (2000)
31. Hovy, E., Lin, C.Y.: Automated text summarization and the SUMMARIST system. In: *Proceedings of a Workshop held at Baltimore, Maryland: October 13-15, 1998*, Association for Computational Linguistics, pp. 197–214 (1998)
32. Reddy, P., Balabantaray, R.C.: Improvisation of the document summarization by combining the IR techniques with code-quantity memory and attention linguistic principles. *Procedia Technology* 6, pp 118–125 (2012)
33. García Hernández, R., Montiel, R., Ledeneva, Y., Rendón, E., Gelbukh, A., Cruz, R.: *Text Summarization by Sentence Extraction Using Unsupervised Learning*. Springer Berlin, Heidelberg, pp. 133–143 (2008)
34. Gillick, D., Favre, B., Hakkani-Tur, D., Bohnet, B., Liuand, Y., Xie, S.: *The ICSI/UTD Summarization System at TAC2009* (2009)