



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Desarrollo de un sistema de clasificación de
temas y resumen automático en el ámbito de la
información periodística

Trabajo Fin de Grado

Grado en Ingeniería Informática

Autor: Marco Canora Alonso

Tutor: Emilio Sanchís Arnal

Cotutor: Fernando García Granada

Curso 2017-2018

Resumen

Este estudio propone un nuevo modelo de resumen de documentos basado en técnicas no supervisadas. Concretamente, uno basado en el etiquetado de palabras utilizando Latent Dirichlet Allocation frente al enfoque clásico basado en Latent Semantic Analysis. Se adecuará la aplicación de dichas técnicas a documentos compuestos por diferentes bloques de contenido de diversa duración. También, se estudia la segmentación en bloques temáticos, que se hará considerando las distancias entre párrafos, representados por vectores continuos obtenidos a través de las palabras que los forman. A continuación, cada uno de los segmentos será resumido. En los experimentos se han empleado tanto programas televisivos como artículos de periódico.

Palabras clave: Resumen de Documentos, Latent Semantic Analysis, Latent Dirichlet Allocation, Segmentación por Tema



Abstract

In this paper we present an approach to document summarization based on unsupervised techniques. We study the adequacy of these techniques to the problem of documents where many topics of different duration are present, in our case the transcriptions of TV programs. The paper compares classical Latent Semantic Analysis approach with a new proposal based on Latent Dirichlet Allocation labeling of words. It is also studied the application of the summarization process to the different segments obtained in a previous process of topic segmentation. The topic segmentation is performed by considering distances between paragraphs, that are represented by means of continuous vectors obtained from the words contained in them. Experiments have been performed over some articles of newspapers, and over some TV programs of political and miscellaneous news.

Keywords: Document Summarization, Latent Semantic Analysis, Latent Dirichlet Allocation, Topic Segmentation

Resum

Aquest estudi proposa un nou model de resum de documents basat en tècniques no supervisades. Concretament basat en l'etiquetatge de paraules Latent Dirichlet Allocation front al mètode clàssic Latent Semantic Analysis. S'adequarà l'aplicació d'aquestes tècniques a documents els quals contenen diversos blocs de continguts de diversa duració. A més, s'estudia la segmentació en blocs temàtics, la qual cosa es farà considerant les distàncies entre paràgrafs, representats per vectors continus obtinguts gràcies a les paraules que els formen. A continuació, cadascun dels segments serà resumit. Als experiments s'han fet servir tant programes televisius com articles de periòdic.

Paraules clau: Resum de Documents, Latent Semantic Analysis, Latent Dirichlet Allocation, Segmentació per Tema

Índice general

Resumen	III
Abstract	V
Resum	VII
Índice general	IX
Índice de figuras	XII
Índice de tablas	XIV
Índice de algoritmos	XVI

1. Introducción	1
1.1. Objetivos	2
1.2. Estructura de la memoria.....	3
2. Estado del arte	5
2.1. Crítica al estado del arte y propuesta.....	6
3. Diseño de la solución	8
3.1. Arquitectura del sistema.....	8
3.2. Diseño detallado.....	10
3.2.1. Procesado del documento.....	10
3.2.2. Construcción de la matriz de entrada	10
3.2.3. Modelado semántico	11
3.2.3.1. Latent Semantic Analysis (LSA).....	11
3.2.3.2. Latent Dirichlet Allocation (LDA).....	13
3.2.4. Segmentación semántica	14
3.2.5. Selección de frases	17
3.2.5.1. Gong & Liu	17
3.2.5.2. Steinberger & Jezek.....	18
3.2.5.3. Murray & Renals & Carletta	19
3.2.5.4. Cross.....	20
3.3. Tecnología utilizada.....	21
4. Pruebas	24
4.1. Corpus	24
4.1.1. Document Understanding Conferences (DUC).....	24
4.1.2. La noche en 24 horas.....	28
4.2. Medidas de evaluación	29



4.2.1.	ROUGE	29
4.2.2.	Pk.....	32
4.3.	Experimentos.....	33
4.3.1.	DUC	34
4.3.2.	La noche en 24 horas.....	35
4.3.2.1.	Aproximaciones sin segmentación.....	35
4.3.2.2.	Segmentación	36
4.3.2.3.	Aproximaciones con segmentación.....	37
5.	Conclusiones.....	40
5.1.	Relación del trabajo desarrollado con los estudios cursados.....	41
6.	Bibliografía.....	42
7.	Tablas de resultados originales	44

Índice de figuras

Figura 1. Esquema del sistema de resumen.....	9
Figura 2. SVD.	11
Figura 3. Matrices C y Ck de ejemplo.	12
Figura 4. Representación geométrica de Dirichlet.	13
Figura 5. Ejemplo de cálculo de distancias entre ventanas consecutivas.....	16
Figura 6. Función de distancia para las ventanas de ejemplo.....	16
Figura 7. El zen de Python.	21
Figura 8. DUC.....	25
Figura 9. Cronología de las "pistas" del TAC.	25
Figura 10. Frecuencia de puntuación de la capacidad de respuesta por procedencia del resumen en DUC 2007.....	27
Figura 11. Frecuencia de la longitud de los segmentos para RTVE.	28
Figura 12. Descripción de RTVE.....	29
Figura 13. Esquema de evaluación del ROUGE.	29
Figura 14. Resumen de los sistemas desarrollados.	34
Figura 15. Gráfico de línea con los Pk medios por tamaño de ventana.	36
Figura 16. Gráfico de línea con la desviación estándar de los Pk por tamaño de ventana.	37

Índice de tablas

Tabla 1. Resumen de los resultados del LSA para DUC.....	34
Tabla 2. Resumen de los resultados del LDA para DUC.	35
Tabla 3. Comparativa entre los resultados del LDA y LSA para DUC.	35
Tabla 4. Resumen de los resultados del LSA para RTVE.....	35
Tabla 5. Resumen de los resultados del LDA para RTVE.	36
Tabla 6. Comparativa entre los resultados del LDA y LSA para RTVE.	36
Tabla 7. Resumen de los resultados del LSA segmentado para RTVE.....	37
Tabla 8. Resumen de los resultados del LDA segmentado para RTVE.	37
Tabla 9. Comparativa entre los resultados segmentados y no segmentados.	38
Tabla 10. Comparativa entre los resultados del LDA y LSA segmentados.	38
Tabla 11. Segmentación.	44
Tabla 12. DUC LSA.....	45
Tabla 13. DUC LDA 5.	46
Tabla 14. DUC LDA 10.	47
Tabla 15. DUC LDA 5.	48
Tabla 16. DUC LDA 40.	49
Tabla 17. DUC LDA 80.	50
Tabla 18. RTVE LSA.....	51
Tabla 19. RTVE LDA 5.	52
Tabla 20. RTVE LDA 10.	53
Tabla 21. RTVE LDA 20.	54
Tabla 22. RTVE LDA 40.	55
Tabla 23. RTVE LDA 80.	56
Tabla 24. RTVE LSA Segmentado.	57
Tabla 25. RTVE LDA 5 Segmentado.	58
Tabla 26. RTVE LDA 10 Segmentado.	59
Tabla 27. RTVE LDA 20 Segmentado.	60
Tabla 28. RTVE LDA 40 Segmentado.	61
Tabla 29. RTVE LDA 80 Segmentado.	62

Índice de algoritmos

Algoritmo 1. Algoritmo generativo del LDA.....	14
Algoritmo 2. Cálculo de ventanas.....	15
Algoritmo 3. Gong & Liu.....	17
Algoritmo 4. Steinberger & Jezek.....	18
Algoritmo 5. Murray & Renals & Carletta.....	19
Algoritmo 6. Cross.....	20
Algoritmo 7. ROUGE-1.....	30
Algoritmo 8. ROUGE-2.....	31
Algoritmo 9. ROUGE-L.....	31
Algoritmo 10. ROUGE-S*.....	32
Algoritmo 11. Pk.....	33

Introducción

El resumen de contenido multimedia es un problema importante en los últimos años. Debido a la gran cantidad de información disponible en la web, es necesario proporcionar al usuario herramientas que le ayuden a procesarla. Hay muchas aplicaciones que podrían beneficiarse de esto, principalmente las relacionadas con la gestión de documentos, tanto escritos, como de audio, o de vídeo. Aunque los principales avances más recientes han sido los relacionados con la Indexación y Recuperación de Información (IR), está cobrando interés la posibilidad de disponer de sistemas que aporten información adicional sobre los documentos recuperados (Tur & De Mori, 2011) (Lloret & Palomar, 2012). Un precedente de ellos podría ser el de los “snippets” en los motores de búsqueda, que expresan el contexto en el que ha aparecido la consulta realizada. Su objetivo es ayudar en la decisión sobre si el resultado es adecuado a la búsqueda o no. Debe tenerse en cuenta que las frases de los “snippets” no son aleatorias, sino que se basa en un mecanismo de optimización que selecciona las frases más representativas para el término buscado.

Los sistemas de resumen empezaron a tener un gran desarrollo a partir de la década de los 2000, sobre todo con la creación de Document Understanding Conference (DUC), en el 2002. Como indica el nombre de la conferencia, en los inicios, los sistemas estaban muy ligados a la comprensión del texto, es decir, saber de qué habla, que se puede representar con un pequeño subconjunto de frases que contiene la semántica del documento. Este tipo de aplicaciones aún está lejos de otras interpretaciones del concepto de resumen, como podría ser resumir una novela o una película, elementos que contienen relaciones temporales entre personajes, distintos acontecimientos y argumento.

Conviene destacar que conferencias como DUC y Text Analysis Conference (TAC) han sido de gran ayuda en el desarrollo de este campo, en las que se evalúan y comparan los enfoques presentados. Para que los resultados sean comparables, se han creado algunos corpus para evaluar los sistemas. Para el resumen automático, los más populares son: DUC y CNN/DailyMail; ampliamente utilizado para entrenar modelos de aprendizaje basados en RNN (Cheng & Lapata, 2016).

Tradicionalmente, las técnicas de resumen se han clasificado en dos categorías: extractivos y abstractivos. Los extractivos detectan las frases más importantes y generan un resumen compuesto de ellas. Por otra parte, los abstractivos pretenden imitar cómo lo haría un humano, generando nuevas frases que pueden no estar en el documento original. A pesar de que el último enfoque es un reto más ambicioso pero también más complejo, trabajos recientes parecen ser prometedores. En el marco de los enfoques extractivos, la mayoría de los sistemas se basan en modelos de aprendizaje no supervisado, es el caso del Análisis Semántico Latente (LSA) o el basado en grafos (Erkan & Radev, 2004). Esto no quita que se puedan utilizar métodos supervisados como Redes Neuronales Recurrente (RNN) o Máquinas Vector Soporte (SVM) (Fuentes et al, 2007).

Existen muchas variables a la hora de escoger las frases del resumen, como escoger las más representativas para cada concepto detectado, o las que contengan mayor número de conceptos, o incluso sólo las que tratan los conceptos más importantes. Todas estas alternativas están ligadas al criterio de resumen ideal que se haya definido. Por ello, el resumen automático es



todavía un problema abierto, en el que existen múltiples para determinar su correctitud, y a la vez diferentes aproximaciones para realizarlo.

Una característica particular que tienen algunos documentos susceptibles de ser resumidos es que su contenido está dividido en secciones, cada una de ellas dedicada a un tema concreto. Por ejemplo, en un magazín de televisión puede haber una parte de noticias de actualidad, y luego diversas secciones con entrevistas o reportajes con temática variopinta. En estos casos puede ser interesante realizar un proceso previo de segmentación, donde cada parte se asocia a un tema diferente con cierta coherencia semántica (Hearst, 1997) (Choi, 2000). Existen líneas de investigación cuyo objetivo es la detección e identificación de tópicos en documentos de texto o de audio. Estos trabajos tienen otras aplicaciones, además de la ya comentada de ayuda a los sistemas de resumen, como puede ser el almacenamiento y la recuperación de información basada en temas, o, por ejemplo, la mejora de los sistemas de reconocimiento de habla, en los que los modelos de lenguaje se adaptan según el tema de que se está hablando.

Debido al avance tecnológico y al interés de la detección de los conceptos principales de los documentos, los sistemas desarrollados en las últimas décadas se han basado en métodos extractivos. Este tipo de aplicaciones puede ser muy útil para actividades periodísticas, donde puede interesar hacer un dossier de prensa sobre una noticia concreta tratada por varios periódicos a lo largo del tiempo. También es de utilidad en aplicaciones turísticas, para extraer de forma automática la información más interesante sobre un destino, o incluso en el ámbito médico, realizando resúmenes de informes. En definitiva, puede ser beneficioso para todas aquellas tareas en las que convenga realizar un primer filtro de una gran cantidad de información.

Otros autores han explorado el resumen considerando pistas de audio como entrada (Furui et al, 2004). Esta aproximación tiene los problemas adicionales de tratar con errores típicos del reconocimiento de voz, por ejemplo, los asociados con la puntuación de las frases. Además, debido a la espontaneidad del habla, es probable que el resumen contenga frases irrelevantes para la explicación del contenido.

En este trabajo, se presenta un enfoque para el resumen de documentos basado en técnicas no supervisadas, en particular Latent Semantic Analysis (LSA) (Ozsoy et al, 2011) y Latent Dirichlet Allocation (LDA) (Blei et al, 2003). Dentro de las técnicas extractivas, pueden clasificarse dentro de las basadas en temas, ya que algunos conceptos pueden ser detectados automáticamente y usados para determinar las frases más relevantes. También se aborda en este trabajo el problema de la segmentación de un documento en función del tema que se trate. Para ello, se ha estudiado una representación continua de las palabras, frases y conjuntos de frases, y se ha definido una medida de distancia que permite encontrar las fronteras entre segmentos de texto.

Otra consideración de este trabajo es que se resumen programas de televisión, en particular uno de noticias. Algunas características de esta tarea generan desafíos específicos para el problema del resumen. El más interesante es que los programas tienen una estructura muy variable, generalmente tratan muchos temas con duraciones diferentes. Se han estudiado dos estrategias de resumen: la primera tiene como entrada la transcripción del programa, mientras que la segunda partes de esta, producto de una previa segmentación del texto.

1.1. Objetivos

El objetivo fundamental del trabajo es el desarrollo de un sistema de resumen automático de documentos basado en métodos no supervisados orientados al descubrimiento de las

características semánticas del texto. Para la consecución de este objetivo se han planteado los siguientes subobjetivos parciales:

- Estudio y análisis de las propuestas existentes para abordar el problema de resumen automático y segmentación por tópicos
- Estudio de las herramientas tecnológicas disponibles para ambos tipos de tareas.
- Estudio de las medidas de evaluación habitualmente usadas para estas tareas, así como los corpus de referencia usados en campañas de evaluación.
- Desarrollo de algoritmos específicos para realización automática de resúmenes basándose en LSA y LDA.
- Desarrollo de algoritmos específicos para la segmentación por tópicos, basándose en una representación continua del texto.
- Creación de un corpus de evaluación y realización de una batería de experimentos.

1.2. Estructura de la memoria

La memoria se divide en cinco capítulos: introducción, estado del arte, diseño de la solución, pruebas y conclusiones. Tras la introducción, se abordará el estado del arte, donde además de exponer las principales formas de abordar el problema de resumir un texto, se realiza una crítica, que justifica el desarrollo del sistema que se propone. El segundo capítulo trata sobre el diseño del sistema, que comienza exponiendo las diferentes piezas lo forman, para posteriormente detallar cada una de ellas. En esta sección también se listan los elementos software en los que se apoya la solución. Conocidas las piezas, se evaluará el rendimiento del sistema. Esto se realiza en el capítulo de pruebas, que se divide en tres bloques: corpus, medidas de evaluación y experimentos realizados. El documento finalizará exponiendo las conclusiones.

Estado del arte

Los principales trabajos de resumen automático son los basados en métodos extractivos, es decir, aquéllos cuyo resumen lo forman frases del texto original. Las técnicas se pueden agrupar en los siguientes métodos:

Estadísticos. Tienen por objetivo asignar pesos a las palabras del documento. Para ello, se suelen utilizar medidas propias del campo de Recuperación de Información (IR), por ejemplo: tf o tf-idf (Gupta et al, 2007). Las frases se puntúan según el peso de sus palabras. Para la selección de frases, se pueden utilizar diversos mecanismos. Una forma simple sería escoger las frases con mayor puntuación, pero tiene el inconveniente de que pueden ser parecidas, ya que estarán compuestas por las palabras de mayor peso. Con el objetivo de escoger frases que representen los distintos temas o aspectos que trata el documento, se pueden aplicar técnicas de “clustering” o de reducción del espacio de representación de los documentos, como es el caso de Latent Semantic Analysis (LSA). Con este mecanismo, basado en la descomposición de matrices a partir de los valores singulares, se pueden asociar las frases a un conjunto de tópicos que subyacen en el documento, y de esa forma escoger una frase de cada uno. Un mecanismo más novedoso es usar Latent Dirichlet Allocation (LDA), que representa cada palabra en función de un conjunto de tópicos obtenidos automáticamente. A partir de esta representación, también se pueden aplicar técnicas de selección de frases, cuyo objetivo sería representar el mayor número de tópicos, o al menos los más relevantes.

Basados en temas. A diferencia del anterior, estas técnicas identifican los temas de un texto y asignan un peso a cada uno (Harabagiu & Lacatusu, 2005). El resumen estará formado por las secciones del documento que traten los temas más relevantes. En este caso es necesario realizar una segmentación y clasificación previa de los temas, para luego generar el resumen a partir de esa estructura.

Basados en grafos. En un grafo, los elementos de un texto se representan mediante nodos, mientras que las aristas los conectan, relacionando los elementos semánticamente. El éxito de estas técnicas radica en descubrir nuevas relaciones en el grafo en base a cierto criterio. Esta técnica se basa en el PageRank de Google y fue propuesta por en (Erkan & Radev, 2004). Proporciona muy buenos resultados y es una de las referencias para comparar nuevas propuestas de sistemas de resúmenes.

Basados en el discurso. Consisten en encontrar la estructura del documento. El resumen se genera a partir de las relaciones entre los distintos componentes (Khan et al, 2005). En este tipo de técnicas se suelen aplicar muchas técnicas de Procesamiento de Lenguaje Natural (NLP), por ejemplo, la resolución de la anáfora. La mayoría de estos sistemas se basan en reglas que representan un conocimiento a priori de lengua, aunque pueden combinarse con métodos estadísticos.

Machine Learning. En la mayoría de las aplicaciones relacionadas con la ingeniería del lenguaje, como la traducción o el reconocimiento del habla, los métodos basados en aprendizaje automático son los que mejores resultados proporcionan. Su problema radica en que necesitan un gran corpus de aprendizaje, es decir, los documentos originales y sus resúmenes. Las primeras aproximaciones utilizaron Hidden Markov Model (HMM) y Support Vector Machine (SVM) (Fuentes et al, 2007). Actualmente, están teniendo mucho éxito las basadas en Recurrent



Neural Network (RNN) (Nallapati et al, 2017). Para conseguir este corpus se suelen utilizar métodos semiautomáticos, que evitan tener que hacer todos los resúmenes de forma manual, aunque se sabe que la calidad de los modelos puede empeorar, ya que depende del corpus con el que se aprenda.

2.1. Crítica al estado del arte y propuesta

Las aportaciones al estado del arte de la sección anterior representan grandes avances en el campo de los resúmenes automáticos y la segmentación. Sin embargo, aún existen ciertos retos que no están del todo resueltos, como es el caso de la representación continua de las palabras mediante vectores, que puede ser de especial interés en tareas de segmentación e identificación de tópicos, y en la selección de frases para elaborar resúmenes. Por otra parte, los principales trabajos de resumen automático están centrados en textos periodísticos escritos, normalmente noticias, de los que se esperan resúmenes de unas pocas frases. Sin embargo, no hay muchos trabajos relacionados con resumir documentos, de texto o audio, con estructura muy variada o con múltiples noticias o secciones

Por ello, la propuesta de este trabajo intenta avanzar tanto en el aprovechamiento de la representación continua de las palabras, como en los mecanismos de obtención automática de clases semánticas. Todo ello orientado a tareas reales usando programas con estructura heterogénea de televisión.

Diseño de la solución

El tercer capítulo trata del diseño de la solución. Está dividido en tres grandes bloques: arquitectura del sistema, diseño detallado y tecnología utilizada. El primer apartado expone en líneas generales el sistema de resumen, hablando del flujo de los datos y de los procesos aplicados. Estos procesos y sus técnicas se describen de forma individual y detallada en el segundo apartado. Por último, en la parte de tecnología utilizada se listan las herramientas software usadas, indicando donde y para qué se han aplicado.

3.1. Arquitectura del sistema

Con el fin de ilustrar el sistema propuesto, se ha elaborado un esquema. Los círculos indican elementos de entrada o salida de los procesos, representados por un rectángulo. Por otra parte, las flechas se usan para relacionar los elementos antes mencionados. Hay de dos tipos: continuas y discontinuas, que representan las opciones de una decisión. Se debe tener en cuenta que los elementos contenidos en las cajas grises no forman parte de la construcción del resumen sino de la evaluación.

El primer paso que aplica el sistema es el de procesado del documento. En esta fase se abstrae el texto original, generando otro con mayor valor semántico. Además, lo hace apto para ser usado por las herramientas software escogidas, ya que algunas no aceptan letras y signos de puntuación del español. A partir del documento procesado se extrae el vocabulario y las frases, elementos utilizados para construir la matriz de entrada de las técnicas de modelado semántico. Opcionalmente, se puede aplicar segmentación para determinar las partes en las que se divide el texto. Por último, se seleccionan las frases, que unidas producen el resumen del sistema.

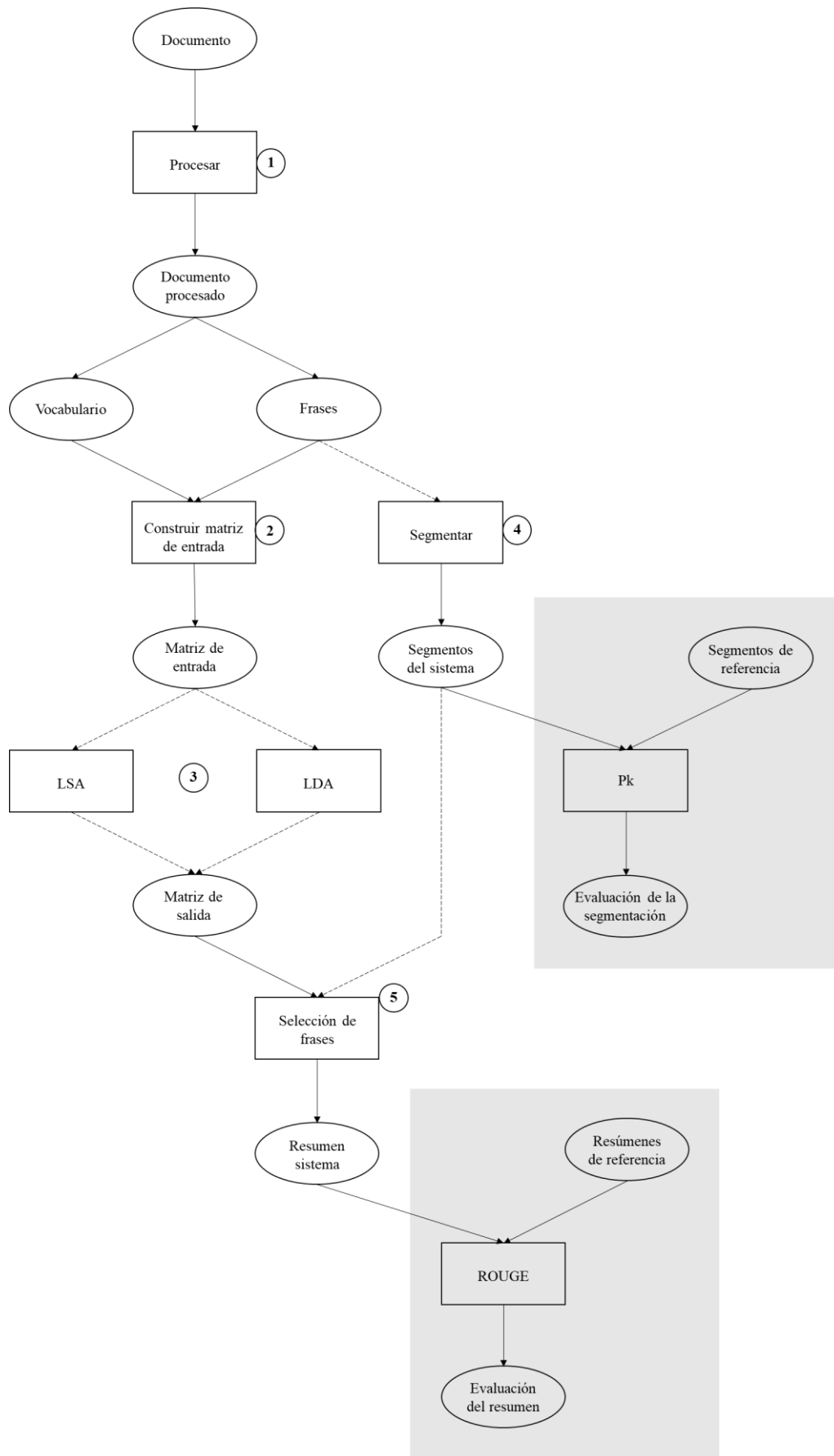


Figura 1. Esquema del sistema de resumen.

3.2. Diseño detallado

El objetivo de este apartado es explicar de forma detallada los procesos del sistema de resumen propuesto. Para ello, se seguirá el orden natural que indica el esquema: procesado del documento, construcción de la matriz de entrada, modelado semántico, segmentación semántica y selección de frases. Estos subapartados terminarán con una justificación de la técnica a utilizar en el sistema, seleccionada de entre todas las propuestas.

3.2.1. Procesado del documento

El primer paso del procesado es dividir el documento en frases. Esta tarea es trivial en el caso del DUC ya que, al tratarse de archivos en formato XML, consiste en extraer el contenido de las etiquetas correspondientes a las frases. Por otra parte, para la noche en 24h, primero se ha concatenado el texto y después se ha dividido por puntos. De esta forma, una frase está compuesta por las palabras que hay entre dos puntos.

Tras esto, a cada frase se le han aplicado una serie de transformaciones, que a continuación enumeramos:

1. Pasar el texto a minúsculas.
2. Dividir la frase en palabras.
3. Eliminar los stopwords, palabras que carecen de significado por si solas, por ejemplo: los conectores o las preposiciones.
4. Aplicar stemming, una técnica que reduce las palabras a su raíz.
5. Descartar los caracteres no alfabéticos.

Obtenidas las frases, se procede a extraer el vocabulario a partir de ellas, es decir, el conjunto de palabras diferentes que aparecen en el documento.

3.2.2. Construcción de la matriz de entrada

Esta fase se apoya en los resultados de la anterior para construir la matriz de incidencia, que sirve de entrada a las técnicas de modelado semántico. Las filas de la matriz son las palabras del vocabulario, mientras que las columnas son las frases. Entre las métricas más populares de relleno, se encuentran:

- Frecuencia. Número de ocurrencias de la palabra en la frase correspondiente.
- Binaria. Si la palabra está en la frase tomará el valor verdadero, en su defecto falso.
- TF-IDF. Contra más frecuente es la palabra en la frase y más infrecuente en el documento, mayor importancia tendrá. Su cálculo se suele separar en dos partes: el Term Frequency (*tf*) que mide la importancia de la palabra en la frase, y el Inverse Document Frequency (*idf*) que mide la inversa de la frecuencia de aparición de la palabra en el conjunto de las frases. Existen diferentes formas de calcular el *tf*, que varían entre ellas principalmente en el mecanismo de normalización. Una de ellas viene dada por la siguiente fórmula:

$$tf(i,j) = \frac{n(i,j)}{\sum_k n(k,j)}$$

donde $n(i, j)$ es el número de ocurrencias de la palabra i en la frase j , mientras que $\sum_k n(k, j)$ el de ocurrencias de todas las palabras en la frase j .

$$idf(i) = \log\left(\frac{|D|}{d_i}\right)$$

$|D|$ es el número de frases del texto de entrada y d_i el número de frases en las aparece la palabra i .

$$tfidf(i, j) = tf(i, j) * idf(i)$$

3.2.3. Modelado semántico

Este apartado tratará sobre el LSA y LDA, las técnicas de modelado semántico usadas en el sistema. Destacar que el LDA históricamente no se ha usado para resumir documentos, razón que hace novedosa la propuesta.

3.2.3.1. Latent Semantic Analysis (LSA)

El Latent Semantic Analysis, habitualmente conocido como LSA, es un método no supervisado que a través de la estadística infiere la representación implícita de la semántica del texto (Landauer et al, 1998). Guarda información acerca de qué palabras se usan juntas y cuáles son comunes entre frases. Cuantas más palabras tengan en común, mayor relación semántica habrá.

LSA aplica Singular Value Decomposition, SVD, una descomposición matricial del álgebra lineal que modela las relaciones entre palabras y frases para hallar patrones.

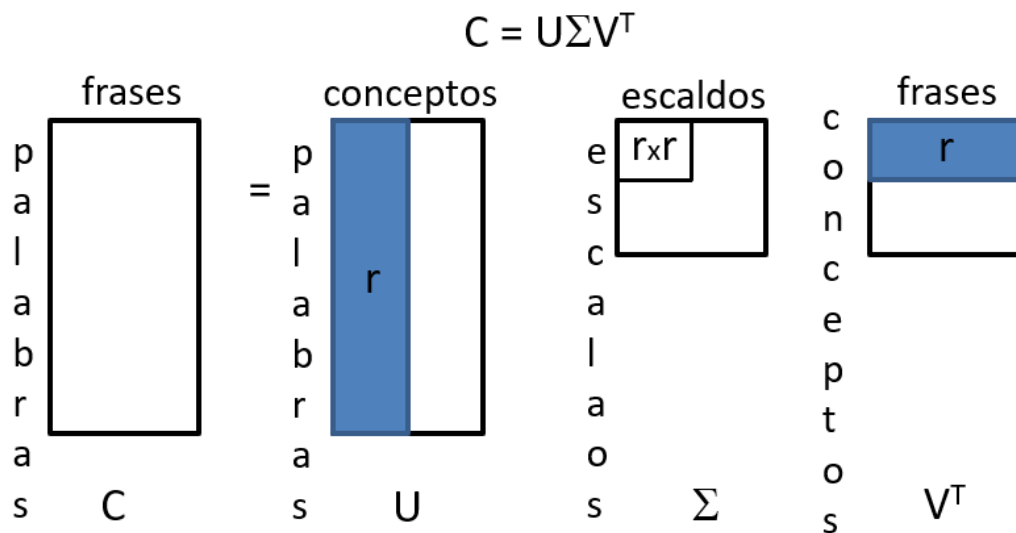


Figura 2. SVD.

C . Es la matriz de entrada de n palabras por m frases.

U . Es la primera matriz que se obtiene de la descomposición. Sus dimensiones son iguales que C . Sus valores representan el peso de cada palabra en los diferentes conceptos.

Σ . Es una matriz rectangular, diagonal y no negativa de dimensiones n por n , siendo los elementos de la diagonal los valores singulares. Cada fila i corresponde al peso del concepto i de U .

V^T . Es una matriz n por n de valores reales que expresa la ocurrencia de las frases en los conceptos.

En el siguiente ejemplo (Martin & Jurafsky, 2009) puede verse el efecto de aplicar LSA a un conjunto de documentos. La matriz inicial C representa la aparición, o no, de cada palabra en los distintos documentos, y la matriz C_2 es la V^T que se obtiene al descomponer la matriz original C considerando sólo dos valores singulares.

C	d_1	d_2	d_3	d_4	d_5	d_6
ship	1	0	1	0	0	0
boat	0	1	0	0	0	0
ocean	1	1	0	0	0	0
wood	1	0	0	1	1	0
tree	0	0	0	1	0	1
C_2	d_1	d_2	d_3	d_4	d_5	d_6
ship	0.85	0.52	0.28	0.13	0.21	-0.08
boat	0.36	0.36	0.16	-0.20	-0.02	-0.18
ocean	1.01	0.72	0.36	-0.04	0.16	-0.21
wood	0.97	0.12	0.20	1.03	0.62	0.41
tree	0.12	-0.39	-0.08	0.90	0.41	0.49

Figura 3. Matrices C y C_k de ejemplo.

El objetivo del SVD es construir una matriz de rango inferior C_k que aproxime C . Reduciendo la dimensionalidad se disminuye el ruido y se consigue que palabras similares se agrupen en localizaciones próximas. Por ejemplo: la semejanza entre d_2 y d_3 en C es 0, mientras que en C_2

$$0.52 * 0.28 + 0.36 * 0.16 + 0.72 * 0.36 + 0.12 * 0.20 - 0.39 * -0.08 \approx 0.52$$

Esto tiene sentido, ya que “*ship*” (barco) y “*boat*” (bote) se parecen.

Para terminar, se comentarán brevemente las limitaciones del LSA:

1. Solo utiliza la información del texto de entrada, dejando a un lado recursos léxicos elaborados manualmente, y por tanto muy orientados a los objetivos de los diseñadores del recurso, como WordNet.
2. No aprovecha el orden de las palabras ni las relaciones semánticas y morfológicas para determinar el significado de las palabras.
3. El rendimiento del SVD decrece drásticamente con el tamaño del texto y con la heterogeneidad de las palabras.

Las dos primeras limitaciones afectan directamente a las palabras polisémicas, aquellas cuyo significado varía según el contexto en el que se utilicen.

3.2.3.2. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation es un modelo probabilístico generativo de un corpus (Blei et al, 2003). Esta técnica supone que los documentos están representados como mezclas aleatorias de temas latentes, donde cada uno se caracteriza por una distribución sobre las palabras. Conviene destacar que es un algoritmo que genera una doble asociación: documentos-temas y temas-palabras. Entre sus aplicaciones, se encuentran:

- Apoyo a la creación de resúmenes automáticos.
- Identificación de temas.
- Segmentación de un texto usando cambios de tema.
- Reducción de la dimensionalidad de los datos (si el número de temas escogido es menor a los que hay en realidad).

Esta técnica parte de una matriz cuyas filas son documentos y las columnas son palabras. Como resultado devuelve dos matrices, la primera de documentos por temas y la segunda de temas por palabras. Estas matrices son fruto de un método de aprendizaje, siendo el más usado es el muestreo de Gibbs.

LDA sigue la distribución Dirichlet, denotada como $Dir(\alpha)$. Pertenece a la familia de las distribuciones de probabilidad continuas multivariantes. Se define como:

$$p(P = \{p_i\} | \alpha_i) = \frac{\prod_i \Gamma(\alpha_i)}{\Gamma(\sum_i \alpha_i)} \prod_i p_i^{\alpha_i - 1}, \text{ donde } \sum_i p_i = 1, p_i \geq 0$$

Dirichlet toma tres parámetros de entrada:

- T . Es un escalar que determina el número de temas.
- α . Un vector de tamaño T (se debe normalizar) que expresa la proporción de cada tema en los documentos.
- β . Contra mayor sea más palabras tendrán los temas.

Para aclarar el impacto de α , se ofrecen representaciones geométricas donde se varia su valor. Contra más oscuro sea el color, mayor densidad habrá en esa zona.

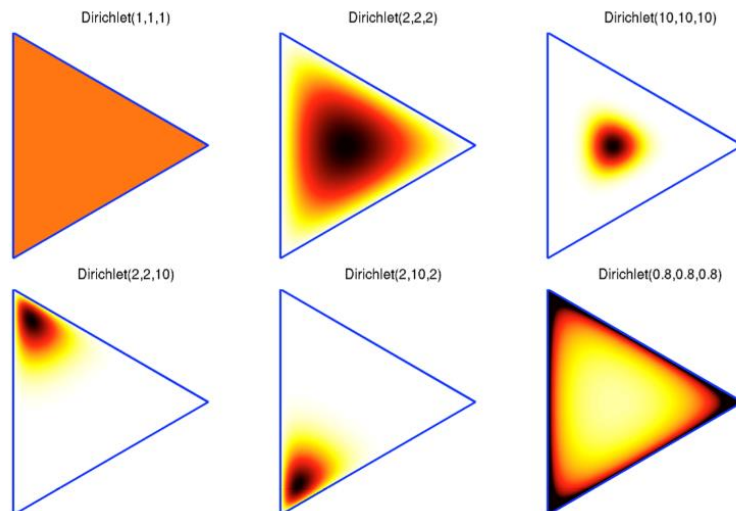


Figura 4. Representación geométrica de Dirichlet.

A continuación, se introduce el algoritmo generativo de Dirichlet. Además, se añade una traza a partir del segundo bucle del algoritmo con el fin de clarificar los pasos que lo definen. Se

asume que las probabilidades de cada palabra en el tema j vienen dadas. El número de palabras y temas escogidos son $d = 2$ y $T = 3$, respectivamente.

Algoritmo generativo del LDA

Desde $j = 1$ hasta T temas

(1) Escoger $\varphi^j \sim \text{Dir}(\beta)$, $\varphi_1^j, \dots, \varphi_V^j$ probabilidad de cada palabra en el tema j

Desde $d = 1$ hasta D documentos

(2) Escoger $\theta^d \sim \text{Dir}(\alpha)$, $\theta_1^d, \dots, \theta_T^d$ probabilidad de cada tema en el documento d

Desde $i = 1$ hasta N_d palabras en el documento d

(3) Escoger $z_i \sim \text{Mult}(\theta^d)$, escoger el tema para la palabra i

(4) Escoger $w_i \sim \text{Mult}(\varphi^{z_i})$, escoger la palabra i

Traza

(2) Escoger $\theta^d \sim \text{Dir}(\alpha)$: $(\frac{1}{3}, \frac{1}{6}, \frac{1}{2})$, que corresponden a los 3 temas

Iteración 1:

(3) Escoger el tema de la palabra 1: **tema 3**

(4) Escoger la palabra 1: **coche**

Iteración 2:

(3) Escoger el tema de la palabra 2: **tema 1**

(4) Escoger la palabra 2: **árbol**

Algoritmo 1. Algoritmo generativo del LDA.

Comentar, que la principal desventaja del LDA es definir el número de temas, ya que determinar este número puede ser complejo.

3.2.4. Segmentación semántica

La mayoría de los idiomas presentan una escritura lineal, de izquierda a derecha, de derecha a izquierda o de arriba abajo. Esto provoca localidad semántica, es decir, las palabras y frases toman sentido gracias a sus adyacentes. Por este motivo las ideas dentro de un texto están bien organizadas. La técnica desarrollada hace uso de los conceptos expuestos para segmentar el texto semánticamente.

Obtenidas las frases, se procede a vincular las palabras que las forman a vectores de números reales (word embeddings), que modelan su significado (Mikolov et al, 2013). Los vectores se obtienen mediante una red neuronal, que utiliza el contexto de las palabras para proyectarlas en

un espacio n-dimensional, donde se agrupan por contenido semántico. En cuanto a la forma de representar frases, se ha optado por sumar los vectores de las palabras que las forman. Un ejemplo ideal de esta representación continua de las palabras en un espacio n-dimensional se muestra a continuación. Si se tiene representada cada palabra con un vector, al restar al vector de Rey el vector de hombre obtendríamos un nuevo vector que sólo contiene las características de un Rey, de forma que al sumarle el vector de Mujer obtenemos un nuevo vector que representa las características de un Rey y de una Mujer, que debe coincidir con el vector que representa la palabra Reina.

$$\text{Rey} - \text{Hombre} + \text{Mujer} = \text{Reina}$$

A continuación, se procede a explicar la búsqueda de los segmentos con coherencia semántica que se ha realizado en el presente trabajo. Este procedimiento se basa en construir ventanas de N frases consecutivas, cuyo punto de inicio puede ser cualquier frase del documento, de modo que donde tantas ventanas como número frases. El objetivo de las ventanas es agrupar frases del mismo tema, consiguiendo una búsqueda más informada y en consecuencia más precisa.

Cálculo de ventanas

Escoger N

Desde $i = 1$ hasta F frases

$$\text{Ventana } i = \text{Media}(frase_i, \dots, frase_{\min(i+N, F)})$$

Tamaño de ventana = 5

Ventana1:	1	2	3	4	5	6	7	8
Ventana2:	1	2	3	4	5	6	7	8
Ventana3:	1	2	3	4	5	6	7	8
Ventana4:	1	2	3	4	5	6	7	8
Ventana5:	1	2	3	4	5	6	7	8
Ventana6:	1	2	3	4	5	6	7	8
Ventana7:	1	2	3	4	5	6	7	8
Ventana8:	1	2	3	4	5	6	7	8

Algoritmo 2. Cálculo de ventanas.

Como puede observarse en el algoritmo, las ventanas a partir de la $N - 1$ no se calculan con N frases, la quinta con 4, la sexta con 3... De no ser así, no se podrían identificar segmentos a partir de la frase $N - 1$.

Tras esto, se calcula la distancia euclídea entre ventanas consecutivas. Siendo P y Q vectores n-dimensionales, $P = (p_1, p_2, \dots, p_n)$ y $Q = (q_1, q_2, \dots, q_n)$, la distancia euclídea se calcula como:



$$d_E(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=0}^n (p_i - q_i)^2}$$

Para hacer más sencillo el ejemplo, se supone que las ventanas son vectores de dos componentes, X e Y .

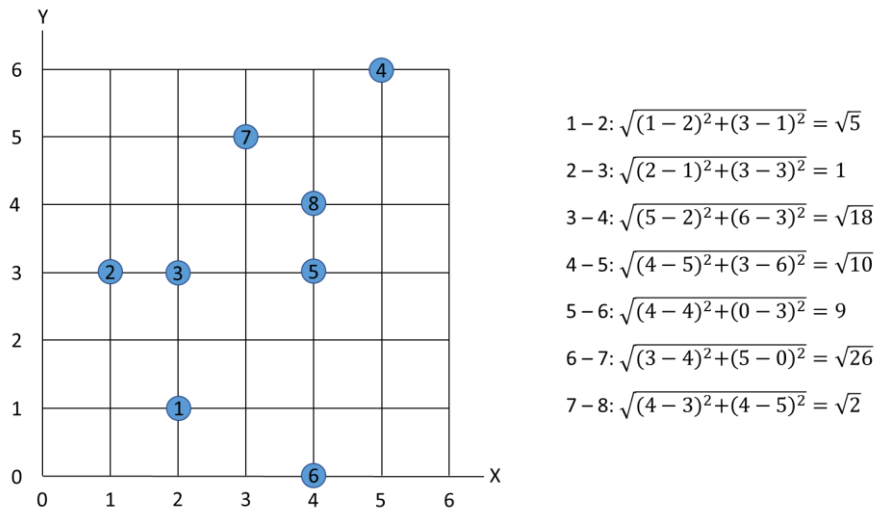


Figura 5. Ejemplo de cálculo de distancias entre ventanas consecutivas.

El último paso es identificar los picos de la función de distancia. Para el caso expuesto, se sitúan en las posiciones 3 y 5, que se corresponde directamente con la numeración de las frases. De esta forma, los segmentos obtenidos son: $[1 - 3[$, $[3 - 5[$ y $[5 - 8[$. Hay que destacar que el intervalo es abierto para la segunda componente, es decir, es excluyente.

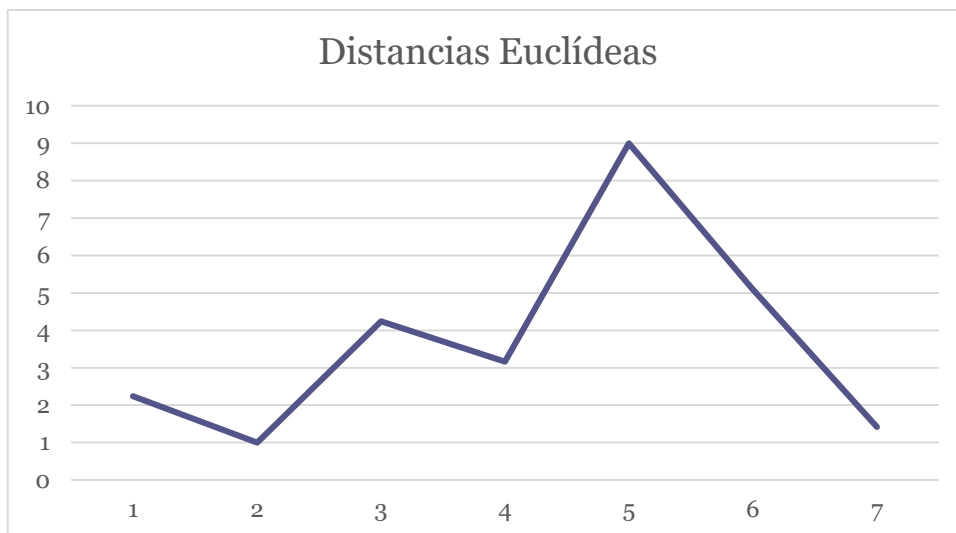


Figura 6. Función de distancia para las ventanas de ejemplo.

Respecto al tamaño de la ventana, se puede pensar que cuanto mayor es N mejores resultados se obtendrán, pero no es totalmente cierto. Este parámetro está relacionado con la naturaleza del texto, si se producen cambios de tema frecuentes es probable que un tamaño menor funcione mejor, ya que se asegura que las frases comparten tema.

3.2.5. Selección de frases

A lo largo de este apartado se expondrán varias técnicas de selección de frases (Ozsoy et al, 2011). La explicación se hará suponiendo el uso de LSA (también se puede usar LDA), e irá acompañada de los cálculos y transformaciones aplicadas sobre las matrices devueltas por el SVD.

3.2.5.1. Gong & Liu

Esta aproximación únicamente usa la matriz V^T [1], que expresa la importancia de los conceptos en las frases. El método es sencillo, para cada concepto se elige la frase con mayor valor (en negrita [2]). Algunas consideraciones:

- Si el número de frases que debe contener el resumen es alto, la selección introducirá algunas menos relevantes.
- Probablemente haya conceptos en los que varias frases tengan valores altos, pero sólo se elige una.
- Se asume que todos los conceptos tienen la misma importancia.

[1]	fras0	fras1	fras2	fras3
con0	0,9784	0,371	0,1109	0,5629
con1	0,3538	0,8315	0,0775	0,0337
con2	0,1357	0,5407	0,7712	0,2674
con3	0,1529	0,2202	0,4659	0,5897



[2]	fras0	fras1	fras2	fras3
con0	0,9784	0,371	0,1109	0,5629
con1	0,3538	0,8315	0,0775	0,0337
con2	0,1357	0,5407	0,7712	0,2674
con3	0,1529	0,2202	0,4659	0,5897

Algoritmo 3. Gong & Liu.

3.2.5.2. Steinberger & Jezek

La aproximación de Steinberger y Jezek trata de solventar las limitaciones del anterior método asignando un peso (longitud) a cada frase [3]. El peso viene determinado por la importancia de las frases en los conceptos, V [1], y por la relevancia de estos, Σ [2]. Además, se vale del parámetro n , que determina el número de conceptos usados en el cálculo.

$$\sqrt{\sum_{j=1}^n V_{ij} * \Sigma_{jj}}$$

[1]	con0	con1	con2	con3
fras0	0,9784	0,3538	0,1357	0,1529
fras1	0,371	0,8315	0,5407	0,2202
fras2	0,1109	0,0775	0,7712	0,4659
fras3	0,5629	0,0337	0,2674	0,5897



[2]	esc0	esc1	esc2	esc3
esc0	0,3569	0	0	0
esc1	0	0,2756	0	0
esc2	0	0	0,2499	0
esc3	0	0	0	0,1177



[3]	con0	con1	con2	con3	long
fras0	0,9784	0,3538	0,1357	0,1529	0,69326017
fras1	0,371	0,8315	0,5407	0,2202	0,70476395
fras2	0,1109	0,0775	0,7712	0,4659	0,50364878
fras3	0,5629	0,0337	0,2674	0,5897	0,52631739

Algoritmo 4. Steinberger & Jezek.

3.2.5.3. Murray & Renals & Carletta

La principal ventaja respecto a la anterior técnica es que permite seleccionar múltiples frases para los conceptos más relevantes. Esta cantidad se calcula extrayendo la parte entera de multiplicar el número de frases, en este caso 4, por Σ [2]. Por último, se seleccionan las n frases con mayor valor en V^T [1], cuyo resultado se encuentra en [3].

[1]	fras0	fras1	fras2	fras3
con0	0,9784	0,371	0,1109	0,5629
con1	0,3538	0,8315	0,0775	0,0337
con2	0,1357	0,5407	0,7712	0,2674
con3	0,1529	0,2202	0,4659	0,5897



[2]	esc0	esc1	esc2	esc3
esc0	0,3569	0	0	0
esc1	0	0,2756	0	0
esc2	0	0	0,2499	0
esc3	0	0	0	0,1177

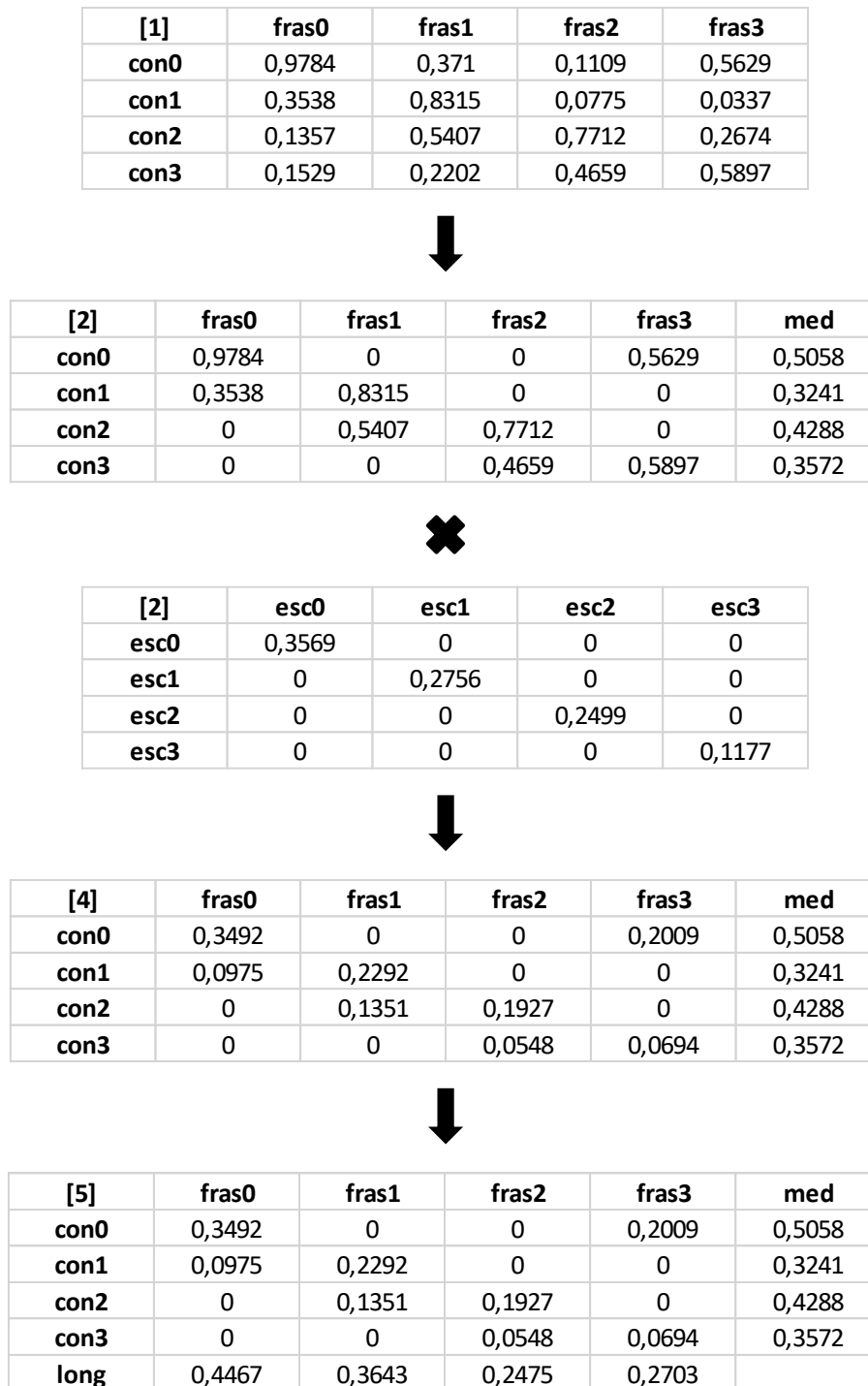


[3]	fras0	fras1	fras2	fras3	num
con0	0,9784	0,371	0,1109	0,5629	1
con1	0,3538	0,8315	0,0775	0,0337	1
con2	0,1357	0,5407	0,7712	0,2674	0
con3	0,1529	0,2202	0,4659	0,5897	0

Algoritmo 5. Murray & Renals & Carletta.

3.2.5.4. Cross

Esta es la técnica que se ha decidido aplicar en el sistema. Cross aplica un preproceso a V^T [1], que consiste en sustituir por 0 todos los valores menores a la media de su fila [2]. Con esto, se consigue descartar las frases poco relevantes. Después, se multiplica la matriz resultante por Σ [3], obteniéndose [4]. El último paso es el cálculo de la longitud, que se obtiene sumando las columnas [5]. Aquellas con mayor valor se añaden al resumen.



Algoritmo 6. Cross.

3.3. Tecnología utilizada

En cuanto al sistema operativo, se ha utilizado Linux, concretamente Ubuntu 12.04 LTS. Respecto al entorno de desarrollo, estaba compuesto por un editor de texto, Sublime Text 3, y una terminal. Ocasionalmente, también se ha usado Jupyter Notebook, herramienta que permite desarrollar de forma interactiva.

El lenguaje de programación escogido para el desarrollo del sistema de resumen ha sido Python. Un lenguaje interpretado de alto nivel y propósito general con una sintaxis que favorece la legibilidad. Es multiparadigma, ya que soporta programación orientada a objetos, programación imperativa y, en menor medida, programación funcional. Python es administrado por la Python Software Foundation y cuenta con una licencia compatible con la GNU General Public License. Las razones por las que se ha escogido son:

- Facilidad de aprendizaje.
- Dispone de una amplia variedad de módulos.
- Es muy popular en NLP.
- Su documentación es muy amplia.
- Tamaño de la comunidad.

```
>>> import this
The Zen of Python, by Tim Peters

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!
```

Figura 7. El zen de Python.

Hay que comentar que se ha utilizado Python a través de Anaconda, una distribución libre de Python y R usada habitualmente en el campo científico. Está orientada a simplificar el despliegue y administración de los paquetes software, que se realiza a través de conda.

A continuación, se enunciarán los paquetes utilizados junto a una breve descripción y para qué se han usado:

- Gensim. Sirve para realizar modelos semánticos no supervisados a partir de texto plano. Es un módulo robusto cuyos algoritmos son eficientes y distribuidos. Se ha utilizado para el entrenamiento del modelo de lenguaje (Word2Vec) y para el LDA.
- NLTK. Plataforma líder para la construcción de programas Python que trabajan con lenguaje natural. Proporciona interfaces sencillas para más de 50 corpus y recursos y un conjunto de bibliotecas clasificación, tokenización, derivación, etiquetado, análisis y razonamiento semántico del texto. Este módulo se ha aplicado principalmente en el procesamiento de los documentos.

- BeautifulSoup. Es un módulo para extraer datos de archivos HTML y XML. Se ha empleado para extraer las frases de los archivos del corpus DUC.
- NumPy. Es el paquete fundamental para la computación científica en Python. Proporciona vectores n-dimensionales y funciones para transformarlos, así como operaciones del algebra lineal como la transformada de Fourier. Los vectores pueden contener cualquier tipo de datos, son genéricos, por lo que la integración con BBDD es sencilla. La parte donde más se ha usado ha sido en el LSA y en los métodos de selección de frases.
- PyRouge. Es un envoltorio para Python del ROUGE.
- Pandas. Expone una nueva estructura de datos, el DataFrame, que se asemeja a una tabla Excel. Se ha utilizado para crear tablas de resultados.
- PeakUtils. Este paquete proporciona utilidades relacionadas con la detección de picos en datos unidimensionales. Incluye entre otros, funcionalidades para estimar las líneas base o encontrar los índices de los picos. Se ha empleado para detectar cambios en la temática de las frases y así segmentar el texto.

Nombre	Versión	Compilación
beautifulsoup4	4.5.3	py35_0
Gensim	0.13.4.1	np112py35_0
Nltk	3.2.2	py35_0
Numpy	1.12.0	py35_0
Pandas	0.23.1	py35h637b7d7_0
PeakUtils	1.1.1	<pip>
Pyrouge	0.1.3	<pip>
Python	3.5.3	1

Por otra parte, como se ha comentado, se ha usado ROUGE en su implementación original para la evaluación de los resúmenes. Esta herramienta está escrita en Perl, por lo que se ha instalado su SDK (Perl Strawberry) para hacerlo funcionar.

Este capítulo está dividido en tres secciones: corpus, medidas de evaluación y experimentos. En el primero se tratarán los conjuntos de datos empleados para evaluar el sistema, es decir, el DUC y La noche en 24 horas. Después se describirán las medidas de evaluación empleadas para los segmentos y los resúmenes, Pk y ROUGE, respectivamente. Por último, se expondrán los resultados obtenidos para las aproximaciones segmentadas y no segmentadas, tratando individualmente el uso del LSA y LDA en cada uno de ellos.

4.1. Corpus

Como se ha comentado, se han usado dos corpus: DUC y La noche en 24 horas. Hay que destacar que para ambos se han aplicado las aproximaciones no segmentadas, mientras que solo en el segundo se ha añadido la segmentación.

4.1.1. Document Understanding Conferences (DUC)

Actualmente hay mucho interés y actividad en construir potentes sistemas de información multipropósito. Entre las agencias involucradas, se encuentran: DARPA (Defense Advanced Research Projects Agency); el programa TIDES (Translingual Information detection Extraction and Summarization), Advanced Question & Answering Program del ARDA (Advanced Research and Development Activity) y el TREC (Text REtrieval Conference) del Instituto Nacional de Estándares y Tecnología (NIST). Estos programas se enfocan en áreas muy concretas de NLP, por lo que requieren sus propios modelos de evaluación.

Dentro del TIDES, y entre otros investigadores interesados en la comprensión de documentos, creció un grupo que fue centrándose en sistemas de resumen y evaluación. Parte del TIDES fue convocado a un taller en otoño del año 2000 para explorar diferentes formas de resumir un conjunto común de documentos. Como añadido, en marzo de ese mismo año comenzaron a desarrollar un sistema de evaluación de resúmenes.

En paralelo al TIDES apareció el DUC (Document Understanding Conferences), unas conferencias a cargo del NIST y patrocinadas por ARDA dedicadas a la evaluación de sistemas de resumen de textos.

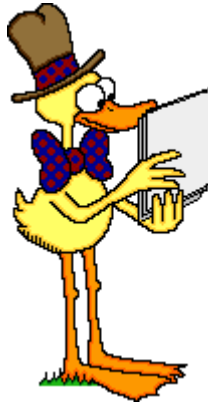


Figura 8. DUC.

En 2008 el DUC pasó a ser parte de las “pistas” del TAC (Text Analysis Conference), concretamente la “pista” de resumen. El TAC son una serie de talleres de evaluación organizados para promover la investigación en NLP y aplicaciones relacionadas. Proporciona grandes colecciones de documentos, procedimientos comunes de evaluación y un foro donde las organizaciones comparten sus resultados. Esta conferencia está formada por conjuntos de tareas conocidas como “pistas”, donde cada una trata subproblemas particulares del NLP con valor para el usuario final.

TAC Track Home Pages										
All Tracks	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Question Answering	2008									
Recognizing Textual Entailment	2008	2009	2010	2011						
Summarization	2008	2009	2010	2011			2014			
Knowledge Base Population		2009	2010	2011	2012	2013	2014	2015	2016	2017
Extraction from Drug Labels										2017

Figura 9. Cronología de las "pistas" del TAC.

TAC está organizado por el Grupo de Recuperación de la División de Acceso a la Información (IAD) del Laboratorio de Tecnología del NIST. Iniciado en 2008, TAC surgió del DUC y de la “pista” de Respuesta a Preguntas del TREC. Está patrocinado por NIST y otras agencias gubernamentales de los EEUU, y supervisado por un Comité Asesor compuesto por representantes del gobierno, la industria y el mundo académico.

Antes de pasar al DUC 2007, cuyo corpus se ha utilizado para evaluar el sistema propuesto, a modo de evolución cronológica, se expondrán los años previos a este.

El primer año que se celebró el DUC fue en 2001, donde varios grupos participaron en la creación de resúmenes de textos periodísticos a partir de uno y varios documentos. El corpus se componía de 30 conjuntos de alrededor de 10 documentos cada uno. Los resúmenes fueron examinados por su cobertura y legibilidad por 10 jueces. La conferencia del 2002 fue similar, como añadido, NIST creó extractos manuales, que sirvieron como línea base para la evaluación automática de los resúmenes.

En la conferencia del 2003 se usó un subconjunto de documentos de los corpus TDT y TREC. Las tareas se centraron en la creación de resúmenes breves, unos de ~100 palabras y otros de ~10. Comentar que los más cortos se crearon considerando eventos, diferentes puntos de vista y a partir de una pregunta.



El DUC del 2004 fue semejante al del 2003 en muchos aspectos, mismo corpus, enfoque y algunas tareas. Aunque las dos primeras tareas fueron iguales, la tercera y cuarta fueron nuevas, ya que exploraron el resumen a partir de una entrada ruidosa producida a través de la traducción automática (del árabe al inglés).

En el DUC del 2005 y 2006, los participantes debían sintetizar una respuesta breve, organizada y fluida a una pregunta. Se usó un subconjunto de 25 a 50 documentos del TREC como corpus. Con esto, se pretendía modelar un sistema real de respuesta a preguntas que no se podían resolver con un solo dato. Los resúmenes generados se compararon con otros 4 de referencia creados por el NIST aplicando ROUGE. Para realizar con éxito la tarea, se debían de combinar técnicas de NLP e IR.

El DUC del 2007 consistió en dos tareas independientes entre sí, donde los participantes escogieron realizar una o ambas.

La primera de ellas es la misma que la del DUC 2006, modelar un sistema de respuesta a preguntas complejas del mundo real, donde las preguntas no se pueden responder indicando un nombre, fecha, cantidad... Dadas unas preguntas y un conjunto de 25 documentos representativos para responderlas, los participantes debían diseñar un sistema que elaborase un resumen bien organizado de no más de 250 palabras que respondiese a las preguntas. Para realizar con éxito la tarea, se debían de combinar técnicas de recuperación de información y de procesado del lenguaje natural, incluyendo métodos como la recuperación de pasajes, compresión de la información y generación de texto fluido.

¿Qué países han firmado el Tratado de Ottawa para la eliminación de minas terrestres antipersona y cuántos lo han ratificado? ¿Qué países se negaron a firmarlo y por qué?
¿Cuán efectivo ha sido el tratado?

Por otra parte, la segunda tarea consistía en producir pequeños resúmenes (aprox. 100 palabras) multidocumento que actualizaran la información de un hilo de noticias. Para ello, se partía de la premisa de que los usuarios se habían leído las noticias previas. El objetivo de esta tarea era la de informar al lector sobre actualizaciones de cierto tema.

Los textos a resumir provenían del AQUAINT, un corpus de noticias en inglés que consta de 1.033.461 documentos tomados del New York Times, la Associated Press y la agencia de noticias Xinhua. El corpus abarca noticias de los años 1999-2000 (1996-2000 para Xinhua), y está distribuido por el Consorcio de Datos Lingüísticos (LDC). Los evaluadores del NIST desarrollaron una lista de 45 temas de interés y asignaron un conjunto de 25 documentos relevantes al tema provenientes del AQUAINT.

Los resúmenes de la primera tarea se evaluaron de diferentes formas, una de ellas fue bajo criterios lingüísticos, dirigidos a evaluar la legibilidad y fluidez del resumen valiéndose únicamente de él mismo. Los criterios se puntuaron por separado en una escala del 1 al 5, donde 5 indica que el resumen es bueno y 1 lo contrario.

- Gramática. El resumen no debe contener fechas, formateos internos del sistema, errores en mayúsculas o frases gramáticamente incorrectas que dificultan la lectura del texto.
- No redundancia. Se debe evitar la repetición de frases completas, hechos o nombres. La repetición de nombres se puede solucionar sustituyéndolos por pronombres o sintagmas nominales.
- Claridad referencial. Debe ser sencillo identificar a quién o a qué hacen referencia los pronombres y sintagmas nominales. Una referencia no será clara si su identidad o relación con la historia es difusa.

- Enfoque. Las frases solo deben contener información relacionada con el resto del texto.
- Estructura y coherencia. El resumen debe estar bien estructurado y organizado, debe construirse manteniendo la coherencia con el tema y con el resto de las frases.

Por otra parte, NIST evaluó manualmente la capacidad de respuesta relativa a la pregunta por responder de cada resumen. La capacidad de respuesta deberá ser medida principalmente en términos de cantidad de información que ayuda a responder la pregunta. La calidad lingüística del texto solo se tendrá en cuenta si interfiere con la capacidad de respuesta.

Criterio	Peso
Gramática	0,05
No redundancia	0,01
Claridad referencial	0,07
Enfoque	0,02
Estructura y coherencia	0,20
Capacidad de respuesta	0,65

Con el objetivo de ejecutar el ROUGE-1.5.5 para calcular el ROUGE-2 y ROUGE-SU4, NIST creó 4 resúmenes de referencia por cada conjunto de documentos.

ROUGE-1.5.5.pl -n 2 -x -m -2 4 -u -c 95 -r 1000 -f A -p 0.5 -t

En total se crearon 4 líneas de resúmenes, que según su procedencia se pueden clasificar en: humanos, línea base simple; las 250 primeras palabras del documento más reciente, línea base general; generados mediante un sistema de resumen genérico de alto rendimiento, y los de los participantes.

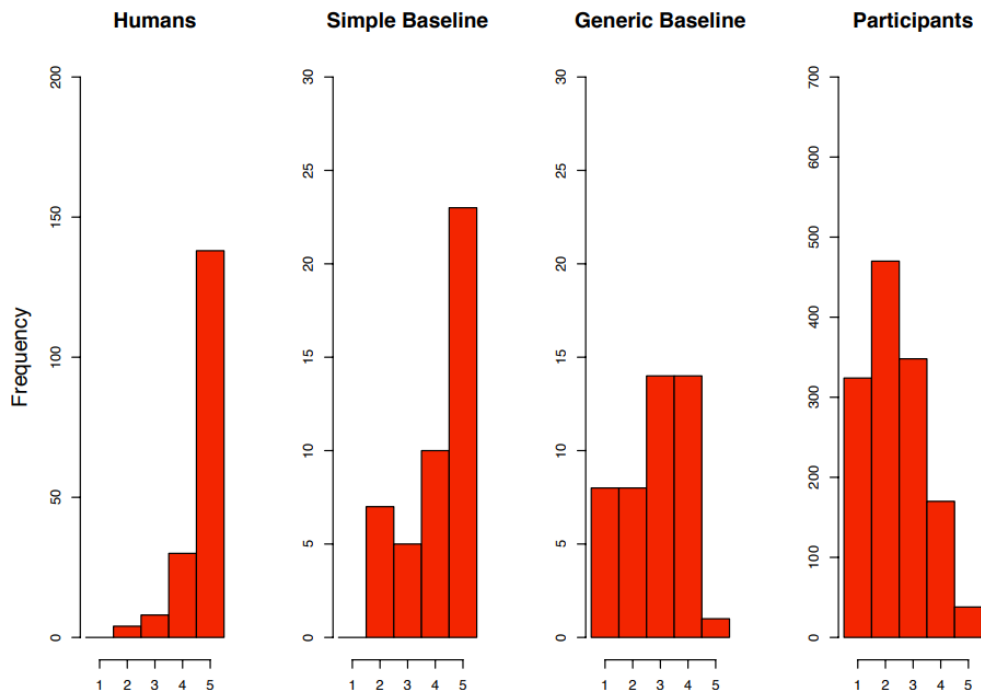


Figura 10. Frecuencia de puntuación de la capacidad de respuesta por procedencia del resumen en DUC 2007.

4.1.2. La noche en 24 horas

La Noche en 24 Horas es un programa de actualidad que se emite de lunes a viernes a partir de las 22:30 en el Canal 24 horas de Radiotelevisión Española. El programa presenta siempre la misma estructura: primero el presentador recibe al entrevistado del día. Los entrevistados suelen ser personajes de la vida política. Acto seguido, y después de la ronda de preguntas individual, ambos se dirigen a la mesa de analistas, donde cuatro contertulios habituales reciben al invitado y continúan debatiendo con él y planteándole cuestiones. Después de ello, el invitado se marcha y el presentador se queda a solas con los contertulios.

A partir de ahí, empieza un bloque en el que se tratan uno a uno todos los temas de la actualidad diaria. Este bloque informativo del programa dura aproximadamente una hora y, con estas dinámicas, es capaz de tratar entre cuatro y cinco noticias al día, tanto nacionales como internacionales.

Al final del programa, el presentador vuelve a establecer conexión con la redacción para que le hagan un repaso rápido de las noticias más importantes del día que no se han dado dentro del análisis. Para finalizar, un periodista anuncia cuáles son los titulares de periódicos que saldrán a la venta al día siguiente.

Se puede destacar que la estructura de estos programas es heterogénea, es decir, junto a secuencias de noticias cortas o titulares hay tramos largos del programa donde se comentan temas políticos o se entrevista a personas. Para realizar los experimentos se han utilizado las transcripciones de algunas partes de los programas. Algunas características de este corpus pueden verse en la siguiente tabla, siendo el eje X las longitudes (número de en frases) de los segmentos y el eje Y la frecuencia.

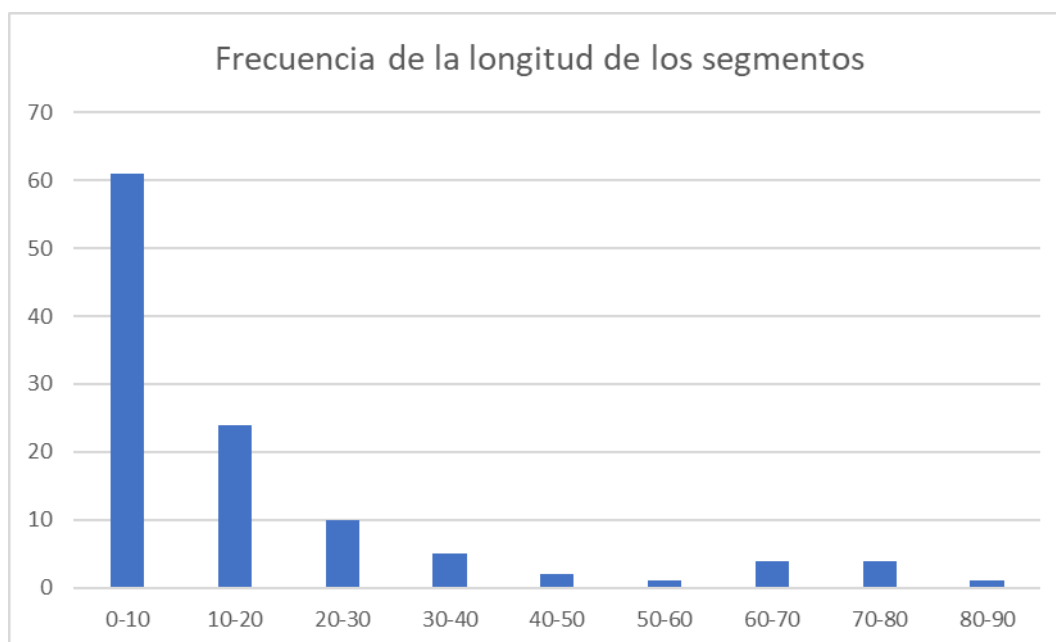


Figura 11. Frecuencia de la longitud de los segmentos para RTVE.

Para la realización de los experimentos se hizo una segmentación manual de los programas, donde se indicaron los cambios de temática. Por otra parte, para evaluar los resúmenes del sistema, fue necesario la elaboración de unos de referencia. Estos, se realizaron manualmente de forma extractiva, y contenían el 20% del documento original.

Número de palabras totales	27881
Número de palabras medias por programa de TV	3983
Número de palabras del programa de TV más corto	2924
Número de palabras del programa de TV más largo	4980

Figura 12. Descripción de RTVE.

4.2. Medidas de evaluación

Este apartado trata sobre las medidas de evaluación. Primero se hablará del ROUGE, que puntúa la calidad del resumen a partir de varias métricas, cuya explicación orbita un ejemplo común. A continuación, le sigue el P_k , encargado de valorar los segmentos del sistema.

4.2.1. ROUGE

ROUGE o Recall-Oriented Understudy for Gusting Evaluation, son un conjunto de métricas que sirven para evaluar un resumen automático (sistema) respecto a un conjunto de resúmenes humanos (referencia).

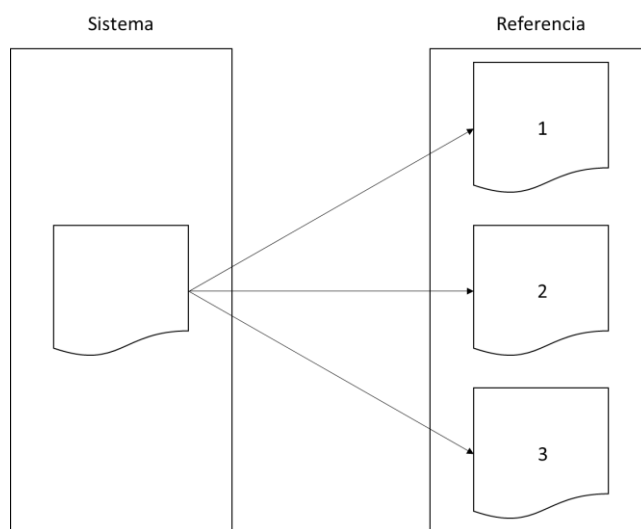


Figura 13. Esquema de evaluación del ROUGE.

Antes de describir las métricas que lo forman, se definen otras que proceden del ámbito de la búsqueda y recuperación de información (IR). En particular las medidas de *Precisión* y *Recall*, y la *F*, que es una puesta en común de ambas, razón por la que se utilizará para evaluar la bondad de los resúmenes.

- $Precisión = \frac{\text{Número de palabras comunes}}{\text{Número de palabras en el resumen sistema}}$
- $Recall = \frac{\text{Número de palabras comunes}}{\text{Número de palabras en el resumen referencia}}$
- $F = 2 * \frac{precisión*recall}{precisión+recall}$

Conocidas las medidas de IR, se procede a definir las que forman el ROUGE. Para ello, se ha elaborado un ejemplo, que por comodidad cuenta con un único resumen de referencia (R) y otro del sistema (S).

Resúmenes
R: <u>rodolfo</u> , <u>el gato negro</u> , <u>está sentado</u>
S: <u>el gato</u> , <u>rodolfo</u> , <u>estaba sentado</u>

ROUGE-N. valora la similitud entre el vocabulario del resumen del sistema respecto al de referencia cuantificando los n-gramas comunes. Por defecto, cuenta con ROUGE-1, ROUGE-2, ROUGE-3 y ROUGE-4. Para simplificar, se explicarán los dos primeros.

ROUGE-1																	
R: { (rodolfo), (el), (gato), (negro), (está), (sentado) } – Tamaño 6																	
S: { (el), (gato), (rodolfo), (estaba), (sentado) } – Tamaño 5																	
<table border="1"> <thead> <tr> <th>Referencia</th> <th>Sistema</th> </tr> </thead> <tbody> <tr> <td>(rodolfo)</td> <td style="background-color: green;"></td> </tr> <tr> <td>(el)</td> <td style="background-color: green;"></td> </tr> <tr> <td>(gato)</td> <td style="background-color: green;"></td> </tr> <tr> <td>(negro)</td> <td style="background-color: red;"></td> </tr> <tr> <td>(está)</td> <td style="background-color: red;"></td> </tr> <tr> <td>(sentado)</td> <td style="background-color: green;"></td> </tr> <tr> <td></td> <td style="text-align: center;">4</td> </tr> </tbody> </table>	Referencia	Sistema	(rodolfo)		(el)		(gato)		(negro)		(está)		(sentado)			4	
Referencia	Sistema																
(rodolfo)																	
(el)																	
(gato)																	
(negro)																	
(está)																	
(sentado)																	
	4																
<table border="1"> <thead> <tr> <th>Precisión</th> <th>Recall</th> <th>F</th> </tr> </thead> <tbody> <tr> <td>$\frac{4}{5} = 0,8$</td> <td>$\frac{4}{6} \approx 0,67$</td> <td>$2 * \frac{0,8 * 0,67}{0,8 + 0,67} \approx 0,73$</td> </tr> </tbody> </table>	Precisión	Recall	F	$\frac{4}{5} = 0,8$	$\frac{4}{6} \approx 0,67$	$2 * \frac{0,8 * 0,67}{0,8 + 0,67} \approx 0,73$											
Precisión	Recall	F															
$\frac{4}{5} = 0,8$	$\frac{4}{6} \approx 0,67$	$2 * \frac{0,8 * 0,67}{0,8 + 0,67} \approx 0,73$															

Algoritmo 7. ROUGE-1.

ROUGE-2

R: { (rodolfo, el), (el, gato), (gato, negro), (negro, está), (está, sentado) } – Tamaño 5

S: { (el, gato), (gato, rodolfo), (rodolfo, estaba), (estaba, sentado) } – Tamaño 4

Referencia	Sistema
(rodolfo,el)	
(el, gato)	
(gato, negro)	
(negro, está)	
(está, sentado)	
	1

Precisión	Recall	F
$\frac{1}{4} = 0,25$	$\frac{1}{5} = 0,2$	$2 * \frac{0,25 * 0,2}{0,25 + 0,2} = 0,2$

Algoritmo 8. ROUGE-2.

Conviene destacar que si se hubiese aplicado stemming, técnica para reducir una palabra a su raíz, palabras como “está” y “estaba” se habrían convertido en la misma, mejorando los resultados obtenidos. Por otra parte, según aumenta el tamaño de los n-gramas se observa un empeoramiento en los resultados, tendencia que también se observa en el sistema desarrollado.

ROUGE-L. Su nombre proviene de la técnica que emplea, la Subsecuencia Común más Larga, también conocida como LCS por su acrónimo en inglés. Gracias a ella se puede conocer la similitud estructural de las oraciones seleccionadas por el sistema respecto a las de referencia.

ROUGE-L

R: rodolfo, el gato negro, está sentado

S: el gato, rodolfo, estaba sentado

Precisión	Recall	F
$\frac{3}{5} = 0,6$	$\frac{3}{6} = 0,5$	$2 * \frac{0,6 * 0,5}{0,6 + 0,5} = 0,54$

Algoritmo 9. ROUGE-L.

En un caso real, los valores de estos resultados son mucho menores. Por lo general, contra más términos contenga el resumen, peor será el ROUGE-L.

ROUGE-W. También se basa en LCS, pero además valora la consecutividad de las palabras. Debido al parecido con ROUGE-L se ha optado por no añadir su cálculo.

ROUGE-S*. Al igual que el ROUGE-N se encarga de medir la similitud del vocabulario. La principal diferencia respecto a su hermano es que utiliza bigramas consecutivos separados por un número variable de términos.

ROUGE-S*

R: { (rodolfo, el), (rodolfo, gato), (rodolfo, negro), (rodolfo, está), (rodolfo, sentado), (el, gato), (el, negro), (el, está), (el, sentado), (gato, negro), (gato, está), (gato, sentado), (negro, está), (negro, sentado), (está, sentado) } – Tamaño 15

S: { (el, gato), (el, rodolfo), (el, estaba), (el, sentado), (gato, rodolfo), (gato, estaba), (gato, sentado), (rodolfo, estaba), (rodolfo, sentado), (estaba, sentado) } – Tamaño 10

Referencia	Sistema
(rodolfo, el)	
(rodolfo, gato)	
(rodolfo, negro)	
(rodolfo, está)	
(rodolfo, sentado)	
(el, gato)	
(el, negro)	
(el, está)	
(el, sentado)	
(gato, negro)	
(gato, está)	
(gato, sentado)	
(negro, está)	
(negro, sentado)	
(está, sentado)	
	4

Precisión	Recall	F
$\frac{4}{10} = 0,4$	$\frac{4}{15} = 0,27$	$2 * \frac{0,4 * 0,27}{0,4 + 0,27} = 0,32$

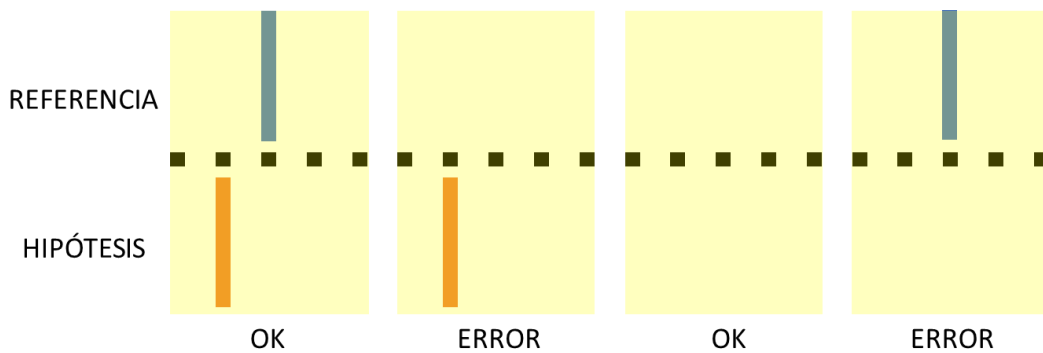
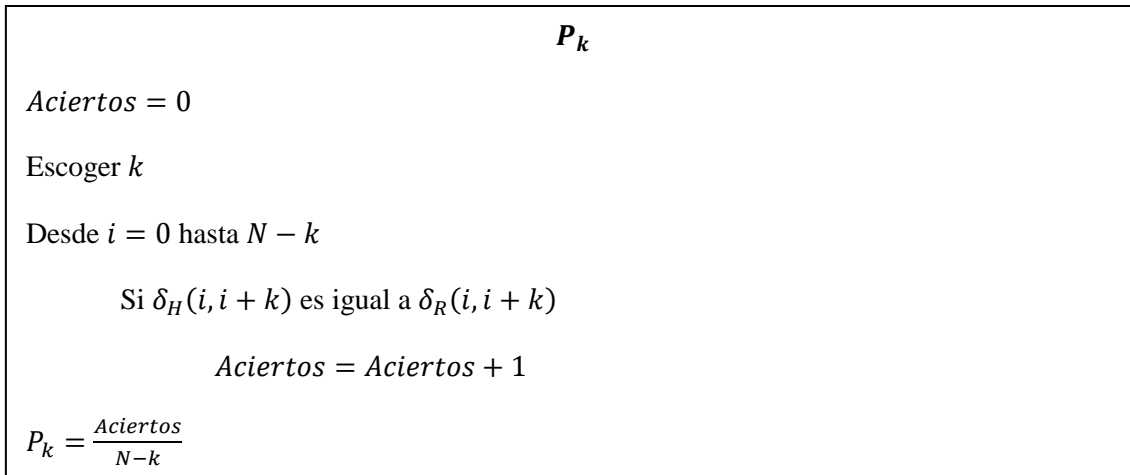
Algoritmo 10. ROUGE-S*.

ROUGE-SU*. Es igual que el ROUGE-S*, pero añade una limitación en el número de términos que separan las palabras.

4.2.2. P_k

P_k es una medida que expresa la probabilidad de error de segmentación. Dados dos puntos en el texto, uno de hipótesis y otro de referencia, calcula la probabilidad media de que la segmentación sea incorrecta. Comentar que es posible añadir un margen de error.

A continuación, se ha incluye el pseudocódigo con el cálculo del P_k , donde N es la longitud del texto; ya sea en palabras, frases u otra unidad, depende del nivel de segmentación, y k es el tamaño de la ventana. La ventana, es decir $\delta(i, j)$, toma el valor 1 cuando incluye un segmento, y 0 en caso contrario.



Algoritmo 11. P_k .

El algoritmo se puede compactar en una fórmula. El símbolo \oplus representa la puerta lógica XOR, que vale 0 cuando los valores son iguales y 1 en caso contrario.

$$P_k = \frac{\sum_{i=0}^{N-k} \delta_H(i, i + k) \oplus \delta_R(i, i + k)}{N - k}$$

4.3. Experimentos

En este apartado se evaluarán los resultados de las técnicas de resumen aplicadas a los corpus del DUC y RTVE. En ambos se analizará el uso del LSA y LDA, mientras que solo en el segundo se añadirá la segmentación. Además, se evaluará la técnica de segmentación desarrollada, cuyos resultados están fuertemente relacionados con la parte segmentada de RTVE.

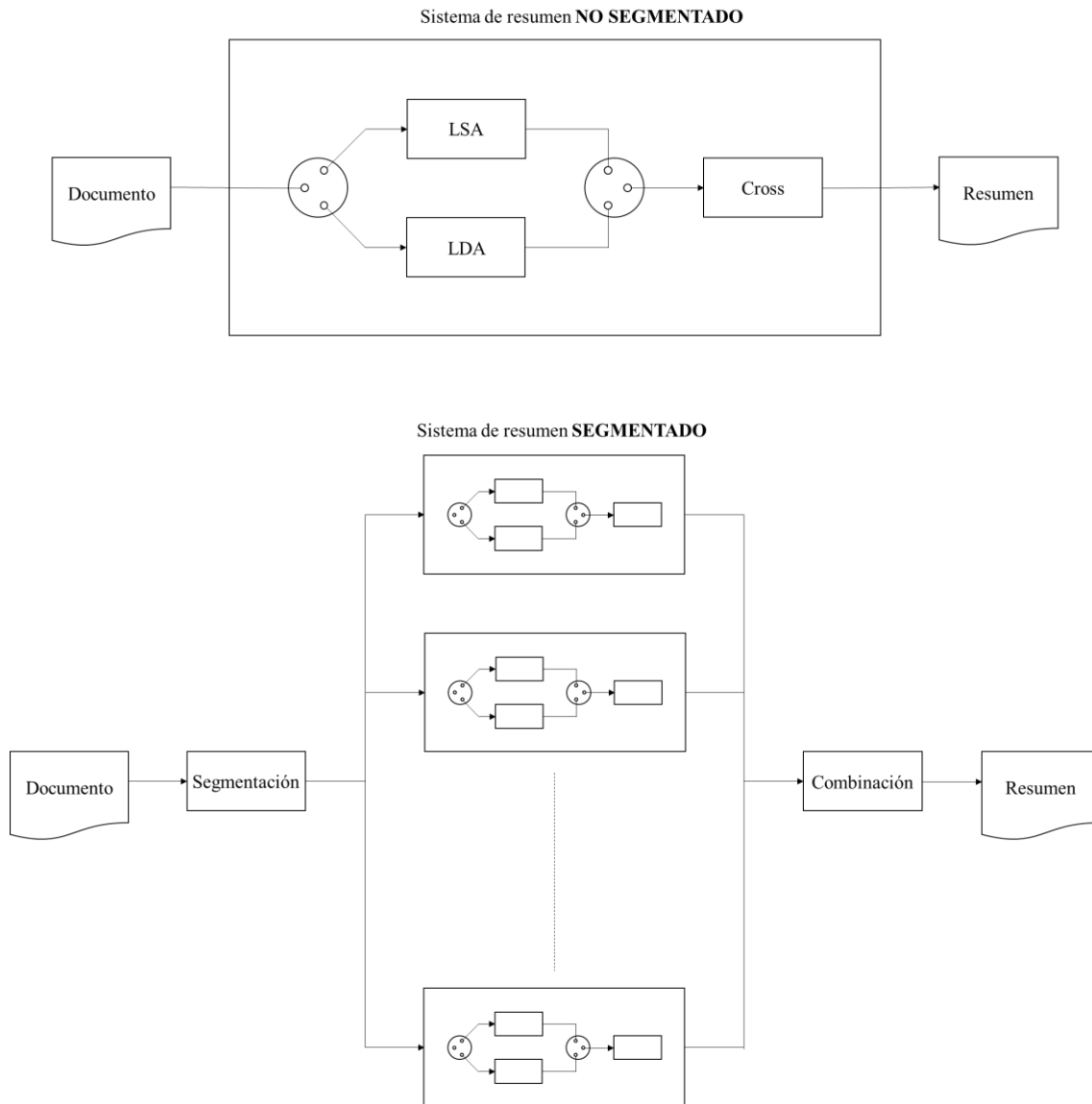


Figura 14. Resumen de los sistemas desarrollados.

Las tablas que se añaden son versiones reducidas y simplificadas de las originales. Comentar que, para cada aproximación y posibles parámetros asociados se han realizado 10 experimentos. Debido a su envergadura, las tablas se pueden encontrar en el anexo de este documento.

4.3.1. DUC

Para este corpus no se han realizado experimentos con segmentación. Al ser un corpus de referencia en la comunidad no se querían edulcorar los resultados añadiendo segmentaciones no consensuadas por la organización.

LSA								
	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
Media	0,3919	0,0771	0,0212	0,0086	0,2595	0,0874	0,1423	0,1443

Tabla 1. Resumen de los resultados del LSA para DUC.

LDA								
	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
5	0,4154	0,0972	0,0305	0,0135	0,2801	0,0945	0,1640	0,1660
10	0,4158	0,0976	0,0307	0,0136	0,2806	0,0949	0,1642	0,1662
20	0,4163	0,0970	0,0303	0,0134	0,2811	0,0948	0,1647	0,1666
40	0,4157	0,0969	0,0302	0,0134	0,2805	0,0946	0,1643	0,1662
80	0,4159	0,0974	0,0305	0,0135	0,2808	0,0947	0,1644	0,1663
Media	0,4158	0,0972	0,0304	0,0135	0,2806	0,0947	0,1643	0,1663
Des. Est.	0,0003	0,0003	0,0002	0,0001	0,0004	0,0002	0,0002	0,0002

Tabla 2. Resumen de los resultados del LDA para DUC.

Como se puede observar, la desviación estándar para los diferentes números de temas es semejante. Con el objetivo de comparar ambas aproximaciones, se ha usado la media del LDA como línea base. De la comparación, se puede extraer que en este caso el LDA es ligeramente superior al LSA, sumando hasta un 2% en campos como: ROUGE-2, ROUGE-L y ROUGE-SU*.

LDA vs LSA							
ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
0,0239	0,0202	0,0092	0,0049	0,0211	0,0073	0,0220	0,0220

Tabla 3. Comparativa entre los resultados del LDA y LSA para DUC.

4.3.2. La noche en 24 horas

Este apartado se ha dividido en tres secciones claramente diferenciadas, en primer lugar, se analizan los resultados de las aproximaciones que no usan segmentación previa, es decir, LSA y LDA sin aditivos. Después, como introducción a las técnicas que sí usan segmentación previa, se evalúan los resultados de la segmentación, determinando el tamaño óptimo de ventana. Por último, como se ha anticipado, se analizarán las técnicas con segmentación, concluyendo empíricamente si la segmentación aporta algún beneficio.

4.3.2.1. Aproximaciones sin segmentación

Tal y como sucedía con el DUC, parece que el número de temas no afecta a los resultados, por ello, se ha usado la media como elemento comparativo al LSA.

LSA								
	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
Media	0,3108	0,1377	0,0970	0,0853	0,2810	0,0824	0,0715	0,0723

Tabla 4. Resumen de los resultados del LSA para RTVE.

LDA								
	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
5	0,3010	0,1143	0,0674	0,0554	0,2709	0,0787	0,0686	0,0693
10	0,3009	0,1146	0,0679	0,0558	0,2707	0,0786	0,0687	0,0695
20	0,3005	0,1146	0,0680	0,0557	0,2704	0,0785	0,0687	0,0694
40	0,3006	0,1144	0,0674	0,0553	0,2698	0,0786	0,0687	0,0694
80	0,2989	0,1115	0,0648	0,0529	0,2681	0,0778	0,0679	0,0686
Media	0,3004	0,1139	0,0671	0,0550	0,2700	0,0784	0,0685	0,0692
Des. Est.	0,0008	0,0012	0,0012	0,0011	0,0010	0,0003	0,0003	0,0003

Tabla 5. Resumen de los resultados del LDA para RTVE.

Se puede apreciar un comportamiento totalmente antagónico al anterior corpus, todas las métricas del LDA son peores en comparación con las del LSA. La degradación es más notable en medidas relacionadas con los n-gramas, es decir, ROUGE-1, ROUGE-2, ROUGE-3 y ROUGE-4, llegando a reducirse hasta en un 3% en el caso del ROUGE-3 y ROUGE-4.

LDA vs LSA							
ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
-0,0104	-0,0238	-0,0299	-0,0303	-0,0110	-0,0040	-0,0030	-0,0030

Tabla 6. Comparativa entre los resultados del LDA y LSA para RTVE.

4.3.2.2. Segmentación

En este apartado se expondrán los resultados de la segmentación. Hay que recordar que la medida usada es el P_k , donde valores menores implican mejores resultados. El primer gráfico es sobre la media, donde el valor mínimo es el obtenido con el tamaño de ventana igual a 25, valiendo aproximadamente 0,42 (58% de precisión).

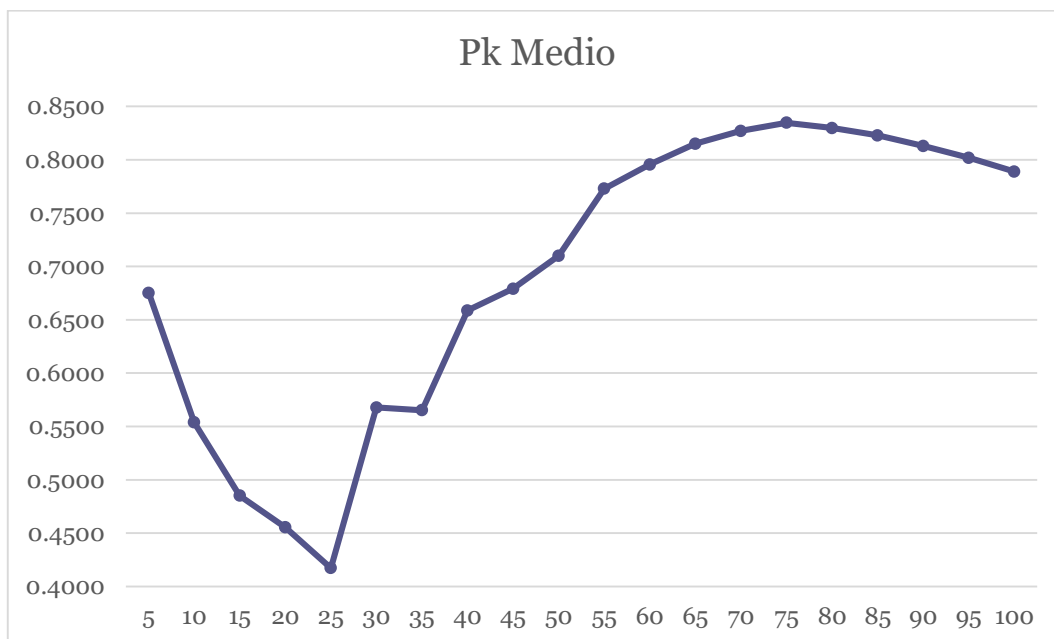


Figura 15. Gráfico de línea con los Pk medios por tamaño de ventana.

Resulta interesante analizar la desviación estándar, que expresa la dispersión en los resultados de los 10 programas que forman el corpus. Como se evidencia, el mínimo se produce con tamaño 65, mientras que el máximo con 35. Para la ventana escogida, la desviación es de aproximadamente 0,17.

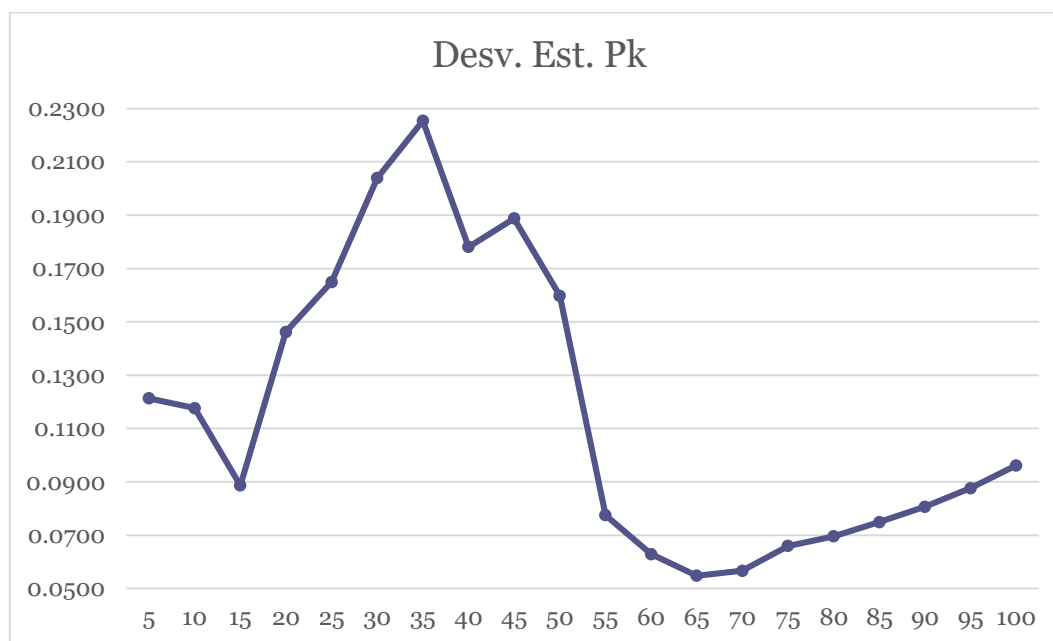


Figura 16. Gráfico de línea con la desviación estándar de los Pk por tamaño de ventana.

4.3.2.3. Aproximaciones con segmentación

Esta sección quizás sea la más interesante, ya que se incluye la segmentación como proceso previo a las técnicas de resumen. Las siguientes tablas comprimen los resultados obtenidos de aplicar LSA y LDA acompañado de la segmentación.

Segmentación + LSA								
	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
Media	0,5639	0,2683	0,1998	0,1799	0,5200	0,1688	0,2895	0,2902

Tabla 7. Resumen de los resultados del LSA segmentado para RTVE.

Segmentación + LDA								
	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
5	0,5338	0,3251	0,2668	0,2462	0,5061	0,1890	0,2417	0,2422
10	0,5315	0,3216	0,2626	0,2422	0,5033	0,1883	0,2400	0,2405
20	0,5286	0,3154	0,2561	0,2356	0,5000	0,1875	0,2382	0,2387
40	0,5272	0,3134	0,2535	0,2333	0,4986	0,1862	0,2360	0,2365
80	0,5256	0,3117	0,2511	0,2309	0,4963	0,1857	0,2345	0,2350
Media	0,5293	0,3174	0,2580	0,2376	0,5009	0,1873	0,2381	0,2386
Desv. Est.	0,0033	0,0057	0,0065	0,0064	0,0039	0,0014	0,0029	0,0029

Tabla 8. Resumen de los resultados del LDA segmentado para RTVE.

Antes de analizar cuál de las dos aproximaciones es mejor, es crucial determinar si la adición de la segmentación aporta algún beneficio. Apoyándose en los resultados se puede afirmar que

el F-score ha aumentado considerablemente, en promedio un 17,67% respecto a las aproximaciones sin segmentación.

Segmentado vs No segmentado								
	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
LSA	0,2531	0,1306	0,1028	0,0946	0,2390	0,0864	0,2180	0,2179
LDA	0,2289	0,2036	0,1909	0,1826	0,2309	0,1089	0,1696	0,1694

Tabla 9. Comparativa entre los resultados segmentados y no segmentados.

Para terminar, se comparan las aproximaciones con segmentación. En la tabla, se observa que ninguna de las alternativas es claramente superior a la otra. En métricas como el ROUGE-2 el LDA es prácticamente un 5% mejor que el LSA, mientras que para el ROUGE-SU* sucede lo contrario. Se concluye que la ganancia y pérdida de escoger una respecto a la otra es significativa.

Segmentación + LDA vs Segmentación + LSA							
ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W-1.2	ROUGE-S*	ROUGE-SU*
-0,0346	0,0491	0,0582	0,0577	-0,0191	0,0185	-0,0514	-0,0516

Tabla 10. Comparativa entre los resultados del LDA y LSA segmentados.

CAPÍTULO 5

Conclusiones

En este trabajo se ha desarrollado un sistema de resumen automático de texto, orientado a tareas donde los documentos contienen diferentes secciones. Para ello, tras estudiar las propuestas existentes, se ha escogido una línea de trabajo basada en temas.

Respecto a las técnicas de resumen, se han escogido dos: LSA y LDA. El motivo es que ambas modelan conceptos subyacentes, característica deseada en este problema. El LDA históricamente no se ha utilizado para resumir, razón por la que se compara respecto al LSA, una técnica tradicional con resultados demostrados.

Por otra parte, se observó que las técnicas de selección de frases no se adaptaban del todo a este tipo de documentos. Para solucionar este problema, se ideó un método de segmentación de texto, que en última instancia clasificaba las frases según el tema al que pertenecían. Este conocimiento hizo más informada la selección de frases, haciendo que las técnicas fuesen apropiadas.

Para poder realizar el proyecto, además de lo ya mencionado, se han tenido que estudiar y adaptar las herramientas tecnológicas necesarias, utilizar un corpus estándar de evaluación y crear uno nuevo para el problema que se quería abordar. Los resultados obtenidos, tanto en la parte de segmentación semántica, como en la de resumen automático, han sido buenos, permitiendo deducir que las técnicas propuestas son apropiadas para el problema.

En cuanto a las dificultades encontradas, han sido varias:

- Sin duda, el mayor problema ha sido manejar el volumen de resultados de los experimentos realizados. Se disponían de dos corpus, dos técnicas de resumen, la posibilidad de aplicar segmentación y varios parámetros a determinar. Además, para que los resultados fuesen fidedignos, se realizaron 10 experimentos por cada una de las combinaciones mencionadas.
- Trabajar con texto es complicado, ya que además de necesitar mucho preprocesado, muchas herramientas (por ejemplo, el ROUGE) están diseñadas para soportar un único idioma, el inglés. Esto choca directamente con la evaluación del sistema, ya que el contenido de uno de los corpus usados está en español. Este problema se solventó adaptando los textos al alfabeto inglés.
- La instalación del ROUGE y la comunicación con Python ha sido tediosa. Como se ha comentado, este módulo está escrito en Perl, y al programar en Python se ha tenido que usar un envoltorio del programa original.
- Ha sido complejo construir las tablas de resultados. El motivo es que ROUGE devuelve un texto plano no estructurado con los resultados de la ejecución correspondientes. Para solucionarlo se ha desarrollado un procedimiento semiautomático que lee estos documentos, filtra los campos de interés y los inserta en una matriz.

Una parte de los resultados de este trabajo se han presentado en un Congreso Internacional (SPECOM 2018) cuyas actas se publican en (Canora et al, 2018).

5.1. Relación del trabajo desarrollado con los estudios cursados

Este trabajo está relacionado con varias de las asignaturas de la carrera, principalmente de la Rama Computación. Por una parte hay una presencia de las asignaturas de programación y algorítmica en el desarrollo de los distintos programas. También están presentes los estudios en procesamiento de información textual, como es la asignatura de Sistemas de Almacenamiento y Recuperación de Información, Aprendizaje Automático y Estadística, que son necesarios para comprender y poder utilizar las herramientas relacionadas con el LSA y el LDA, o la representación continua de las palabras.

Bibliografía

- Tur, G., & De Mori, R. (2011). Spoken language understanding: Systems for extracting semantic information from speech. John Wiley & Sons. (Tur & De Mori, 2011)
- Lloret, E., Palomar, M.: Text summarisation in progress: a literature review. *Artificial Intelligence Review* 37(1), 1{41 (2012). (Lloret & Palomar, 2012)
- Cheng, J., & Lapata, M. (2016). Neural Summarization by Extracting Sentences and Words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. (Cheng & Lapata, 2016)
- Erkan, G., Radev, D.R.: Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Int. Res.* 22(1), 457{479 (Dec 2004). (Erkan & Radev, 2004)
- Fuentes, M., Alfonseca, E., Rodriguez, H.: Support vector machines for query-focused summarization trained and evaluated on pyramid data. In: *Proceedings of the 45th Annual Meeting of the ACL, 2007*. (Fuentes et al, 2007)
- Hearst, M. A. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, 23(1), 33-64. (Hearst, 1997)
- Choi, F. Y. (2000, April). Advances in domain independent linear text segmentation. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* (pp. 26-33). Association for Computational Linguistics. (Choi, 2000)
- Furui, S., Kikuchi, T., Shinnaka, Y., Hori, C.: Speech-to-text and speech-to-speech summarization of spontaneous speech. *IEEE Transactions on Speech and Audio Processing* 12(4), 401{408 (2004). (Furui et al, 2004)
- Ozsoy, M. G., Alpaslan, F. N., & Cicekli, I. (2011). Text summarization using latent semantic analysis. *Journal of Information Science*, 37(4), 405-417. (Ozsoy et al, 2011)
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022. (Blei et al, 2003)
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2-3), 259-284. (Landauer et al, 1998)
- Martin, J. H., & Jurafsky, D. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson/Prentice Hall. (Martin & Jurafsky, 2009)
- Gupta, S., Nenkova, A., & Jurafsky, D. (2007, June). Measuring importance and query relevance in topic-focused multi-document summarization. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions* (pp. 193-196). Association for Computational Linguistics. (Gupta et al, 2007)

Marco Canora, Fernando García-Granada, Emilio Sanchis and Encarna Segarra. “An approach to automatic summarization of television programs”. Lecture Notes in Artificial Intelligence. Speech and Computer. 20th International Conference, SPECOM 2018, Leipzig, Germany, September 18–22, 2018. (Canora et al, 2018)

Harabagiu, S., & Lacatusu, F. (2005, August). Topic themes for multi-document summarization. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 202-209). ACM. (Harabagiu & Lacatusu, 2005)

Khan, A. U., Khan, S., & Mahmood, W. (2005, February). MRST: A New Technique For Information Summarization. In WEC (2)(pp. 249-252). (Khan et al, 2005)

Nallapati, R., Zhai, F., Zhou, B.: Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA. pp. 3075-3081 (2017). (Nallapati et al, 2017)

T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS 2013. (Mikolov et al, 2013)



Tablas de resultados originales

	Segmentación									
	5	10	15	20	25	30	35	40	45	50
1029	0,7879	0,6995	0,6011	0,5464	0,4831	0,3237	0,3452	0,2331	0,2722	0,3137
1104	0,6726	0,5963	0,4742	0,4615	0,2118	0,5859	0,5337	0,5745	0,6175	0,6629
1023	0,6197	0,3431	0,5530	0,7323	0,6557	0,8291	0,8214	0,8131	0,8725	0,7938
1103	0,7322	0,6581	0,5546	0,4643	0,4247	0,5654	0,3445	0,6716	0,5126	0,7062
1117	0,7840	0,6369	0,4868	0,3810	0,1620	0,1387	0,1439	0,6457	0,6721	0,7009
1112	0,5316	0,4224	0,5242	0,6126	0,5806	0,6415	0,6618	0,7030	0,7462	0,7917
1028	0,7051	0,5613	0,4396	0,4356	0,4569	0,5313	0,5829	0,6264	0,6554	0,6744
1027	0,7935	0,6480	0,5287	0,2249	0,4756	0,6289	0,6494	0,6711	0,6944	0,7194
1204	0,4200	0,4414	0,3071	0,3630	0,2231	0,7760	0,7500	0,7391	0,8273	0,8190
1106	0,7056	0,5314	0,3824	0,3333	0,5000	0,6581	0,8200	0,9103	0,9214	0,9185
Media	0,6752	0,5539	0,4852	0,4555	0,4174	0,5678	0,5653	0,6588	0,6792	0,7101
Desv. Est.	0,1213	0,1177	0,0887	0,1462	0,1649	0,2039	0,2253	0,1781	0,1887	0,1598

	Segmentación									
	55	60	65	70	75	80	85	90	95	100
1029	0,6959	0,7552	0,8188	0,8571	0,8828	0,8780	0,8729	0,8673	0,8611	0,8544
1104	0,7110	0,7619	0,8098	0,8228	0,8170	0,8108	0,8042	0,7971	0,7895	0,7813
1023	0,8587	0,8506	0,8415	0,8312	0,8194	0,8060	0,7903	0,7719	0,7500	0,7234
1103	0,7249	0,7446	0,7654	0,7874	0,8107	0,8354	0,8491	0,8442	0,8389	0,8333
1117	0,7321	0,7664	0,7745	0,7629	0,7500	0,7356	0,7195	0,7013	0,6806	0,6567
1112	0,8396	0,8736	0,8983	0,9244	0,9521	0,9506	0,9490	0,9474	0,9456	0,9437
1028	0,6946	0,7160	0,7389	0,7632	0,7755	0,7676	0,7591	0,7500	0,7402	0,7295
1027	0,7463	0,7752	0,8065	0,8403	0,8772	0,8716	0,8654	0,8586	0,8511	0,8427
1204	0,8100	0,8000	0,7889	0,7765	0,7625	0,7467	0,7286	0,7077	0,6833	0,6545
1106	0,9154	0,9120	0,9083	0,9043	0,9000	0,8952	0,8900	0,8842	0,8778	0,8706
Media	0,7728	0,7956	0,8151	0,8270	0,8347	0,8298	0,8228	0,8130	0,8018	0,7890
Desv. Est.	0,0775	0,0629	0,0548	0,0567	0,0660	0,0696	0,0749	0,0807	0,0876	0,0961

Tabla 11. Segmentación.

DUC LSA												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,3943	0,3844	0,3892	0,0783	0,0763	0,0773	0,0220	0,0214	0,0217	0,0087	0,0084	0,0086
Exp 2	0,3960	0,3856	0,3906	0,0766	0,0746	0,0756	0,0240	0,0234	0,0237	0,0111	0,0108	0,0110
Exp 3	0,3961	0,3874	0,3916	0,0797	0,0780	0,0788	0,0224	0,0220	0,0222	0,0091	0,0090	0,0091
Exp 4	0,4041	0,3966	0,4003	0,0872	0,0856	0,0864	0,0257	0,0254	0,0255	0,0104	0,0103	0,0104
Exp 5	0,3979	0,3912	0,3944	0,0767	0,0756	0,0761	0,0204	0,0202	0,0203	0,0082	0,0081	0,0082
Exp 6	0,3998	0,3899	0,3947	0,0780	0,0761	0,0770	0,0188	0,0184	0,0186	0,0077	0,0075	0,0076
Exp 7	0,3994	0,3908	0,3949	0,0779	0,0763	0,0771	0,0197	0,0194	0,0195	0,0077	0,0076	0,0077
Exp 8	0,3964	0,3877	0,3919	0,0763	0,0746	0,0754	0,0217	0,0212	0,0214	0,0083	0,0082	0,0082
Exp 9	0,3905	0,3819	0,3860	0,0712	0,0696	0,0704	0,0174	0,0171	0,0172	0,0058	0,0057	0,0058
Exp 10	0,3892	0,3826	0,3858	0,0772	0,0760	0,0766	0,0221	0,0218	0,0219	0,0094	0,0093	0,0093
Media	0,3964	0,3878	0,3919	0,0779	0,0763	0,0771	0,0214	0,0210	0,0212	0,0087	0,0085	0,0086

DUC LSA												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2612	0,2547	0,2579	0,0634	0,1390	0,0870	0,0214	0,1376	0,1410	0,1468	0,1395	0,1430
Exp 2	0,2640	0,2573	0,2606	0,0643	0,1409	0,0882	0,0234	0,1379	0,1413	0,1471	0,1398	0,1433
Exp 3	0,2626	0,2570	0,2597	0,0638	0,1404	0,0877	0,0220	0,1390	0,1419	0,1470	0,1409	0,1438
Exp 4	0,2640	0,2592	0,2615	0,0642	0,1420	0,0884	0,0254	0,1448	0,1473	0,1522	0,1467	0,1493
Exp 5	0,2642	0,2599	0,2620	0,0642	0,1418	0,0883	0,0202	0,1440	0,1462	0,1507	0,1459	0,1482
Exp 6	0,2658	0,2593	0,2625	0,0641	0,1406	0,0880	0,0184	0,1398	0,1432	0,1491	0,1417	0,1452
Exp 7	0,2658	0,2602	0,2630	0,0640	0,1409	0,0880	0,0194	0,1413	0,1442	0,1494	0,1432	0,1462
Exp 8	0,2595	0,2539	0,2565	0,0625	0,1375	0,0859	0,0212	0,1368	0,1397	0,1450	0,1388	0,1417
Exp 9	0,2565	0,2511	0,2537	0,0620	0,1366	0,0853	0,0171	0,1352	0,1380	0,1431	0,1371	0,1399
Exp 10	0,2597	0,2555	0,2575	0,0632	0,1397	0,0869	0,0218	0,1382	0,1404	0,1448	0,1401	0,1423
Media	0,2623	0,2568	0,2595	0,0636	0,1399	0,0874	0,0210	0,1395	0,1423	0,1475	0,1414	0,1443

Tabla 12. DUC LSA.

DUC LDA 5												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,4181	0,4086	0,4132	0,0980	0,0957	0,0968	0,0312	0,0304	0,0308	0,0140	0,0136	0,0138
Exp 2	0,4224	0,4120	0,4170	0,0990	0,0964	0,0976	0,0306	0,0297	0,0301	0,0135	0,0131	0,0133
Exp 3	0,4209	0,4111	0,4159	0,0984	0,0960	0,0972	0,0305	0,0298	0,0301	0,0133	0,0130	0,0131
Exp 4	0,4204	0,4105	0,4153	0,0980	0,0956	0,0968	0,0299	0,0291	0,0295	0,0132	0,0129	0,0130
Exp 5	0,4195	0,4097	0,4144	0,0970	0,0946	0,0957	0,0300	0,0292	0,0296	0,0129	0,0125	0,0127
Exp 6	0,4209	0,4112	0,4159	0,0983	0,0959	0,0971	0,0310	0,0302	0,0306	0,0138	0,0134	0,0136
Exp 7	0,4210	0,4113	0,4160	0,0997	0,0974	0,0985	0,0320	0,0312	0,0316	0,0145	0,0142	0,0144
Exp 8	0,4195	0,4093	0,4143	0,0982	0,0957	0,0969	0,0306	0,0298	0,0302	0,0135	0,0132	0,0134
Exp 9	0,4220	0,4121	0,4169	0,1001	0,0977	0,0989	0,0320	0,0312	0,0316	0,0144	0,0140	0,0142
Exp 10	0,4203	0,4106	0,4153	0,0981	0,0957	0,0969	0,0312	0,0304	0,0308	0,0139	0,0136	0,0138
Media	0,4205	0,4106	0,4154	0,0985	0,0961	0,0972	0,0309	0,0301	0,0305	0,0137	0,0133	0,0135

DUC LDA 5												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2825	0,2761	0,2792	0,0689	0,1518	0,0947	0,0304	0,1589	0,1624	0,1684	0,1609	0,1644
Exp 2	0,2831	0,2761	0,2795	0,0689	0,1513	0,0946	0,0297	0,1613	0,1651	0,1714	0,1632	0,1671
Exp 3	0,2833	0,2767	0,2799	0,0686	0,1511	0,0943	0,0298	0,1602	0,1638	0,1698	0,1621	0,1657
Exp 4	0,2830	0,2763	0,2795	0,0684	0,1505	0,0940	0,0291	0,1605	0,1642	0,1704	0,1625	0,1662
Exp 5	0,2822	0,2757	0,2789	0,0680	0,1497	0,0935	0,0292	0,1591	0,1627	0,1688	0,1610	0,1647
Exp 6	0,2841	0,2777	0,2808	0,0690	0,1520	0,0948	0,0302	0,1606	0,1642	0,1703	0,1626	0,1662
Exp 7	0,2845	0,2779	0,2811	0,0692	0,1524	0,0951	0,0312	0,1609	0,1645	0,1705	0,1628	0,1665
Exp 8	0,2831	0,2763	0,2796	0,0686	0,1509	0,0942	0,0298	0,1595	0,1633	0,1695	0,1615	0,1652
Exp 9	0,2853	0,2786	0,2818	0,0691	0,1521	0,0950	0,0312	0,1623	0,1660	0,1722	0,1642	0,1680
Exp 10	0,2839	0,2775	0,2806	0,0689	0,1517	0,0947	0,0304	0,1604	0,1640	0,1700	0,1623	0,1659
Media	0,2835	0,2769	0,2801	0,0688	0,1513	0,0945	0,0301	0,1604	0,1640	0,1701	0,1623	0,1660

Tabla 13. DUC LDA 5.

DUC LDA 10												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,4185	0,4088	0,4135	0,0985	0,0961	0,0972	0,0316	0,0308	0,0312	0,0144	0,0141	0,0143
Exp 2	0,4220	0,4124	0,4170	0,0988	0,0964	0,0976	0,0311	0,0304	0,0307	0,0141	0,0137	0,0139
Exp 3	0,4229	0,4128	0,4177	0,0996	0,0972	0,0984	0,0312	0,0304	0,0308	0,0137	0,0133	0,0135
Exp 4	0,4220	0,4126	0,4171	0,0990	0,0967	0,0978	0,0306	0,0298	0,0302	0,0133	0,0129	0,0131
Exp 5	0,4190	0,4088	0,4137	0,0985	0,0960	0,0972	0,0309	0,0301	0,0305	0,0136	0,0132	0,0134
Exp 6	0,4185	0,4083	0,4132	0,0969	0,0944	0,0956	0,0297	0,0289	0,0293	0,0129	0,0125	0,0127
Exp 7	0,4204	0,4106	0,4153	0,0983	0,0959	0,0971	0,0309	0,0301	0,0305	0,0133	0,0130	0,0132
Exp 8	0,4225	0,4129	0,4175	0,1002	0,0978	0,0990	0,0315	0,0307	0,0311	0,0143	0,0139	0,0141
Exp 9	0,4219	0,4120	0,4168	0,0981	0,0957	0,0969	0,0305	0,0297	0,0301	0,0133	0,0129	0,0131
Exp 10	0,4213	0,4111	0,4160	0,1003	0,0978	0,0990	0,0327	0,0319	0,0323	0,0148	0,0144	0,0146
Media	0,4209	0,4110	0,4158	0,0988	0,0964	0,0976	0,0311	0,0303	0,0307	0,0138	0,0134	0,0136

DUC LDA 10												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2834	0,2768	0,2800	0,0692	0,1522	0,0950	0,0308	0,1591	0,1627	0,1687	0,1610	0,1646
Exp 2	0,2846	0,2781	0,2813	0,0691	0,1522	0,0950	0,0304	0,1611	0,1647	0,1708	0,1631	0,1667
Exp 3	0,2853	0,2785	0,2818	0,0692	0,1521	0,0950	0,0304	0,1615	0,1652	0,1714	0,1635	0,1672
Exp 4	0,2845	0,2784	0,2813	0,0690	0,1520	0,0948	0,0298	0,1618	0,1652	0,1711	0,1638	0,1672
Exp 5	0,2832	0,2764	0,2797	0,0689	0,1515	0,0946	0,0301	0,1595	0,1633	0,1695	0,1614	0,1652
Exp 6	0,2824	0,2756	0,2789	0,0689	0,1513	0,0946	0,0289	0,1586	0,1623	0,1686	0,1605	0,1643
Exp 7	0,2831	0,2766	0,2798	0,0688	0,1513	0,0945	0,0301	0,1602	0,1638	0,1699	0,1621	0,1658
Exp 8	0,2838	0,2774	0,2805	0,0690	0,1519	0,0948	0,0307	0,1622	0,1658	0,1718	0,1641	0,1677
Exp 9	0,2850	0,2784	0,2816	0,0692	0,1522	0,0950	0,0297	0,1615	0,1651	0,1713	0,1634	0,1671
Exp 10	0,2846	0,2778	0,2811	0,0695	0,1528	0,0955	0,0319	0,1606	0,1644	0,1706	0,1625	0,1663
Media	0,2840	0,2774	0,2806	0,0691	0,1519	0,0949	0,0303	0,1606	0,1642	0,1704	0,1625	0,1662

Tabla 14. DUC LDA 10.

DUC LDA 20												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,4211	0,4110	0,4159	0,0981	0,0957	0,0969	0,0304	0,0296	0,0300	0,0135	0,0131	0,0133
Exp 2	0,4201	0,4098	0,4148	0,0979	0,0955	0,0967	0,0306	0,0298	0,0302	0,0133	0,0130	0,0131
Exp 3	0,4220	0,4116	0,4166	0,0992	0,0967	0,0979	0,0317	0,0309	0,0313	0,0142	0,0138	0,0140
Exp 4	0,4171	0,4075	0,4121	0,0964	0,0941	0,0952	0,0296	0,0288	0,0292	0,0130	0,0126	0,0128
Exp 5	0,4214	0,4119	0,4165	0,0985	0,0962	0,0973	0,0309	0,0302	0,0305	0,0137	0,0133	0,0135
Exp 6	0,4215	0,4120	0,4166	0,0974	0,0951	0,0962	0,0305	0,0297	0,0301	0,0132	0,0129	0,0131
Exp 7	0,4235	0,4130	0,4180	0,0995	0,0970	0,0982	0,0310	0,0302	0,0306	0,0138	0,0134	0,0136
Exp 8	0,4223	0,4127	0,4174	0,0987	0,0963	0,0974	0,0309	0,0301	0,0305	0,0139	0,0135	0,0137
Exp 9	0,4236	0,4136	0,4185	0,0989	0,0964	0,0976	0,0310	0,0302	0,0306	0,0137	0,0133	0,0135
Exp 10	0,4214	0,4114	0,4163	0,0977	0,0953	0,0964	0,0304	0,0296	0,0300	0,0133	0,0130	0,0131
Media	0,4214	0,4115	0,4163	0,0982	0,0958	0,0970	0,0307	0,0299	0,0303	0,0135	0,0132	0,0134

DUC LDA 20												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2840	0,2773	0,2805	0,0690	0,1518	0,0948	0,0296	0,1609	0,1646	0,1708	0,1628	0,1666
Exp 2	0,2843	0,2774	0,2807	0,0689	0,1513	0,0946	0,0298	0,1599	0,1637	0,1699	0,1618	0,1656
Exp 3	0,2839	0,2770	0,2804	0,0688	0,1513	0,0945	0,0309	0,1612	0,1650	0,1713	0,1631	0,1670
Exp 4	0,2841	0,2776	0,2807	0,0689	0,1517	0,0947	0,0288	0,1585	0,1621	0,1682	0,1605	0,1641
Exp 5	0,2845	0,2781	0,2812	0,0691	0,1521	0,0949	0,0302	0,1621	0,1657	0,1716	0,1641	0,1676
Exp 6	0,2854	0,2791	0,2821	0,0691	0,1520	0,0949	0,0297	0,1611	0,1646	0,1705	0,1631	0,1666
Exp 7	0,2854	0,2784	0,2818	0,0695	0,1526	0,0954	0,0302	0,1620	0,1658	0,1722	0,1639	0,1678
Exp 8	0,2833	0,2770	0,2800	0,0688	0,1513	0,0945	0,0301	0,1614	0,1649	0,1710	0,1633	0,1669
Exp 9	0,2854	0,2787	0,2819	0,0690	0,1518	0,0948	0,0302	0,1620	0,1657	0,1719	0,1639	0,1676
Exp 10	0,2846	0,2779	0,2811	0,0690	0,1516	0,0947	0,0296	0,1608	0,1645	0,1706	0,1627	0,1664
Media	0,2845	0,2778	0,2811	0,0690	0,1518	0,0948	0,0299	0,1610	0,1647	0,1708	0,1629	0,1666

Tabla 15. DUC LDA 5.

DUC LDA 40												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,4195	0,4099	0,4145	0,0971	0,0948	0,0959	0,0301	0,0294	0,0297	0,0135	0,0132	0,0133
Exp 2	0,4214	0,4112	0,4161	0,0982	0,0957	0,0969	0,0304	0,0296	0,0300	0,0134	0,0130	0,0132
Exp 3	0,4203	0,4105	0,4152	0,0988	0,0964	0,0976	0,0311	0,0303	0,0307	0,0137	0,0134	0,0135
Exp 4	0,4232	0,4132	0,4180	0,0990	0,0966	0,0977	0,0311	0,0303	0,0307	0,0140	0,0136	0,0138
Exp 5	0,4213	0,4120	0,4165	0,0982	0,0959	0,0970	0,0304	0,0297	0,0300	0,0133	0,0130	0,0132
Exp 6	0,4202	0,4107	0,4153	0,0976	0,0953	0,0964	0,0307	0,0300	0,0303	0,0138	0,0135	0,0136
Exp 7	0,4202	0,4107	0,4153	0,0976	0,0953	0,0964	0,0307	0,0300	0,0303	0,0138	0,0135	0,0136
Exp 8	0,4215	0,4112	0,4161	0,0981	0,0956	0,0969	0,0304	0,0296	0,0300	0,0132	0,0128	0,0130
Exp 9	0,4182	0,4088	0,4133	0,0979	0,0956	0,0967	0,0313	0,0305	0,0309	0,0139	0,0136	0,0138
Exp 10	0,4209	0,4117	0,4161	0,0981	0,0959	0,0970	0,0301	0,0293	0,0297	0,0132	0,0129	0,0131
Media	0,4207	0,4110	0,4157	0,0981	0,0957	0,0969	0,0306	0,0299	0,0302	0,0136	0,0132	0,0134

DUC LDA 40												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2839	0,2774	0,2805	0,0688	0,1515	0,0945	0,0294	0,1601	0,1636	0,1696	0,1620	0,1655
Exp 2	0,2835	0,2768	0,2801	0,0688	0,1514	0,0945	0,0296	0,1610	0,1647	0,1710	0,1629	0,1667
Exp 3	0,2838	0,2772	0,2804	0,0690	0,1518	0,0947	0,0303	0,1603	0,1639	0,1700	0,1622	0,1659
Exp 4	0,2850	0,2784	0,2816	0,0692	0,1524	0,0951	0,0303	0,1625	0,1662	0,1723	0,1644	0,1681
Exp 5	0,2845	0,2782	0,2813	0,0690	0,1519	0,0948	0,0297	0,1615	0,1649	0,1709	0,1634	0,1669
Exp 6	0,2843	0,2779	0,2810	0,0690	0,1520	0,0949	0,0300	0,1601	0,1636	0,1696	0,1621	0,1656
Exp 7	0,2843	0,2779	0,2810	0,0690	0,1520	0,0949	0,0300	0,1601	0,1636	0,1696	0,1621	0,1656
Exp 8	0,2838	0,2769	0,2803	0,0687	0,1510	0,0943	0,0296	0,1604	0,1643	0,1706	0,1624	0,1662
Exp 9	0,2826	0,2762	0,2793	0,0685	0,1508	0,0941	0,0305	0,1597	0,1632	0,1691	0,1616	0,1651
Exp 10	0,2832	0,2770	0,2800	0,0688	0,1516	0,0945	0,0293	0,1613	0,1647	0,1705	0,1632	0,1666
Media	0,2839	0,2774	0,2805	0,0689	0,1516	0,0946	0,0299	0,1607	0,1643	0,1703	0,1626	0,1662

Tabla 16. DUC LDA 40.

DUC LDA 80												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,4218	0,4114	0,4164	0,1002	0,0977	0,0989	0,0317	0,0308	0,0313	0,0142	0,0138	0,0140
Exp 2	0,4215	0,4111	0,4161	0,0991	0,0967	0,0979	0,0312	0,0303	0,0307	0,0136	0,0133	0,0134
Exp 3	0,4201	0,4097	0,4147	0,0992	0,0967	0,0979	0,0308	0,0300	0,0304	0,0138	0,0134	0,0136
Exp 4	0,4182	0,4083	0,4131	0,0974	0,0950	0,0961	0,0305	0,0297	0,0301	0,0136	0,0132	0,0134
Exp 5	0,4226	0,4124	0,4174	0,0992	0,0967	0,0979	0,0309	0,0301	0,0305	0,0135	0,0132	0,0133
Exp 6	0,4233	0,4133	0,4182	0,0996	0,0971	0,0983	0,0312	0,0304	0,0308	0,0138	0,0134	0,0136
Exp 7	0,4205	0,4105	0,4154	0,0983	0,0959	0,0970	0,0317	0,0310	0,0313	0,0145	0,0141	0,0143
Exp 8	0,4235	0,4134	0,4183	0,0983	0,0959	0,0971	0,0303	0,0295	0,0299	0,0132	0,0128	0,0130
Exp 9	0,4198	0,4093	0,4144	0,0977	0,0951	0,0964	0,0303	0,0295	0,0299	0,0132	0,0129	0,0131
Exp 10	0,4202	0,4101	0,4150	0,0981	0,0956	0,0968	0,0306	0,0298	0,0302	0,0137	0,0133	0,0135
Media	0,4212	0,4109	0,4159	0,0987	0,0962	0,0974	0,0309	0,0301	0,0305	0,0137	0,0133	0,0135

DUC LDA 80												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2838	0,2768	0,2802	0,0690	0,1516	0,0947	0,0308	0,1611	0,1649	0,1713	0,1630	0,1669
Exp 2	0,2829	0,2759	0,2793	0,0689	0,1513	0,0946	0,0303	0,1609	0,1648	0,1711	0,1629	0,1667
Exp 3	0,2839	0,2770	0,2804	0,0689	0,1514	0,0946	0,0300	0,1605	0,1643	0,1706	0,1624	0,1662
Exp 4	0,2834	0,2767	0,2799	0,0689	0,1515	0,0946	0,0297	0,1590	0,1627	0,1688	0,1609	0,1646
Exp 5	0,2855	0,2786	0,2819	0,0692	0,1521	0,0950	0,0301	0,1614	0,1651	0,1714	0,1633	0,1671
Exp 6	0,2861	0,2794	0,2827	0,0693	0,1524	0,0952	0,0304	0,1620	0,1657	0,1719	0,1639	0,1676
Exp 7	0,2846	0,2779	0,2811	0,0692	0,1521	0,0950	0,0310	0,1598	0,1634	0,1696	0,1617	0,1654
Exp 8	0,2841	0,2773	0,2806	0,0688	0,1512	0,0944	0,0295	0,1614	0,1651	0,1714	0,1634	0,1671
Exp 9	0,2843	0,2773	0,2807	0,0689	0,1513	0,0946	0,0295	0,1598	0,1637	0,1700	0,1617	0,1656
Exp 10	0,2850	0,2782	0,2815	0,0690	0,1517	0,0947	0,0298	0,1604	0,1642	0,1704	0,1624	0,1661
Media	0,2843	0,2775	0,2808	0,0690	0,1516	0,0947	0,0301	0,1606	0,1644	0,1706	0,1626	0,1663

Tabla 17. DUC LDA 80.

RTVE LSA												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2043	0,6472	0,3092	0,0949	0,2991	0,1433	0,0698	0,2185	0,1052	0,0618	0,1935	0,0931
Exp 2	0,2115	0,6733	0,3204	0,0962	0,3120	0,1464	0,0690	0,2254	0,1051	0,0610	0,2004	0,0931
Exp 3	0,2101	0,6680	0,3182	0,1008	0,3261	0,1533	0,0768	0,2497	0,1169	0,0689	0,2253	0,1050
Exp 4	0,2054	0,6461	0,3102	0,0894	0,2778	0,1347	0,0589	0,1802	0,0884	0,0505	0,1542	0,0758
Exp 5	0,1992	0,6323	0,3015	0,0821	0,2649	0,1246	0,0563	0,1835	0,0857	0,0501	0,1633	0,0762
Exp 6	0,1966	0,6281	0,2980	0,0796	0,2653	0,1220	0,0550	0,1877	0,0848	0,0482	0,1658	0,0744
Exp 7	0,2098	0,6645	0,3174	0,1012	0,3230	0,1534	0,0721	0,2306	0,1093	0,0634	0,2041	0,0963
Exp 8	0,2084	0,6674	0,3160	0,0913	0,3027	0,1396	0,0650	0,2185	0,0997	0,0563	0,1915	0,0866
Exp 9	0,1990	0,6367	0,3018	0,0785	0,2629	0,1204	0,0515	0,1776	0,0795	0,0450	0,1565	0,0697
Exp 10	0,2081	0,6620	0,3151	0,0916	0,2998	0,1397	0,0624	0,2074	0,0955	0,0543	0,1824	0,0833
Media	0,2052	0,6526	0,3108	0,0906	0,2934	0,1377	0,0637	0,2079	0,0970	0,0559	0,1837	0,0853

RTVE LSA												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1884	0,5953	0,2848	0,0484	0,2958	0,0829	0,2185	0,3881	0,0707	0,0397	0,3902	0,0714
Exp 2	0,1901	0,6066	0,2880	0,0490	0,3034	0,0841	0,2254	0,4217	0,0745	0,0417	0,4237	0,0753
Exp 3	0,1922	0,6121	0,2911	0,0514	0,3143	0,0880	0,2497	0,4217	0,0760	0,0427	0,4236	0,0768
Exp 4	0,1846	0,5807	0,2788	0,0455	0,2752	0,0778	0,1802	0,3911	0,0722	0,0406	0,3931	0,0729
Exp 5	0,1788	0,5689	0,2707	0,0476	0,2930	0,0816	0,1835	0,3756	0,0671	0,0377	0,3777	0,0679
Exp 6	0,1768	0,5668	0,2683	0,0461	0,2868	0,0791	0,1877	0,3781	0,0661	0,0370	0,3801	0,0668
Exp 7	0,1891	0,5995	0,2861	0,0505	0,3098	0,0866	0,2306	0,4100	0,0746	0,0419	0,4120	0,0753
Exp 8	0,1867	0,5995	0,2834	0,0478	0,2981	0,0821	0,2185	0,4166	0,0727	0,0406	0,4186	0,0734
Exp 9	0,1786	0,5733	0,2711	0,0445	0,2753	0,0764	0,1776	0,3865	0,0677	0,0379	0,3885	0,0684
Exp 10	0,1899	0,6058	0,2878	0,0499	0,3077	0,0857	0,2074	0,4161	0,0737	0,0413	0,4181	0,0745
Media	0,1855	0,5908	0,2810	0,0481	0,2959	0,0824	0,2079	0,4006	0,0715	0,0401	0,4026	0,0723

Tabla 18. RTVE LSA.

RTVE LDA 5												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1985	0,6290	0,3003	0,0745	0,2348	0,1124	0,0432	0,1358	0,0651	0,0353	0,1108	0,0531
Exp 2	0,1970	0,6243	0,2981	0,0736	0,2323	0,1112	0,0431	0,1353	0,0649	0,0351	0,1104	0,0529
Exp 3	0,1994	0,6311	0,3016	0,0785	0,2467	0,1184	0,0470	0,1471	0,0708	0,0386	0,1206	0,0581
Exp 4	0,1975	0,6254	0,2988	0,0737	0,2325	0,1113	0,0431	0,1352	0,0649	0,0351	0,1103	0,0529
Exp 5	0,2012	0,6375	0,3045	0,0765	0,2404	0,1154	0,0455	0,1419	0,0685	0,0380	0,1186	0,0572
Exp 6	0,2011	0,6356	0,3040	0,0783	0,2453	0,1180	0,0480	0,1493	0,0722	0,0404	0,1253	0,0606
Exp 7	0,2003	0,6339	0,3030	0,0748	0,2356	0,1129	0,0430	0,1348	0,0647	0,0354	0,1110	0,0533
Exp 8	0,1986	0,6294	0,3005	0,0764	0,2416	0,1154	0,0450	0,1422	0,0680	0,0369	0,1164	0,0556
Exp 9	0,1992	0,6307	0,3013	0,0768	0,2424	0,1160	0,0460	0,1443	0,0693	0,0381	0,1193	0,0573
Exp 10	0,1971	0,6247	0,2982	0,0743	0,2347	0,1122	0,0433	0,1361	0,0652	0,0353	0,1108	0,0531
Media	0,1990	0,6302	0,3010	0,0757	0,2386	0,1143	0,0447	0,1402	0,0674	0,0368	0,1153	0,0554

RTVE LDA 5												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1789	0,5663	0,2706	0,0455	0,2762	0,0778	0,1358	0,3799	0,0683	0,0383	0,3819	0,0690
Exp 2	0,1773	0,5613	0,2681	0,0453	0,2753	0,0776	0,1353	0,3767	0,0677	0,0380	0,3786	0,0684
Exp 3	0,1803	0,5704	0,2727	0,0466	0,2826	0,0797	0,1471	0,3848	0,0694	0,0389	0,3868	0,0701
Exp 4	0,1783	0,5640	0,2696	0,0456	0,2767	0,0780	0,1352	0,3773	0,0679	0,0381	0,3793	0,0686
Exp 5	0,1807	0,5721	0,2733	0,0466	0,2827	0,0797	0,1419	0,3843	0,0692	0,0388	0,3864	0,0700
Exp 6	0,1808	0,5710	0,2734	0,0470	0,2842	0,0803	0,1493	0,3853	0,0697	0,0391	0,3873	0,0704
Exp 7	0,1799	0,5690	0,2720	0,0460	0,2793	0,0788	0,1348	0,3808	0,0687	0,0385	0,3828	0,0694
Exp 8	0,1794	0,5682	0,2714	0,0460	0,2793	0,0787	0,1422	0,3800	0,0683	0,0383	0,3820	0,0690
Exp 9	0,1778	0,5629	0,2689	0,0464	0,2813	0,0794	0,1443	0,3840	0,0691	0,0388	0,3860	0,0699
Exp 10	0,1778	0,5632	0,2690	0,0453	0,2750	0,0775	0,1361	0,3766	0,0676	0,0379	0,3786	0,0683
Media	0,1791	0,5668	0,2709	0,0460	0,2793	0,0787	0,1402	0,3810	0,0686	0,0385	0,3830	0,0693

Tabla 19. RTVE LDA 5.

RTVE LDA 10												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2002	0,6348	0,3030	0,0804	0,2537	0,1215	0,0505	0,1588	0,0762	0,0425	0,1336	0,0641
Exp 2	0,1979	0,6268	0,2994	0,0739	0,2330	0,1115	0,0431	0,1353	0,0649	0,0351	0,1103	0,0529
Exp 3	0,1973	0,6255	0,2986	0,0736	0,2323	0,1112	0,0431	0,1353	0,0649	0,0351	0,1104	0,0529
Exp 4	0,2007	0,6355	0,3036	0,0791	0,2495	0,1194	0,0488	0,1535	0,0736	0,0408	0,1282	0,0615
Exp 5	0,2000	0,6337	0,3026	0,0766	0,2419	0,1157	0,0439	0,1382	0,0662	0,0354	0,1112	0,0533
Exp 6	0,1973	0,6250	0,2985	0,0732	0,2309	0,1105	0,0431	0,1352	0,0649	0,0351	0,1103	0,0529
Exp 7	0,1993	0,6315	0,3016	0,0766	0,2423	0,1158	0,0453	0,1430	0,0683	0,0372	0,1176	0,0562
Exp 8	0,2003	0,6341	0,3030	0,0749	0,2355	0,1130	0,0439	0,1370	0,0660	0,0367	0,1144	0,0552
Exp 9	0,1977	0,6260	0,2990	0,0744	0,2350	0,1124	0,0433	0,1361	0,0652	0,0353	0,1107	0,0531
Exp 10	0,1983	0,6289	0,3001	0,0759	0,2403	0,1147	0,0456	0,1442	0,0688	0,0374	0,1185	0,0565
Media	0,1989	0,6302	0,3009	0,0759	0,2394	0,1146	0,0450	0,1416	0,0679	0,0371	0,1165	0,0558

RTVE LDA 10												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1806	0,5721	0,2732	0,0466	0,2829	0,0797	0,1588	0,3910	0,0703	0,0394	0,3930	0,0710
Exp 2	0,1781	0,5638	0,2694	0,0453	0,2748	0,0775	0,1353	0,3768	0,0678	0,0380	0,3788	0,0685
Exp 3	0,1774	0,5621	0,2684	0,0453	0,2751	0,0775	0,1353	0,3786	0,0679	0,0381	0,3805	0,0686
Exp 4	0,1810	0,5732	0,2738	0,0469	0,2849	0,0803	0,1535	0,3900	0,0701	0,0393	0,3919	0,0708
Exp 5	0,1795	0,5682	0,2715	0,0461	0,2801	0,0789	0,1382	0,3865	0,0695	0,0390	0,3885	0,0702
Exp 6	0,1773	0,5612	0,2681	0,0453	0,2750	0,0775	0,1352	0,3770	0,0678	0,0380	0,3790	0,0685
Exp 7	0,1795	0,5685	0,2716	0,0459	0,2788	0,0786	0,1430	0,3844	0,0691	0,0387	0,3864	0,0698
Exp 8	0,1788	0,5658	0,2704	0,0464	0,2815	0,0794	0,1370	0,3792	0,0684	0,0384	0,3812	0,0692
Exp 9	0,1779	0,5630	0,2690	0,0454	0,2758	0,0777	0,1361	0,3783	0,0680	0,0382	0,3803	0,0687
Exp 10	0,1798	0,5698	0,2720	0,0460	0,2796	0,0787	0,1442	0,3828	0,0686	0,0385	0,3848	0,0693
Media	0,1790	0,5668	0,2707	0,0459	0,2788	0,0786	0,1416	0,3825	0,0687	0,0386	0,3844	0,0695

Tabla 20. RTVE LDA 10.

RTVE LDA 20												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1978	0,6266	0,2993	0,0752	0,2380	0,1137	0,0454	0,1432	0,0684	0,0373	0,1179	0,0563
Exp 2	0,2007	0,6350	0,3035	0,0768	0,2420	0,1159	0,0448	0,1412	0,0676	0,0369	0,1163	0,0556
Exp 3	0,1980	0,6271	0,2995	0,0751	0,2384	0,1135	0,0444	0,1412	0,0670	0,0363	0,1159	0,0549
Exp 4	0,2000	0,6332	0,3025	0,0766	0,2421	0,1157	0,0455	0,1436	0,0686	0,0375	0,1183	0,0565
Exp 5	0,1997	0,6334	0,3022	0,0792	0,2524	0,1198	0,0481	0,1540	0,0728	0,0392	0,1260	0,0594
Exp 6	0,1976	0,6258	0,2989	0,0743	0,2353	0,1123	0,0438	0,1384	0,0660	0,0360	0,1138	0,0543
Exp 7	0,1997	0,6318	0,3020	0,0765	0,2406	0,1154	0,0465	0,1453	0,0700	0,0386	0,1204	0,0581
Exp 8	0,1972	0,6244	0,2983	0,0737	0,2328	0,1113	0,0433	0,1360	0,0652	0,0353	0,1106	0,0531
Exp 9	0,1989	0,6300	0,3009	0,0773	0,2445	0,1168	0,0459	0,1448	0,0692	0,0373	0,1179	0,0563
Exp 10	0,1970	0,6246	0,2981	0,0739	0,2332	0,1115	0,0433	0,1362	0,0653	0,0351	0,1104	0,0529
Media	0,1987	0,6292	0,3005	0,0758	0,2399	0,1146	0,0451	0,1424	0,0680	0,0369	0,1168	0,0557

RTVE LDA 20												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1786	0,5656	0,2701	0,0459	0,2788	0,0785	0,1432	0,3814	0,0685	0,0384	0,3834	0,0692
Exp 2	0,1807	0,5715	0,2733	0,0461	0,2797	0,0789	0,1412	0,3849	0,0694	0,0389	0,3869	0,0701
Exp 3	0,1773	0,5609	0,2681	0,0452	0,2743	0,0773	0,1412	0,3789	0,0680	0,0382	0,3808	0,0687
Exp 4	0,1798	0,5694	0,2720	0,0462	0,2805	0,0790	0,1436	0,3874	0,0696	0,0390	0,3893	0,0703
Exp 5	0,1807	0,5727	0,2733	0,0466	0,2837	0,0798	0,1540	0,3878	0,0694	0,0389	0,3897	0,0701
Exp 6	0,1774	0,5620	0,2683	0,0459	0,2789	0,0786	0,1384	0,3793	0,0680	0,0382	0,3813	0,0687
Exp 7	0,1790	0,5660	0,2706	0,0461	0,2795	0,0789	0,1453	0,3843	0,0694	0,0389	0,3863	0,0701
Exp 8	0,1765	0,5586	0,2670	0,0451	0,2736	0,0772	0,1360	0,3773	0,0678	0,0380	0,3792	0,0685
Exp 9	0,1800	0,5701	0,2723	0,0462	0,2809	0,0791	0,1448	0,3853	0,0692	0,0388	0,3873	0,0699
Exp 10	0,1778	0,5634	0,2690	0,0454	0,2757	0,0777	0,1362	0,3765	0,0676	0,0379	0,3785	0,0683
Media	0,1788	0,5660	0,2704	0,0459	0,2786	0,0785	0,1424	0,3823	0,0687	0,0385	0,3843	0,0694

Tabla 21. RTVE LDA 20.

RTVE LDA 40												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,2000	0,6328	0,3025	0,0784	0,2485	0,1186	0,0482	0,1525	0,0728	0,0403	0,1277	0,0608
Exp 2	0,1982	0,6283	0,3000	0,0764	0,2414	0,1154	0,0445	0,1401	0,0671	0,0358	0,1128	0,0540
Exp 3	0,1980	0,6267	0,2995	0,0738	0,2325	0,1114	0,0432	0,1357	0,0651	0,0353	0,1107	0,0531
Exp 4	0,2013	0,6370	0,3045	0,0774	0,2437	0,1168	0,0455	0,1425	0,0685	0,0376	0,1176	0,0566
Exp 5	0,1972	0,6248	0,2983	0,0750	0,2373	0,1133	0,0456	0,1440	0,0688	0,0374	0,1183	0,0565
Exp 6	0,1974	0,6247	0,2986	0,0728	0,2301	0,1100	0,0421	0,1324	0,0634	0,0341	0,1070	0,0513
Exp 7	0,1975	0,6254	0,2987	0,0741	0,2344	0,1119	0,0433	0,1368	0,0653	0,0354	0,1118	0,0534
Exp 8	0,2003	0,6335	0,3030	0,0794	0,2516	0,1200	0,0478	0,1516	0,0722	0,0400	0,1272	0,0604
Exp 9	0,1995	0,6317	0,3018	0,0752	0,2374	0,1136	0,0434	0,1365	0,0654	0,0354	0,1111	0,0533
Exp 10	0,1980	0,6262	0,2994	0,0747	0,2357	0,1128	0,0434	0,1364	0,0654	0,0354	0,1111	0,0533
Media	0,1987	0,6291	0,3006	0,0757	0,2393	0,1144	0,0447	0,1409	0,0674	0,0367	0,1155	0,0553

RTVE LDA 40												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1803	0,5702	0,2726	0,0466	0,2828	0,0797	0,1525	0,3876	0,0697	0,0391	0,3896	0,0704
Exp 2	0,1791	0,5672	0,2709	0,0460	0,2794	0,0787	0,1401	0,3831	0,0688	0,0386	0,3851	0,0695
Exp 3	0,1780	0,5629	0,2691	0,0455	0,2761	0,0779	0,1357	0,3800	0,0685	0,0384	0,3820	0,0692
Exp 4	0,1792	0,5668	0,2710	0,0465	0,2818	0,0795	0,1425	0,3871	0,0698	0,0391	0,3891	0,0705
Exp 5	0,1772	0,5614	0,2681	0,0454	0,2759	0,0777	0,1440	0,3791	0,0680	0,0381	0,3811	0,0687
Exp 6	0,1763	0,5580	0,2666	0,0454	0,2752	0,0776	0,1324	0,3776	0,0680	0,0381	0,3796	0,0687
Exp 7	0,1774	0,5620	0,2684	0,0458	0,2780	0,0783	0,1368	0,3760	0,0675	0,0379	0,3780	0,0682
Exp 8	0,1788	0,5653	0,2704	0,0469	0,2848	0,0803	0,1516	0,3850	0,0694	0,0389	0,3870	0,0701
Exp 9	0,1794	0,5678	0,2713	0,0458	0,2780	0,0784	0,1365	0,3840	0,0691	0,0388	0,3860	0,0698
Exp 10	0,1782	0,5636	0,2695	0,0454	0,2756	0,0777	0,1364	0,3796	0,0684	0,0384	0,3816	0,0691
Media	0,1784	0,5645	0,2698	0,0459	0,2788	0,0786	0,1409	0,3819	0,0687	0,0385	0,3839	0,0694

Tabla 22. RTVE LDA 40.

RTVE LDA 80												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1986	0,6286	0,3004	0,0749	0,2361	0,1131	0,0433	0,1360	0,0653	0,0351	0,1103	0,0529
Exp 2	0,1970	0,6236	0,2980	0,0735	0,2321	0,1110	0,0422	0,1328	0,0636	0,0341	0,1071	0,0513
Exp 3	0,1976	0,6260	0,2990	0,0745	0,2353	0,1125	0,0434	0,1368	0,0654	0,0354	0,1119	0,0534
Exp 4	0,1999	0,6326	0,3023	0,0745	0,2344	0,1124	0,0428	0,1340	0,0644	0,0348	0,1091	0,0524
Exp 5	0,1977	0,6262	0,2991	0,0755	0,2388	0,1140	0,0454	0,1432	0,0684	0,0372	0,1175	0,0561
Exp 6	0,1986	0,6286	0,3003	0,0742	0,2332	0,1120	0,0432	0,1348	0,0650	0,0349	0,1082	0,0524
Exp 7	0,1965	0,6224	0,2972	0,0727	0,2297	0,1098	0,0424	0,1336	0,0639	0,0348	0,1099	0,0525
Exp 8	0,1946	0,6161	0,2943	0,0714	0,2252	0,1077	0,0418	0,1311	0,0629	0,0341	0,1070	0,0513
Exp 9	0,1986	0,6292	0,3005	0,0748	0,2361	0,1129	0,0441	0,1388	0,0665	0,0366	0,1155	0,0552
Exp 10	0,1969	0,6234	0,2978	0,0728	0,2301	0,1100	0,0417	0,1312	0,0628	0,0339	0,1066	0,0511
Media	0,1976	0,6257	0,2989	0,0739	0,2331	0,1115	0,0430	0,1352	0,0648	0,0351	0,1103	0,0529

RTVE LDA 80												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,1788	0,5656	0,2703	0,0455	0,2759	0,0778	0,1360	0,3802	0,0684	0,0384	0,3822	0,0691
Exp 2	0,1763	0,5582	0,2666	0,0450	0,2732	0,0770	0,1328	0,3768	0,0678	0,0380	0,3788	0,0685
Exp 3	0,1774	0,5617	0,2683	0,0460	0,2791	0,0786	0,1368	0,3798	0,0682	0,0383	0,3818	0,0689
Exp 4	0,1784	0,5640	0,2697	0,0453	0,2747	0,0775	0,1340	0,3812	0,0687	0,0386	0,3832	0,0694
Exp 5	0,1781	0,5640	0,2694	0,0455	0,2765	0,0779	0,1432	0,3809	0,0684	0,0383	0,3829	0,0691
Exp 6	0,1791	0,5668	0,2708	0,0456	0,2764	0,0781	0,1348	0,3820	0,0687	0,0385	0,3840	0,0694
Exp 7	0,1753	0,5550	0,2651	0,0452	0,2746	0,0774	0,1336	0,3714	0,0668	0,0375	0,3734	0,0675
Exp 8	0,1741	0,5511	0,2633	0,0451	0,2733	0,0771	0,1311	0,3708	0,0667	0,0374	0,3728	0,0674
Exp 9	0,1791	0,5673	0,2709	0,0458	0,2784	0,0784	0,1388	0,3777	0,0679	0,0381	0,3797	0,0686
Exp 10	0,1763	0,5580	0,2666	0,0454	0,2751	0,0776	0,1312	0,3735	0,0671	0,0377	0,3755	0,0678
Media	0,1773	0,5612	0,2681	0,0454	0,2757	0,0778	0,1352	0,3774	0,0679	0,0381	0,3794	0,0686

Tabla 23. RTVE LDA 80.

RTVE LSA Segmentado												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,5747	0,5618	0,5661	0,2760	0,2684	0,2712	0,2064	0,1998	0,2024	0,1871	0,1805	0,1832
Exp 2	0,5674	0,5633	0,5632	0,2634	0,2617	0,2616	0,1948	0,1935	0,1935	0,1756	0,1739	0,1742
Exp 3	0,5601	0,5588	0,5574	0,2666	0,2624	0,2636	0,2000	0,1953	0,1970	0,1798	0,1746	0,1767
Exp 4	0,5642	0,5526	0,5575	0,2656	0,2581	0,2614	0,1971	0,1905	0,1934	0,1774	0,1709	0,1738
Exp 5	0,5751	0,5698	0,5701	0,2818	0,2759	0,2777	0,2108	0,2049	0,2071	0,1897	0,1832	0,1858
Exp 6	0,5787	0,5650	0,5697	0,2789	0,2714	0,2741	0,2078	0,2019	0,2041	0,1878	0,1818	0,1842
Exp 7	0,5563	0,5662	0,5594	0,2614	0,2638	0,2618	0,1959	0,1964	0,1956	0,1772	0,1770	0,1767
Exp 8	0,5669	0,5605	0,5612	0,2786	0,2709	0,2736	0,2089	0,2014	0,2044	0,1887	0,1807	0,1840
Exp 9	0,5902	0,5665	0,5756	0,2880	0,2744	0,2799	0,2176	0,2061	0,2109	0,1975	0,1860	0,1909
Exp 10	0,5636	0,5586	0,5594	0,2609	0,2570	0,2581	0,1917	0,1882	0,1894	0,1724	0,1687	0,1700
Media	0,5697	0,5623	0,5639	0,2721	0,2664	0,2683	0,2031	0,1978	0,1998	0,1833	0,1777	0,1799

RTVE LSA Segementado												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,5286	0,5163	0,5205	0,1310	0,2469	0,1707	0,1998	0,2848	0,2882	0,3004	0,2855	0,2888
Exp 2	0,5232	0,5191	0,5193	0,1285	0,2464	0,1684	0,1935	0,2884	0,2873	0,2948	0,2891	0,2880
Exp 3	0,5148	0,5126	0,5119	0,1253	0,2405	0,1642	0,1953	0,2910	0,2904	0,2980	0,2917	0,2911
Exp 4	0,5224	0,5114	0,5160	0,1277	0,2410	0,1667	0,1905	0,2856	0,2909	0,3007	0,2863	0,2916
Exp 5	0,5327	0,5270	0,5277	0,1311	0,2497	0,1713	0,2049	0,2934	0,2932	0,3025	0,2941	0,2939
Exp 6	0,5349	0,5218	0,5264	0,1325	0,2494	0,1726	0,2019	0,2879	0,2914	0,3040	0,2886	0,2921
Exp 7	0,5141	0,5224	0,5166	0,1265	0,2479	0,1670	0,1964	0,2913	0,2846	0,2853	0,2920	0,2853
Exp 8	0,5211	0,5149	0,5157	0,1284	0,2443	0,1677	0,2014	0,2889	0,2899	0,3011	0,2896	0,2906
Exp 9	0,5455	0,5230	0,5318	0,1356	0,2505	0,1753	0,2061	0,2922	0,3011	0,3217	0,2929	0,3017
Exp 10	0,5175	0,5136	0,5140	0,1252	0,2396	0,1640	0,1882	0,2770	0,2780	0,2863	0,2778	0,2788
Media	0,5255	0,5182	0,5200	0,1292	0,2456	0,1688	0,1978	0,2881	0,2895	0,2995	0,2887	0,2902

Tabla 24. RTVE LSA Segmentado.

RTVE LDA 5 Segmentado												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,7431	0,4202	0,5353	0,4605	0,2589	0,3305	0,3798	0,2133	0,2724	0,3510	0,1970	0,2517
Exp 2	0,7403	0,4191	0,5337	0,4526	0,2548	0,3251	0,3728	0,2096	0,2676	0,3442	0,1934	0,2469
Exp 3	0,7412	0,4149	0,5307	0,4459	0,2483	0,3182	0,3621	0,2016	0,2584	0,3326	0,1852	0,2373
Exp 4	0,7438	0,4207	0,5360	0,4579	0,2573	0,3286	0,3770	0,2115	0,2702	0,3485	0,1953	0,2497
Exp 5	0,7399	0,4154	0,5307	0,4509	0,2515	0,3221	0,3688	0,2055	0,2633	0,3409	0,1899	0,2433
Exp 6	0,7444	0,4204	0,5359	0,4595	0,2576	0,3293	0,3796	0,2124	0,2717	0,3505	0,1960	0,2508
Exp 7	0,7439	0,4205	0,5358	0,4550	0,2556	0,3264	0,3738	0,2097	0,2679	0,3451	0,1934	0,2472
Exp 8	0,7425	0,4193	0,5344	0,4558	0,2556	0,3266	0,3759	0,2105	0,2691	0,3471	0,1942	0,2484
Exp 9	0,7412	0,4190	0,5338	0,4485	0,2520	0,3218	0,3664	0,2058	0,2628	0,3371	0,1892	0,2417
Exp 10	0,7403	0,4163	0,5314	0,4518	0,2518	0,3225	0,3714	0,2066	0,2648	0,3435	0,1909	0,2447
Media	0,7421	0,4186	0,5338	0,4538	0,2543	0,3251	0,3728	0,2086	0,2668	0,3440	0,1924	0,2462

RTVE LDA 5 Segmentado												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,7064	0,3994	0,5088	0,1826	0,1995	0,1901	0,2133	0,1622	0,2433	0,5054	0,1626	0,2438
Exp 2	0,7027	0,3977	0,5065	0,1817	0,1985	0,1892	0,2096	0,1608	0,2412	0,4998	0,1612	0,2417
Exp 3	0,7014	0,3926	0,5022	0,1814	0,1960	0,1879	0,2016	0,1570	0,2370	0,4994	0,1574	0,2375
Exp 4	0,7045	0,3983	0,5075	0,1823	0,1989	0,1896	0,2115	0,1631	0,2447	0,5078	0,1635	0,2452
Exp 5	0,7030	0,3946	0,5042	0,1817	0,1967	0,1884	0,2055	0,1587	0,2391	0,5016	0,1591	0,2396
Exp 6	0,7079	0,3996	0,5095	0,1842	0,2009	0,1917	0,2124	0,1623	0,2439	0,5074	0,1627	0,2444
Exp 7	0,7060	0,3990	0,5084	0,1819	0,1984	0,1892	0,2097	0,1619	0,2429	0,5049	0,1622	0,2434
Exp 8	0,7022	0,3964	0,5052	0,1794	0,1955	0,1865	0,2105	0,1607	0,2412	0,5022	0,1611	0,2417
Exp 9	0,7021	0,3969	0,5056	0,1815	0,1980	0,1888	0,2058	0,1623	0,2434	0,5056	0,1627	0,2439
Exp 10	0,7017	0,3943	0,5035	0,1821	0,1973	0,1889	0,2066	0,1598	0,2405	0,5039	0,1602	0,2410
Media	0,7038	0,3969	0,5061	0,1819	0,1980	0,1890	0,2086	0,1609	0,2417	0,5038	0,1613	0,2422

Tabla 25. RTVE LDA 5 Segmentado.

RTVE LDA 10 Segmentado												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,7342	0,4146	0,5286	0,4420	0,2483	0,3171	0,3598	0,2020	0,2580	0,3315	0,1860	0,2377
Exp 2	0,7436	0,4183	0,5340	0,4546	0,2546	0,3255	0,3735	0,2091	0,2674	0,3447	0,1928	0,2467
Exp 3	0,7459	0,4189	0,5349	0,4604	0,2569	0,3288	0,3766	0,2098	0,2687	0,3473	0,1933	0,2477
Exp 4	0,7334	0,4153	0,5288	0,4407	0,2477	0,3163	0,3588	0,2014	0,2573	0,3309	0,1856	0,2372
Exp 5	0,7443	0,4174	0,5334	0,4557	0,2538	0,3251	0,3735	0,2078	0,2663	0,3446	0,1915	0,2456
Exp 6	0,7424	0,4183	0,5337	0,4557	0,2552	0,3263	0,3743	0,2093	0,2678	0,3456	0,1932	0,2472
Exp 7	0,7396	0,4166	0,5315	0,4475	0,2508	0,3206	0,3647	0,2043	0,2612	0,3360	0,1881	0,2406
Exp 8	0,7433	0,4176	0,5333	0,4538	0,2536	0,3245	0,3724	0,2080	0,2662	0,3431	0,1915	0,2452
Exp 9	0,7417	0,4139	0,5301	0,4499	0,2494	0,3202	0,3685	0,2040	0,2620	0,3413	0,1890	0,2427
Exp 10	0,7363	0,4120	0,5271	0,4372	0,2427	0,3114	0,3536	0,1958	0,2515	0,3253	0,1800	0,2313
Media	0,7405	0,4163	0,5315	0,4497	0,2513	0,3216	0,3676	0,2051	0,2626	0,3390	0,1891	0,2422

RTVE LDA 10 Segmentado												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,6963	0,3931	0,5012	0,1785	0,1944	0,1856	0,2020	0,1607	0,2413	0,5016	0,1611	0,2418
Exp 2	0,7053	0,3966	0,5063	0,1840	0,2000	0,1911	0,2091	0,1604	0,2413	0,5049	0,1608	0,2418
Exp 3	0,7073	0,3970	0,5070	0,1837	0,1988	0,1903	0,2098	0,1599	0,2404	0,5043	0,1602	0,2409
Exp 4	0,6948	0,3931	0,5007	0,1780	0,1946	0,1854	0,2014	0,1594	0,2390	0,4959	0,1597	0,2395
Exp 5	0,7062	0,3958	0,5059	0,1829	0,1977	0,1895	0,2078	0,1599	0,2408	0,5061	0,1602	0,2413
Exp 6	0,7035	0,3962	0,5055	0,1817	0,1974	0,1887	0,2093	0,1614	0,2426	0,5061	0,1617	0,2431
Exp 7	0,6966	0,3923	0,5005	0,1786	0,1941	0,1854	0,2043	0,1596	0,2400	0,5009	0,1600	0,2405
Exp 8	0,7043	0,3956	0,5052	0,1840	0,1996	0,1909	0,2080	0,1596	0,2402	0,5037	0,1599	0,2407
Exp 9	0,7046	0,3930	0,5034	0,1835	0,1975	0,1898	0,2040	0,1574	0,2380	0,5039	0,1578	0,2385
Exp 10	0,6945	0,3883	0,4969	0,1800	0,1941	0,1863	0,1958	0,1566	0,2365	0,4989	0,1569	0,2370
Media	0,7013	0,3941	0,5033	0,1815	0,1968	0,1883	0,2051	0,1595	0,2400	0,5026	0,1598	0,2405

Tabla 26. RTVE LDA 10 Segmentado.

RTVE LDA 20 Segmentado												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,7354	0,4131	0,5277	0,4396	0,2456	0,3144	0,3574	0,1995	0,2554	0,3291	0,1836	0,2351
Exp 2	0,7411	0,4163	0,5316	0,4480	0,2500	0,3200	0,3659	0,2039	0,2611	0,3372	0,1877	0,2405
Exp 3	0,7367	0,4130	0,5280	0,4359	0,2434	0,3116	0,3502	0,1955	0,2503	0,3214	0,1793	0,2296
Exp 4	0,7375	0,4138	0,5288	0,4344	0,2422	0,3102	0,3513	0,1956	0,2507	0,3234	0,1799	0,2307
Exp 5	0,7387	0,4137	0,5289	0,4482	0,2499	0,3200	0,3658	0,2039	0,2611	0,3372	0,1878	0,2406
Exp 6	0,7385	0,4125	0,5279	0,4460	0,2478	0,3178	0,3646	0,2024	0,2596	0,3361	0,1864	0,2392
Exp 7	0,7406	0,4160	0,5313	0,4513	0,2522	0,3227	0,3688	0,2059	0,2635	0,3398	0,1894	0,2426
Exp 8	0,7378	0,4165	0,5309	0,4428	0,2486	0,3175	0,3607	0,2023	0,2585	0,3326	0,1864	0,2382
Exp 9	0,7343	0,4109	0,5256	0,4339	0,2419	0,3098	0,3507	0,1955	0,2504	0,3214	0,1791	0,2294
Exp 10	0,7345	0,4108	0,5254	0,4354	0,2416	0,3099	0,3522	0,1951	0,2504	0,3233	0,1790	0,2298
Media	0,7375	0,4136	0,5286	0,4416	0,2463	0,3154	0,3587	0,1999	0,2561	0,3301	0,1839	0,2356

RTVE LDA 20 Segmentado												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,6968	0,3913	0,4999	0,1822	0,1977	0,1891	0,1995	0,1589	0,2393	0,5015	0,1592	0,2398
Exp 2	0,7029	0,3946	0,5040	0,1824	0,1975	0,1891	0,2039	0,1586	0,2387	0,5006	0,1590	0,2392
Exp 3	0,6947	0,3894	0,4978	0,1788	0,1935	0,1853	0,1955	0,1575	0,2375	0,4993	0,1579	0,2380
Exp 4	0,6952	0,3899	0,4984	0,1803	0,1950	0,1869	0,1956	0,1574	0,2374	0,4990	0,1578	0,2379
Exp 5	0,7004	0,3922	0,5014	0,1812	0,1957	0,1876	0,2039	0,1582	0,2384	0,5021	0,1585	0,2388
Exp 6	0,6991	0,3904	0,4997	0,1803	0,1941	0,1864	0,2024	0,1573	0,2374	0,5025	0,1576	0,2379
Exp 7	0,7006	0,3934	0,5025	0,1822	0,1976	0,1890	0,2059	0,1604	0,2414	0,5062	0,1608	0,2419
Exp 8	0,6986	0,3945	0,5027	0,1830	0,1995	0,1903	0,2023	0,1603	0,2405	0,5009	0,1607	0,2410
Exp 9	0,6947	0,3888	0,4972	0,1786	0,1928	0,1849	0,1955	0,1564	0,2359	0,4972	0,1568	0,2364
Exp 10	0,6945	0,3883	0,4967	0,1799	0,1939	0,1861	0,1951	0,1563	0,2358	0,4981	0,1567	0,2363
Media	0,6978	0,3913	0,5000	0,1809	0,1957	0,1875	0,1999	0,1581	0,2382	0,5007	0,1585	0,2387

Tabla 27. RTVE LDA 20 Segmentado.

RTVE LDA 40 Segmentado												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,7400	0,4127	0,5284	0,4450	0,2465	0,3163	0,3612	0,1998	0,2566	0,3323	0,1837	0,2360
Exp 2	0,7319	0,4082	0,5225	0,4255	0,2356	0,3024	0,3393	0,1878	0,2410	0,3110	0,1721	0,2209
Exp 3	0,7342	0,4074	0,5227	0,4339	0,2394	0,3077	0,3491	0,1925	0,2475	0,3208	0,1768	0,2274
Exp 4	0,7395	0,4136	0,5292	0,4419	0,2456	0,3149	0,3588	0,1991	0,2555	0,3305	0,1833	0,2353
Exp 5	0,7368	0,4110	0,5262	0,4403	0,2442	0,3133	0,3578	0,1982	0,2544	0,3290	0,1822	0,2339
Exp 6	0,7406	0,4106	0,5271	0,4438	0,2445	0,3145	0,3583	0,1972	0,2538	0,3299	0,1816	0,2337
Exp 7	0,7449	0,4165	0,5329	0,4547	0,2528	0,3241	0,3730	0,2072	0,2657	0,3441	0,1910	0,2450
Exp 8	0,7381	0,4100	0,5260	0,4388	0,2422	0,3115	0,3528	0,1945	0,2502	0,3246	0,1790	0,2302
Exp 9	0,7387	0,4122	0,5279	0,4413	0,2445	0,3140	0,3580	0,1980	0,2544	0,3298	0,1823	0,2343
Exp 10	0,7419	0,4134	0,5295	0,4441	0,2454	0,3153	0,3613	0,1993	0,2562	0,3335	0,1840	0,2365
Media	0,7387	0,4116	0,5272	0,4409	0,2441	0,3134	0,3569	0,1974	0,2535	0,3285	0,1816	0,2333

RTVE LDA 40 Segmentado												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,7011	0,3909	0,5005	0,1803	0,1939	0,1863	0,1998	0,1573	0,2374	0,5035	0,1576	0,2379
Exp 2	0,6896	0,3845	0,4923	0,1767	0,1901	0,1826	0,1878	0,1544	0,2330	0,4942	0,1548	0,2335
Exp 3	0,6931	0,3844	0,4932	0,1782	0,1906	0,1837	0,1925	0,1533	0,2323	0,4964	0,1537	0,2327
Exp 4	0,7000	0,3913	0,5008	0,1814	0,1956	0,1877	0,1991	0,1570	0,2372	0,5006	0,1574	0,2377
Exp 5	0,6973	0,3889	0,4980	0,1801	0,1937	0,1861	0,1982	0,1562	0,2359	0,4994	0,1566	0,2364
Exp 6	0,7015	0,3888	0,4991	0,1807	0,1932	0,1863	0,1972	0,1557	0,2361	0,5047	0,1560	0,2366
Exp 7	0,7078	0,3956	0,5062	0,1837	0,1981	0,1901	0,2072	0,1592	0,2403	0,5068	0,1596	0,2407
Exp 8	0,6971	0,3870	0,4966	0,1797	0,1925	0,1854	0,1945	0,1560	0,2366	0,5043	0,1564	0,2371
Exp 9	0,6984	0,3895	0,4990	0,1800	0,1936	0,1861	0,1980	0,1556	0,2354	0,4985	0,1559	0,2359
Exp 10	0,7007	0,3903	0,5000	0,1816	0,1950	0,1875	0,1993	0,1564	0,2363	0,5015	0,1568	0,2368
Media	0,6987	0,3891	0,4986	0,1802	0,1936	0,1862	0,1974	0,1561	0,2360	0,5010	0,1565	0,2365

Tabla 28. RTVE LDA 40 Segmentado.

RTVE LDA 80 Segmentado												
	ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-4		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,7378	0,4069	0,5233	0,4362	0,2391	0,3082	0,3493	0,1912	0,2466	0,3208	0,1755	0,2264
Exp 2	0,7358	0,4097	0,5251	0,4340	0,2404	0,3087	0,3500	0,1936	0,2487	0,3219	0,1779	0,2287
Exp 3	0,7382	0,4077	0,5239	0,4364	0,2391	0,3081	0,3523	0,1927	0,2484	0,3240	0,1772	0,2285
Exp 4	0,7430	0,4149	0,5311	0,4462	0,2477	0,3177	0,3615	0,2005	0,2573	0,3332	0,1848	0,2371
Exp 5	0,7411	0,4092	0,5259	0,4436	0,2435	0,3136	0,3580	0,1962	0,2528	0,3288	0,1802	0,2322
Exp 6	0,7384	0,4098	0,5258	0,4359	0,2403	0,3090	0,3508	0,1931	0,2484	0,3221	0,1772	0,2280
Exp 7	0,7378	0,4081	0,5243	0,4397	0,2417	0,3112	0,3547	0,1947	0,2508	0,3264	0,1790	0,2307
Exp 8	0,7373	0,4085	0,5243	0,4407	0,2427	0,3122	0,3554	0,1956	0,2516	0,3266	0,1796	0,2311
Exp 9	0,7403	0,4101	0,5266	0,4423	0,2434	0,3132	0,3564	0,1958	0,2522	0,3279	0,1801	0,2320
Exp 10	0,7408	0,4091	0,5257	0,4457	0,2443	0,3148	0,3605	0,1973	0,2543	0,3322	0,1817	0,2343
Media	0,7390	0,4094	0,5256	0,4401	0,2422	0,3117	0,3549	0,1951	0,2511	0,3264	0,1793	0,2309

RTVE LDA 80 Segmentado												
	ROUGE-L			ROUGE-W-1.2			ROUGE-S*			ROUGE-SU*		
	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F	Average_R	Average_P	Average_F
Exp 1	0,6971	0,3843	0,4943	0,1790	0,1904	0,1841	0,1912	0,1524	0,2319	0,5005	0,1528	0,2324
Exp 2	0,6942	0,3866	0,4955	0,1799	0,1933	0,1859	0,1936	0,1553	0,2350	0,4991	0,1557	0,2355
Exp 3	0,6954	0,3839	0,4934	0,1791	0,1907	0,1841	0,1927	0,1531	0,2323	0,4997	0,1535	0,2328
Exp 4	0,7020	0,3919	0,5017	0,1808	0,1947	0,1870	0,2005	0,1577	0,2382	0,5037	0,1581	0,2387
Exp 5	0,6997	0,3863	0,4964	0,1806	0,1923	0,1857	0,1962	0,1544	0,2344	0,5040	0,1548	0,2349
Exp 6	0,6958	0,3860	0,4953	0,1792	0,1918	0,1848	0,1931	0,1551	0,2350	0,5015	0,1554	0,2355
Exp 7	0,6981	0,3860	0,4960	0,1806	0,1925	0,1859	0,1947	0,1543	0,2343	0,5026	0,1547	0,2348
Exp 8	0,6970	0,3860	0,4955	0,1809	0,1931	0,1863	0,1956	0,1550	0,2349	0,5027	0,1553	0,2354
Exp 9	0,6998	0,3876	0,4977	0,1805	0,1929	0,1860	0,1958	0,1553	0,2356	0,5040	0,1556	0,2360
Exp 10	0,7009	0,3870	0,4973	0,1826	0,1947	0,1879	0,1973	0,1540	0,2339	0,5032	0,1544	0,2344
Media	0,6980	0,3866	0,4963	0,1803	0,1926	0,1857	0,1951	0,1547	0,2345	0,5021	0,1550	0,2350

Tabla 29. RTVE LDA 80 Segmentado.