Resumidor de textos basado en marcadores discursivos

Facultad de Ingeniería – Universidad de la República

José Enrique González
Roberto Clavell

Tutor: Juan José Prada

Resumen

La generación de grandes cantidades de información electrónica que se genera en el mundo es una de las características de las últimas décadas. Alrededor de 1500 millones GB de información se producen anualmente de los cuales un 90% se almacenan en dispositivos ópticos y magnéticos. Esto ha provocado un gran interés en la investigación y desarrollo de sistemas de búsqueda y recuperación cada vez más eficientes que ayuden a los usuarios a organizar, buscar y comprender la información. Los sistemas de generación automática de resúmenes de texto colaboran en la realización de éstas tareas.

En este proyecto se realizó un estudio sobre diferentes criterios empleados por resumidores automáticos de textos basados en extracción. Se propuso un conjunto de métodos para realizar un resumen y se hizo énfasis en el análisis de la coherencia, dado que mediante la técnica de extracción de frases completas empleadas en su creación se puede introducir cierto grado de inconsistencias en el nuevo texto.

Los marcadores discursivos juegan un papel importante en un texto, ya que estos guían, ordenan y añaden cierta estructura a éste. Se pueden clasificar de acuerdo a la cantidad de segmentos textuales que relacionan: unarios (un segmento), binarios (dos segmentos), compuestos (más de dos segmentos). Basándose en la ocurrencia de estos marcadores se realiza un análisis e implementación de una solución a las posibles incoherencias del resumen.

Se realizó un prototipo el cual permite realizar resúmenes de textos y obtener dos tipos de éstos: un resumen al cual se le aplican técnicas de coherencia basadas en las ocurrencias de marcadores discursivos binarios y otro al que no se le aplican estas técnicas. De esta forma, se verifica la intención de la propuesta, brindar una mejor calidad y legibilidad al resumen.

Es importante destacar, que en el 100% de las pruebas realizadas se obtuvo un resultado mayor al 50% de coincidencias entre el resumen manual y el resumen de sistema. Además, el 45% de las pruebas superó el 70% de coincidencias con el resumen manual, y dentro de este grupo, el 61.1% superó el 75% de coincidencias.

En el 100% de las pruebas se obtuvieron resultados positivos en la resolución de los problemas de cohesión y coherencia, que se presentaron por la ocurrencia de marcadores discursivos (binarios). Este es el principal aporte de la propuesta.

1. Introducción

En este capítulo se presenta una descripción del problema, la motivación y objetivos del trabajo, el cronograma de actividades realizadas durante el desarrollo del proyecto, y por último, la organización de este informe.

1.1. Descripción del problema

El problema que se plantea es realizar un resumidor automático de textos basado en marcadores discursivos.

Resumir es reducir la complejidad y largo de un texto, reteniendo algunas cualidades esenciales de éste. La meta es encontrar un subconjunto del documento original, que sea indicativo de su contenido.
El término resumidor automático, se refiere a sistemas computacionales (no necesariamente) capaces de reducir el tamaño de un texto manteniendo las ideas y el orden en que se presentan las oraciones. Una de las formas de realizar un resumen automático es seleccionando un subconjunto de oraciones del texto a resumir, estas oraciones "deberían" contener las ideas principales del texto; otra forma es la de interpretar el texto fuente y generar nuevo texto para el resumen.

Por otro lado, el término marcador discursivo se refiere a un conjunto de unidades que establecen relaciones entre segmentos textuales. Su cometido es guiar y ordenar los procesos de interpretación asociados a la comprensión de un texto.

Un problema de la mayoría de los resumidores automáticos de texto actuales, es la posible falta de coherencia producida por la extracción de oraciones. Una de las causas de esta incoherencia en los resúmenes proviene porque no se respeta las funciones que cumplen los marcadores discursivos en el texto.

1.2. Motivación

Los grandes avances en la tecnología en los últimos años, han dado lugar al crecimiento del volumen de la información electrónica. Son muchos los libros, revistas, periódicos y textos en general que se manejan en forma electrónica. La cantidad de medios diferentes para acceder a esta información que se genera cada día en el mundo es una de las características de las últimas décadas.

Investigaciones y estudios demuestran que la información que se produce anualmente en el mundo es aproximadamente 1500 millones de GB, de los cuales el 90% del total se almacena en medios ópticos y magnéticos.

Esta gran cantidad de información que se produce, transmite y almacena ha provocado un gran interés en la investigación y desarrollo de sistemas de búsqueda y recuperación cada vez más eficientes que ayuden a los usuarios a organizar, buscar y comprender la información. Los sistemas de generación automática de resúmenes de texto colaboran en la realización de éstas tareas.

Tener un resumen del texto permite conocer de antemano cuales son los temas que son tratados y de esta forma saber si su contenido es de interés, ahorrando mucho tiempo de lectura.

1.3. Objetivos

Los objetivos de este proyecto son: el estudio de técnicas de resumen automático basado en extracción de frases, la investigación de las propiedades y funciones de los marcadores discursivos, y la realización de una propuesta de las técnicas a utilizar para confeccionar el resumen.

El objetivo principal de este trabajo, es el solucionar el problema de coherencia en los resúmenes teniendo en cuenta las propiedades y funciones de los marcadores discursivos, es decir, basados en la presencia de éstos. Claro está que la no ocurrencia de marcadores discursivos en el texto a resumir deja sin sentido este aporte.

1.4. Cronograma

A continuación se presenta el cronograma donde se muestran los tiempos y dedicación para cada actividad a lo largo del proyecto.

Mes	Mayo		Junio		Julio		Agosto		Setiembre		Octubre		Noviembre		Diciembre		Enero		Febrero		Marzo		Abril	
1	X	X	X	X	X	X	X	X																
2							X	X	X	X	X	X	X	X	X	X		X	X					
3													X	X	X		X		X		X			
4			X	X			X	X	X			X				X	X	X	X	X	X	X	X	

Figura 1 - Cronograma

1- Estudio de bibliografías. Investigación de marcadores discursivos y etiquetado de texto. Técnicas de resumen a partir de un texto etiquetado.
2- Diseño e implementación del componente resumen.
3- Evaluación y validación de resultados.
4- Elaboración de informes.

1.5. Organización del informe

El presente documento se divide en 6 capítulos en los cuales se desarrollan todas las etapas que conforman este proyecto. A continuación se realiza una breve reseña de los 5 capítulos restantes.

- Capítulo 2: se presenta el estudio e investigación de trabajos de diferentes autores sobre resumidores automáticos de textos. El capítulo se divide en dos principales temas: ponderación de las oraciones y técnicas de aplicación de coherencia al resumen. De aquí se obtuvieron las principales heurísticas para la ponderación de las oraciones.
- Capítulo 3: se presentan las principales decisiones tomadas para la resolución del problema. Se divide en tres grandes partes: interpretación del texto fuente, ponderación de las oraciones y aplicación de coherencia al resumen basado en marcadores discursivos.
- Capítulo 4: se describe el sistema a nivel de diseño e implementación.
- Capítulo 5: se describen las pruebas realizas en cada una de las etapas y los resultados obtenidos.
- Capítulo 6: se presentan las conclusiones y algunas ideas de trabajo a futuro.

2. Estado del arte

Este capítulo comienza con una breve descripción de algunos conceptos básicos necesarios para comprender el trabajo (sección). En las secciones y se presentan investigaciones realizadas por diferentes autores relacionados a resúmenes automáticos de textos. Para la realización de este proyecto se tomaron las propuestas y definiciones de dichos autores. Cabe destacar que las ideas contenidas en la sección se obtuvieron del documento "Anexo 1 – Trabajos relacionados".

2.1. Conceptos básicos

A continuación se definirán conceptos importantes que se presentan a lo largo del documento y es esencial que el lector esté en conocimiento de ellos para entender las principales ideas.

Una **oración** es la menor unidad del habla que transmite un mensaje completo por sí misma, por lo que se dice que tiene sentido completo e independencia sintáctica. Las oraciones comienzan con mayúscula y terminan con un punto. El punto se utiliza para dar fin a una oración e indicar una pausa entre las distintas ideas que se expresan.

Un **párrafo** está conformado por un conjunto de oraciones. Se define como parte de un escrito que se considera con unidad suficiente para poder separarlo mediante una pausa que se indica con el punto aparte. Es una unidad del texto escrito en la cual se desarrolla determinada idea que presenta una información de manera organizada y coherente.

Los **marcadores discursivos** son términos cuyo cometido es ordenar y guiar al texto. Establecen cohesión y coherencia entre las oraciones que conforman el texto (su definición y estudio se presenta en la sección).

Se define un **resumen por extracción** como un conjunto de oraciones que son seleccionadas de un texto origen, conformando un nuevo texto y manteniendo el orden original de las oraciones.

La **cohesión** es una característica de todo texto bien formado, consiste en que las diferentes frases están conectadas entre sí mediante diversos procedimientos lingüísticos que permiten que cada frase sea interpretada en relación con las demás

La **coherencia** es una propiedad de los textos bien formados que permite concebirlos como entidades unitarias, de manera que las diversas ideas secundarias aportan información relevante para llegar a la idea principal, o tema, de forma que el lector pueda encontrar el significado global del texto.

El **score** o **ponderación** es el puntaje o valor que se le da a las oraciones de acuerdo a diferentes métodos. Sinónimo de **peso**.

Las **oraciones candidatas**, son oraciones propuestas por los métodos ponderadores para que conformen el resumen final por ser las que han obtenido mejor **ponderación**, pero aún están sujetas al análisis de coherencia utilizando los **marcadores discursivos**.

Las **oraciones marcadas para resumen**, son oraciones que serán incluidas en el resumen final luego de realizado el estudio de coherencia.

2.2. Marcadores discursivos

El concepto de marcadores discursivos es muy importante para este trabajo, por lo cual se trata de explicar en forma breve y concisa (citando a los diferentes autores) sus propiedades para dar una idea general al lector.

Prada[1] define a los marcadores como:

> "*Los marcadores discursivos son un conjunto de términos que establecen relación entre segmentos textuales. Su cometido es fundamentalmente el de guiar y ordenar los procesos de interpretación asociados a la comprensión de un texto. Los marcadores discursivos son operadores que añaden estructura al texto. Son en general operadores binarios, si bien en algunos casos la aridad de la relación es mayor que dos.*"

Estas definiciones y propiedades son fundamentales para este trabajo pues:

- Por su propia definición se realiza el estudio de los marcadores discursivos para aplicar coherencia y cohesión al texto (resumen).
- En presente documento, cuando se habla de argumentos de un marcador se hace referencia a los "segmentos textuales" que menciona el autor (argumentos de un marcador también definidos por dicho autor).
- Agrupa los marcadores en categorías según su aridad y/o cantidad de argumentos: unarios, binarios o compuestos. Ejemplos:
    - Unarios: "por su parte".
    - Binarios: "adicionalmente", "sin embargo", "en consecuencia".
    - Compuestos: "por un lado", "por otro lado", "en primer lugar".

Estas propiedades se utilizaron en la aplicación y se nombrarán a lo largo del documento.

A continuación se mencionan las ideas de dos autores en los cuales podemos apreciar que todas las definiciones convergen a una misma idea: los marcadores discursivos guían y ordenan el texto.

Según Zorraquino[2], con el término "Marcadores del Discurso" se alude a:

> "*las unidades lingüísticas invariables, que no desempeñan una función sintáctica dentro del ámbito de la predicación oracional y poseen un cometido coincidente en el discurso: el de guiar, de acuerdo con sus distintas posibilidades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación*".

Poblete[3] define cohesión y marcadores discursivos como:

> "*la cohesión es un concepto semántico que se refiere a las relaciones de significado que se dan en el discurso. Estas relaciones se establecen cuando la interpretación de un elemento depende de otro dentro del discurso. Los marcadores discursivos son uno de los medios lingüísticos que permiten la cohesión de las unidades supraoracionales, es decir, los marcadores discursivos son usados con un fin: cohesionar el discurso.*"

En los siguientes ejemplos se presentan algunos segmentos de textos que contienen marcadores discursivos de distinta aridad.

Ejemplo 1 - Marcador unario

"***Por su parte**, los de Avellaneda esperaban y se acomodaban para aprovechar la contra.*"
Diario Clarín – Argentina – 20/01/2006

Ejemplo 2 - Marcador binario

"*Las prácticas cotidianas de las organizaciones internacionales que trabajan en el terreno ofrecen una ayuda concreta a miles de niños y adultos que viven en situaciones intolerables. **No obstante**, la ayuda prestada no tiene que sustituir la responsabilidad del Estado, de la sociedad civil y de las comunidades.*"
http://www.paginadigital.com.ar/articulos/2006/2006prim/noticias/pobreza-290106.asp

Ejemplo 3 - Marcador compuesto

"*Los telescopios milimétricos son llamados así porque captan radiación con longitud de onda de alrededor de un milímetro. Esta radiación corresponde a la transición entre las ondas de radio, con longitud de ondas de decenas de centímetros, y el infrarrojo lejano, con longitudes de onda de decenas o centenares de micras. Las ondas milimétricas, que corresponden también a microondas de alta frecuencia, son emitidas principalmente por gas y polvo frío. **Por un lado**, si el gas está suficientemente frío para estar constituido por moléculas, estas giran alrededor de alguno de sus ejes y cuando disminuye este movimiento de rotación emiten ondas milimétricas. **Por otro lado**, el mismo gas o el polvo cuando se encuentran a temperaturas alrededor de diez grados por encima del cero absoluto (o sea -263 grados centígrados) brilla en ondas milimétricas.*"
http://www.inaoep.mx/~rincon/unifrio.html

2.3. Resúmenes de textos

En esta sección se presenta una clasificación de los resúmenes y las distintas técnicas de ponderación de oraciones utilizadas por diferentes autores.

2.3.1. Tipos de resúmenes

Existen en la actualidad dos métodos para realizar resúmenes automáticos de textos. Estos métodos son: abstracción y extracción. El primero genera un nuevo texto a partir de un análisis del texto fuente, mientras que el segundo es una colección de oraciones o frases extraídas del texto fuente.
Hovy[4] menciona que los resúmenes pueden ser clasificados de acuerdo a tres características:

- Entrada: características del texto fuente.
  - Tamaño del fuente:
    - Documento simple: un solo documento de entrada.
    - Multi-documento: el resumen es obtenido de varios textos, y normalmente se usa cuando los textos de entrada están temáticamente relacionados.
  - Especificidad:

        - Dominio específico: se realiza un resumen de dominio específico cuando el texto de entrada pertenece a un dominio simple restringido.
        - Dominio general: se deriva de textos de cualquier dominio, y no se puede tomar las asunciones anteriores.
    - Genero y escala:
        - Géneros típicos son: artículos de diarios, novelas, historias cortas, etc. La escala puede ser desde largos de libros y largos de párrafos.

- Salida: características del resumen como un texto.
    - Derivación:
        - Extracto: Un extracto es una colección de pasajes (desde una simple palabra, hasta párrafos enteros) extraídos del texto fuente.
        - Abstracto: es un nuevo texto generado el cual se produce luego de un análisis del texto fuente.
    - Coherencia:
        - Fluente: todas las oraciones son coherentes y una sigue a otra de acuerdo a una estructura de discurso coherente.
        - No fluente: las oraciones se forman de palabras individuales o porciones de texto que no forman oraciones o párrafos gramáticamente correctos.
    - Parcialidad:
        - Neutral: refleja el contenido del texto de entrada.
        - Evaluativo: incluye cierta inclinación del sistema, tal como manifestación de opiniones e inclusión de un material si y otro no.
    - Convencionalidad:
        - Fijo: creado para un uso específico, lector (o clase de lector), y situación.
        - Flotante: es creado para distintos tipos de lectores para una variedad de propósitos.

- Propósito: características del uso del resumen.
    - Audiencia:
        - Genérico: se le da igual importancia a todos los temas principales del texto fuente.
        - Orientados a consulta: favorece temas o aspectos específicos del texto en respuesta al deseo del usuario.
    - Uso:
        - Indicativo: provee solo una reseña del asunto o tema principal del texto.
        - Informativo: refleja el contenido del texto original y permite al lector describir lo que se encontraba en dicho texto.

La idea es construir un resumidor automático basado en extracción de oraciones cuya entrada es un documento simple (un solo documento). Se tomará como entrada documentos del género periodístico ya que éstos contienen una rica gama de marcadores discursivos, los cuales serán utilizados para obtener coherencia en el resumen producido. Estos resúmenes serán de *uso informativo*, ya que reflejarán el contenido del texto original y permitirán al lector describir las ideas principales de dicho texto; *uso informativo* es una característica dentro de la clase *propósito* (según Hovy[4]).

2.3.2. Ponderadores

Dado que este trabajo se basa en los resúmenes por extracción de frases, el estudio sobre resumidores de otros autores se concentra en los que utilizan dicha técnica.

La extracción es la aproximación más fácil al problema de la generación automática de resúmenes de textos, puesto que no es necesario generar nuevo texto. El problema se reduce a la identificación de los elementos significativos del texto fuente, habitualmente frases u oraciones, y a la selección de los mismos.
Enumeraremos algunas ventajas de la extracción de frases:

- Bajo costo.
- No se necesitan recursos adicionales con conocimientos del dominio ya que no hay que generar texto.
- Como se trata de un análisis superficial del texto es muy robusto.
- Posee gran independencia del dominio e incluso del género de los documentos, por lo que es fácilmente aplicable a contextos de propósito general.

Básicamente la extracción de oraciones o frases consta de dos fases: análisis y síntesis. En la fase de análisis se procesan cada una de las oraciones o frases del texto fuente y se mide su relevancia para asignarles un puntaje. En la fase de síntesis se extraen las oraciones o frases con mejor puntuación y se colocan en el mismo orden que aparecen en el texto fuente para conformar el resumen.
Existen en la actualidad muchos trabajos sobre generación de resúmenes por extracción de los cuales surgen una gran cantidad de métodos que conformarían la fase de análisis. Estos métodos se pueden clasificar en:

- Estadísticos: incluyen la frecuencia de aparición de ciertas palabras.
- Posicionales: se basan en la posición que ocupa la frase dentro del documento.
- Formato y largo: buscan ciertas características en la escritura del texto.

A continuación se describirán los diferentes métodos empleados para la fase de análisis en los resumidotes, así como una propuesta (en algunos casos) de su utilización en este trabajo.

Métodos Estadísticos

2.3.2.1. Palabras Clave

Éste es uno de los métodos mas conocidos y mas nombrado por los distintos autores. Sigue la hipótesis de que las palabras que aparecen frecuentemente en el documento son relevantes. No se deberá tener en cuenta la existencia de aquellas palabras que aparecen con gran frecuencia en los documentos y no acarrean significado cuya función es la de enlazar otras palabras. Ejemplo de este tipo de palabras son los artículos, conjunciones, determinantes, pronombres, preposiciones, etc.

Mateo[5] utiliza un método de puntuación el cual comienza creando una tabla que recoge todas las palabras categorizadas en adjetivos y verbos léxicos, asignando a cada palabra una puntuación proporcional a su frecuencia de aparición. De esta tabla inicial se eliminan las palabras de baja frecuencia, que son aquellas cuya frecuencia es menor al promedio de frecuencia de todas las palabras de la tabla. Luego se asigna una puntuación a cada frase comparando todas las palabras de dicha oración con todas las palabras de la tabla creada. Para cada coincidencia se suma el valor de la palabra en la tabla a la frase y luego se normaliza dividiendo la suma total de la frase entre el número de palabras de la frase. Por último, se eliminan las frases menos puntuadas, y las salidas de este módulo son: la tabla con las puntuaciones de cada frase y una lista de frases candidatas.

Acero[6] propone extraer las M palabras mas importantes del texto y comprueba cuantas de esas palabras se encuentran en cada frase. De esta forma se asigna mayor peso a las frases que contengan mayor número de palabras claves del texto. Para obtener las M palabras mas relevantes, se utiliza el método tf.idf (producto entre la frecuencia del término y la inversa de la frecuencia en el documento).

En Maña[7], se utilizan dos métodos para encontrar las palabras mas relevantes o significativas del texto: tf (frecuencia del término) y tf.idf (producto entre la frecuencia del término y la inversa de la frecuencia en el documento). Para puntuar la frase se presentan también dos enfoques: los que puntúan la aparición de grupos de palabras claves en la frase y los que tienen en cuenta la aparición por separado de las palabras claves en la frase.

Luhn[8], emplea la frecuencia de los términos (tf) y luego define grupos de palabras (conjunto de palabras claves separadas por un máximo de otras cuatro palabras) para puntuar las frases. Si una frase contiene más de un grupo, se elige el grupo con mayor cantidad de palabras significativas. La puntuación final de la frase se obtiene dividiendo el cuadrado del número de palabras claves del grupo entre el número total de palabras del mismo.

Un posible método para calcular la ponderación de una oración de acuerdo a sus palabras claves es mediante el siguiente método:

Primero se calcula la frecuencia (fi) de cada palabra (sustantivo o nombre) en el documento, y luego se halla su peso normalizado (pi) como:

$$pi = fi / \sum j(fj)$$

con $\sum j(fj)$ igual a la sumatoria de la frecuencia de todas las palabras (sustantivo o nombre).

Luego se determina el peso de la oración como:

$$Pj = \sum i\ (nij * pi) / \text{cantidad de palabras de la oración } j$$

con *nij* igual a la cantidad de apariciones de la palabra i en la frase j.

2.3.2.2. Palabras del título

Normalmente las palabras del título que aparecen en el texto denotan importancia de la frase u oración donde ocurren. Este método es muy usado por distintos autores para realizar la ponderación de las oraciones.

En Mateo[5], se menciona que el título de un texto suele estar fuertemente relacionado con su contenido y a menudo constituye el mejor resumen del mismo. Presenta un algoritmo que identifica en primer lugar las palabras del título, luego pondera las frases del documento conforme a su aparición, normalizando luego los valores obtenidos.

Edmundson[9] fue el primero en implementar esta heurística para un sistema resumidor. Para esto se asignan pesos positivos a las palabras significativas que forman parte del título. La puntuación final de la frase se calcula sumando los pesos de los términos que incluye.

Un posible método para calcular la ponderación de las oraciones basado en las ideas expuestas es:

$$p_i = \text{Total de palabras}^1 \text{ del título en oración } i / \text{Total palabras de la oración } i$$

Donde pi es la ponderación de la oración i y con *palabras*[1] se hace referencia a sustantivos y nombres.

2.3.2.3. Palabras de encabezados

Al igual que las palabras del título, la aparición de palabras de encabezados en el texto, denotan importancia de las frases donde aparecen. Éste método no es tan usado como el de palabras del título, pero igualmente es nombrado e implementado en diferentes trabajos. Los autores mencionan éste método como un complemento del anterior.
En Maña[7] se menciona que las palabras que aparecen en los títulos, subtítulos y encabezamientos pueden ser buenas indicadoras del contenido de los mismos. Se pondera de la misma forma que las palabras del título, pero en la experiencia menciona que los resultados que se obtienen son peores que cuando utilizan sólo palabras del título.
En Georgantopoulos[10] se sugiere que las palabras de encabezados y subtítulos suministran gran cantidad de información acerca del contenido del texto. Se pondera al igual que las palabras del título y se expone que es posible agregar al resumen los títulos, subtítulos y encabezados para crear un resumen más coherente y estructurado.

Un posible método para calcular la ponderación de las oraciones basado en las ideas expuestas es:

p_i= *Total de palabras[1] del encabezado j en oración i / Total palabras de la oración i*

Donde pi es la ponderación de la oración i, dicha oración pertenece a la sección del encabezado j y *palabras[1]* hace referencia a sustantivos y nombres.

2.3.2.4. Palabras indicadoras

Existen en los diferentes textos ciertas palabras o frases que pueden indicar la presencia de información importante y las frases en donde éstas aparecen deberían ser incluidas en el resumen.
Mateo[5] indica que palabras como "importante", "esencial", o "para concluir" son palabras indicativas. Uno de los principales inconvenientes de este método es la dependencia de estas palabras con el género del documento fuente. Se usa un conjunto de 130 palabras indicadoras que son encontradas por el algoritmo en el texto para puntuar las frases que las contienen.
Edmundson[9] construye dos listas de palabras denominadas *bonus* y *stigma*, donde en la primera lista se incluyen las palabras (comparativos, superlativos o adverbios de conclusión) que puntúan positivamente las frases que las contienen y en la segunda las palabras (anáforas, o expresiones que indican especulación o tienen carácter evasivo) que puntúan negativamente. La puntuación final de una frase se obtiene sumando los puntos de las palabras indicadoras que incluye.

Para implementar esta técnica no alcanza solamente con identificar las palabras *bonus y stigma* dentro del texto pues las palabras tienen una fuerte relación con el estilo y contexto del documento en estudio, por lo que quizás globalizar las palabras *bonus y stigma* para todo tipo de documento no sea lo más correcto
Por ejemplo: la palabra "conclusión" es incluida por todos los autores en la lista de palabras *bonus* aunque en el texto fuente se presente algo del estilo como:

"*…., entendemos que no es una buena conclusión.*", lo cual genera cierta ambigüedad para tomar "conclusión" como palabra *bonus*, la cual puede ser importante en un documento, pero no así en otro.

Una posible solución a este problema es utilizar sistemas de exploración contextual para desambiguar las palabras pertenecientes a las lista *bonus y stigma.* El Método de

Exploración Contextual (MEC) desarrollado por el grupo LaLIC[1] y dirigido por Jean Pierre Desclés[16] provee el marco necesario para identificar información semántica específica contenida en los textos. El método se apoya en la hipótesis de que en todo texto aparecen determinadas unidades lingüísticas que ayudan a levantar indeterminaciones semánticas y a encontrar relaciones entre segmentos de texto.

2.3.2.5. Nombres propios

Los nombres de personas, lugares, organizaciones, etc., proporcionan información y su inclusión en el resumen puede aumentar la cantidad de datos para el lector.
Santos[11] menciona que la utilización de nombres propios por parte de los autores de los documentos es más común si se trata de un texto de noticias.
Mateo[5] simplemente puntúa las frases del documento que contienen nombres propios, normalizando los valores obtenidos.

Un posible método para calcular la ponderación de las oraciones basado en las ideas expuestas es asignar un valor fijo *w* (parametrizable) a todos los nombres, luego el peso p_i de la oración *i* será:

p_i= *(Total de nombres en la oración i * w) / Total palabras de la oración i.*

2.3.2.6. Consulta de usuario

Distintos usuarios pueden desear orientar o enfocar sus resúmenes a temas distintos, esto se llama realizar un resumen personalizado. Es por esto que uno de los métodos utilizados es el de consulta de usuario, en donde éste ingresa ciertos términos de acuerdo a sus pretensiones. Luego, estos términos son usados para la confección del resumen.
Normalmente el usuario ingresa un conjunto de palabras las cuales ponderarán las frases u oraciones donde se encuentran.
Acero[6] orienta la elección de frases de tal forma que se elijan aquellas que tengan mayor similitud con las preferencias del usuario. Dado un modelo de usuario se obtiene la información con respecto a los pesos que el usuario a asignado a sus categorías y a sus términos personales. También se extrae del modelo, los términos que representan cada categoría así como los términos que el usuario haya definido. Con toda esta información se calcula la similitud entre el modelo y la frase asignando un peso a dicha frase.
Mateo[5] permite usar una consulta opcional definida como una serie de palabras clave que introduce el usuario y sirve para guiar el proceso de síntesis.

Una posible propuesta es asignar un valor fijo *w* (parametrizable) a todas las palabras ingresadas por el usuario, luego el peso p_i de la oración *i* será:

p_i= *(Total de palabras[1] de usuario en oración i * w) / Total palabras de la oración i.*

Donde *palabras[1]* hace referencia a sustantivos y nombres.

Métodos Posicionales

2.3.2.7. Extracción según posición en el documento

Otro método utilizado para ponderar las frases tiene en cuenta la posición de las oraciones dentro del párrafo, y la posición de éste último dentro del documento fuente. Como se verá a continuación los autores presentan diferentes técnicas para implementar dicho método, variando entre técnicas "estáticas" (Ej. La primer oración del primer párrafo es más importante) y "dinámicas" (Ej. Se usa una función para determinar el peso de una oración según su posición en el documento).

[1] Universidad de Paris IV, Francia

Mateo[5] menciona que los párrafos iniciales y finales de un documento suelen contener gran riqueza semántica, siendo menos importantes los párrafos centrales. El algoritmo puntúa las frases de acuerdo a una función lineal de pendiente configurable que asigna puntuación máxima a la primera frase.
Hovy[4] menciona que luego de estudiar resúmenes hechos por personas, creó el OPP (Optimal Position Policy) el cual es una lista rankeada que indica en que posiciones ordinales en el texto las oraciones con tópicos importantes tienden a aparecer.
Acero[6] da básicamente mayor peso a las primeras N frases del texto. En dominios periodísticos, el título y las primeras frases de un texto dan una idea aproximada al lector del contenido del texto que va a leer. Afirma que un valor típico de N es cinco, asignando los siguientes pesos a las cinco primeras frases: Frase 1 = 1.0, Frase 2 = 0.99, Frase 3 = 0.98, Frase 4 = 0.95, Frase 5 = 0.90.
Según Maña[7], otros autores también tienen en cuenta la posición de la frase dentro del párrafo, valorándose en mayor medida las frases inicial y final. También se da preferencias a frases que se encuentran a continuación de encabezamientos que incluyen palabras como: "conclusiones", "resumen" o "discusión".

Métodos según formato y largo de oraciones

2.3.2.8. Tipografía del texto

Una de las formas que los autores en sus textos tienen para resaltar información relevante es proporcionándole distinto formato a dicho texto. Normalmente esto es indicado por la presencia de frases en negrita, cursiva, mayúsculas, etc.
Kupiec[12] se refiere en particular a palabras que están escritas en mayúscula como son las siglas. Una restricción que establece es que la palabra en cuestión no debe ser inicial en la oración, y además, debe aparecer varias veces y no ser una abreviación de una unidad de medida.
Mateo[5] y Georgantopoulos[10], mencionan que los autores utilizan tipografía y formato del texto para resaltar frases o palabras importantes. Sus algoritmos comprueban la presencia de palabras en mayúsculas, negritas y subrayadas, puntuando las frases que las contienen.

Un posible método para implementar lo expuesto es definir un conjunto de posibles formatos de texto (negrita, cursiva, subrayado, mayúscula, etc.) y a cada uno asignarle un valor fijo w (parametrizable). El peso de la oración i se calcula como:

*Pi=Σj (wj * cantidad palabras con formato j en la oración i) / total de palabras orac i.*

2.3.2.9. Largo de las oraciones

Algunos autores como Kupiec[12], tienen en cuenta el largo de las oraciones del documento, ya que afirman que en los resúmenes realizados por humanos no se tienen en cuenta las oraciones cortas. Por dicha razón, dicho autor elimina las oraciones que no superan un valor prefijado de palabras.

Esta técnica se contrapone a los métodos descritos anteriormente, ya que en estos últimos se considera que una oración tiene importancia o no para el extracto según su ponderación, sin importar la cantidad de palabras que incluye.

2.3.2.10. Combinación de los métodos

Algunos autores como Acero[6] y Mateo[5] realizan una combinación lineal con los resultados de los diferentes métodos. Para cada frase u oración se calcula la ponderación final según la siguiente ecuación:

$$a_1m_1 + a_2m_2 + \ldots + a_nm_n$$

donde a_i es el peso que se le da al método i y m_i es el valor obtenido por la oración para el método i. En caso de que se quiera dar mas peso al método i, simplemente se aumenta el valor a_i para dicho método.

2.3.3. Cohesión y Coherencia basada en marcadores discursivos

A continuación se mencionan trabajos donde se aplica cohesión y coherencia al texto, algunos usando marcadores discursivos.

Mateo[5] menciona que si la frase inmediatamente anterior a la frase que contiene el marcador discursivo no pertenece al extracto, entonces quita ambas frases. Esta idea es un poco vaga ya que no se realiza un estudio de los marcadores del discurso, toma que la frase inmediatamente anterior es el alcance de todos los marcadores discursivos pero esto no siempre es así.
En su trabajo, Santos[11], menciona solamente a los marcadores discursivos que inician una oración. Indica que marcadores como "Porque", "Además" y "Adicionalmente", son considerados como indicadores de la presencia de información adicional, no esencial para el resumen. La idea que presenta el autor para trabajar con los marcadores discursivos es muy genérica y escasa, ya que la sugerencia que hace para realizar el tratamiento de éstos marcadores es igual para todos, sin tener en cuenta los tipos, alcance, categorías, etc. Tampoco tiene en cuenta la ponderación de las frases que contienen dichos marcadores.
Marcu[13] adopta una teoría donde postula 25 relaciones que "atan" cláusulas (que forman una estructura jerárquica del texto) para darle coherencia al resumen. Estas relaciones son tales como: "pero", "sin embargo", "en orden de", "porque", "luego", etc., -todas ellas son marcadores discursivos- las cuales tienen un componente principal (núcleo) y uno subordinado (el satélite). Estas relaciones pueden estar anidadas recursivamente y un texto será coherente si todas sus cláusulas pueden ser unidas, primero bajo subárboles locales y luego bajo árboles mas grandes bajo estas relaciones. Para producir el resumen, Marcu descarta el material menos saliente, en orden, recorriendo en forma top-down la estructura del discurso y solo siguiendo los links de satélites.

3. Resumidor de textos basado en marcadores discursivos

Este capítulo consta de dos partes, la primera contiene un análisis del tratamiento de los marcadores discursivos basado en el trabajo de Prada[1] junto a una propuesta de la selección de oraciones para el resumen (sección). En la segunda parte se describen las distintas etapas de procesamiento del texto fuente que resultará en el resumen final (sección). Las etapas son las siguientes:

- Reconocimiento del texto
  - Lectura del texto
  - Lectura de marcadores discursivos
  - Armado de la estructura
- Resumen de texto según ponderación de las oraciones
- Resumen de texto aplicando coherencia

3.1. Análisis del tratamiento de marcadores discursivos

Las oraciones que comprenden el texto fuente estarán en alguna de las siguientes categorías o clasificación:

- Oración candidata: oración que por su ponderación está dentro del conjunto de las N oraciones con mejor ponderación, donde N es la cantidad de oraciones para el resumen final, calculado como: porcentaje de resumen * cantidad de oraciones del texto / 100.
- Oración no candidata: oración que no pertenece a la clasificación anterior.
- Oración marcada para resumen: oración que formará parte del resumen final.

Se define "Lista de oraciones candidatas" como el conjunto de oraciones candidatas de un texto.

3.1.1. Oraciones que no comienzan con un marcador discursivo

Este tipo de oraciones se podrán incluir en el extracto final pues no presentan incoherencias al agregarse al resumen final. Se agregará la oración al resumen si y sólo si el porcentaje de resumen lo permite, es decir, que no se sobrepase el número de oraciones que conformarán el extracto.

3.1.2. Oraciones que comienzan con un marcador discursivo

Este tipo de oraciones requieren de un análisis de coherencia y cohesión dado que de no tenerse en cuenta dicho análisis, existe una alta probabilidad de que el extracto quede incoherente y que no tenga una buena legibilidad.

Este conjunto de oraciones se puede separar en diferentes clases según el tipo de marcador del discurso con el cual comiencen: unarias, binarias y compuestas como se presenta en el estudio de Prada[1].

Sólo se tomarán en cuenta los marcadores discursivos que comienzan una oración, pues el estudio de marcadores dentro de una oración no es necesario ya que el resumen se realiza por extracción de oraciones enteras.

Marcadores discursivos unarios

Las oraciones que contengan este tipo de marcador se incluirán en el resumen, ya que la presencia o no en el resumen de dichas oraciones no influirán en la coherencia y cohesión. Su inclusión en el resumen dependerá únicamente de la ponderación obtenida de acuerdo a los métodos de ponderación y del porcentaje de resumen que el usuario haya ingresado.

En el Ejemplo 4 se observa la ocurrencia del marcador discursivo unario "Por su parte", la oración que lo contiene será incluida en el resumen dependiendo únicamente de la ponderación que ha obtenido.

Ejemplo 4

"***Por su parte**, Chela ratificó su buen momento y, jugando un tenis de altísimo vuelo, resolvió fácil su compromiso ante el belga Vliegen.*"

Diario Clarín – Argentina – 20/01/2006

Marcadores binarios

Para las oraciones que contienen este tipo de marcador se debe diferenciar si el alcance del marcador en estudio incluye simplemente oraciones (conecta oraciones dentro del mismo párrafo) o párrafos (conecta párrafos, la oración que comienza con el marcador discursivo es la primera párrafo).

Nota: cuando se explica el alcance de los marcadores se presentan las siguientes etiquetas:

- <Pi> - comienzo del párrafo i
- <\Pi> - fin del párrafo i
- <Oi> - comienzo de la oración i
- <\Oi> - fin de la oración i
- MD – marcador discursivo
- ARGi – argumento i de un marcador discursivo (oración o párrafo)

Caso 1: Alcance intra-párrafo

Las oraciones que contienen un marcador discursivo binario al comienzo y se encuentran dentro del párrafo se pueden representar de la forma:

<P>….<O>ARG1<\O><O>MD, ARG2<\O>…..<\P>

donde **ARG1** y **ARG2** son las oraciones incluidas en el alcance del marcador del discurso. Los puntos suspensivos representan que puede haber otras oraciones dentro del párrafo las cuales no participan en el alcance del marcador en estudio.

Ejemplo 5

"**<P><O1>***En otros tiempos, felizmente, esto era un desafío para la Unión Europea.***<\O1><O2>***Pero lo que la UE quiere y lo que el gobierno quiere son lo mismo.***<\O2><O3>***"Es esencial que hagamos balance de la creciente demanda de desplazamientos", dice el informe del gobierno, "con nuestro objetivo de proteger el medio ambiente" [13].***<\O3><O4>***Hasta hace poco, teníamos una política de reducir la demanda de desplazamientos.***<\O4><O5>*****Ahora**, aunque no se ha anunciado de ninguna forma, esa política ya no existe.***<\O5>….<\P>**"

En el Ejemplo 5 se puede apreciar la ocurrencia del marcador discursivo binario "Ahora". El argumento 1 de dicho marcador es la oración inmediata anterior (O4) y el argumento 2 es la oración que lo contiene (O5).

En las proposiciones 1, 2 y 3, el argumento1 y argumento2 son O4 y O5 en el Ejemplo 5 respectivamente.

Proposición 1

Si el argumento1 es una oración marcada para resumen (ya fue estudiada), se marcará el argumento2 para el resumen solo si el porcentaje de resumen lo permite; en otro caso no se incluirá la oración en estudio para el resumen y se continuará el estudio de la siguiente oración candidata.

Proposición 2

Si el argumento1 (O1) es una oración candidata, se marcará el argumento1 (O1) y el argumento2 para el resumen solo si el porcentaje de resumen lo permite; en otro caso no se incluirá ninguna de las dos oraciones en el resumen y se continuará el estudio de la siguiente oración candidata.

Proposición 3

Si el argumento1 es una oración que no esta incluida en la lista de oraciones candidatas se calculará el promedio de los scores de ambas oraciones (argumento1 y argumento2); en caso que dicho valor sea mayor que el score mas bajo de las oraciones candidatas, se marcarán ambos argumentos como oraciones para el resumen; en caso de que sea menor, no se tomarán éstas oraciones para el resumen final. Si se decide incluir alguna oración, se deberá chequear si el porcentaje de resumen lo permite, y en caso negativo no se marca para resumen ninguna de las dos oraciones.

Nota: Para las tres proposiciones anteriores, antes de marcar el argumento1 como parte del resumen final, se debe realizar el estudio de coherencia sobre este argumento.

Caso 2: Alcance entre-párrafos

Las oraciones que comienzan con un marcador discursivo binario y "conectan párrafos" se pueden representar de la siguiente forma:

<P>ARG1 <\P><P><O>MD, …<\O>….<\P>.

ARG2

En este caso los argumentos o alcance del marcador del discurso son dos párrafos contiguos como define Prada[1].

Ejemplo 6

"**<P1><O1>***Pero claramente hubo la intención de que las caricaturas fueran una provocación.* **<\O1><O2>***Fueron tan absurdas, que lo que lo único que causaron fue una reacción.***<\O2><\P1>**

<P2><O1>*Además**, este no es el momento más adecuado para recalentar la vieja basura de Samuel Huntington sobre "el choque de civilizaciones".***<\O1><O2>***Irán tiene nuevamente un gobierno clerical.***<\O2><O3>***Lo mismo ocurre, para todo fin práctico, en Irak (donde supuestamente no iban a usar su democracia para elegir a un gobierno religioso, pero eso es lo que pasa cuando uno se pone a derrocar dictadores).* **<\O3><\P2>**"

En el Ejemplo 6 se puede apreciar la ocurrencia del marcador discursivo binario "Además". El argumento 1 de dicho marcador es el párrafo inmediato anterior (P1) y el argumento 2 es el párrafo que contiene a la oración que comienza con el marcador (P2).

En las proposiciones 4, 5 y 6, el argumento1 y argumento2 son P1 y P2 en el Ejemplo 6 respectivamente.

Proposición 4

Si el argumento1 (párrafo anterior) ya tiene oraciones marcadas para resumen, se marcará la oración en estudio para resumen y las oraciones candidatas del argumento2 (párrafo actual) siempre y cuando el porcentaje del resumen lo permita. Si el porcentaje no lo permite, se intentará marcar al menos la oración en estudio.

Proposición 5

Si el argumento1 tiene oraciones candidatas se marcarán estas oraciones y las oraciones candidatas del argumento2 (incluyendo la oración en estudio) para que conformen el extracto (si el porcentaje de resumen ingresado lo permite). Si el porcentaje de resumen no lo permite, se intentará marcar al menos una de las oraciones candidatas del argumento1 y la oración en estudio.

Proposición 6

Si ninguna de las oraciones pertenecientes al argumento1 esta incluida en la lista de oraciones candidatas se analizará la posibilidad de incluir la oración con mayor score dentro del argumento1 (O1 ó O2 dentro de P1 en el Ejemplo 6). Se calculará el promedio entre las oraciones candidatas del argumento2 con la oración de mayor score del argumento1. Si dicho valor es mayor que el menor score de las oraciones candidatas, se marcará la oración del argumento1 y las oraciones candidatas del argumento2 para el resumen (si el porcentaje de resumen ingresado lo permite). Si el promedio no supera el score de la peor de las candidatas, no se marcará la oración en estudio ni la oración de mejor score del argumento1 para el resumen final.

En caso de que no se pueda incluir la oración del argumento1 por superar el porcentaje de resumen, se intentará no tomar para el resumen alguna de las candidatas del argumento 2 (comenzando por las de menor score) y se verá si con esta alternativa se puede incluir las oraciones antes mencionadas.

Nota: Siempre que se decida marcar oraciones de un argumento como parte del resumen final, se deberá realizar el estudio de coherencia de cada una de ellas.

Marcadores compuestos

Otro tipo de marcadores discursivos son los compuestos, los cuales relacionan más de dos segmentos de texto. Para este tipo de marcadores se pueden diferenciar dos casos de acuerdo a la posición de los argumentos que lo conforman. Esto es, los argumentos se encuentran en un mismo párrafo o esparcidos en diferentes párrafos (no necesariamente consecutivos).

Nota: se agregan los siguientes tags en la representación del alcance de los marcadores:

- MA – marcador de apertura
- MC – marcador de continuidad
- MF – marcador de fin o cierre

Caso 1: Alcance intra-párrafos

Este caso se puede representar de la siguiente forma:

<P><O>….</O><O> MA,….</O><O>MC,….</O> (<O>MC,….</O>)* …. <O>MF, ….</O><O>…</O></P>

Donde MA es la representación de un marcador de apertura ("en primer lugar", "primeramente", "por un lado"), MC es la representación de los marcadores de continuidad ("en segundo lugar", "en tercer lugar", etc.) y MF es la representación de los marcadores de cierre ("finalmente", "por otro lado"). La aparición del carácter "*" indica la posible aparición de mas marcadores de continuidad.

Como se puede apreciar éstos marcadores están fuertemente relacionados, por lo que el análisis será más complejo.

Ejemplo 7

"**<P><O1>***Los telescopios milimétricos son llamados así porque captan radiación con longitud de onda de alrededor de un milímetro***<\O1>***. **<O2>***Esta radiación corresponde a la transición entre las ondas de radio, con longitud de ondas de decenas de centímetros, y el infrarrojo lejano, con longitudes de onda de decenas o centenares de micras***<\O2>***. **<O3>***Las ondas milimétricas, que corresponden también a microondas de alta frecuencia, son emitidas principalmente por gas y polvo frío***<\O3>***. **<O4>*****Por un lado***, *si el gas está suficientemente frío para estar constituido por moléculas, estas giran alrededor de alguno de sus ejes y cuando disminuye este movimiento de rotación emiten ondas milimétricas***<\O4>***. **<O5>*****Por otro lado***, *el mismo gas o el polvo cuando se encuentran a temperaturas alrededor de diez grados por encima del cero absoluto (o sea -263 grados centígrados) brilla en ondas milimétricas***<\O5>***.<\P>**"

En el Ejemplo 7 se detecta la ocurrencia del marcador compuesto de apertura "Por un lado" y el marcador compuesto de cierre "Por otro lado". No hay ocurrencia de marcadores de continuidad.

Para las proposiciones 7, 8 y 9, la oración que contiene el marcador de apertura es O4 y la oración que contiene el marcador de cierre es O5.

Proposición 7

Si la oración en estudio comienza con un marcador de apertura (Ej.: "En primer lugar"), se buscarán los sucesivos argumentos (Ej.: "En segundo lugar", "En tercer lugar") dentro del párrafo hasta encontrar el argumento que contenga el marcador del discurso de cierre.

Cuando se tengan identificadas todas las oraciones que conforman la estructura del marcador compuesto en estudio, se analizará la posibilidad de ingresar al extracto final dichas oraciones. Si todas las oraciones pertenecen a la lista de oraciones candidatas, simplemente se marcarán para el resumen, en otro caso se realizará el promedio de los score de las oraciones involucradas y si es mayor que el score de la peor oración de la lista de oraciones candidatas, se marcarán todas para el resumen.

Proposición 8

Si la oración en estudio comienza con un marcador de continuidad, se deberá buscar el marcador de apertura en oraciones anteriores y otros marcadores de continuidad (si existen) en oraciones anteriores y posteriores, por último se buscará el marcador de cierre en oraciones posteriores.

Luego de identificadas dichas oraciones se realiza el mismo estudio que en el caso anterior.

Proposición 9

Si la oración en estudio comienza con un marcador de cierre se buscarán en las oraciones anteriores (dentro del párrafo) los marcadores de apertura y continuidad que conforman la estructura del marcador compuesto. Una vez identificadas dichas oraciones se realizará el mismo estudio que se menciona en la proposición 7.

Caso 1: Alcance entre-párrafos

Este caso se puede representar de la siguiente forma:

<P><O> MA,…<\O>…<\P> ((<P>…<\P>)* <P><O>MC,…<\O>…<\P>)* <P><O>MF, ….<\O>…<\P>

La aparición del primer carácter "*" indica la posible ocurrencia de mas párrafos, y el segundo carácter "*" indica la posible ocurrencia de mas marcadores de continuidad.

Ejemplo 8

"**<P1><O1>*En primer lugar***, *se ha reformulado el procesamiento de manera de realizar los procesos de crítica de la información entrada a medios magnéticos y análisis de consistencia de la información, conjuntamente para Montevideo y el interior urbano.* **<\O1><\P1>**

<P2><O1>*En segundo lugar*, *se ha reformulado el proceso de entrada de la información a medios magnéticos, cambiándose el sistema anterior de entrada masiva de la información con digitadores especializados a una entrada de la información con personal que realiza al mismo tiempo la crítica y el análisis, lográndose una mejora en la calidad de este procedimiento y la reducción de errores en esta etapa***<\O1>**.**<O2>** *Este cambio en el procesamiento determinó un tiempo inicial de acomodamiento a las nuevas funciones que redundó en mayores plazos a los anteriores, que ya se ha normalizado***<\O2>**. **<\P2>**

<P3><O1>*En tercer lugar*, *las tareas relativas al Censo de Población requirieron que, fundamentalmente en los meses de marzo a junio pasados, personal de experiencia de la Encuesta Continua de Hogares pasara a apoyar puntualmente las tareas censales, determinando una merma en los recursos de la ECH durante dichos meses***<\O1>**.**<O2>** *Es importante destacar que el INE mantuvo, aun durante los períodos de mayor demanda de personal del Censo de Población, el relevamiento en forma continua, evitando tener que discontinuar las series de la ECH***<\O2>**.**<O3>** *Si bien el relevamiento se mantuvo, las tareas posteriores de procesamiento en oficina se vieron resentidas***<\O3>**. **<\P3>**"
Opinión, 4 al 10 de octubre de 1996

En el Ejemplo 8 se pueden apreciar los marcadores compuestos "en primer lugar", "en segundo lugar" y "en tercer lugar", los 3 en distintos párrafos. El primero es un marcador de apertura y los dos siguientes son marcadores de continuidad. No hay marcadores de fin o cierre.

Para las proposiciones 10, 11 y 12 se tendrá en cuenta el ejemplo recién presentado, donde en O1 del párrafo 1 (P1) se encuentra el marcador de apertura "en primer lugar", y los marcadores de continuidad "en segundo lugar" y "en tercer lugar" se encuentran en O1 del párrafo 2 (P2) y O1 del párrafo 3 (P3) respectivamente.

Proposición 10

Si la oración en estudio comienza con un marcador discursivo de apertura (Ej.: "En primer lugar") – O1 en P1 en el Ejemplo 8-, se buscarán los sucesivos argumentos (Ej.: "En segundo lugar", "en tercer lugar"), primero en el mismo párrafo y luego en los siguientes párrafos hasta encontrar el argumento que contenga el marcador del discurso de cierre – O1 en P2 y O1 en P3 en el Ejemplo 8-.

Una vez identificados todos los argumentos, se estudiará la posibilidad de incluirlos en el extracto final. Para esto, se calculará el promedio de todas las oraciones incluidas en todos los argumentos (párrafos enteros u oraciones simples).

Si el promedio es mayor que el peor score de la lista de oraciones candidatas, se incluirán todas las oraciones en estudio en el extracto final.

En caso de que el promedio sea menor al score de la peor candidata se quitarán de a una las oraciones en orden ascendente según score (no incluye las oraciones que contengan los marcadores pertenecientes a la estructura compuesta) pertenecientes a los argumentos que sean párrafos. Cuando el promedio de score de las oraciones restantes supere el score de la peor oración de la lista de oraciones, se marcarán dichas oraciones como parte del extracto final. Si con esta alternativa no se supera el score de la peor oración candidata no se tendrán en cuenta las oraciones en estudio para el resumen.

Proposición 11

Si la oración comienza con un marcador de continuidad – O1 en P2 o O1 en P3 para el Ejemplo 8-, se deberá buscar el marcador de apertura en los párrafos anteriores – O1 en P1 en el Ejemplo 8- y otros marcadores de continuidad (si existen) en oraciones y párrafos anteriores y posteriores – dependiendo del caso, O1 en P2 ó O1 en P3 para el Ejemplo 8-, por último se buscará el marcador de cierre en oraciones y párrafos posteriores. Luego de identificadas las oraciones que componen el alcance de los marcadores encontrados, se realiza el mismo estudio que en el caso anterior.

Proposición 12

Si la oración en estudio comienza con un marcador de cierre se buscarán los marcadores de apertura y continuidad en los párrafos anteriores – O1 en P1 y O1 en P2 junto con O1 en P3 respectivamente en el Ejemplo 8-. Luego de identificadas las oraciones que componen el alcance de los marcadores encontrados, se realiza el mismo estudio que en la proposición 10.

Nota: Siempre que se decida marcar oraciones de un argumento como parte del resumen final, se deberá realizar el estudio de coherencia de cada una de ellas. Al ingresar las oraciones al resumen, se debe tener en cuenta el porcentaje de resumen ingresado por el usuario.

3.2. Etapas del resumen.

La representación de las etapas de procesamiento a las cuales es sometido el texto fuente se puede apreciar en la Figura 2:

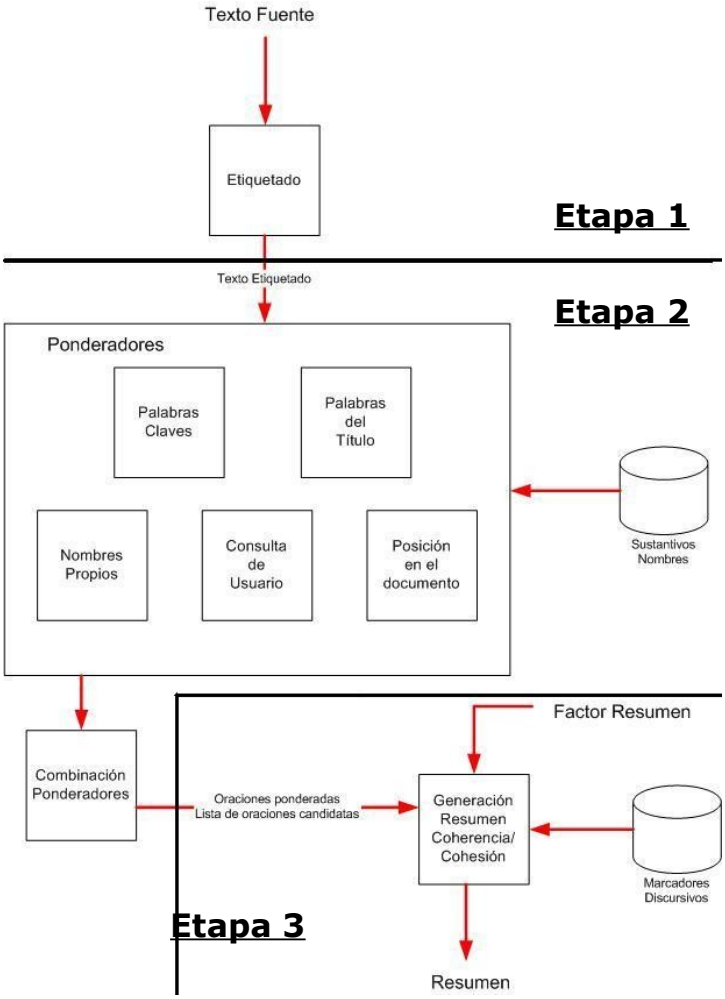

Figura 2 – Etapas del proceso

En primera etapa se realiza el reconocimiento del texto (oraciones, párrafos, marcadores discursivos, etc.). En la segunda etapa se realiza la ponderación de las oraciones (se le da cierto puntaje a cada una de ellas de acuerdo a un conjunto de métodos). Por último, en la tercera etapa se intenta dar coherencia al resumen mediante el uso de los marcadores discursivos que aparecen en el texto.
A continuación se describirán las diferentes etapas y las decisiones que se tomaron para este trabajo.

3.2.1. Reconocimiento del texto.

La principal entrada al resumidor es el texto en sí. Es por esto que la lectura e interpretación de dicho texto juega un papel fundamental e imprescindible en el sistema. La condición que debe cumplir esta entrada es la de ser texto plano.

El primer problema que se nos plantea es el reconocimiento de dicho texto, una forma de interpretar la entrada y poder reconocer palabras, marcadores discursivos, oraciones, párrafos, signos de puntuación, etc. Se evaluó la utilización de la herramienta Freeling[15] pero debido a que se debía "levantar" o interpretar el texto etiquetado por dicha herramienta y a problemas de instalación, se optó por usar como herramienta de reconocimiento de texto un generador de analizadores lexicográficos llamado JLex[14] que permite el reconocimiento e interpretación del texto. Dicha herramienta ejecuta junto con Java y ya había sido utilizada por el grupo anteriormente.

Inicialmente se realizaron un conjunto de pruebas para corroborar la posible utilización de JLex[14]. Estas pruebas dieron resultados alentadores por lo cual se decidió utilizar esta herramienta para la lectura del texto fuente.

3.2.1.1. Lectura del texto

Se toma una oración como una secuencia de caracteres que comienza con una letra mayúscula y termina con un punto (carácter '.'). Un párrafo se toma como un conjunto de oraciones seguidas del carácter 'enter' (carácter '\n'), o sea, la ocurrencia de un "punto y aparte".

El analizador lexicográfico (archivo con extensión 'lex') reconoce oraciones y las almacena en una colección como un string (texto) junto a un número de oración y número de párrafo para luego ser procesadas (armado de la estructura). Los elementos de la colección de oraciones son de la forma:

[número de oración, número de párrafo, texto de la oración]

El número de párrafo de las oraciones aumenta con la aparición del carácter 'enter' ('\n').

Uno de los problemas que se tuvo durante esta implementación fue la falta de reconocimiento de algunos caracteres, por lo que las oraciones quedaban cortadas e inentendibles. Este problema fue reapareciendo a medida que se realizaban pruebas del reconocedor de textos con distintos archivos de entrada, por lo cual se fue complementando el reconocedor para los demás caracteres.

Otro problema fue el de la asignación del número de párrafo, ya que la aparición de dos caracteres 'enter' ('\n') seguidos, llevaba al aumento del número de párrafo en dos unidades, por lo que la siguiente oración reconocida (primera oración del siguiente párrafo) quedaría con número de párrafo incorrecto. También se dan casos donde la aparición de un carácter punto ('.') no indica la finalización de una oración (abreviaciones, formatos de hora, etc.), lo que lleva al rearmado de las oraciones.

Existirá una etapa de post-reconocimiento o procesamiento del texto, que se ocupa de reconocer sustantivos, nombres, solucionar el problema del número de párrafo en las oraciones, realizar el rearmado de las oraciones, etc., esto se verá en la sección .

3.2.1.2. Lectura de marcadores discursivos.

Dado que el resumidor de textos intenta darle coherencia al resumen basándose en la función que cumplen los marcadores discursivos, será necesario localizar en el texto las apariciones de los distintos marcadores.
Al igual que en la sección , para el reconocimiento de los marcadores, también se utiliza JLex[14].

Se tiene un conjunto inicial de marcadores discursivos el cual puede ser ampliado por el usuario. Todos los marcadores que forman este conjunto son incluidos en un nuevo archivo 'lex' que es armado dinámicamente cada vez que el usuario agrega o quita marcadores del conjunto.

Para este trabajo, solo se toman en cuenta ocurrencias de marcadores que inician una oración, dado que la extracción de oraciones para el resumen se hace de forma completa (se extraen oraciones enteras) y la aparición de un marcador discursivo en medio de una oración no influirá en la coherencia del resumen. Es por esto que se dejan de lado los marcadores discursivos que no comiencen una oración (Ejemplo 11).

En la mayoría de los casos puede necesitarse que el marcador discursivo este seguido del carácter ',' ("coma") para desambiguar posibles situaciones en donde una palabra o un conjunto de éstas puede no estar cumpliendo la función de marcador discursivo. La necesidad de este carácter puede ser definida por el usuario y ya se encuentra definida en el conjunto inicial de marcadores discursivos para cada uno de ellos basándose en el

trabajo de Prada[1]. En dicho trabajo se menciona también que hay marcadores que no presentan ambigüedad, por lo tanto la única condición para que sean reconocidos es que comiencen una oración (ejemplo: "No obstante").

Ejemplo 9

"*Además de tener buenas condiciones físicas, tiene una mentalidad brillante…*"

a b

Podemos ver en el Ejemplo 9 que la ocurrencia de la palabra "Además" no esta cumpliendo la función de marcador discursivo con la oración anterior, sino que actúa como conjunción de las dos proposiciones (a y b).

Cuando se reconoce un marcador, este se agrega a una colección el número de oración, número de párrafo (se lleva un contador de oraciones y párrafos) y cantidad de argumentos (valor obtenido del conjunto de marcadores discursivos). O sea, la colección de marcadores tendrá elementos de la forma:

[número de oración, número de párrafo, cantidad de argumentos]

Cuando finaliza el proceso de reconocimiento de marcadores, se tiene una colección con todas las ocurrencias de estos en el texto en estudio. Mas adelante esta colección es procesada junto al conjunto de oraciones reconocidas en la sección de forma de asignar los marcadores reconocidos a las oraciones a las cuales pertenecen.

Ejemplo 10

"***Pero,*** *la explosión de mayor violencia, se dio recién cuando llegaron a la esquina de Sarandí y Misiones.*"
Diario La República – Montevideo, Uruguay – 05/11/2005

En este caso se agrega a la colección de marcadores un marcador con número de oración y número de párrafo igual a los contadores (de oraciones y párrafos) utilizados y un valor de la cantidad de argumentos del marcador reconocido (en este caso la cantidad es 2 - binario). Si la oración presentada es la número 4 del párrafo 5, entonces se agrega a la colección de marcadores la siguiente elemento: [4,5,2].

Ejemplo 11

"*Estima,* ***además,*** *que su tecnología permite construir sistemas para generar hasta 500 megavatios. Dependiendo de los ríos se pueden instalar turbinas flotantes en serie a una distancia adecuada entre ellas.*"
http://www.paginadigital.com.ar/articulos/2005/2005terc/noticias9/energia-barata-280106 - 28/01/2006.

En este caso el marcador intraoracional "además" no es reconocido por el analizador lexicográfico ya que no está al comienzo de la oración, por lo tanto, no será ingresado en la colección de marcadores discursivos. Se puede ver en este ejemplo que se deja de lado hacer un análisis de incisos (Ejemplo 11), donde el término "además" cumple la función de marcador discursivo.

3.2.1.3. Armado de la estructura.

Luego de la lectura del texto y de los marcadores discursivos mediante los analizadores lexicográficos, se procede al armado de la estructura (etapa de post-reconocimiento o procesamiento). Dicho armado consiste en crear un texto formado por un conjunto de párrafos, estos últimos formados por un conjunto de oraciones. Las oraciones contarán

además con un conjunto de palabras las cuales se distinguirán entre nombres propios o sustantivos, también se distinguirá entre oraciones que comienzan con marcador discursivo o no.
A medida que se van formando las oraciones se van creando los párrafos. A éstos se le asocia un número único secuencial que los identifica y el conjunto de oraciones que los componen (oraciones con igual número de párrafo corresponden al mismo párrafo). Luego de haber creado todos los párrafos, se crea el texto formado por el conjunto de párrafos reconocidos.
En la Figura 3, se puede apreciar como queda conformada la estructura luego del procesamiento:

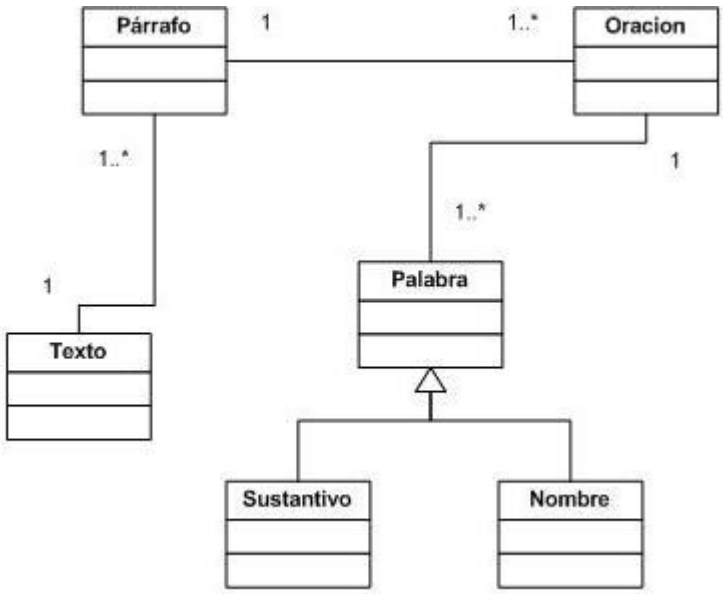

Figura 3 – Estructura del texto

La estructura anterior es fundamental para todo el proceso de resumen y para el proceso de coherencia, siendo además, el resultado (salida) de esta etapa.
En este proceso de armado del texto -armado de la estructura- se parte de la colección de oraciones y de la colección de marcadores que resultan de los puntos de Lectura del texto (sección) y Lectura de marcadores discursivos (sección). Se analizan una a una las oraciones interpretando que las oraciones que tienen igual número de párrafo pertenecerán al mismo párrafo del texto.
Para el posterior procesado del texto (confección del resumen) se asume que si el primer párrafo del texto (párrafo con número 1) tiene una única oración, entonces se lo toma como título.
En los siguientes puntos se detalla la etapa de post-reconocimiento o procesamiento para solucionar algunos "problemas" introducidos en la etapa de lectura y para identificar los nombres y sustantivos de cada oración de la estructura.

- ***Números de párrafo.***

Uno de los problemas que se genera en la lectura del texto (sección) es la posibilidad de obtener números de párrafo erróneos. Este problema surge ya que en el analizador lexicográfico se aumenta el número de párrafo cada vez que se encuentra un carácter enter ('\n'). Lo que se realiza para dar una solución es aumentar de a una unidad el número de párrafo cada vez que se produce un cambio de dicho número en las oraciones que se están procesando.
Ejemplo: si se procesó una oración con número de párrafo 5 y la siguiente oración que se procesa tiene número de párrafo 7 (existen dos 'enters' entre ambas oraciones), se corrige el número de la última oración para que tenga el número 6.

Esta corrección se hace para tener números de párrafo consecutivos de forma que se refleje lo más posible la realidad, y porque serán utilizados en módulos que necesitan tener el correcto número de párrafo.

- ***Rearmado de oraciones.***

El rearmado de oraciones soluciona algunos casos donde la aparición del carácter '.' no indica el fin de una oración. Estos casos fueron sucediendo cuando se utilizaron diferentes textos fuentes para el testeo del analizador lexicográfico.
A continuación se describen algunas situaciones que se descubrieron y forzaron a encontrar una solución al problema:

1 - Abreviaciones

Ejemplo 12

"*Acto seguido, un pequeño grupo, que se reunió en el lugar, comenzó a entonar consignas agresivas hacia el "régimen imperialista y el presidente de EEUU,* ***George W. Bush****", y tras unos pocos minutos se les sumaron varias decenas de mujeres, niños, jóvenes, muchos de los cuales traían sus rostros cubiertos.*"
Diario La República – Montevideo, Uruguay – 02/11/2005

Aquí se puede ver la aparición de un nombre abreviado. La salida del analizador lexicográfico devuelve dos oraciones: **"** "*Acto seguido, un pequeño grupo, que se reunió en el lugar, comenzó a entonar consignas agresivas hacia el "régimen imperialista y el presidente de EEUU, George W.*" **y** **"** *Bush", y tras unos pocos minutos se les sumaron varias decenas de mujeres, niños, jóvenes, muchos de los cuales traían sus rostros cubiertos.***"**

La decisión que se toma es que si la oración anterior a la que se está analizando termina con una palabra de una sola letra (un caracter), la oración en estudio se "unirá" a la oración anterior. Al realizar esta "unión", se recalcula el número total de palabras de la oración formada.

Ejemplo 13

"*Entre tanto, hoy a las 12:30 horas, los detenidos concurrirán ante el juez de* ***1er. turno****, doctor Fernández Lecchini, quien se hizo presente en el lugar de los hechos y constató los destrozos ocasionados por los revoltosos.*"
Diario La República – Montevideo, Uruguay – 02/11/2005

Aquí se presenta un ejemplo de abreviación en donde el analizador lexicográfico devuelve dos oraciones: **"***Entre tanto, hoy a las 12:30 horas, los detenidos concurrirán ante el juez de 1er.***"** **y** **"***turno, doctor Fernández Lecchini, quien se hizo presente en el lugar de los hechos y constató los destrozos ocasionados por los revoltosos.***"**

La decisión que se toma aquí es que si la oración que se está analizando comienza con una palabra en letra minúscula (primer carácter de la primera palabra en minúscula) o si la oración comienza con un símbolo de puntuación "," o ";" (para tener en cuenta casos como: "*…fondo, etc.; que…*"), entonces la oración en estudio se "unirá" a la oración anterior. Al realizar esta "unión", se recalcula el número total de palabras de la oración formada.
Un posible problema que puede surgir en este tipo de abreviaciones se da en el caso que la palabra "turno" (en el Ejemplo 13) aparezca en mayúscula. En estos casos no es posible distinguir fácilmente que estamos frente a una abreviación, por lo que se tomará como dos oraciones distintas.

2 - Formatos de hora

Ejemplo 14

"*Por su parte, Buquebus emitió un comunicado en el que explica que el buque Eladia Isabel había partido desde el puerto de Colonia del Sacramento a las* ***19.30 horas****."*
Diario La República – Montevideo, Uruguay – 21/01/2006

Se puede observar la ocurrencia de un carácter '.' (punto) el cual no indica la finalización de una oración. El analizador lexicográfico devuelve dos oraciones: **"***Por su parte, Buquebus emitió un comunicado en el que explica que el buque Eladia Isabel había partido desde el puerto de Colonia del Sacramento a las 19.* **" y "***30 horas.***".**

La decisión que se toma es que si la oración en estudio comienza con un dígito, ésta se "une" a la oración anterior. Al realizar esta "unión", se recalcula el número total de palabras de la oración formada. Esta solución se contrapone a oraciones que efectivamente comienzan con un número, pero vemos mas lejana esta posibilidad.

3 - Puntos suspensivos.

Ejemplo 15

"*Fue un militante por la vida que demostró con hechos en más de una ocasión su apoyo y solidaridad con procesos de transformación en el continente, y con el bolivariano especialmente, por lo que es considerado un amigo de nuestro pueblo (...), queremos darle el adiós a alguien que ha dejado una huella indeleble en nuestros corazones.*"
Diario La República – Montevideo, Uruguay – 21/01/2006.

La aparición de los tres puntos seguidos ("…") lleva a que el tokenizador reconozca más de una oración. La primer oración reconocida va desde el inicio de la oración (palabra "Fue") hasta el primer punto. Luego se reconocen dos oraciones formadas nada mas que por un punto (".") y por último una oración desde el segundo paréntesis (")") hasta el punto final.
Una de las decisiones que se toma es que si una oración consta de un punto (podría tener espacios antes), se unirá a la oración anterior. En el Ejemplo 15, los puntos suspensivos se unirán a la oración anterior.
La otra decisión es que si una oración consta únicamente de un punto, la siguiente oración se unirá a ésta si y solo si no comienza con una letra mayúscula. En el Ejemplo 15, la oración "), queremos darle el adiós a alguien que ha dejado una *huella indeleble en nuestros corazones.*", se unirá a la anterior.
En caso de que la siguiente oración comience con mayúscula esta no se unirá al anterior pues se la considerará como una nueva oración.

Nota: Todas estas decisiones son tomadas sólo si la oración en estudio no es la primera oración del párrafo. Se asume que no es posible "unir" una oración (primera del párrafo) con la última oración del párrafo anterior.

- ***Reconocimiento de marcadores discursivos.***

A medida que se realiza el análisis de las oraciones, una de las tareas a realizar es el chequeo de los marcadores discursivos, o sea, verificar si la oración en estudio comienza con un marcador discursivo o no.
Para esto, se toma el número de oración y número de párrafo de la oración en estudio y se busca la aparición de un marcador con igual número de oración y párrafo en la colección de marcadores discursivos creada en la etapa de lectura de marcadores (sección). Esta búsqueda se realiza en este momento para mantener la correlación entre los números de oración y párrafo, ya que el reconocimiento de marcadores discursivos por parte de JLex[14] tiene el mismo defecto que el reconocimiento de

oraciones y párrafos (números no consecutivos de párrafo), y además para que la "unión" de oraciones no influya en este proceso.
En caso de encontrar un marcador, se marca la oración en estudio dentro de la estructura para indicar que comienza con marcador, y se le asigna como dato a la oración un valor igual a la cantidad de argumentos del marcador (unario, binario, compuesto). Estos valores serán de suma importancia para el posterior proceso de resumen del texto, ya que juegan un papel determinante en la etapa de coherencia del resumen.

Ejemplo 16

"***Pero,*** *la explosión de mayor violencia, se dio recién cuando llegaron a la esquina de Sarandí y Misiones.*"
Diario La República – Montevideo, Uruguay – 05/11/2005

Si suponemos que esta oración es la número 4 del párrafo 5 entonces la colección de oraciones generada en la lectura del texto contendría este elemento: [4, 5, "Pero, la explosión de mayor violencia, se dio recién cuando llegaron a la esquina de Sarandí y Misiones."]; y la colección de marcadores generada en la etapa de lectura de marcadores discursivos contendría este elemento: [4, 5, 2] (oración 4, párrafo 5, marcador de 2 argumentos).
Cuando se estudia esta oración y se chequea si comienza con marcador, se obtiene un resultado positivo, marcando la oración como que comienza con marcador y asignándole también la cantidad de argumentos de éste (en este caso 2).

Ejemplo 17

"*El pesquero de bandera argentina no sufrió muchos daños y contaba con una tripulación de 29 marinos.*"
Diario La República – Montevideo, Uruguay – 21/01/2006

Si suponemos que esta oración es la número 2 del párrafo 5, entonces la colección de oraciones generada en la etapa de lectura del texto contendría este elemento: [2, 5, "El pesquero de bandera argentina no sufrió muchos daños y contaba con una tripulación de 29 marinos."].
Cuando se estudia esta oración y se chequea si comienza con marcador discursivo, se obtiene resultado negativo, por lo que se marca la oración como que no comienza con marcador.

- ***Reconocimiento de nombres y sustantivos.***

Una vez creado el conjunto de párrafos que conformarán el texto y los conjuntos de oraciones que forman los párrafos, se procede al reconocimiento de los nombres y sustantivos de cada oración.
Se tiene una base de datos con más de 5900 nombres y más de 6600 sustantivos[1], donde se chequeará si cada palabra del texto se corresponde con un sustantivo o nombre de esta base de datos.
Se toma como convención que los nombres en el texto deben aparecer con la primera letra en mayúscula.
Para cada oración del texto se realiza el siguiente algoritmo:

```
Para cada palabra
   Si la palabra comienza con mayúscula
      La busco en la base de nombres
      Si la encontré
         Es_nombre = true
```

[1] Los nombres y sustantivos se fueron obtuviendo de diferentes fuentes: otros proyectos de grado, Internet, ingresados manualmente a través del sistema.

```
                    Si puedo unir al nombre anterior (nombre compuesto se toma
            como único nombre)
                Concateno con nombre anterior
            Si no
                Creo una nueva palabra (nombre) en la oración
            Fin si
        Si no
            Es_nombre = false
    fin si
    Si Es_nombre = false
      busco la palabra en la base de sustantivos
      Si es sustantivo
        Creo una nueva palabra (sustantivo) en la oración
      Fin si
    fin si
fin para
```

Se toma cada palabra (desde un carácter espacio (‘ ’) a otro) y se le quitan los caracteres que no sean letras (por ejemplo: a la palabra "*hospital?*" se le quita el carácter ‘?’) quedando la palabra "*hospital*", esto es para poder compararla con un nombre o sustantivo de la base de datos.
Cada nombre o sustantivo de cada oración tendrá almacenado la cantidad de ocurrencias dentro de la oración. Cuando se identifica un nombre o sustantivo, se debe verificar si ya esta incluida en la colección de palabras de la oración, si no está, se agrega a la colección de palabras con cantidad de ocurrencias igual a uno; en otro caso, se incrementa en uno la cantidad de ocurrencias de la palabra en la oración.
Cabe destacar que primero se chequea si la palabra corresponde a un nombre y luego a un sustantivo, por lo que en el caso de palabras que son nombres y sustantivos, la aparición de una de estas palabras con la primer letra en mayúscula se tomará como nombre.

Para el reconocimiento de nombres se toma en cuenta la posible ocurrencia de dos o mas nombres seguidos sin signos de puntuación entre ellos (Ej. nombre y apellido de una persona), en cuyo caso, son tomados como uno solo.
También se asume como convención que si aparecen en la oración nombres separados por algunas palabras que los unen, se está frente a un único nombre. Las palabras que se toman como "uniones" de nombres son: "de", "De", "del", "Del" y abreviaciones de nombres, o sea, una letra en mayúscula seguida de un punto.
Se debe tener en cuenta que para que se reconozca en el texto la unión de nombres, deben aparecer en la base de datos de nombres cada nombre por separado (los siguientes ejemplos explican esto último). Cabe destacar que al momento de calcular la cantidad de palabras de una oración se contabilizan los nombres compuestos como una única palabra, o sea, se calcula como: total de palabras de la oración menos cantidad de uniones de nombres. A continuación se presentan una serie de ejemplos donde se muestran los problemas antes mencionados.

Ejemplo 18 (nombres)

"*El ministro del Interior, José Díaz, repudió los hechos y anunció que aplicará la ley "con todo rigor".*"
Diario La República – Montevideo, Uruguay – 05/11/2005

En este caso se reconocen dos nombres en la oración: "Interior" y "José Díaz". Se puede notar que se unen los nombres "José" y "Díaz" (cada uno debe aparecer en la base de datos de nombres) por encontrarse separados nada mas que por el carácter espacio (" "). En cambio, el nombre "Interior" es separado del nombre "José" ya que se encuentra entre ellos el carácter de puntuación "," (coma).
La cantidad de palabras de la oración es 17 = 18 (total de palabras) – 1 (uniones de nombres).

Ejemplo 19 (nombres)

"*Hubo quienes "grafitearon" consignas antiimperialistas, contrarias al presidente norteamericano* ***George W. Bush*** *e impactaron "bombas de alquitrán" y de pintura roja en cuanta pared o muro por donde pasaban, incluyendo a una dependencia de la Armada.*"
Diario La República – Montevideo, Uruguay – 05/11/2005

En este ejemplo se reconoce el nombre "George W. Bush" como un único nombre, se realiza la unión de "George", "W." y "Bush" ya que se toma "W." como una abreviación de un nombre.

Ejemplo 20 (nombres)

"*Por su parte, Buquebus emitió un comunicado en el que explica que el buque Eladia Isabel había partido desde el puerto de Colonia del Sacramento a las 19.30 horas.*"
Diario La Republica – Montevideo, Uruguay – 21/01/2006

En esta oración se reconocen los siguientes nombres: "Buquebus", "Eladia Isabel" y "Colonia del Sacramento". Podemos ver la "unión" de los nombres "Eladia" e "Isabel" (ambos deben aparecer en la base de datos). También se unen los nombres "Colonia" y "Sacramento" (ambos deben aparecer en la base de datos) junto a la palabra "del" por la convención mencionada anteriormente.

Ejemplo 21 (sustantivos y nombres)

"*Lejos de las imágenes de Copacabana en Río de Janeiro, lejos de los clichés de playas paradisíacas del Caribe o de las imágenes brumosas de las postales del Machu Pichu en Perú, 98 millones de seres humanos duermen cada día en la calle.*"
http://www.paginadigital.com.ar/articulos/2006 - 29/01/06

Estas son las palabras que son reconocidas por el sistema, también se muestra la cantidad de ocurrencias de cada una.

SUSTANTIVO: imágenes	Cantidad: 2
NOMBRE: Río de Janeiro	Cantidad: 1
SUSTANTIVO: playas	Cantidad: 1
NOMBRE: Caribe	Cantidad: 1
SUSTANTIVO: postales	Cantidad: 1
NOMBRE: Machu Pichu	Cantidad: 1
NOMBRE: Perú	Cantidad: 1
SUSTANTIVO: seres	Cantidad: 1
SUSTANTIVO: humanos	Cantidad: 1
SUSTANTIVO: día	Cantidad: 1
SUSTANTIVO: calle	Cantidad: 1

Figura 4 – Reconocimiento de nombres y sustantivos

El reconocimiento de nombres y sustantivos juega un papel muy importante dentro del sistema, ya que es la base para el funcionamiento del resumidor. Más adelante se explicará el rol de estas palabras.

Problemas detectados.

- Problema 1 - Ejemplo de nombres

"*…un par de horas antes que otra marcha repudiara la presencia del presidente de EEUU George Bush en Mar del Plata.*"
Diario La República – Montevideo, Uruguay – 05/11/2005

En este ejemplo surge el problema de la aparición contigua de nombres que no están formando un nombre único. Se reconoce el nombre "EEUU George Bush" el cual debería tomarse como dos nombres: "EEUU" y "George Bush".

- Problema 2 - Ejemplo de sustantivos 1

""*Vino corriendo uno, me dio un empujón que me tiró al piso…*""
Diario La República – Montevideo, Uruguay – 05/11/2005

En este ejemplo se reconoce la palabra "Vino" como sustantivo, pero se puede observar que dicha palabra en este caso actúa como verbo.
Dicho problema surge con ocurrencias de palabras que son sustantivos y verbos (en el texto cumplen una única función), las cuales aunque estén cumpliendo la función de verbos, son reconocidas como sustantivos. Se necesitaría un estudio del contexto donde se encuentra la palabra para decidir que función esta cumpliendo o utilizar un "Tagger" que incluya una solución al problema.

3.2.2. Resumen de texto

En el punto se explica como obtener un resumen por ponderación de un texto. Las mismas oraciones que conforman el resumen por ponderación serán las oraciones candidatas a formar el resumen por coherencia lo cual se explica en el punto .

3.2.2.1. Resumen según ponderación de las oraciones.

Como se presenta en el capítulo 2, la técnica de resumen utilizada es la de extracción de oraciones. En esta técnica se ponderan las oraciones del texto de acuerdo a un conjunto de heurísticas o métodos. En este trabajo se utilizan los siguientes:

- Ponderación por nombres
- Ponderación por posición
- Ponderación por tf (frecuencia del término)
- Ponderación por palabras del título
- Ponderación por consulta de usuario

La ponderación final de cada oración se obtiene como una combinación lineal de los resultados de todos los métodos aplicados. Los coeficientes de esta combinación lineal pueden ser manipulados por el usuario, de esta forma éste podrá darle más o menos "fuerza" a cada uno de los métodos. Se tiene un porcentaje (valor entre 0 y 100) para cada método (inicialmente todos son 100), los cuales podrán ser alterados por el usuario. Así se podrá por ejemplo "anular" algunos de los métodos asignándole a éstos porcentaje 0.

Luego de haberse calculado la ponderación final de cada oración, son seleccionadas para el resumen final las oraciones que hayan obtenido la mayor ponderación de acuerdo a un porcentaje ingresado por el usuario (resumen por ponderación).

A continuación se describe cada uno de los métodos utilizados en este trabajo.

- ***Ponderación por nombres.***

Dicha heurística pondera las oraciones que contengan nombres propios. Para ello, se asigna un valor fijo w a todos los nombres, luego el peso p_i de la oración i será:

$$p_i = (\text{Total de nombres en oración i} * w) / \text{Total palabras de la oración i}$$

La heurística itera en todos los párrafos y para cada oración debe tomar la cantidad de nombres propios que posee y la cantidad de palabras en total.

Ejemplo 22

"Los Spurs siguen con su racha de victorias, ya que venían de derrotar 95-92 a Milwaukee."
Diario Clarín, Argentina – 21/01/2006

Para esta oración se reconocen dos nombres: "Spurs" y "Milwaukee", ambos con cantidad 1, la cantidad de palabras de la oración es 16. El cálculo de la ponderación para esta oración es el siguiente:

Ponderación = (2 * w) / 16 , por lo que la ponderación sería w/8 = 0.125*w.

Ejemplo 23

"Perdía 1-0 con un gol de Valdemarín, pero Patiño (2) y Loeschbor lo dieron vuelta."
Diario Clarín, Argentina – 21/01/2006

En esta oración se reconocen tres nombres: "Valdemarín", "Patiño" y "Loeschbor". La cantidad de palabras en la oración es 15. El cálculo de la ponderación para esta oración es el siguiente:

Ponderación = (3*w) / 15 ~ 0.2*w.

Ejemplo 24

"*Por su parte el jefe de la Prefectura del puerto de Buenos Aires, José Romero, declaró ayer a Clarín.com que "están investigando las causas del accidente."*
Diario La República – Montevideo, Uruguay – 21/01/2006

En este ejemplo se reconocen los siguientes nombres: "Prefectura", "Buenos Aires" y "José Romero". Notar que los nombres compuestos se toman como un nombre único a la hora de ponderar por lo tanto la oración tendría la siguiente ponderación:

Ponderación = (3*w) / 24 ~ 0.125*w.

Se puede notar que la ponderación por nombres de la oración aumenta a medida que: aumenta la cantidad de nombres de la oración y/o disminuye la cantidad total de palabras de la oración.

- ***Ponderación por posición.***

Este método se basa en los trabajos de Mateo[5] y Acero[6] descriptos en la sección . Dicha heurística pondera las oraciones del primer y último párrafo por considerar que son importantes ya que el primero proporciona un panorama general del texto y el último proporciona conclusiones o resultados. Esta idea es expresada por muchos autores.
Para cada una de las oraciones de un párrafo (primero y último), se les da una ponderación de acuerdo a la siguiente expresión:

$$pond_i = w-((numOra-1) * descuento)$$

donde numOra es el número de la oración, w y descuento son valores fijos, hasta que $pond_i$ sea igual a cero o hasta que se terminen las oraciones del párrafo.
Notar que para la primer oración del párrafo el valor de la ponderación es igual a w.

Ejemplo 25

"El barco Eladia Isabel de la empresa Buquebus recibió en la madrugada de ayer un fuerte impacto por parte de un barco pesquero de bandera argentina denominado "Depemás 5". Los daños no fueron importantes, pero Buquebus realizará una investigación al respecto. Altas fuentes consultadas indicaron que el accidente se produjo por una "mala maniobra" del pesquero. Llama la atención de expertos que el pesquero argentino se pusiera delante de la proa del barco de pasajeros."
Diario La República, Uruguay – 21/01/2006

Siendo este el primer párrafo del texto el cual consta de 4 oraciones la ponderación para cada una de las oraciones según este método es la siguiente:

Oración 1: ponderación = w
Oración 2: ponderación = w - descuento
Oración 3: ponderación = w – (2 * descuento)
Oración 4: ponderación = w - (3 * descuento)

- ***Ponderación por palabras del título.***

Esta heurística pondera con mayor valor a las oraciones que contengan palabras del título. Como se mencionó en la sección se considera que las palabras que se encuentran en el título son importantes y por tanto darán mayor importancia a las oraciones que las contengan (solo se toman en cuenta sustantivos y nombres). Esta heurística pierde validez en caso de que el texto a resumir no contenga título.
Se asigna un valor fijo w a todas las palabras[1] del título, luego la ponderación p_i de la oración i será:

p_i = Total de palabras[1] del título en oración i * w / Total palabras de la oración i

Donde *palabras*[1] hace referencia a sustantivos y nombres.

Ejemplo 26

Si tenemos un texto con el titulo: "Ginóbili, lesionado en la victoria de los Spurs.". Y la siguiente oración del texto:

"En el segundo parcial, Ginóbili pudo romper el cero y convirtió un de dos en dobles y uno de cuatro en tiros libres."
Diario Clarín, Argentina – 21/01/2006

La cantidad de palabras de la oración es 23 y se reconoce en la oración el nombre "Ginóbili" (aparece en el título), por lo que la ponderación de la oración será:

Ponderación = (1 * w) / 23 = 0,044*w

Ejemplo 27

Si tenemos un texto con el título: "El pesquero argentino embistió al Eladia Isabel.". Y la oración:

"Un pesquero habría rozado al Eladia Isabel, de Buquebus."
Diario La República, Uruguay – 21/01/2006

La cantidad de palabras en esta oración es de 9, se reconocen 2 palabras del título: un sustantivo ("pesquero") y un nombre ("Eladia Isabel"). La ponderación por palabras del título para esta oración será:

Ponderación = (2 * w) / 9 = 0,22*w

Podemos notar que la ponderación por título de la oración aumenta a medida que: aumenta en la oración la cantidad de palabras del título, disminuye la cantidad total de palabras en la oración en estudio.

- ***Ponderación por consulta de usuario***

En esta heurística se permite al usuario indicar su preferencia a la hora de hacer el resumen. Para esto se le permite ingresar un conjunto de palabras[1] (nombres y sustantivos) para "guiar" el resumen a ciertos temas de su interés. Se asigna un valor fijo w a todas las palabras[1] ingresadas por el usuario, luego el peso p_i de la oración i será:

p_i = (Total de palabras[1] de usuario en oración i * w) / Total palabras de la oración i

Donde *palabras[1]* hace referencia a sustantivos y nombres.
En caso de que los sustantivos o nombres ingresados por el usuario no estén en la base de datos, el sistema los ingresa en sus respectivas tablas para que sean reconocidas la próxima vez que se realice un resumen. No serán tenidos en cuenta para el resumen que se esta realizando, sino que se debe volver a abrir el archivo fuente (archivo a resumir) para reconocer en la estructura las nuevas palabras ingresadas. Esto último es debido a que si el nombre o sustantivo no estaba ingresado en la base de datos, entonces no pudo haber sido reconocido cuando se abrió el texto fuente (ocasión en que se construye la estructura).
Cuando el usuario ingresa sustantivos o nombres que no estén en la base de datos, estas se ingresan en la base pero quedan pendientes de la aprobación de un supervisor o usuario administrador. Esto se hace para evitar el ingreso de palabras malintencionadas o malformadas.
Esta heurística pierde validez en caso de que el usuario no ingrese preferencias para el resumen.

Ejemplo 28

"Llama la atención de expertos que el pesquero argentino se pusiera delante de la proa del barco de pasajeros."
Diario La Republica, Uruguay – 21/01/2006

La cantidad total de palabras de la oración es 19. En caso de que el usuario haya ingresado las palabras "pesquero" y "barco" como preferencias, la ponderación sería la siguiente:

Ponderación = (2 * w) / 19 ~ 0,105*w

Ejemplo 29

"Al declarar, varios de los "muy" jóvenes detenidos no tenían idea de porqué participaron de este hecho."
Diario La Republica, Uruguay – 05/11/2005

Esta oración tiene un total de 17 palabras. Si el usuario ingresa como preferencias la palabra "jóvenes", la ponderación seria la siguiente:

Ponderación = (1 * w) / 17 ~ 0,059*w

Podemos notar que la ponderación por consulta de usuario aumenta a medida que: aumenta en la oración la cantidad de palabras que pertenecen al conjunto de palabras de preferencias del usuario y/o disminuye la cantidad total de palabras en la oración.

Si el usuario ingresa nombres compuestos, estos se tomarán por separados para ingresarlos en la base de datos (en caso de que no estén) y también para realizar la ponderación por consulta de usuario.

Ejemplo 30

"*Por su parte el jefe de la Prefectura del puerto de Buenos Aires, José Romero, declaró ayer a Clarín.com que "están investigando las causas del accidente".*"
Diario La República – Montevideo, Uruguay – 21/01/2006

En este ejemplo, si el usuario ingresa en la consulta el nombre compuesto "José Romero", la ponderación será la siguiente:

Ponderación = (2 * w) / 24 ~ 0,083*w, donde el número 2 (dos) corresponde a los nombres "José" y "Romero" por separado.

Notar que para una oración que sólo contenga el nombre "José" la consulta de usuario también ponderará dicho nombre aunque su ocurrencia no este junto al nombre "Romero".

- ***Frecuencia de las palabras o palabras claves***

Esta heurística se conoce con el nombre de tf (frecuencia de la palabra). Su función es la de darle mayor ponderación a las oraciones que contengan palabras[1] (nombres y sustantivos) cuya frecuencia en el texto es alta.

Se procede de la siguiente forma, en primer lugar se calcula la frecuencia de aparición de cada nombre y sustantivo en el texto como:

F_i = cantidad ocurrencias de la palabra[1] i en el texto / cantidad palabras texto

Donde la cantidad de palabras en el texto es la suma de los nombres y sustantivos reconocidos en todo el texto (sin tener en cuenta el título).

Luego, para calcular la ponderación de cada oración se realiza el siguiente cálculo:

P_i = Sumatoria (cantidad apariciones de palabra[1] p en oración i * F_i)

Donde *palabra[1]* hace referencia a sustantivos y nombres.

Ejemplo 31 (frecuencia de palabra)

Si tenemos un texto con un total de 276 palabras, y una palabra p aparece 8 veces en todo el texto, entonces el valor F_p para la palabra[1] p será de 8/276 = 0.029.

Ejemplo 32 [1]

"Llama la atención de expertos que el pesquero argentino se pusiera delante de la proa del barco de pasajeros."
Diario La Republica, Uruguay – 21/01/2006

[1]Los valores de w mostrados son reales (calculados por el sistema cuando se toma como texto fuente la noticia donde se encuentra la oración del ejemplo).

Si tenemos los siguientes valores de w para las distintas palabras (sustantivos y nombres):

atención – 0.0042
expertos – 0.0042
pesquero – 0.0333
proa – 0.0083
barco – 0.0417
pasajeros – 0.0375

La ponderación del método tf (frecuencia del término) para esta oración será:

1*0.0042 + 1*0.0042 + 1*0.0333 + 1*0.0083 + 1*0.0417 + 1*0.0375 = **0,1292**

Ejemplo 33 [1]

"*El barco Eladia Isabel de la empresa Buquebus recibió en la madrugada de ayer un fuerte impacto por parte de un barco pesquero de bandera argentina denominado "Depemás 5".*"
Diario La Republica, Uruguay – 21/01/2006

Si tenemos los siguientes valores de w para las distintas palabras (sustantivos y nombres):

barco – 0.0417
Eladia Isabel – 0.0292
empresa – 0.0125
Buquebus – 0.0333
madrugada – 0.0083
ayer – 0.0125
fuerte – 0.0083
impacto – 0.0083
parte – 0.0125
pesquero – 0.0333
bandera – 0.0083

Aclaración: Los valores de w mostrados anteriormente son reales (calculados por el sistema cuando se toma como texto fuente la noticia donde se encuentra la oración del ejemplo).

La ponderación del método tf (frecuencia del término) para esta oración será:

2*0.0417 + 1*0.0292 + 1*0.0125 + 1*0.0333 + 1*0.0083 + 1*0.0125 + 1*0.0083 + 1*0.0083 + 1*0.0125 + 1*0.0333 + 1*0.0083 = **0,2499**

Podemos notar que la ponderación por el método tf aumenta cuando crece en la oración la cantidad de palabras que tienen alta frecuencia en el texto.

- ***Combinación de los métodos***

Luego de aplicar cada uno de los métodos antes descritos se calcula la ponderación final de las oraciones. Para cada oración del texto se calcula esta ponderación como una combinación lineal de los resultados de cada método utilizando los porcentajes definidos por el usuario.

[1] Los valores de w mostrados son reales (calculados por el sistema cuando se toma como texto fuente la noticia donde se encuentra la oración del ejemplo).

En este momento se calculó un resumen ponderado del texto fuente. Las oraciones que componen el resumen son las *N* oraciones que han obtenido la mayor ponderación, donde *N* se calcula como:

N = (total de oraciones del texto) * (porcentaje de resumen) / 100

Estas N oraciones colocadas en el mismo orden que en el texto original son las que componen el resumen ponderado, también conforman el conjunto de "oraciones candidatas" que serán la base para el resumen con coherencia.

Por ejemplo, si tenemos un texto con 100 oraciones y el usuario ingresa un porcentaje de 45%, entonces se marcarán las 45 oraciones con mayor ponderación (en orden descendente por ponderación) para el resumen ponderado. Además, estas mismas 45 oraciones serán marcadas como oraciones candidatas para el resumen con coherencia. La oración en la posición 45 según el orden descrito, será la "peor candidata" (la oración candidata con la peor ponderación).

3.2.2.2. Resumen de texto aplicando coherencia

- ***Introducción***

Hasta ahora se han descrito las decisiones tomadas para poder realizar un resumen ponderado, pero se puede observar que con este tipo de resúmenes es posible que se encuentren problemas de coherencia en el texto.
La técnica que se describe e implementa para solucionar dicho problema está basada en la función que cumplen los marcadores discursivos dentro un texto (guiar y ordenar el texto). Si bien no todos los problemas de coherencia del resumen son solucionados con esta técnica, se contribuye con el fin de mantener la coherencia del resumen.
Dado que estos marcadores relacionan segmentos de un texto, se los puede distinguir de acuerdo a la cantidad de segmentos que relacionan.
Este trabajo se concentra en los marcadores del tipo binario (unen dos segmentos del texto) por considerar que son los que mas abundan en los textos en español y a que los marcadores discursivos compuestos (unen más de dos segmentos del texto) son más difíciles de tratar ya que estos segmentos pueden no ser contiguos, incluso se dan casos donde los marcadores de apertura o cierre (que conforman el marcador compuesto) están implícitos, como se explica en Prada[1].

- ***Aplicando coherencia***

Para realizar el estudio de coherencia se comienza por las oraciones mejor puntuadas ("candidatas") propuestas por la etapa de ponderación de las oraciones. Para esto se crea una colección ordenada descendentemente (ver Figura 5) según ponderación de todas las oraciones, identificando que las primeras N oraciones son las "candidatas", donde N fue calculado en el punto "Combinación de los métodos" de la sección .

Nota 1: Las N oraciones "candidatas" para este módulo coinciden con las N oraciones que forman parte del resumen por ponderación.

La colección se representa en la Figura 5:

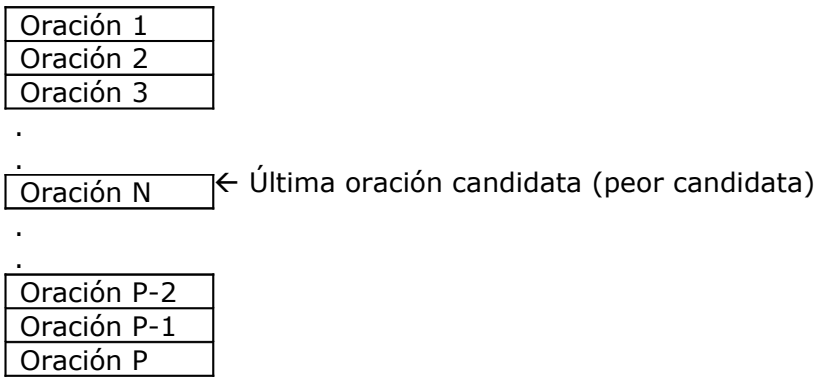

Figura 5 – Oraciones ordenadas por ponderación

Con P igual al total de oraciones del texto fuente y N el lugar de la última oración candidata.

Dada la colección ordenada por ponderación, se estudian una a una las oraciones mientras no se haya completado el resumen y no se termine de recorrer la colección. Cuando se decide que una oración formará parte del resumen final, se la marca para resumen.
Para la oración en estudio se verifica si comienza con marcador discursivo, en caso negativo la oración es marcada para el resumen final; en caso afirmativo se debe contemplar el tipo de marcador, en donde para los tipos unario y compuesto se marca también la oración para el resumen final. En caso de comenzar con un marcador de tipo binario se procede como lo indican los siguientes puntos.

Caso de marcador binario con argumentos contenidos en un mismo párrafo

Este caso se presenta cuando la oración en estudio comienza con marcador y no es la primera del párrafo. Por lo tanto, haciendo referencia a Prada[1], consideramos que el primer argumento es la oración anterior y el segundo argumento es la oración que contiene el marcador discursivo. Para mantener la coherencia del texto, si se decide ingresar la oración que contiene el marcador, entonces se debe incluir también la oración anterior; en caso de no poder ingresar esta última, tampoco se ingresará la oración en estudio. Dentro de este caso se pueden dar las siguientes situaciones:

- La oración anterior ya fue marcada para resumen. Aquí solo se marca la oración en estudio para resumen final.
- La oración anterior es candidata y no esta marcada para resumen. Aquí se decide marcar ambas oraciones para resumen final.
- La oración anterior no es candidata y no esta marcada para el resumen final. Aquí se calcula el promedio de la ponderación de ambas oraciones. En caso de ser mayor dicho valor que la peor ponderación de las oraciones candidatas, ambas oraciones se marcan para resumen final, en caso contrario no se marcan ninguna de las dos oraciones.

En todas las situaciones se ingresarán las oraciones sólo si el porcentaje de resumen lo permite.

Ejemplo 34

"Lo mejor de River llegó siempre desde la derecha y sobre todo, con las subidas de Alvarez. ***De hecho,*** *la más peligrosa del primer tiempo fue tras un centro bajo del chileno, que ni Patiño ni Oberman llegaron a empujar en el área."*
Diario Clarín, Argentina – 21/01/2006

Al analizar la oración que comienza con el marcador discursivo "De hecho", se puede estar en alguna de las tres situaciones anteriores, esto depende de la ponderación que haya obtenido la oración anterior a esta última. Se puede apreciar claramente que ambas oraciones están relacionadas mediante el marcador discursivo, por lo que marcar para el resumen la segunda oración y no la primera, introduciría cierto grado de incoherencia al resumen, puesto que no se sabría que se estaba hablando de "*las subidas de Álvarez*".

Caso de marcador con argumentos en distintos párrafos.

Este caso se presenta cuando la oración en estudio comienza con marcador (binario) y es la primera del párrafo. Por lo tanto, haciendo referencia a Prada[1], consideramos que el primer argumento es el párrafo anterior y el segundo argumento es el párrafo que contiene a la oración en estudio. Para mantener la coherencia del texto, si se decide ingresar la oración que contiene el marcador, entonces se debe incluir al menos una oración del párrafo anterior; en caso de no poder ingresar esta última, tampoco se ingresará la oración en estudio. Dentro de dicho caso se pueden dar las siguientes situaciones:

- En el párrafo anterior hay alguna oración marcada para resumen. En este caso, la oración que contiene al marcador discursivo es ingresada al resumen y se intenta marcar también las oraciones candidatas del párrafo de la oración en estudio; para cada una de estas candidatas, se realiza el estudio de coherencia.
- En el párrafo anterior no hay oraciones marcadas para resumen pero si hay oraciones candidatas. Aquí se intenta marcar las oraciones candidatas del párrafo anterior (argumento 1), en caso de marcar alguna, se marca la oración en estudio para resumen y se intenta marcar las oraciones candidatas del párrafo actual (argumento 2). Para cada una de las oraciones candidatas de ambos párrafos, se realiza el estudio de coherencia. En caso de no marcar ninguna oración del párrafo anterior (argumento 1), la oración que contiene el marcador discursivo tampoco será marcada para resumen.
- En el párrafo anterior no hay oraciones candidatas ni oraciones marcadas para resumen. Aquí se intenta marcar la oración con mejor ponderación del párrafo anterior. Para esto se calcula el promedio de ponderación entre esta oración y las oraciones candidatas del párrafo actual (argumento 2); si el promedio es mejor que la peor ponderación de las oraciones candidatas, se marca la oración en estudio y la mejor oración del párrafo anterior para el resumen. Luego se intentan marcar para resumen las oraciones candidatas del párrafo actual. Para cada una de estas, se debe realizar el estudio de coherencia. En caso de no poder incluir la oración con mejor ponderación del párrafo anterior, no se marca la oración en estudio para el resumen.

En todas las situaciones se ingresarán las oraciones sólo si el porcentaje de resumen lo permite.

Ejemplo 35

"Por su parte, Buquebus emitió un comunicado en el que explica que el buque Eladia Isabel había partido desde el puerto de Colonia del Sacramento a las 19.30 horas. A las 22.20 horas, cuando se encontraba a sólo 6 kilómetros del puerto de Buenos Aires, y por motivos que se desconocen y "son objeto de investigación", un buque pesquero lo impactó en uno de sus laterales. La colisión fue leve, ocasionando solo algunos daños menores en la bodega y algunos sistemas eléctricos.

En consecuencia, *se trabajó en forma conjunta con Prefectura Naval Argentina y dos remolcadores se encargaron de trasladar a ambas embarcaciones hacia el Puerto de Buenos Aires. Afortunadamente, los sistemas eléctricos fueron reparados durante el viaje, lo que posibilitó que el buque amarrara en puerto con sus sistemas propios."*
Diario La República, Uruguay – 21/01/2006

Al analizar la oración que comienza con el marcador discursivo "En consecuencia", se puede estar en alguna de las tres situaciones anteriores, esto depende de la ponderación que hayan obtenido las oraciones del párrafo anterior a esta última. Se puede apreciar claramente que ambos párrafos están relacionados mediante el marcador discursivo, por lo que marcar para el resumen la primera oración del segundo párrafo y ninguna oración del primero, introduciría incoherencias al resumen.
En este ejemplo se da que la primer oración del primer párrafo también comienza con marcador, pero al ser un marcador discursivo unario, se trata la oración como una oración que no comienza con marcador.

Caso combinado

Para una oración en estudio se puede dar que se combinen los dos casos anteriormente descritos. Por ejemplo, se esta estudiando una oración que comienza con un marcador discursivo. Al estudiar la oración anterior, ésta comienza con marcador y además es la primera del párrafo, por lo cual se debe estudiar el párrafo anterior también. Otro ejemplo se presenta cuando se realiza el estudio de coherencia de una oración que comienza con marcador y es la primera del párrafo, al realizar el estudio de las oraciones del párrafo anterior, alguna comienza con marcador discursivo.
En estas situaciones se hace recursión para resolver la combinación de los casos.

Nota 2: Cuando finaliza el estudio de coherencia de una oración, se debe tener en cuenta la cantidad X de oraciones no candidatas que fueron marcadas para resumen. Si esta cantidad es mayor a cero, se selecciona como peor oración candidata a la oración que se encuentra en la posición (posición de peor candidata actual - X) de la lista ordenada de oraciones candidatas (Figura 5).

En la Figura 6 se presenta lo que se describió en la Nota 2 para el caso de que se marque **una** (X=1) oración no candidata para resumen final.

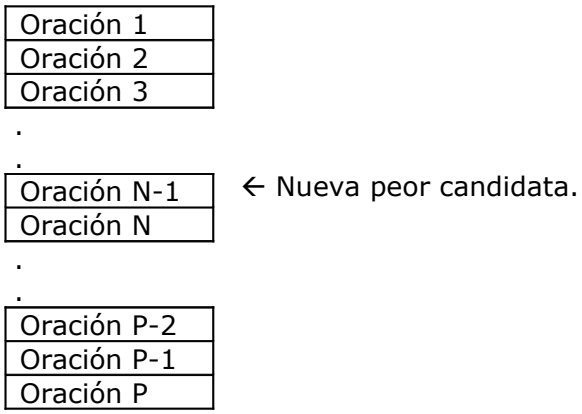

Figura 6 – Cálculo de la nueva oración "peor candidata"

La posición de la nueva peor candidata será: N – X = N - 1

Nota 3: Siempre que se decida marcar oraciones de un argumento como parte del resumen final, se deberá realizar el estudio de coherencia de cada una de ellas. Al ingresar las oraciones al resumen, se debe tener en cuenta que la cantidad de oraciones marcadas para resumen no sobrepase el porcentaje ingresado por el usuario.

3.3. Mantenimiento de la Base de Datos

A continuación se describirán las principales funcionalidades que presenta el sistema para que el usuario pueda mantener y manipular la información de la base de datos.
Los datos que se pueden manipular son:

- Nombres: el usuario podrá ingresar y borrar nombres. Por nombres se entiende que pueden ser nombre de personas, regiones, países, compañías, etc.
- Sustantivos: el usuario puede ingresar y borrar sustantivos.
- Marcadores discursivos: el usuario puede ingresar y borrar marcadores discursivos. En el ingreso, los datos que se piden son: el nombre del marcador, el tipo (unario, binario o compuesto) y si lleva coma o no. Esto último significa si el marcador debe tener un carácter coma (",") luego de su ocurrencia o no.
- Parámetros: el usuario puede modificar los parámetros. Los parámetros son los coeficientes de los ponderadores que se toman en cuenta a la hora de calcular la ponderación total de una oración.

Cuando un usuario ingresa una palabra (sustantivo, nombre o marcador discursivo), dicha palabra se ingresa a la base de datos como "no validada" y podrá ser utilizada por la sesión actual. Al cerrar la sesión las palabras ingresadas estarán inhabilitadas para las siguientes sesiones hasta que el usuario administrador las marque como "revisada" o "validada". Las formas para ingresar una palabra son:

- Directamente por la pantalla destinada al ingreso.
- Cuando se realiza un resumen y se lo quiere personalizar (ponderación por consulta de usuario). En este caso se ingresan solo los sustantivos y nombres que no estén en la base de datos.

Para controlar las palabras que ingresan los diferentes usuarios, un usuario administrador puede "validar" dichas palabras. El usuario administrador tiene una pantalla que le muestra todas las palabras no revisadas y puede aceptarlas (pasando a ser "validada") o si no esta de acuerdo tiene la posibilidad de borrarlas. El usuario administrador (root) es el único que tiene permiso para "validar" las palabras ingresadas y para borrar palabras de la base de datos.
Por más información ver el documento "Manual de usuario".

4. Implementación

En este capítulo se describe a nivel de diseño y decisiones de implementación los principales componentes de la aplicación. En la sección se presenta una breve descripción de la arquitectura utilizada, en las secciones , y se describen cada una de las capas que componen dicha arquitectura (Acceso a datos, Lógica y Presentación respectivamente). En la sección se describe la base de datos utilizada y por último, en la sección se detallada la estructura del texto.

4.1. Decisiones de trabajo – Introducción

La arquitectura del sistema es claramente una arquitectura de tres capas las cuales se mencionan a continuación:

- Acceso a datos (DAO): las clases que están incluidas en esta capa tienen como principal objetivo realizar funciones de comunicación de información con la base de datos (MySQL).
- Lógica: las clases que se incluyen en esta capa tienen como función implementar la lógica de la aplicación (rearmar oraciones, identificar nombres y sustantivos, ponderar oraciones, seleccionar oraciones para el resumen, aplicar coherencia al resumen, etc.).
- Presentación: en esta capa las clases se encargan de implementar la interfaz gráfica o salida visual con la cual interactúa el usuario.

En la Figura 7 se presenta el diseño del sistema estructurado en tres capas junto a sus clases que las implementan.

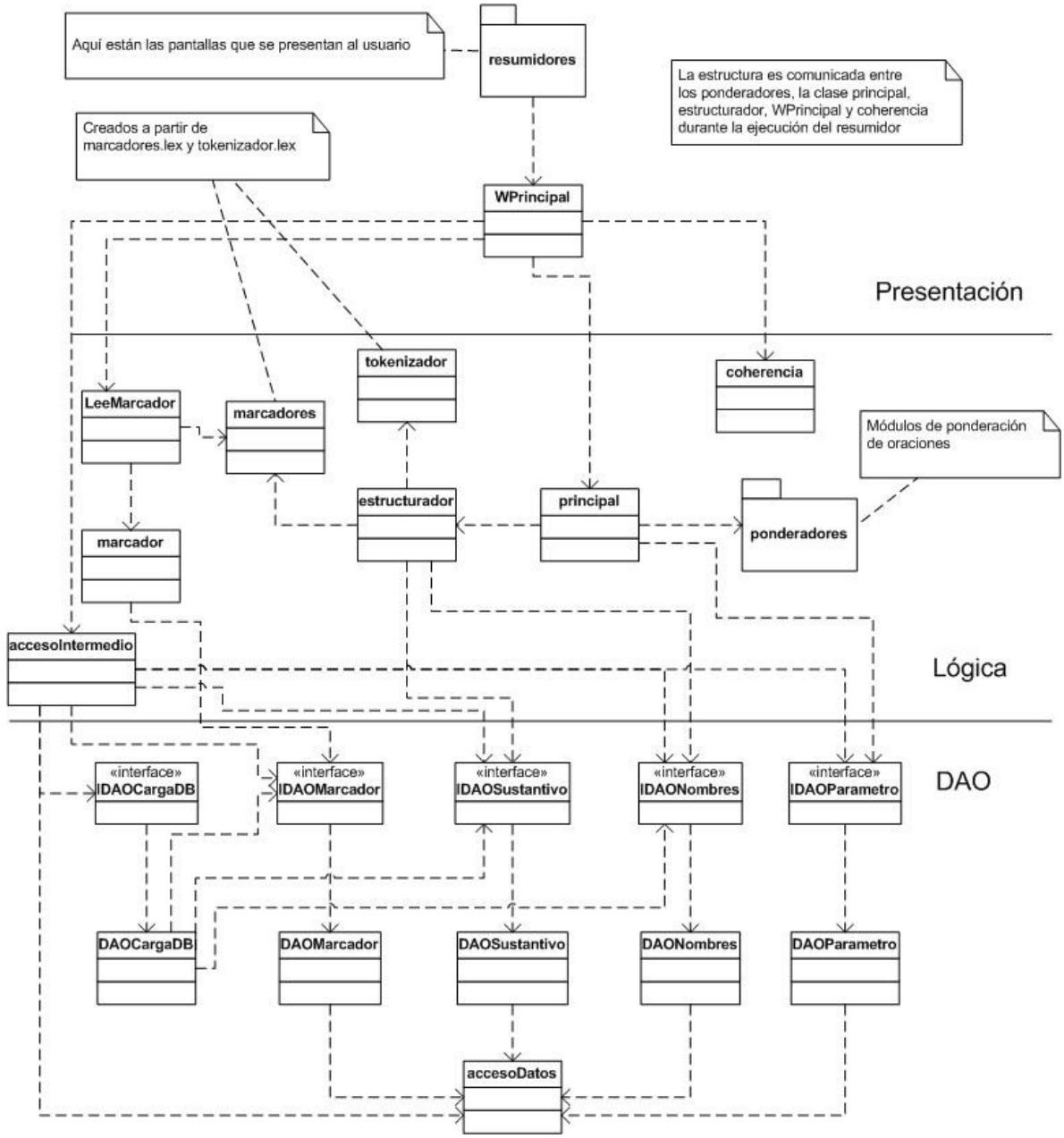

Figura 7 – Diseño del Sistema

A continuación se describirán con más detalle cada una de las tres capas con sus clases y principales funciones u objetivos.

4.2. Capa de acceso a datos (DAO)

Como ya se mencionó, dicha capa se encarga del acceso a la base de datos. Su principal objetivo es permitir la comunicación entre la base de datos y la lógica de la aplicación. Algunas funcionalidades pueden ser: realizar una conexión, acceder a una determinada tabla, etc.
En la Figura 8 se presenta la capa DAO y sus clases en más detalle:

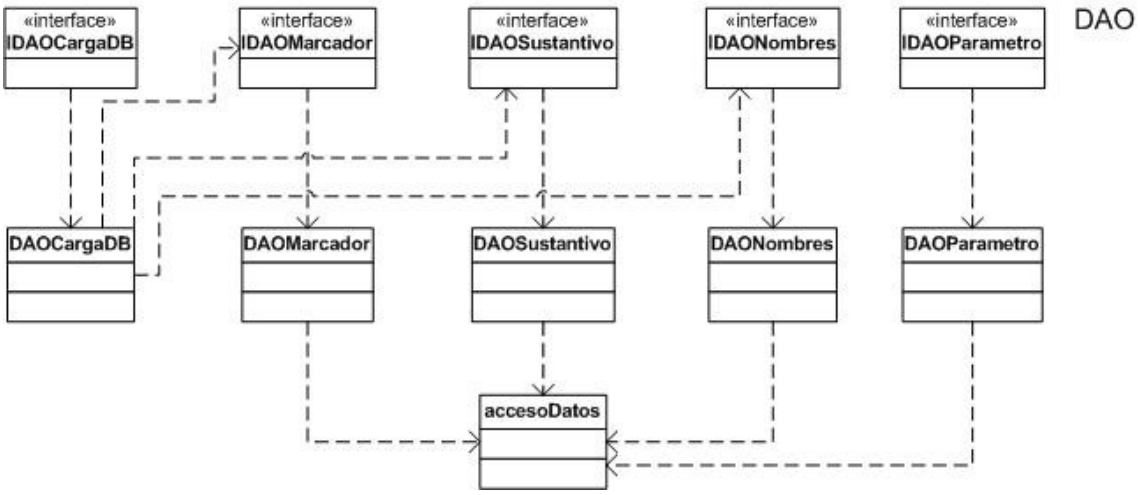

Figura 8 – Acceso a Datos (DAO)

A continuación se describen las diferentes clases y sus principales funcionalidades.

4.2.1. accesoDatos

Dicha clase es la encargada de realizar la conexión y desconexión con la base de datos. La conexión se realiza solo una vez cuando comienza la ejecución del programa y no se realiza la desconexión hasta que no se de fin a la aplicación. Otras funciones que realiza son: consultas (querys), y ejecuciones (execute) sobre la base de datos.
Esta clase se implementó con el pattern "Singleton", para que exista sólo una instancia y se acceda en forma sincrónica y secuencial a la base de datos. En dicha clase se centralizan todas las funciones de acceso a la base de datos. "Recibe ordenes" de las siguiente clases: DAOMarcador, DAOSustantivo, DAONombre y DAOParametros que se describirán en los siguientes puntos.

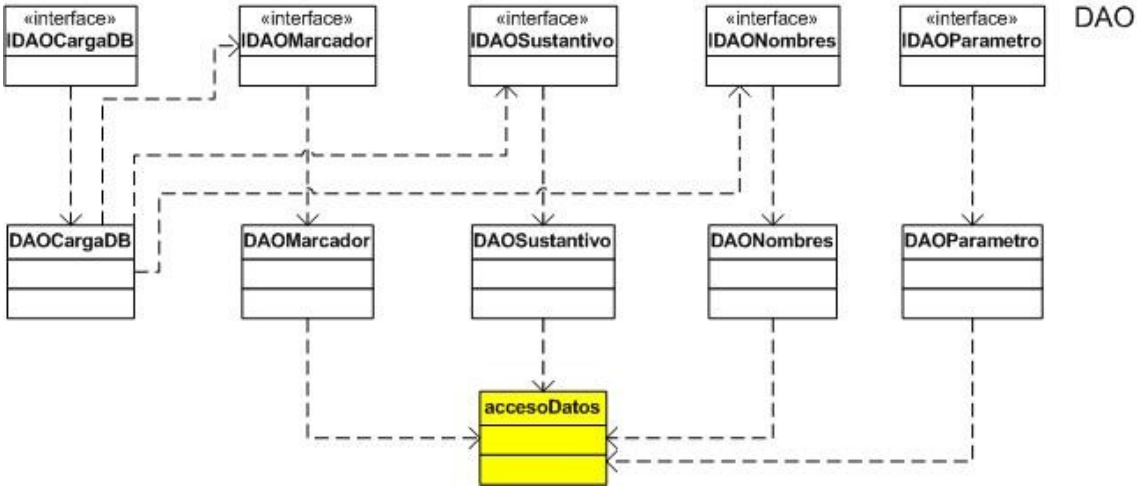

Figura 9 - accesoDatos

4.2.2. DAOMarcador (IDAOMarcador)

Esta clase se encarga de todas las funcionalidades referidas a los marcadores discursivos, es decir, tiene como función ingresar marcadores, actualizar, borrar,

consultar, etc. En dicha clase se arma el "script", comando o función que se quiere realizar en la base de datos. Es decir, si se quiere ingresar un marcador en la base de datos, se realiza el comando (string) y se llama a la función "execute" de accesoDatos que efectivamente hace el ingreso.
Las funciones de dicha clase implementan la interfaz correspondiente (IDAOMarcador).

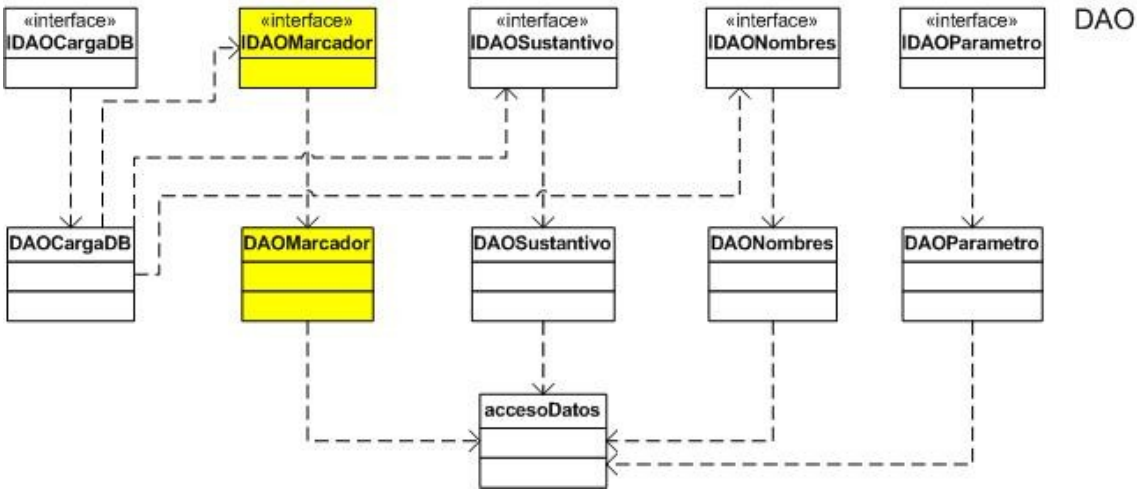

Figura 10 - DAOMarcador

4.2.3. DAOSustantivo (IDAOSustantivo)

Esta clase tiene las mismas funcionalidades que la anterior, pero se centraliza en los sustantivos. Las funciones que posee son las mismas que la anterior y también invoca a las funciones de accesoDatos para que las acciones de dichas funciones se vean reflejadas en la base de datos.
La interfase IDAOSustantivo es la interfaz de dicha clase, o sea, esta clase implementa dicha interfaz.

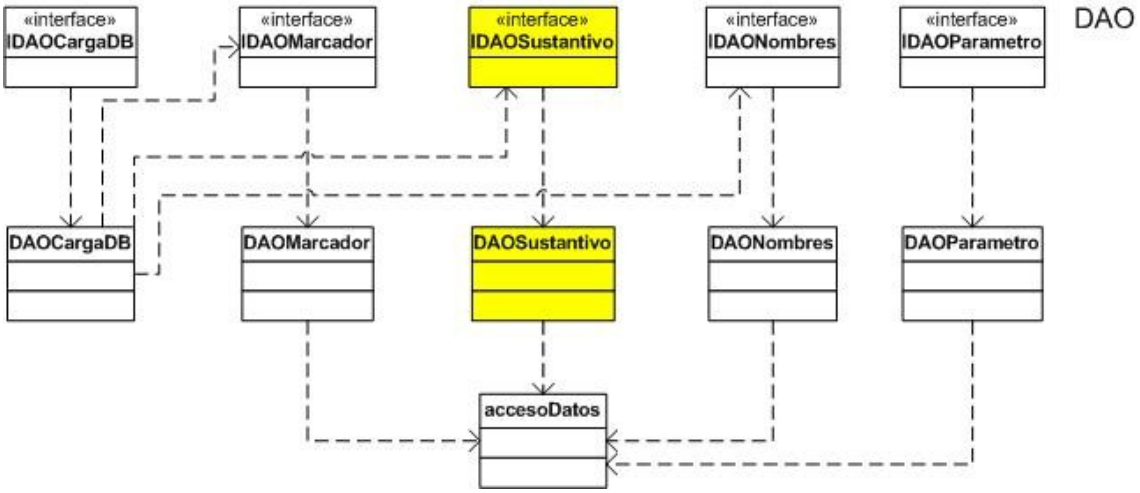

Figura 11 - DAOSustantivo

4.2.4. DAONombre (IDAONombre)

Es similar a las dos anteriores ya que tiene las mismas funciones pero se centraliza en los nombres. También invoca a las funciones de accesoDatos para realizar su cometido: acceder a la base de datos.

Dicha clase implementa a la interfaz IDAONombre, o sea, el acceso a dicha clase se realiza a través de la interfaz mencionada.

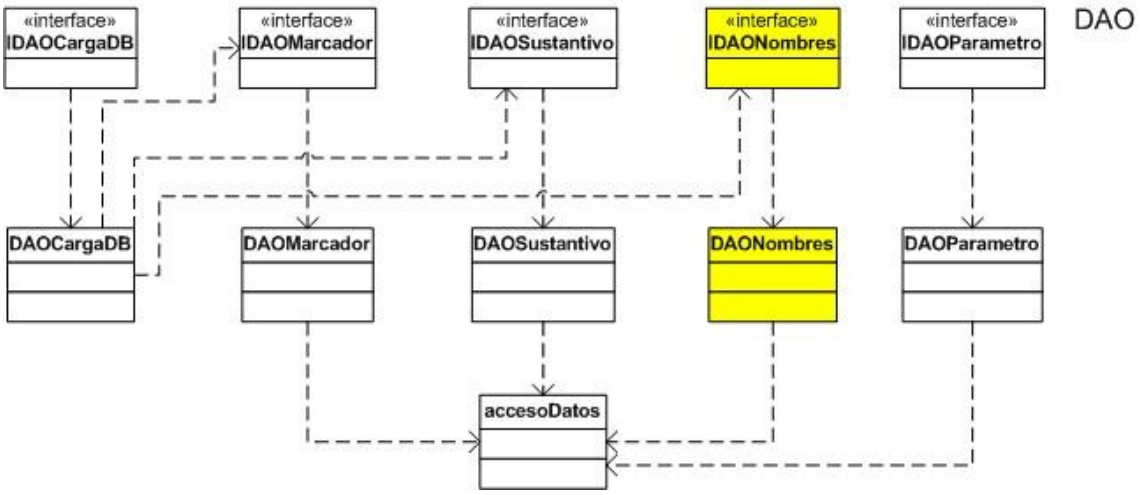

Figura 12 - DAONombre

4.2.5. DAOCargaDB (IDAOCargaDB)

Esta clase tiene como cometido ingresar/descargar de la base de datos los marcadores discursivos, sustantivos y nombres.

Para realizar la carga de la base se leen tres archivos con extensión txt donde residen dichas palabras y se ingresan en la base de datos (carga). Para esto utiliza las clases antes mencionadas (DAONombre, DAOSustantivo, DAOMarcador y accesoDatos).

Para respaldar o guardar la base de datos (realizar la descarga), se realiza el proceso inverso: se utiliza un archivo vacío con extensión txt para cada tipo de palabras (marcadores, sustantivos y nombres), o sea, se descargan todas las palabras desde la base de datos al correspondiente archivo.

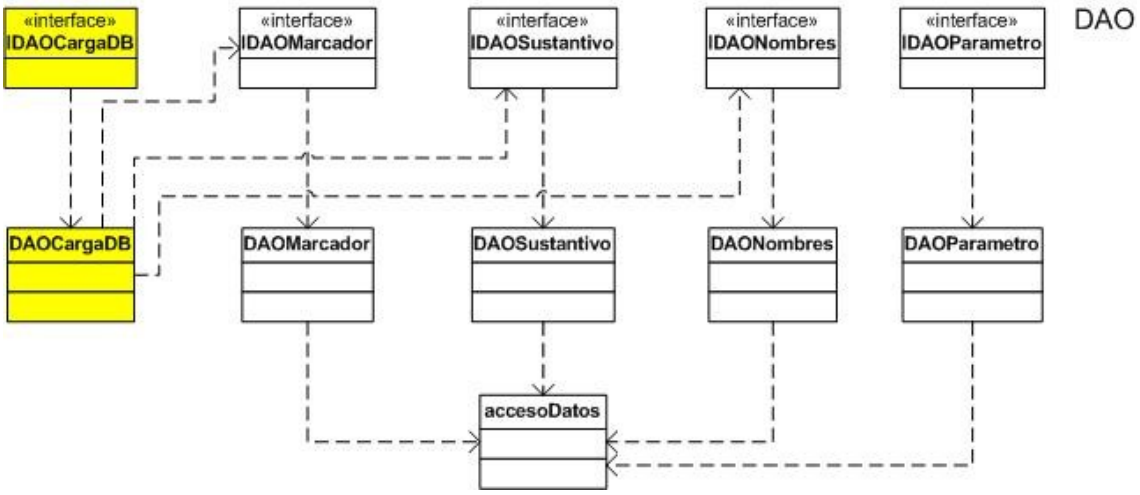

Figura 13 - DOACargaDB

4.2.6. DAOParametro (IDAOParametro)

A través de esta clase se puede tener acceso y manipular los parámetros que residen en la base de datos. Hay un total de 5 parámetros en la base de datos, uno por cada tipo de ponderación (posición, título, frecuencia de la palabra, consulta de usuario, y nombres). El valor de cada parámetro (entre 0 y 100) nos indica cuanto influye el ponderador correspondiente en la ponderación total de una oración. Por ejemplo: si todos los parámetros tienen un valor de 100 y el de posición tiene un valor de 50, significa que en la ponderación total de cada oración se tomará la ponderación por posición a la mitad (50%) y el resto de las ponderaciones se tomarán al 100%.
Esta clase implementa la interfase IDAOParametro, o sea, el acceso a dicha clase se realiza a través de la interfaz mencionada.

Las clases de la lógica que se comunican con esta clase son:

- accesoIntermedio (sección): es el que da la orden de actualización de los parámetros y también los pide para mostrarlos al usuario.
- Principal (sección): pide los parámetros a la hora de calcular la ponderación total para cada oración.

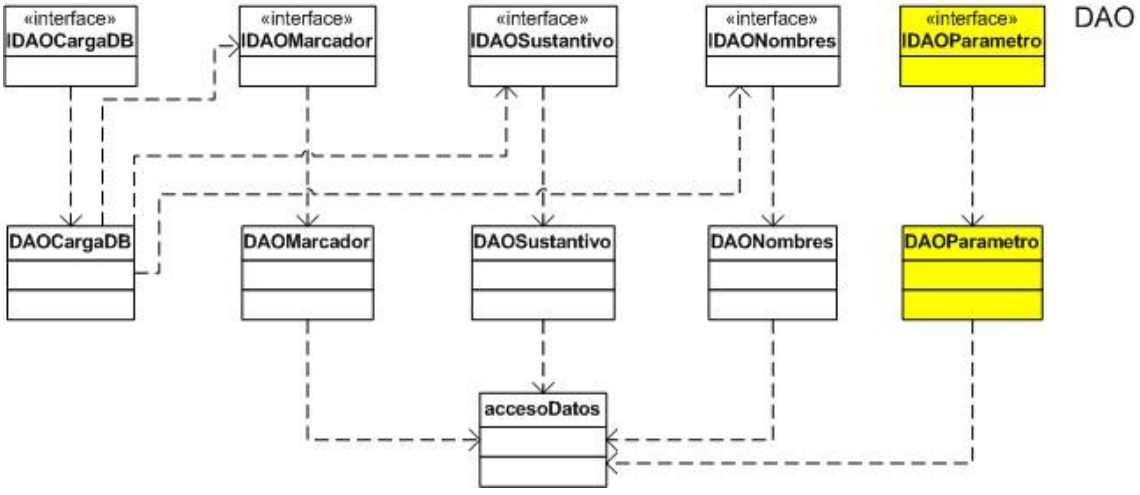

Figura 14 - DAOParametro

4.3. Lógica

Las clases que componen esta capa se encargan de realizar cálculos (ponderación de las oraciones, entre otras), manipular objetos (creación y baja), etc. Esta capa realiza comunicación ascendente y descendente, o sea, con la presentación (capa superior) y con DAO (capa inferior) respectivamente. Atiende "pedidos" solicitados de la "presentación" (interfaz de usuario, descrita en la sección), realiza cálculos, y en caso de ser necesario, se comunica con "DAO", para obtener información de la base de datos.

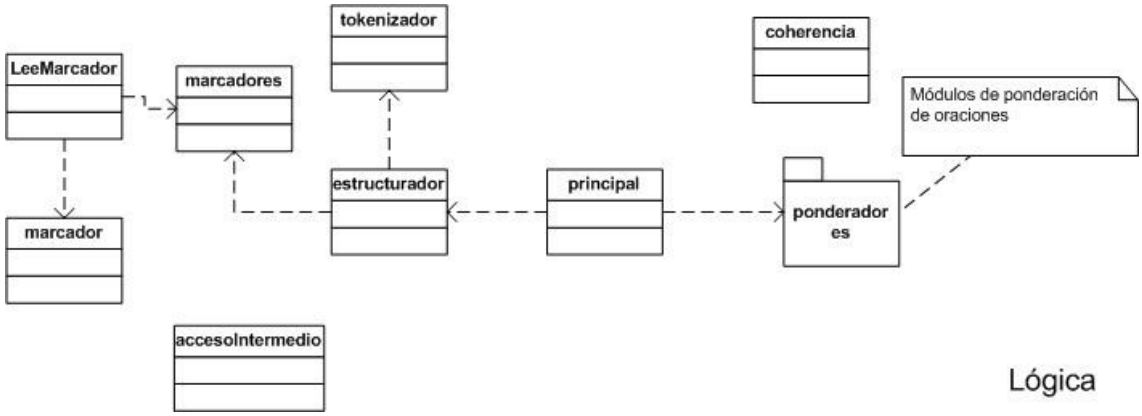

Figura 15 – Lógica de la aplicación

A continuación se describirán las clases que la componen.

4.3.1. accesoIntermedio

En esta clase se realizan cálculos que no necesariamente necesitan datos de la base de datos, por ejemplo se utiliza para el armado de un string de consulta de usuario, etc.; estos cálculos son productos de un "pedido" proveniente de la "presentación". También puede llegar a comunicarse con DAO para acceder a la base de datos cuando recibe, por ejemplo, el pedido de ingresar o dar de baja una palabra (marcador, sustantivo o nombre). Se comunica con la clase IDAOCargaDB para indicarle que se necesita cargar o descargar la base de datos. También se comunica con IDAOParametro para indicarle que el usuario está actualizando alguno de los parámetros.

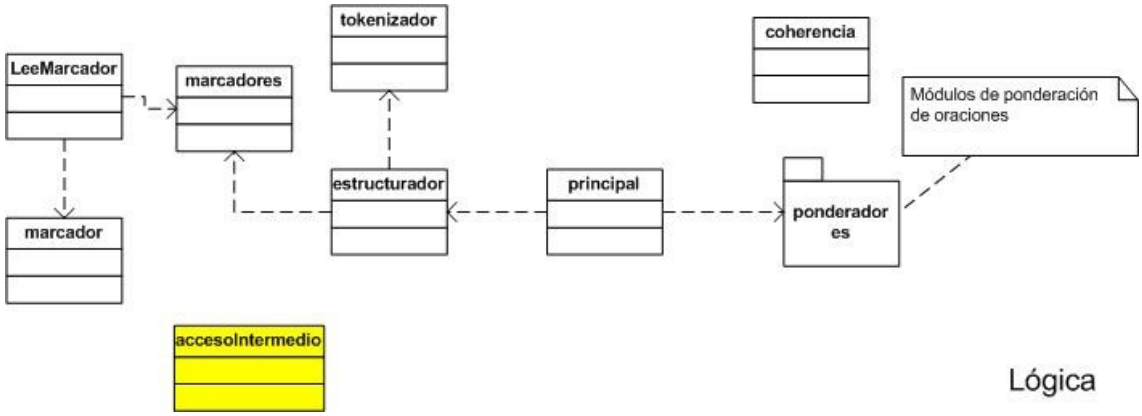

Figura 16 - accesoIntermedio

4.3.2. Marcador

La clase marcador se encarga de crear el archivo JLex para el reconocimiento de los marcadores discursivos en el texto fuente (texto a resumir), los pasos para crearlo son:

- Crea un "string" de código que se divide a su vez en:
  - Encabezado: es la parte estática del archivo JLex, donde se definen variables, se importan librerías, etc.
  - Cuerpo: es la parte "dinámica" del archivo JLex ya que es la parte donde se incluyen todos los marcadores de la base de datos, y estos varían de una petición a otra de la creación del archivo.
- Se crea el archivo JLex (marcadores.lex).
- Se incluye el string armado dentro del archivo creado.

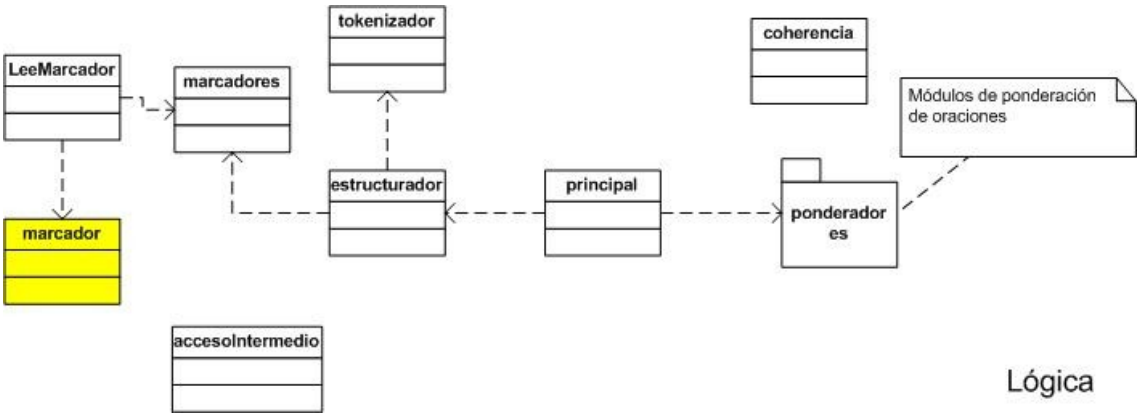

Figura 17 - marcador

4.3.3. LeeMarcador

Esta clase es la que se encarga de ejecutar el pedido o comando proveniente de la "presentación" para crear el archivo marcadores.lex. Este pedido se realiza en los siguientes tres casos: ingreso de marcadores, baja de marcadores y carga de la base de datos.
Para cumplir su función, delega el pedido a la clase descrita en la sección (marcador.java) para la creación del archivo JLex. Luego ejecuta un archivo de extensión "bat" el cual realiza lo siguiente:

- Compila el archivo Lex (se crea el archivo marcadores.java, que es la traducción del archivo JLex a código Java).
- Se compila marcadores.java.

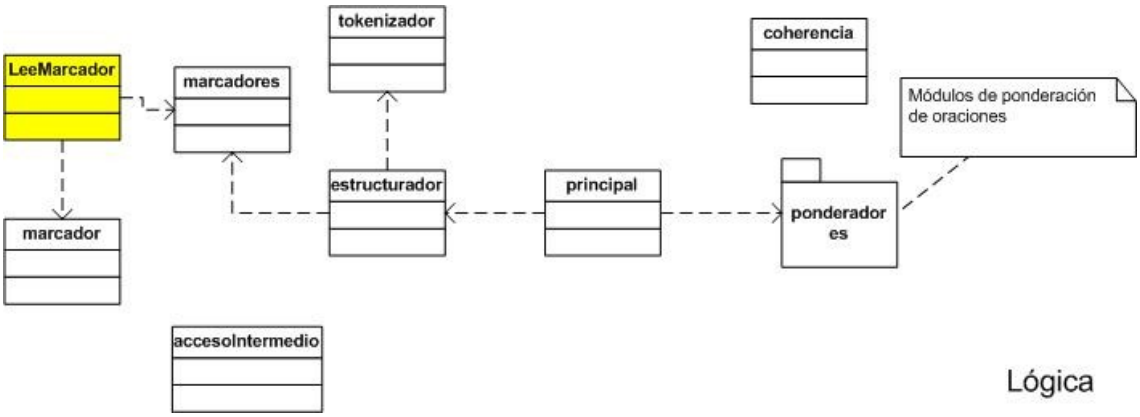

Figura 18 - LeeMarcador

4.3.4. Marcadores

Esta clase es creada por la clase LeeMarcadores cuando realiza la compilación del archivo JLex. Es la encargada de reconocer los marcadores que se encuentran en el texto de entrada. Es manipulada desde la clase "estructurador", la cual pide el reconocimiento de los marcadores (se describe en la sección).

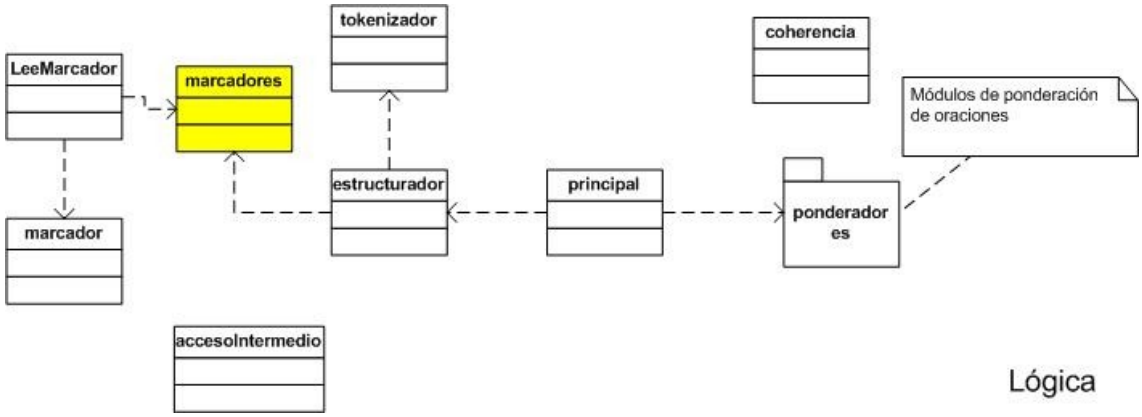

Figura 19 - marcadores

4.3.5. Tokenizador

Esta clase es el producto de la compilación del archivo tokenizador.lex (explicado en la sección) y se encarga de la lectura del texto. Se reconocen los párrafos y dentro éste las oraciones que los componen. La salida es una colección de oraciones.
A diferencia de marcadores.java, esta clase es estática (no se modifica) ya que se reconocen estructuras estáticas como oraciones y párrafos y no se usa la base de datos como en el caso anterior. El reconocimiento del texto es producto del "pedido" realizado por la clase "estructurador" que se describirá a continuación.

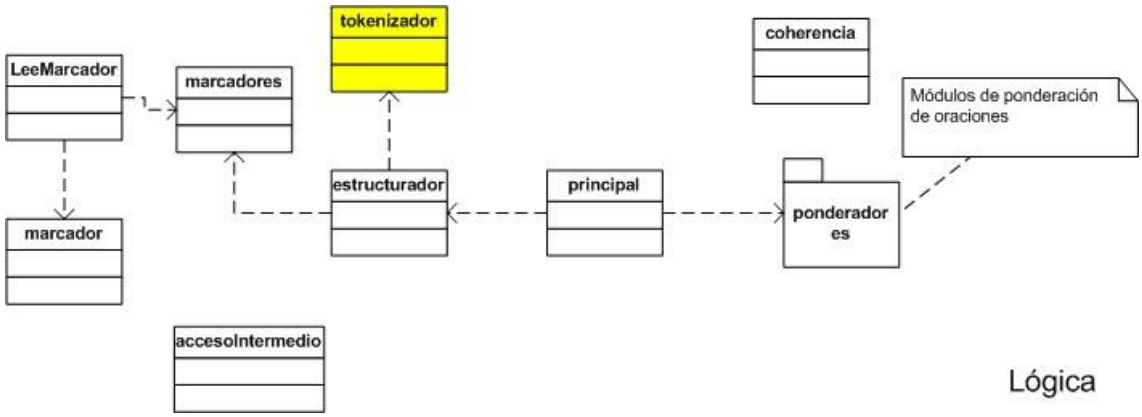

Figura 20 - tokenizador

4.3.6. Estructurador

Esta clase realiza un post-reconocimiento o procesamiento de las oraciones ya reconocidas, o sea, realiza el armado de la estructura y resuelve problemas (descritos en la sección) generados en la etapa de lectura del texto (sección).
Las principales funciones son:

- Realizar el pedido a tokenizador.java para que identifique los párrafos y oraciones en el texto de entrada.
- Realizar el pedido a marcadores.java para que identifique a los marcadores junto con su tipo en el texto de entrada.
- Armar la estructura (descrita en la sección). En dicho armado, "soluciona" algunos inconvenientes que se producen con la utilización de la herramienta JLex[14], como por ejemplo, la frase "*www.fing.edu.uy*", la arma como una única frase, ya que la clase tokenizador la reconoce como cuatro frases diferentes (causa de las apariciones del carácter '.'). Asigna números correctos de párrafos y marcadores discursivos a las oraciones.
- Identificar dentro de cada oración que conforma la estructura del texto, los nombres propios y los sustantivos. Por cada nombre o sustantivo crea una clase palabra asociándola a la oración a la cual pertenece (ver Figura 3).

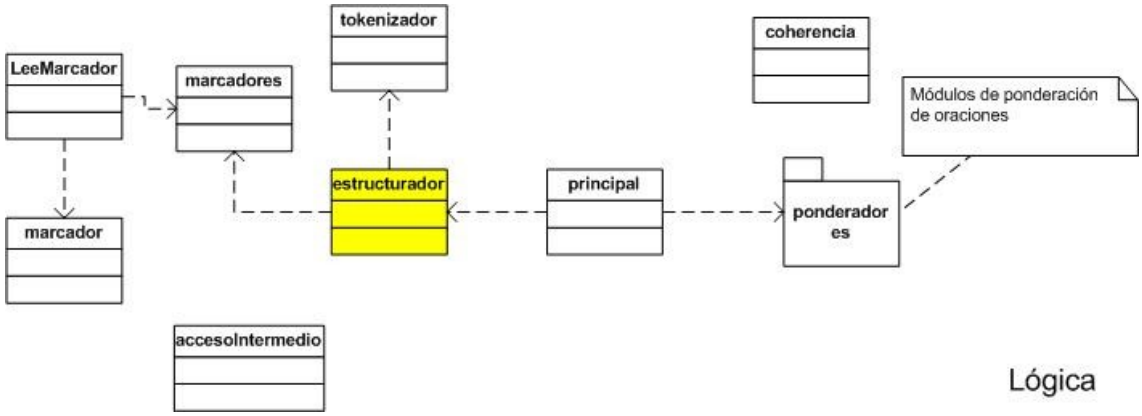

Figura 21 - estructurador

4.3.7. Principal

Esta clase se encarga de tomar la estructura creada por la clase estructurador y "pide" a los distintos módulos ponderadores (sección) que ponderen las oraciones de la estructura. Luego de que cada módulo calcule su ponderación correspondiente, calcula la ponderación total de cada oración como una combinación lineal de las anteriores.

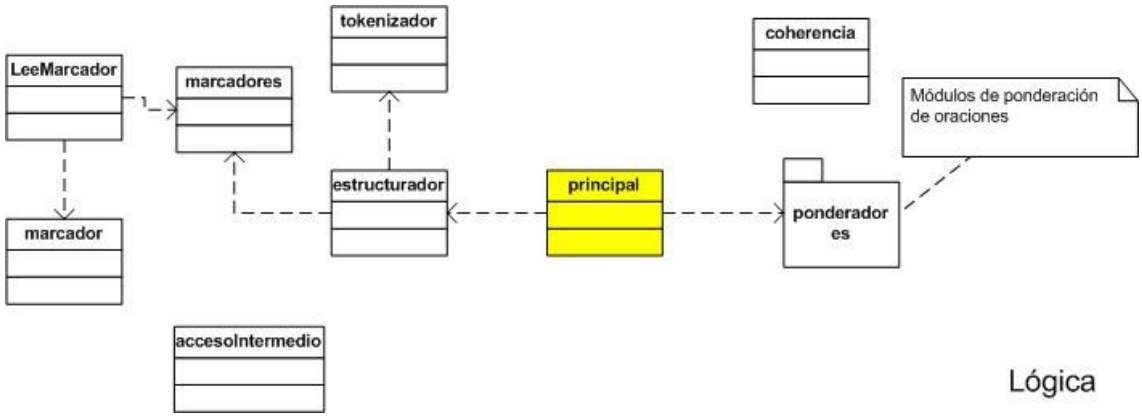

Figura 22 - Principal

4.3.8. Ponderadores

En este package se encuentran los módulos que implementan las diferentes heurísticas vistas en la sección . A continuación se describirán brevemente cada uno de los módulos o clases pertenecientes a dicho package.
Ponderación por:

- Nombres propios: este módulo pondera las oraciones por la cantidad de nombres propios que posea.
- Palabras del título: dicho módulo pondera a las oraciones que posean palabras que se encuentran en el título.
- Por posición en el documento: pondera por la posición de la oración dentro del párrafo, y a este por la posición dentro del documento. Se pondera el primer y el último párrafo del documento.
- Frecuencia de la palabra: primeramente se calcula un valor a cada palabra del texto según su frecuencia dentro del texto, dicho valor será igual a la cantidad de apariciones de la palabra (nombres y sustantivos) en el texto sobre (dividido) el total de palabras (nombres y sustantivos) en el texto. Luego, para cada oración en la que aparecen dichas palabras, se calcula la ponderación como la sumatoria del producto de la frecuencia de la palabra por la cantidad de veces que aparece en la oración.
- Personalización del texto: se pondera cada oración según la cantidad de palabras que contenga y que además estén incluidas en la colección de palabras previamente ingresadas por el usuario.

Todos los módulos ponderadores tienen como entrada la estructura (explicada en la sección) y el resultado es la estructura con las oraciones ponderadas.

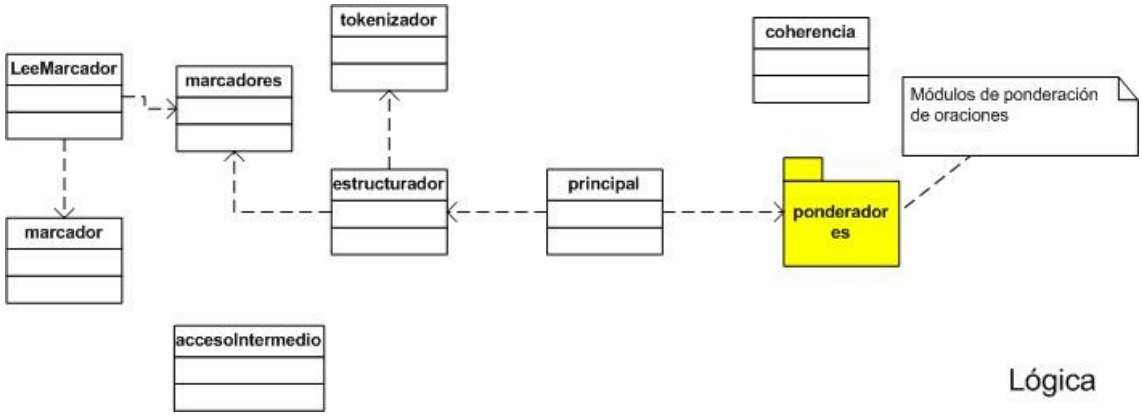

Figura 23 - ponderadores

4.3.9. Coherencia

Esta es la clase que se encarga de aplicarle coherencia al resumen propuesto por los módulos ponderadores utilizando marcadores discursivos.
En este módulo se estudia la incorporación o no de una oración que comienza con un marcador discursivo. Se debe diferenciar si la oración es la primera del párrafo o si es interior a él.
El algoritmo de coherencia basado en marcadores discursivos consta básicamente de tres partes:

- Procedimiento principal.
  Recorre la lista de oraciones ordenadas descendentemente por ponderación. Para una oración chequea si ya fue marcada para resumen, en caso afirmativo continúa con la siguiente oración; en otro caso verifica si comienza con marcador binario. En caso negativo se marca la oración para resumen, y en caso positivo se pasa a uno de los dos siguientes puntos.

- Oración que comienza con marcador binario y no es la primera del párrafo (intra-párrafo).
  Este procedimiento analiza la posible inclusión de la oración en estudio en el resumen. Es un procedimiento recursivo ya que se debe tener en cuenta la posible ocurrencia de un marcador discursivo binario en el primer argumento (oración anterior a la oración en estudio). Notar que se puede llegar a combinar este método con el que se describe a continuación (argumento 1 comienza con marcador binario y es la primer oración del párrafo).
- Oración que comienza con marcador binario y es la primera del párrafo (entre-párrafos).
  Este procedimiento analiza la posible inclusión de la oración en estudio en el resumen. Dado que la oración es la primera del párrafo, se estudia la inclusión de alguna oración del párrafo anterior, y la inclusión de oraciones del párrafo actual (párrafo que contiene la oración en estudio). Es un procedimiento recursivo ya que se debe tener en cuenta la posible ocurrencia de un marcador discursivo binario en cualquiera de las oraciones que se analiza su inclusión en el resumen. Notar que se puede llegar a combinar este método con el anterior (cuando se analiza una oración que comienza con marcador binario y no es la primera oración del párrafo).

El ingreso de cualquier oración al resumen esta condicionado por la cantidad de oraciones que formarán el resumen final.
Por más detalles del algoritmo ver el punto "Aplicando coherencia" de la sección .

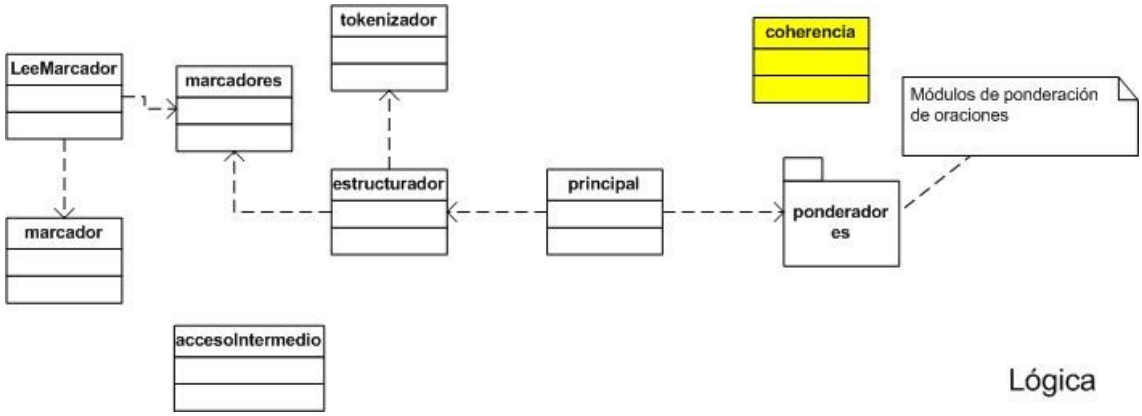

Figura 24 - coherencia

4.4. Presentación

En dicha capa se presentan las clases que implementan la interfaz gráfica implementada con Swing (Java), que permite que los usuarios puedan interactuar con el sistema. Las clases que implementan la interfaz están contenidas en el package "resumidores", mientras que la clase WPrincipal, es la clase "intermediaria" entre la presentación y la lógica de la aplicación.

La Figura 25 muestra la relación entre el package y la clase intermediaria (WPrincipal).

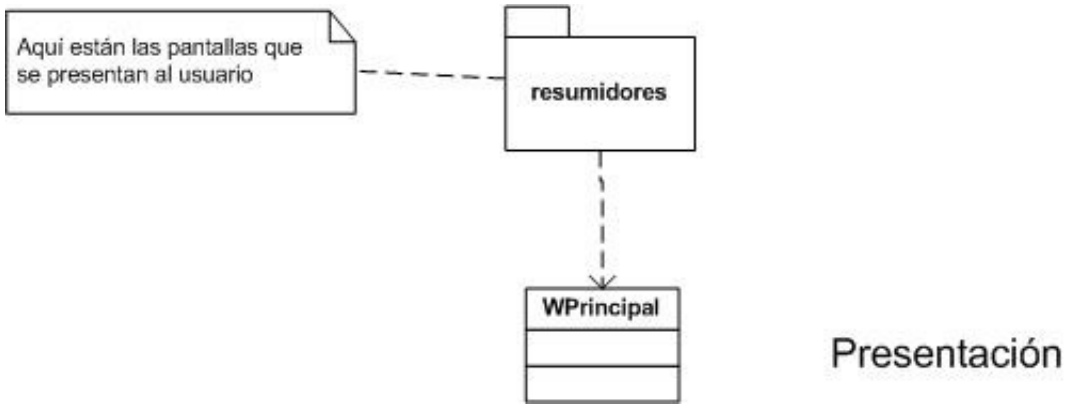

Figura 25 - Presentación

A continuación se describe la clase WPrincipal y el package "resumidores".

4.4.1. WPrincipal

Esta clase es la encargada de la comunicación entre la capa de "presentación" y la lógica de la "aplicación". Es la que se encarga de hacer "pedidos" o "peticiones" a la capa "lógica".
Algunos de los pedidos que realiza son:

- Conexión con la base de datos (se realiza por única vez mientras se ejecuta la aplicación)
- Pedido de datos (sustantivos, nombres, marcadores)
  - Dar de alta
  - Dar de baja
  - Consulta
- Modificación de parámetros
- Carga/descarga la base de datos desde/hacia archivos
- Ponderar las oraciones
- Aplicar coherencia al resumen
- Guardar a archivo el resumen
- Obtener el texto fuente etiquetado (archivo XML)

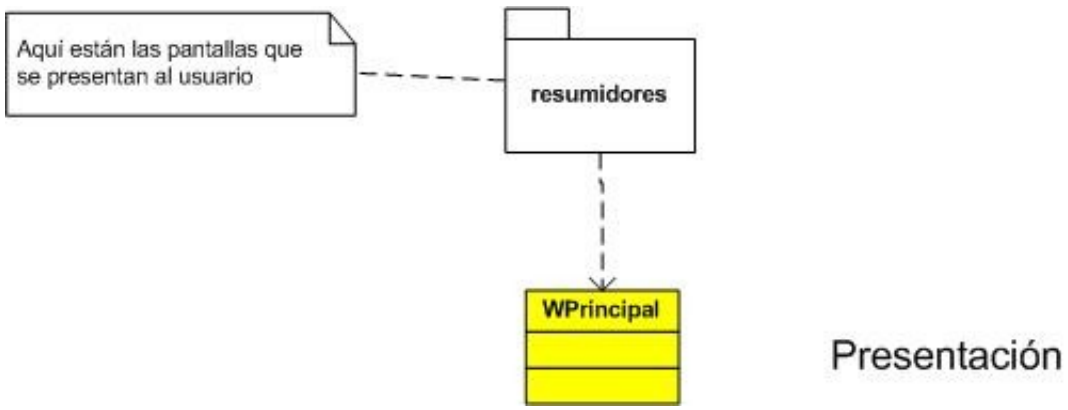

Figura 26 - WPrincipal

4.4.2. Package resumidores

En este package se encuentran las clases que implementan las "pantallas" o interfaz para usuario. Dichas pantallas son el medio de comunicación entre los usuarios y el sistema. La comunicación entre la interfaz y la lógica se realiza por medio de la clase WPrincipal antes descrita. Algunas funcionalidades que se presentan son:

- Con respecto a la base:
  - Conectarse a la base de datos
  - Cargar la base
  - Descargar la base
  - Ingreso y baja de palabras (marcadores, sustantivos o nombres)
  - Modificación de parámetros
- Con respecto al resumen de textos:
  - Abrir archivo
  - Realizar resumen
  - Guardar resumen (por coherencia o sólo por ponderación)
- Con respecto al texto fuente:
  - Guardar el texto etiquetado (archivo XML)

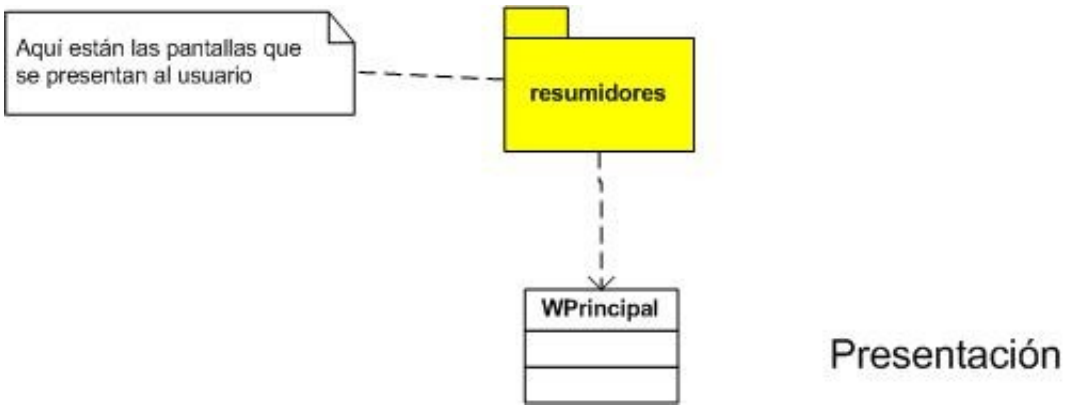

Figura 27 - resumidores

4.5. Base de datos

En este proyecto se utiliza como base de datos MySql versión 4.1. La misma tiene las ventajas de tener un fácil manejo, además, es un software de libre acceso (freeware).
Como se vio en puntos anteriores, la aplicación está implementada con una arquitectura de tres capas, donde la implementación de las clases es independiente de la base de datos que se manipule, por lo cual el programa puede adaptarse fácilmente a otro manejador de base de datos.

La base de datos consta de cuatro tablas:

- Tabla de nombres
- Tabla de sustantivos
- Tabla de marcadores
- Tabla de parámetros

En estas tablas se almacenan los datos necesarios para el funcionamiento del resumidor.

4.6. Composición de la estructura

Como ya se ha mencionado, principalmente en la lógica se realizan cálculos tales como el armado del texto; es aquí donde se conforma la estructura que se describe a continuación.

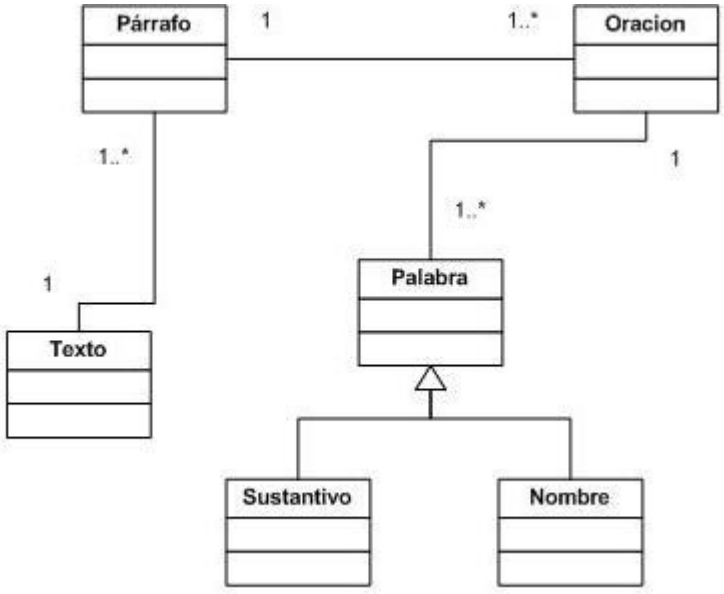

Figura 28 - estructura

A medida que se realiza el reconocimiento del texto de entrada, se van conformando los párrafos, las oraciones y las palabras (sustantivos y nombres), armando así la "estructura" que se presenta en la Figura 28.
La misma se compone de:

- Una clase "texto": dicha clase es única ya que el resumen se puede hacer de a un texto por vez. Esta clase contiene una colección de párrafos.
- Clase "párrafo": cada uno es representado con un número que coincide con el número de párrafo en el texto original y contiene una colección de oraciones que coinciden con las oraciones del párrafo en el texto orinal.
- Clase "oración": cada una contiene un número que la identifica, dicho número coincide con el número dentro del párrafo en el texto. También es identificada con el número de párrafo al que pertenece. Tiene otros campos como por ejemplo, si empieza con marcador o no, un campo por cada tipo de ponderación a la que es sujeta, un campo de ponderación total, otro que representa la cantidad de palabras que contiene, etc. La mayoría se utilizan al momento de ponderar la oración o en la etapa de coherencia del resumen. Cada oración contiene una colección de palabras que son sustantivos o nombres.
- Clase "palabra": es una clase abstracta que contiene el string o nombre que representa la palabra y la cantidad de ocurrencias que tiene en la oración a la cual pertenece. Las clases herederas son "sustantivo" y "nombre", que representan el "tipo" de palabra.

5. Testeo y Resultados

En este capítulo se describen las diferentes pruebas realizadas. En primer lugar, pruebas del reconocedor de textos (sección). En segundo lugar, pruebas de los algoritmos de coherencia junto con dos casos de estudio para comprobar el funcionamiento de dichos algoritmos (sección). Por último, se presentan las diferentes pruebas del resumidor realizadas sobre textos de noticias obtenidos de Internet (sección).

5.1. Pruebas de la interpretación de textos

Se realizaron las pruebas con diferentes textos fuentes de noticias obtenidas en Internet. En algunos casos se notó que cuando el sistema interpretaba el texto, es decir, creaba la estructura, "cortaba" las oraciones por no reconocer algunos caracteres como por ejemplo el carácter "&" o vocales con tilde. Para solucionar este problema, se ingresaron los caracteres no detectados en el archivo lex (tokenizador.lex).

También se detectaron problemas con el carácter punto (".") en casos donde la ocurrencia de dicho carácter no indicaba la finalización de una oración. Éste y otros problemas relacionados, se solucionan en la etapa de post-reconocimiento descrita en la sección .

En cada caso de la sección se presenta como un ítem ("Cantidad de uniones") la cantidad de casos solucionados donde la ocurrencia de un punto no indica la finalización de una oración.

El usuario puede acceder a un archivo con extensión txt (Salida.txt) creado por el sistema al momento en que se selecciona un texto fuente. En dicho archivo se imprime en forma ordenada (ascendente) según el número de párrafo y número de oración, cada una de las oraciones reconocidas (pertenecientes a la estructura) mostrando los siguientes datos:

- Número de párrafo y número de oración.
- Comienza con marcador: indica si la oración comienza con marcador, true en caso afirmativo, false en otro caso.
- Cantidad palabras: es la cantidad total de palabras que componen la oración.
- Lista de Sustantivos: imprime cada uno de los sustantivos reconocidos dentro de la oración.
- Lista de Nombres: imprime cada uno de los nombres reconocidos dentro de la oración.

De esta manera, el usuario tiene una forma sencilla de verificar si el sistema "reconoce" los datos que componen el texto fuente. Por ejemplo: puede verificar si las oraciones están completas, o sea, si no se presenta algún "error conocido" de los que se describen en la sección . Se destaca una vez más, que este punto se hubiera solucionado con la utilización de un etiquetador (tagger) que al menos segmentara en oraciones y párrafos. Fue nuestra decisión la de construir este módulo con JLex[14].

También puede "enriquecer" la base de datos con nombres y sustantivos al verificar cuales sustantivos y nombres se están reconociendo para cada oración, y en caso de no reconocer alguno, el usuario puede ingresarlo a la base de datos.

Dicho archivo también es útil al momento en que se realiza un resumen, ya que ofrece al usuario la posibilidad de ver la ponderación obtenida de cada oración (ponderación total y de cada método). De esta forma, puede "elegir" las distintas oraciones a través de la consulta de usuario o dando distintos valores a los coeficientes de los

ponderadores, para realizar las pruebas o testeos con mayor ponderación en las oraciones deseadas.

Por último, en el archivo de salida se imprimen las principales decisiones que toma el sistema para cada oración en estudio al momento de aplicarle coherencia al resumen, o sea, cómo se comporta ante una oración que comienza con marcador o una oración común según el porcentaje de resumen, etc.

5.2. Pruebas de coherencia

Para el correcto testeo del algoritmo de coherencia se hicieron distintas pruebas. Se modificaron algunos textos para tratar de abarcar la mayor cantidad de casos posibles y así verificar la correctitud del algoritmo. Se realizaron pruebas de casos intraoracionales, intra-párrafos y casos combinados. En la sección se presentan dos textos, uno real y otro modificado, para describir con ejemplos las decisiones que toma el sistema cuando se enfrenta a alguno de los casos que se presentan en la sección .

Nota: Cuando se menciona "párrafo actual" se hace referencia al párrafo que contiene la oración en estudio. Cuando se menciona "párrafo anterior" se hace referencia al párrafo anterior al que contiene la oración en estudio. Cuando se menciona "oración x" se hace referencia a la oración número x (comenzando desde la primer oración del párrafo que la contiene). Cuando se menciona "párrafo z" se hace referencia al párrafo número z (comenzando desde el primer párrafo del texto).

5.2.1. Casos de Prueba

A continuación se presentan los diferentes casos de prueba del algoritmo de coherencia, es decir, se describen las distintas situaciones que pueden suceder cuando el sistema se enfrenta a un marcador discursivo.

Casos intra-párrafos (la oración en estudio no es la primera del párrafo)

Oración en estudio: oración que comienza con marcador.

- **Situación 1**: la oración anterior no comienza con marcador.
    - La oración anterior no es candidata (se calcula el promedio de ponderaciones de ambas oraciones y se compara con la ponderación de la peor candidata).

      Resultado: *se marcan las dos oraciones para el resumen (mejor promedio) o no se marca ninguna (peor promedio). En caso de marcar oraciones no candidatas se modifica la peor ponderación como se explica en la Nota 2 del punto "Aplicando coherencia" de la sección .*
    - La oración anterior es candidata

      Resultado: *se marcan las dos oraciones para el resumen.*
    - La oración anterior ya fue marcada para el resumen.

      Resultado: *se marca la oración en estudio para el resumen.*
- **Situación 2**: la oración anterior comienza con marcador.

  Resultado: *se está en un caso combinado.*

Casos entre-párrafos

Oración en estudio: primera oración del párrafo, comienza con marcador discursivo binario.

- **Situación 1:**
  Párrafo anterior: hay candidatas no marcadas que no comienzan con marcador.
  Párrafo actual: hay candidatas no marcadas y otras marcadas y no comienzan con marcador.
  Resultado: *se marcan todas las candidatas del párrafo anterior, se marca la oración en estudio y las candidatas del párrafo actual. Las oraciones que ya estaban marcadas no se toman en cuenta.*

- **Situación 2:**
  Párrafo anterior: hay oraciones marcadas para resumen.
  Párrafo actual: hay una candidata no marcada para resumen que comienza con marcador y otras oraciones marcadas.
  Resultado: *se marca la oración en estudio para resumen, la candidata del párrafo actual que comienza con marcador no se marca ya que al calcular el promedio con la oración anterior (no candidata), éste es peor que la peor ponderación (aquí se da un caso combinado).*

- **Situación 3:**
  Párrafo anterior: hay oraciones marcadas para resumen.
  Párrafo actual: hay una candidata no marcada que no comienza con marcador y otras oraciones marcadas.
  Resultado: *se marca la oración en estudio para resumen, la candidata del párrafo actual también se marca.*

- **Situación 4:**
  Párrafo anterior: no hay candidatas, ni marcadas para resumen, el promedio de la mejor oración con las candidatas del párrafo actual es mejor que la peor ponderación.
  Párrafo actual: hay oraciones candidatas no marcadas, oraciones marcadas y no comienzan con marcador.
  Resultado: *se calcula el promedio de la mejor oración del párrafo anterior con las candidatas del actual, como es mejor que la peor ponderación, se marca la oración en estudio y la oración del párrafo anterior para resumen. También se marca para resumen una candidata no marcada del párrafo actual. Al marcar oraciones no candidatas se modifica la peor ponderación como se explica en la Nota 2 del punto "Aplicando coherencia" de la sección .*

- **Situación 5:**
  Párrafo anterior: no hay candidatas, ni marcadas para resumen, el promedio de la mejor oración con las candidatas del párrafo actual es peor que la peor ponderación.
  Párrafo actual: hay oraciones candidatas no marcadas, oraciones marcadas y no comienzan con marcador.
  Resultado: *no se marca la oración en estudio para resumen ya que el promedio de la mejor oración del párrafo anterior con las candidatas del actual es peor que la peor ponderación, tampoco se marca la oración del párrafo anterior ni la candidata no marcada del párrafo actual.*

Casos combinados[1]

- **Situación 1:**

  La oración 2 del párrafo 5 comienza con marcador, la oración 1 también comienza con marcador y es candidata.

  Párrafo 4 (anterior): la oración 1 es candidata no marcada para resumen, comienza con marcador y es la primera del párrafo 4.

  Párrafo 3: tiene oraciones marcadas para resumen.

  Resultado: *se marcan para resumen final la oración 1 y 2 del párrafo 5, y la oración 1 del párrafo 4.*

- **Situación 2:**

  La oración 2 del párrafo 5 comienza con marcador, la oración 1 también comienza con marcador y es candidata.

  Párrafo 4 (anterior): la oración 1 es **no** candidata, no esta marcada para resumen y comienza con marcador.

  Párrafo 3: tiene oraciones marcadas para resumen.

  Resultado: *se realiza el promedio de las candidatas del párrafo 5 con la oración 1 del párrafo 4 que es la mejor ponderada no candidata, dicho promedio es mejor que la peor ponderación por lo cual se estudian las oraciones del párrafo 3. El párrafo 3 tiene oraciones marcadas para resumen, por lo tanto, se marcan para resumen final la oración 1 y 2 del párrafo 5, y la oración 1 del párrafo 4. Al marcar oraciones no candidatas se modifica la peor ponderación como se explica en la Nota 2 del punto de la sección .*

- **Situación 3:**

  La oración 1 del párrafo 5 comienza con marcador.

  Párrafo anterior: no tiene oraciones marcadas para resumen ni candidatas; la oración 2 comienza con marcador y es la mejor ponderada del párrafo, la oración anterior (oración 1 del párrafo 4) no es candidata y no comienza con marcador.

  Resultado: *se realiza el promedio entre las candidatas del párrafo 5 y la oración 2 del párrafo 4; el promedio es mejor que la peor candidata por lo que se analiza la oración anterior (oración 1 del párrafo 4). Esta oración (1) tampoco es candidata por lo cual se calcula el promedio entre la oración 2 y la oración 1 del párrafo 4, el promedio es peor que la peor de las candidatas por lo cual todas las oraciones en estudio se rechazan (no se marca ninguna para resumen).*

- **Situación 4:**

  La oración 1 del párrafo 5 comienza con marcador.

  Párrafo anterior: no tiene oraciones marcadas para resumen, tiene una candidata que comienza con marcador (oración 2) y la oración anterior (oración 1) no es candidata y no comienza con marcador.

  Resultado: *se trata de ingresar la única candidata del párrafo 4, como ésta comienza con marcador, se intenta ingresar la oración anterior (oración 1 del párrafo 4); se realiza el promedio entre la oración 1 del párrafo 5, la oración 2 del párrafo 4 y la oración 1 del párrafo 4, el promedio es mejor que la peor ponderación por lo que se marcan para resumen las tres oraciones estudiadas. Al marcar oraciones no candidatas se modifica la peor ponderación como se explica en la Nota 2 de la sección .*

[1] Los números de oraciones y párrafos utilizados en las distintas situaciones son números reales del caso de prueba, se muestran para simplificar la descripción de la situación.

5.2.2. Ejemplos: Casos de Estudio

A continuación se presenta con dos ejemplos las decisiones a nivel de coherencia que toma el sistema para verificar si ante una oración que comienza con marcador, ésta se toma en cuenta para formar parte del resumen, intentando mantener la cohesión y coherencia.

Con los ejemplos se pretende describir en forma práctica alguna de las situaciones descritas en la sección .

Para cada una de las oraciones del ejemplo se muestran los siguientes datos:

- Número de párrafo.
- Número de oración.
- Si comienza con marcador.
- Total de palabras
- Valor total de ponderación.
- Si es una oración candidata o no.

CASO 1: REAL

"*Los más pobres entre los pobres.*

Noventa y ocho millones de indigentes viven en ciudades o en suburbios de América Latina. No verlos es imposible, ignorarles está al alcance de todos. Noventa y ocho millones de personas representan la suma de los habitantes de Inglaterra, Holanda, Bélgica, Austria, Finlandia y Suiza. Pero más insoportable resulta pensar que el 50% de ellos son niños. El equivalente a la población de España y Dinamarca juntas. Si todas estas personas se alinearan cogidas de la mano formarían una fila humana que daría más de dos vueltas a nuestro planeta.

En América Latina, el 18,5% de la población vive en situación de extrema pobreza, a lo que se suma el 42% en situación de "simple" pobreza, es decir un total de 319 millones de pobres. El equivalente a toda la población de EEUU y Australia.

Detrás de estas cifras hay personas con nombres y apellidos, niños mal alimentados, mal vestidos, menos limpios, menos mimados, protegidos y queridos... pero que tienen al nacer el mismo potencial que el resto. Sin embargo, son más vulnerables y están más expuestos a todo tipo de abuso y explotación. Tanto los niños y niñas de aquí como los de allí están igualmente sujetos a la Convención de Naciones Unidas sobre los Derechos del Niño, la realidad pone en evidencia una desproporción que es indispensable borrar.

A consecuencia del devastador tsunami que asoló las costas asiáticas hace ya más de un año, la inmensa respuesta de generosidad de miles de europeos llena de esperanza. Demuestra que los seres humanos somos naturalmente sensibles y estamos dispuestos a movilizarnos para paliar las tragedias vividas por otros seres humanos. Sin embargo, las personas donan y continuarán donando si están convencidas de que sirve para algo y que la ayuda llega a quien la necesita. Es precisamente en este punto en el que la responsabilidad y los resultados de las organizaciones humanitarias adquiere su importancia. Las ayudas deben responder con eficacia a necesidades concretas ya se trata de crisis mediatizadas o crisis olvidadas.

Si utilizáramos las cantidades generosas de los españoles obtenidas tras el tsunami para ofrecer una comida diaria de 60 céntimos de euro a cada uno de los 98 millones de personas sin hogar de América Latina, los recursos serían consumidos en un día y medio.

Por ello, toda respuesta asistencialista no resulta viable. Para conseguir cambios reales hace falta cambiar mentalidades, modificar radicalmente las relaciones de los poderes económicos, conseguir por ejemplo una equidad real en las relaciones comerciales Norte-Sur o la supresión de las patentes sobre medicamentos esenciales. Pero ante todo, una respuesta realista debe inscribir la "erradicación de la pobreza" en el primer lugar de la agenda internacional. En vista de las promesas no cumplidas, a pesar de las múltiples cumbres internacionales salpicadas de buenas intenciones, los dirigentes del planeta no pasarán a la acción hasta que la sociedad civil se movilice y les obligue a actuar.

Algunos intelectuales explican la existencia de estos 98 millones de indigentes con eslóganes como: "este número es el reflejo del problema estructural", "es un desequilibrio generado por la mala distribución de la

riqueza" "son consecuencias incontrolables de los regímenes y dictaduras", "de los intereses económicos, políticos y estratégicos de los países industrializados", "de la corrupción endémica de los gobiernos locales o nacionales".

Sin embargo, ésta es una situación inaceptable, una violación constante de los derechos fundamentales de todo ser humano. La necesidad de cambiar las cosas, de pelearnos por las personas en situación de pobreza en América Latina y por los miles de niños y mujeres que sufren esta situación es un compromiso y la responsabilidad de todos.

Las prácticas cotidianas de las organizaciones internacionales que trabajan en el terreno ofrecen una ayuda concreta a miles de niños y adultos que viven en situaciones intolerables. No obstante, la ayuda prestada no tiene que sustituir la responsabilidad del Estado, de la sociedad civil y de las comunidades.

Lejos de las imágenes de Copacabana en Río de Janeiro, lejos de los clichés de playas paradisíacas del Caribe o de las imágenes brumosas de las postales del Machu Pichu en Perú, 98 millones de seres humanos duermen cada día en la calle."

Para asignarle más "fuerza" a algunas oraciones, se ingresan las siguientes palabras en la consulta de usuario: "Estado" como nombre, y los siguientes sustantivos: "abuso", "tipo", "explotación", "derechos", "humanos", "violación" y "sociedad".

El porcentaje de resumen es el 50%, por lo tanto, la cantidad de oraciones que componen el resumen final son: 27 (total de oraciones) * 50 (porcentaje de resumen) / 100 = 13.

Los coeficientes de los ponderadores se utilizaron todos al 100%.

A continuación se presenta un análisis de cada oración del texto fuente:

Título (párrafo 1 para el sistema): Los más pobres entre los pobres.

Párrafo 2: compuesto por 6 oraciones

"*Noventa y ocho millones de indigentes viven en ciudades o en suburbios de América Latina.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
2	1	No	14	0.37150884	si

"*No verlos es imposible, ignorarles está al alcance de todos.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
2	2	no	10	0.13561799	no

"*Noventa y ocho millones de personas representan la suma de los habitantes de Inglaterra, Holanda, Bélgica, Austria, Finlandia y Suiza.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
2	3	No	20	0.82797754	si

"*Pero más insoportable resulta pensar que el 50% de ellos son niños.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
	4	No	12	0.12370788	no

"*El equivalente a la población de España y Dinamarca juntas.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
2	5	No	10	0.51494384	si

"*Si todas estas personas se alinearan cogidas de la mano formarían una fila humana que daría más de dos vueltas a nuestro planeta.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
2	6	No	23	0.11179777	no

<u>Párrafo 3:</u> compuesto por 2 oraciones

"*En América Latina, el 18,5% de la población vive en situación de extrema pobreza, a lo que se suma el 42% en situación de "simple" pobreza, es decir un total de 319 millones de pobres.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
3	1	No	34	0.30865827	si

"*El equivalente a toda la población de EEUU y Australia.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
3	2	No	10	0.43932584	si

<u>Párrafo 4:</u> compuesto por 3 oraciones

"*Detrás de estas cifras hay personas con nombres y apellidos, niños mal alimentados, mal vestidos, menos limpios, menos mimados, protegidos y queridos... pero que tienen al nacer el mismo potencial que el resto.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
4	1	No	33	0.10674157	no

"*Sin embargo, son más vulnerables y están más expuestos a todo tipo de abuso y explotación.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
4	2	Si	16	0.39185393	si

"*Tanto los niños y niñas de aquí como los de allí están igualmente sujetos a la Convención de Naciones Unidas sobre los Derechos del Niño, la realidad pone en evidencia una desproporción que es indispensable borrar.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
4	3	No	31	0.24972814	si

<u>Párrafo 5:</u> compuesto por 5 oraciones

"*A consecuencia del devastador tsunami que asoló las costas asiáticas hace ya más de un año, la inmensa respuesta de generosidad de miles de europeos llena de esperanza.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	1	No	28	0.06179775	no

"*Demuestra que los seres humanos somos naturalmente sensibles y estamos dispuestos a movilizarnos para paliar las tragedias vividas por otros seres humanos.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	2	No	22	0.25485188	si

"*Sin embargo, las personas donan y continuarán donando si están convencidas de que sirve para algo y que la ayuda llega a quien la necesita.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	3	Si	25	0.050561797	no

"*Es precisamente en este punto en el que la responsabilidad y los resultados de las organizaciones humanitarias adquiere su importancia.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	4	No	20	0.04494382	no

"*Las ayudas deben responder con eficacia a necesidades concretas ya se trata de crisis mediatizadas o crisis olvidadas.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	5	No	18	0.03932584	no

<u>Párrafo 6:</u> compuesto por 1 oración

"*Si utilizáramos las cantidades generosas de los españoles obtenidas tras el tsunami para ofrecer una comida diaria de 60 céntimos de euro a cada uno de los 98 millones de personas sin hogar de América Latina, los recursos serían consumidos en un día y medio.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
6	1	No	44	0.19713993	si

<u>Párrafo 7:</u> compuesto por 4 oraciones

"*Por ello, toda respuesta asistencialista no resulta viable.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
7	1	No	8	0.016853932	no

"*Para conseguir cambios reales hace falta cambiar mentalidades, modificar radicalmente las relaciones de los poderes económicos, conseguir por ejemplo una equidad real en las relaciones comerciales Norte-Sur o la supresión de las patentes sobre medicamentos esenciales.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
7	2	No	36	0.06179775	no

"*Pero ante todo, una respuesta realista debe inscribir la "erradicación de la pobreza" en el primer lugar de la agenda internacional.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
7	3	No	21	0.061797753	no

"*En vista de las promesas no cumplidas, a pesar de las múltiples cumbres internacionales salpicadas de buenas intenciones, los dirigentes del planeta no pasarán a la acción hasta que la sociedad civil se movilice y les obligue a actuar.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
7	4	No	39	0.10746183	no

Párrafo 8: compuesto por 1 oración

"*Algunos intelectuales explican la existencia de estos 98 millones de indigentes con eslóganes como: "este número es el reflejo del problema estructural", "es un desequilibrio generado por la mala distribución de la riqueza" "son consecuencias incontrolables de los regímenes y dictaduras", "de los intereses económicos, políticos y estratégicos de los países industrializados", "de la corrupción endémica de los gobiernos locales o nacionales"."*

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
8	1	No	62	0.14044942	no

Párrafo 9: compuesto por 2 oraciones

"*Sin embargo, ésta es una situación inaceptable, una violación constante de los derechos fundamentales de todo ser humano."*

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
9	1	Si	18	0.26716605	si

"*La necesidad de cambiar las cosas, de pelearnos por las personas en situación de pobreza en América Latina y por los miles de niños y mujeres que sufren esta situación es un compromiso y la responsabilidad de todos.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
9	2	No	37	0.25630125	si

Párrafo 10: compuesto por 2 oraciones

"*Las prácticas cotidianas de las organizaciones internacionales que trabajan en el terreno ofrecen una ayuda concreta a miles de niños y adultos que viven en situaciones intolerables.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
10	1	No	27	0.07865168	no

"*No obstante, la ayuda prestada no tiene que sustituir la responsabilidad del Estado, de la sociedad civil y de las comunidades."*

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
10	2	Si	21	0.34189406	si

Párrafo 11: compuesto por 1 oración

"*Lejos de las imágenes de Copacabana en Río de Janeiro, lejos de los clichés de playas paradisíacas del Caribe o de las imágenes brumosas de las postales del Machu Pichu en Perú, 98 millones de seres humanos duermen cada día en la calle.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
11	1	No	40	0.5404494	si

En la Figura 29 se presentan las oraciones candidatas ordenadas descendentemente por ponderación total:

Párrafo	Oración	Ponderación
2	3	0.82797754
11	1	0.5404494
2	5	0.51494384
3	2	0.43932584
4	2	0.39185393
2	1	0.37150884
10	2	0.34189406
3	1	0.30865827
9	1	0.26716605
9	2	0.25630125
5	2	0.25485188
4	3	0.24972814
6	1	0.19713993

Figura 29 – Oraciones candidatas para el caso 1

De la tabla se desprende que la oración 1 del párrafo 6 tiene la **peor ponderación** de las oraciones candidatas.

El estudio de coherencia comienza por la oración 3 del párrafo 2 por ser la más ponderada, luego por la segunda mejor ponderada y así sucesivamente hasta alcanzar el porcentaje de resumen. Para cada oración en estudio se realiza el estudio de coherencia como se describe en la sección .

Estudio de la oración 3 del párrafo 2:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 1.

Estudio de la oración 1 del párrafo 11:

No comienza con marcador, por lo tanto se marca para resumen

Oraciones marcadas hasta el momento: 2.

Estudio de la oración 5 del párrafo 2:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 3.

Estudio de la oración 2 del párrafo 3:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 4.

Estudio de la oración 2 del párrafo 4:

Comienza con marcador, la oración es interior al párrafo, por lo cual se estudia la oración anterior: oración 1 del mismo párrafo.

La oración 1 del párrafo 4 no esta macada para resumen y tampoco es candidata por lo tanto se realiza el promedio con la oración que contiene el marcador: 0.39185393 + 0.10674157 / 2 = 0.24929775. Este promedio supera a la ponderación de la peor candidata que es 0.19713993, por lo tanto se marcan ambas oraciones para el resumen. La nueva peor ponderación es 0.24972814 que pertenece a la ponderación de la penúltima oración candidata.

Oraciones marcadas hasta el momento: 6.

Estudio de la oración 1 del párrafo 2:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 7.

Estudio de la oración 2 del párrafo 10:

Comienza con marcador, la oración es interior al párrafo, por lo cual se estudia la oración anterior: oración 1 del mismo párrafo.

La oración 1 del párrafo 10 no esta macada para resumen y tampoco es candidata por lo tanto se realiza el promedio con la oración que contiene el marcador: 0.34189406 + 0.07865168 / 2 = 0.21027288. Este promedio NO supera a la ponderación de la peor oración candidata que es 0.24972814, por lo tanto no se marca ninguna de las dos oraciones para resumen. Al no marcar ninguna oración, no se modifica el valor de la peor ponderación.

Oraciones marcadas hasta el momento: 7.

Estudio de la oración 1 del párrafo 3:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 8.

Estudio de la oración 1 del párrafo 9:

Comienza con marcador, la oración es la primera del párrafo, por lo cual se estudian las oraciones del párrafo anterior: el párrafo 8 no tiene oraciones candidatas, por lo cual se toma la oración con mayor ponderación de este párrafo (la oración 1), y se calcula el promedio entre las candidatas del párrafo 9 y la oración 1 del párrafo 8. El promedio es: 0.26716605 + 0.25630125 + 0.14044942 / 3 = 0.22130, este promedio No supera a la ponderación de la peor oración candidata que es 0.24972814, por lo tanto no se marca ninguna de las oraciones para resumen.

Oraciones marcadas hasta el momento: 8.

Estudio de la oración 2 del párrafo 9:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 9.

Estudio de la oración 2 del párrafo 5:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 10.

Estudio de la oración 3 del párrafo 4:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 11.

Estudio de la oración 1 del párrafo 6:

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 12.

Estudio de la oración 1 del párrafo 8:

Dicha oración no pertenece a la lista de oraciones candidatas. El estudio de las oraciones candidatas se terminó, por lo tanto se continúa con las oraciones no candidatas por ponderación descendente.

No comienza con marcador, por lo tanto se marca para resumen.

Oraciones marcadas hasta el momento: 13.

El total de oraciones marcadas para resumen es 13 (equivalente al 50% de las oraciones, porcentaje de resumen), por lo tanto se terminó el estudio por coherencia → FIN del estudio.

A continuación se presentará una comparación entre el resultado del resumen por ponderación (oraciones candidatas) y el resumen por coherencia, sobre todo en las oraciones que contienen marcadores y fueron candidatas.

La oración 2 del párrafo 4 comienza con el marcador binario "Sin embargo". En el resumen por ponderación se marca para resumen dicha oración, pero no se marca la

oración inmediata anterior. Por esto, dicha oración quedaría "conectada" a las oraciones del párrafo anterior (párrafo 3), aquí se presentaría una clara falta de coherencia. En el resumen por coherencia, cuando se estudia dicha oración, se decide marcarla para resumen junto con la oración inmediata anterior, por lo tanto se mantiene la coherencia del resumen.

Para la oración 2 del párrafo 10 sucede algo similar. Dicha oración comienza con el marcador "No obstante"; en el resumen por ponderación ésta se marca para resumen, pero no se marca la oración inmediata anterior por lo que se presenta cierta incoherencia ya que la oración que contiene el marcador quedaría conectada al párrafo 9. El resumen por coherencia NO marca ninguna de las dos oraciones para resumen, por lo tanto se mantiene la coherencia del resumen.

La oración 1 del párrafo 9 comienza con marcador y es la primera del párrafo, por lo tanto esta fuertemente conectada a lo expuesto en el párrafo anterior (párrafo 8).

El resumen por ponderación marca dicha oración como parte del resumen final, pero no marca ninguna oración del párrafo 8 ni del párrafo 7, por lo tanto la oración que contiene el marcador discursivo, queda "conectada" con el párrafo 6, por lo que se identifica cierta incoherencia en el resumen. El resumen por coherencia, al enfrentarse a este caso, no incluye la oración para resumen final por lo tanto se mantiene la coherencia en el resumen.

CASO 2: NO REAL

"*…*

Los agresores, muchos de ellos enmascarados, destrozaron todo a su paso, "a pedradas" y "patadas" parabrisas de autos, vidrieras de bancos, dependencias públicas, comercios, restoranes de comida rápida y de la Bolsa de Valores capitalina, además de agredir a cuanta persona se les cruzaba. Hubo quienes "grafitearon" consignas antiimperialistas, contrarias al presidente norteamericano e impactaron "bombas de alquitrán" y de pintura roja en cuanta pared o muro por donde pasaban, incluyendo a una dependencia. Por lo tanto, varios propietarios de comercios y de vehículos reaccionaron ante los desmanes que minutos después fueron controlados por la Policía. El ministro del Interior, José Díaz, repudió los hechos y anunció que aplicará la ley "con todo rigor".

Mientras tanto, sobre las 16:00 horas, comenzaron a marchar -al son del tamboril por la peatonal hacia el sur con la intención de hacerlo por toda la ciudad, para luego desplazarse hasta la sede de la embajada norteamericana, mientras que a su paso continuaban entonando las "agresivas" consignas. Todo venía "bien" hasta que "voló" una piedra. En su camino, la turba agredía a cuanta persona se les cruzaba, además de realizar pintadas en muros y paredes, como, por ejemplo, "muerte al capitalismo".

Adicionalmente, apenas pasadas las 17:00 horas comenzaron a nuclearse en la Plaza Matriz los manifestantes que deseen oír su voz de denuncia y resistencia, los cuales fueron convocadas por al menos una organización denominada ALCArajo!, la cual en su sitio web (www.alcarajo.entodaspartes.org) promocionaba la concentración. Un pequeño grupo, que se reunió en el lugar, comenzó a entonar consignas agresivas hacia el "régimen imperialista y el presidente de EEUU Bush", y tras unos pocos minutos se les sumaron varias decenas de mujeres, niños, jóvenes, muchos de los cuales traían sus rostros cubiertos. Que era en respuesta de la profunda disconformidad hacia la IV Cumbre de las Roberto Américas, que culmina hoy en la ciudad argentina de Mar del Plata, y a la presencia del primer mandatario estadounidense en la misma: "Los poderosos se encuentran para fortalecer la globalización neoliberal que beneficia a las potencias y sus corporaciones. No al ALCA y a la militarización de nuestros pueblos!.

….."

Esto es parte del documento "*Marcha anti Bush desató el caos y violencia en Ciudad Vieja*" obtenido del Diario "La República". El texto se modificó para realizar pruebas del algoritmo de coherencia y solamente se muestran tres párrafos.

El porcentaje de resumen utilizado fue de 80%, los coeficientes se tomaron al 100% y no se utilizó la consulta de usuario.

<u>Párrafo 3</u>: compuesto por 4 oraciones

"*Los agresores, muchos de ellos enmascarados, destrozaron todo a su paso, "a pedradas" y "patadas" parabrisas de autos, vidrieras de bancos, dependencias públicas, comercios, restoranes de comida rápida y de la Bolsa de Valores capitalina, además de agredir a cuanta persona se les cruzaba.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
3	1	No	42	0.118041575	si

"*Hubo quienes "grafitearon" consignas antiimperialistas, contrarias al presidente norteamericano e impactaron "bombas de alquitrán" y de pintura roja en cuanta pared o muro por donde pasaban, incluyendo a una dependencia*".

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
3	2	No	32	0.07746479	si

"*Por lo tanto, varios propietarios de comercios y de vehículos reaccionaron ante los desmanes que minutos después fueron controlados por la Policía.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
3	3	Si	22	0.12259923	si

"*El ministro del Interior, José Díaz, repudió los hechos y anunció que aplicará la ley "con todo rigor".*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
3	4	No	17	0.2634631	si

<u>Párrafo 4</u>: compuesto por 3 oraciones

"*Mientras tanto, sobre las 16:00 horas, comenzaron a marchar -al son del tamboril por la peatonal hacia el sur con la intención de hacerlo por toda la ciudad, para luego desplazarse hasta la sede de la embajada norteamericana, mientras que a su paso continuaban entonando las "agresivas" consignas.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
4	1	Si	48	0.07394366	si

"*Todo venía "bien" hasta que "voló" una piedra.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
4	2	No	8	0.0070422534	no

"*En su camino, la turba agredía a cuanta persona se les cruzaba, además de realizar pintadas en muros y paredes, como, por ejemplo, "muerte al capitalismo".*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
4	3	No	26	0.03521126	no

<u>Párrafo 5</u>: compuesto por 4 oraciones

"*Adicionalmente, apenas pasadas las 17:00 horas comenzaron a nuclearse en la Plaza Matriz los manifestantes que deseen oír su voz de denuncia y resistencia, los cuales fueron convocadas por al menos una organización denominada ALCArajo!, la cual en su sitio web (www.alcarajo.entodaspartes.org) promocionaba la concentración.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	1	Si	45	0.15226917	si

"*Un pequeño grupo, que se reunió en el lugar, comenzó a entonar consignas agresivas hacia el "régimen imperialista y el presidente de EEUU Bush", y tras unos pocos minutos se les sumaron varias decenas de mujeres, niños, jóvenes, muchos de los cuales traían sus rostros cubiertos.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	2	No	45	0.14303601	si

"*Que era en respuesta de la profunda disconformidad hacia la IV Cumbre de las Roberto Américas, que culmina hoy en la ciudad argentina de Mar del Plata, y a la presencia del primer mandatario estadounidense en la misma: "Los poderosos se encuentran para fortalecer la globalización neoliberal que beneficia a las potencias y sus corporaciones.*"

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	3	No	52	0.14030334	si

"*No al ALCA y a la militarización de nuestros pueblos!"."*

Párrafo	Oración	Comienza marcador	Total palabras	Ponderación total	Candidata
5	4	No	10	0.0035211267	no

En la Figura 30 se presentan las oraciones candidatas ordenadas descendentemente por ponderación total:

Párrafo	Oración	Ponderación
2	1	0.48568076
6	4	0.27284822
3	4	0.2634631
6	1	0.25391236
7	4	0.21370532
6	3	0.2113794
12	1	0.19577464
6	2	0.19561815
7	3	0.17150463
5	1	0.15226917
2	2	0.15112677
7	5	0.14632238
5	2	0.14303601
8	2	0.14208221
5	3	0.14030334
10	1	0.13108653
3	3	0.12259923
8	3	0.12046521
3	1	0.118041575
11	1	0.08450705
9	1	0.08450703
3	2	0.07746479
7	6	0.07746479
4	1	0.07394366
7	1	0.07042254

Figura 30 – Oraciones candidatas para el caso 2

El estudio de coherencia comienza con la oración mas ponderada (en el ejemplo, la oración 1 del párrafo 2), luego continúa por la siguiente mejor ponderada y así sucesivamente hasta cumplir con el porcentaje ingresado por el usuario o hasta que se termine el estudio de las oraciones.

A continuación se presenta el estudio de la oración 1 del párrafo 5, por ser ésta la oración que genera el estudio o situación que se quiere ejemplificar.

Estudio de la oración 1 del párrafo 5:

La oración comienza con marcador binario y es la primera del párrafo, por lo tanto se debe estudiar la inclusión en el resumen de las oraciones del párrafo anterior (párrafo 4).

El párrafo 4 contiene una oración candidata (ver Figura 30) que es la que se trata de incluir en el resumen. Dicha oración comienza con marcador discursivo binario y no ha sido marcada para resumen (aún no se le ha hecho el estudio de coherencia), por lo tanto se debe estudiar la inclusión en el resumen de alguna oración del párrafo anterior (párrafo 3).

El párrafo 3 ya tiene una oración marcada para resumen (oración 4) por lo tanto se marca la oración 1 del párrafo 4 y se intenta marcar las demás candidatas de dicho párrafo (la única candidata es la oración 1 que ya fue marcada).

A continuación se marca la oración 1 del párrafo 5 y se intenta marcar las demás candidatas de dicho párrafo.

En conclusión, en el estudio de la oración 1 del párrafo 5 se abarcaron 3 párrafos, y se marcaron para resumen 4 oraciones (oración 1 del párrafo 4 y las oraciones 1, 2 y 3 del párrafo 5) manteniéndose así la coherencia en el resumen.

5.3. Pruebas y resultados del resumidor

A continuación se muestran los resultados obtenidos con los diferentes textos.

La metodología fue la siguiente: se buscaron textos de noticias en Internet y se entregaron a dos personas ajenas al grupo (Maestras de Educación Primaria), a las cuales se les solicitó realizar resúmenes por extracción de oraciones enteras, sin mencionar ninguna de las técnicas que utiliza el sistema para resumir (técnicas de ponderación y coherencia).

De un total de 30 textos se repartieron 20 a cada una de las dos personas (10 de estos textos fueron entregados a ambas personas para realizar una comparación entre ambos resúmenes y el resumen del sistema), obteniéndose un total de 40 resúmenes (casos de prueba).

Luego, para cada texto, se contaron las oraciones que forman parte del resumen realizado por la persona para calcular el porcentaje de resumen a ingresar en el sistema (cantidad de oraciones que seleccionó la persona * 100 / total de oraciones del texto). Con este dato se realizó el resumen automático del texto (sistema) y se contó la cantidad de oraciones coincidentes -seleccionadas para el resumen- entre el resumen manual y el resumen del sistema. Luego, se realizó un chequeo de las palabras (nombres y sustantivos) que no se tomaron en cuenta por no estar en la base de datos, las mismas se ingresaron y se realizó el resumen nuevamente. Finalmente, se comparó éste último con el resumen humano y se contaron las oraciones coincidentes.

A continuación se presentan algunos Casos de estudio, donde se muestran los siguientes datos:

- Título del texto: es el título del texto al que se le realizó la prueba.
- Origen: Es la fuente u origen de donde se obtuvo el texto.
- Total de oraciones: total de oraciones que componen el texto fuente.
- Cantidad de uniones: es la cantidad de oraciones "arregladas" por alguno de los problemas descritos en el punto "Rearmado de oraciones" en la sección .

- Oraciones propuestas en resumen manual: cantidad de oraciones que comprenden un resumen manual.
- Porcentaje resumen: es el valor de porcentaje utilizado para realizar el resumen en el sistema.

Los coeficientes de los ponderadores se utilizaron todos en 100%.

Como ya se mencionó, el sistema genera dos tipos de resúmenes: por ponderación y por coherencia. En los casos que se presentan a continuación, cuando se mencione la cantidad de coincidencias, se mencionará la cantidad de coincidencias entre el resumen humano y el resumen del sistema. En caso de que el resumen por ponderación y por coherencia no tengan la misma cantidad de coincidencias, se mencionará la comparación por separado.

En esta sección se analizan 5 casos, el texto completo de cada uno y su respectivo resumen del sistema se encuentran en el capítulo (Apéndice). El resto de los casos se encuentran en el Anexo 2.

Caso 1

Título del texto: "El pesquero argentino embistió al Eladia Isabel".

Origen: Diario La República.

Total de oraciones: 40

Cantidad de uniones: 4

Oraciones propuestas en resumen manual: 18

Porcentaje de resumen: 46%

Cantidad de oraciones coincidentes: 8 – 44.5%

La poca coincidencia entre el resumen humano y el del sistema puede ser debida a una excesiva ponderación por posición. Dicha ponderación da un valor a la primera oración de 0.3 y resta para las siguientes oraciones del párrafo el valor 0.05 (según fórmula propuesta en el punto "Ponderación por posición" de la sección). Se cambiaron los valores a 0.4 y 0.1 respectivamente, obteniéndose el mismo resultado por lo cual se constató que el problema continuaba y se cambiaron nuevamente los valores a 0.15 para la primera oración del párrafo y restando de a 0.02 para el resto de las oraciones del párrafo. Los nuevos datos obtenidos fueron:

Cantidad de oraciones coincidentes: 10 – 55.6%

La coincidencia de oraciones fue mejor, superando el 50%. Los valores para la ponderación por posición se mantuvieron para el resto de las pruebas.

Caso propuesto por la segunda persona:

Oraciones propuestas en resumen manual: 23

Porcentaje de resumen: 58%

Cantidad de oraciones coincidentes: 15 – 65.3%

Caso 2

Título del texto: "Una marcha anti Bush desató el caos y violencia en Ciudad Vieja."

Origen: Diario "La República"

Total de oraciones: 28

Cantidad de uniones: 6

Oraciones propuestas en resumen manual: 11
Porcentaje de resumen: 40%

Cantidad de oraciones coincidentes: 6 – 54.6%.
Las coincidencias superan el 50%

Caso 3

Título del texto: "Y el Corto agarró y se fue."
Origen: Diario "La República"
Total de oraciones: 24
Cantidad de uniones: 7
Oraciones propuestas en resumen manual: 12
Porcentaje de resumen: 50%

Cantidad de oraciones coincidentes: 9 – 75%
En este caso podemos notar la gran coincidencia entre el resumen propuesto por una persona y el resumen que propone el sistema.
Luego de ingresar sustantivos y nombres que no se encontraban en la base de datos se obtuvieron nuevos resultados:

Cantidad de oraciones coincidentes: 11 – 91.7%
Se observa que luego de ingresar las palabras faltantes las coincidencias aumentaron hasta alcanzar casi el 100%

Caso 4

Título del texto: "La tiranía de Israel sobre Estados Unidos."
Origen: Página Digital – Noticias y Artículos
Total de oraciones: 55
Cantidad de uniones: 0
Oraciones propuestas en resumen manual: 25
Porcentaje de resumen: 46%

Cantidad de oraciones coincidentes: 14 – 56%
La cantidad de oraciones coincidentes supera el 50%

Luego de ingresado los nombres y sustantivos que faltaban en la base de datos se obtuvo el mismo resultado.

Caso 5

Título del texto: "El último leño a la hoguera."
Origen: Página Digital – Noticias y Artículos

Total de oraciones: 51
Cantidad de uniones: 0
Oraciones propuestas en resumen manual: 30
Porcentaje de resumen: 59

Cantidad de oraciones coincidentes: 15 – 50%

Caso propuesto por la segunda persona:
Oraciones propuestas en resumen manual: 39
Porcentaje de resumen: 77%

Cantidad de oraciones coincidentes por ponderación: 32 – 82.1%
Cantidad de oraciones coincidentes por coherencia: 33 – 84.6%
Aquí el resumen por ponderación y el resumen por coherencia dieron diferentes valores de coincidencia con el resumen propuesto por la persona. Ambos obtuvieron una coincidencia mayor al 80%

Casos realizados por ambas personas

En este punto, se presentan los casos compartidos por ambas personas, comparando las coincidencias entre ellos y comparando los resultados de cada uno con el resumen automático. Se realizaron un total de 10 casos en común, con los cuales se pretende observar si la forma de resumir de ambas personas es homogénea.

En la Figura 31 se presenta una tabla con los resultados obtenidos la cual tiene las siguientes columnas:

- Nº Caso: es el número de caso que se entregó a ambas personas, se encuentran en la sección o en el Anexo 2.
- Total Oraciones: es el total de oraciones del texto fuente.
- Cantidad coincidencias: es la cantidad de coincidencias -de oraciones seleccionadas- entre ambos resúmenes manuales.
- Estimación-Porcentaje: es una estimación de la cantidad de coincidencias en caso de que ambas personas hubieran seleccionado la misma cantidad de oraciones (tomando el mayor porcentaje). El cálculo de la estimación para el caso 1 en la Figura 31 es: 12 * 23 / 18 = 15.3 coincidencias (este valor sería la cantidad de coincidencias en caso de que la persona que seleccionó menor porcentaje de oraciones hubiera seleccionado el porcentaje de la otra persona). El segundo valor de la columna es el porcentaje de coincidencias de la estimación.
- Cantidad-Porcentaje Resumen Persona 1: es la cantidad de oraciones seleccionadas para el resumen y porcentaje de éste (utilizado para el resumidor automático).
- Coincidencias Persona 1: es la cantidad de oraciones coincidentes entre el resumen de la persona 1 y el resumidor.
- Cantidad-Porcentaje Resumen Persona 2: es la cantidad de oraciones seleccionadas para el resumen y porcentaje de éste (utilizado para el resumidor automático).
- Coincidencias Persona 2: es la cantidad de oraciones coincidentes entre el resumen de la persona 2 y el resumidor.

En caso de que las coincidencias entre el resumen por ponderación y resumen con coherencia sean distintas, se presentan los datos en dos renglones, primero los datos por ponderación y luego coherencia (columnas 5 y 7 para cada persona).

Nº Caso	Total Oraciones	Cantidad Coincidencias	Estimación-Porcentaje	Cantidad-Porcentaje Resumen Persona 1	Coincidencias (cantidad-porcentaje) Persona 1	Cantidad-Porcentaje Resumen Persona 2	Coincidencias (cantidad-porcentaje) Persona 2
1	40	12	**15.3-66.5%**	23 – 58%	15 – 65.3%	18 – 45%	10 – 55.6%
5	51	22	**28.6-73.3%**	30 - 59%	15 – 50%	39 - 77%	32 – 82.1% 33 – 84.6%
6	27	14	**14.8-78%**	18 - 67%	13 – 72.2%	19 - 70%	14 – 73.7%
7	44	14	**14 - 54%**	26 - 60%	21 – 80.8%	26 - 60%	18 – 69.2%
9	67	22	**23.9-64.6%**	37 - 55%	22 – 59.6%	34 - 51%	18 – 53%
10	27	15	**16.8-88.4%**	17 – 63%	11 – 64.7% 12 – 70.6%	19 – 71%	12 – 63.2%
17	18	7	**8.2-58.6%**	12 – 67%	8 – 66.7%	14 – 78%	11 – 78.6%
18	27	9	**11.6-64.4%**	14 – 52%	8 – 57.1%	18 – 67%	12 – 66.7%
19	32	10	**16.4-71.3%**	14 – 44%	7 – 50%	23 – 72%	17 – 74%
20	30	17	**19.4-80.8%**	21 – 70%	13 – 61.9%	24 – 80%	19 – 79.2%

Figura 31 – Casos compartidos

Tomando el porcentaje de estimación se obtiene un promedio de 70% de coincidencias entre los resúmenes de ambas personas. Este valor nos indica que el estilo de resumen de ambas personas es relativamente homogéneo, por lo que se tiene cierto margen de diferencia al momento de evaluar los resultados del resumidor.

En la sección se mencionan las principales observaciones y conclusiones que se obtienen de los resultados.

5.3.1. Conclusiones de las pruebas

Como conclusión se puede mencionar que de las 40 pruebas, el 100% superó el 50% de coincidencias (dos casos obtuvieron 50% de coincidencias). Además, el 45% de las pruebas superó el 70% de coincidencias con el resumen manual, y dentro de este grupo, el 61.1% superó el 75% de coincidencias.

El valor de coincidencias crece con el aumento del porcentaje de resumen. O sea, cuanto más grande es el porcentaje de resumen (el resumen se compone por más oraciones) mayor es el valor de coincidencias, lo cual es razonable.

En la gráfica de la Figura 32, se verifica lo expuesto en el caso de los resúmenes por coherencia:

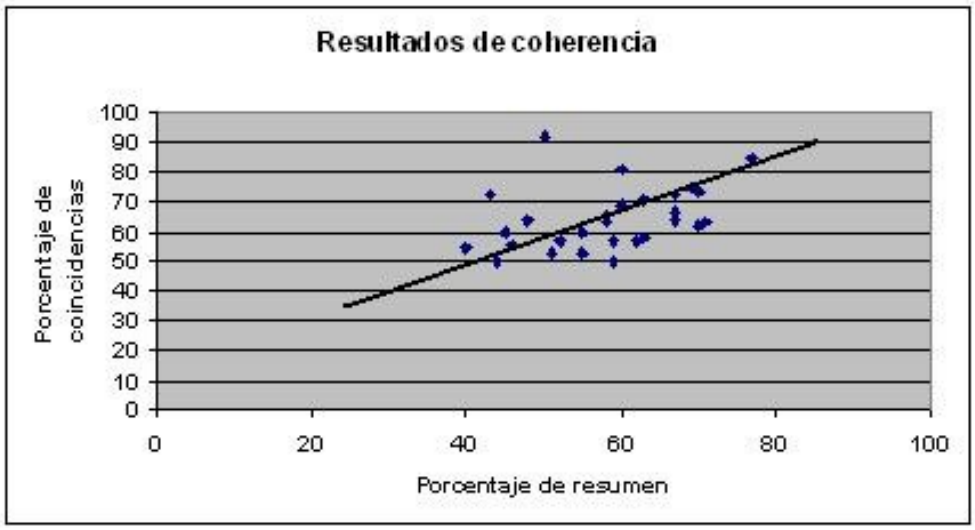

Figura 32 – Gráfico de Coincidencias vs. Porcentaje (Coherencia)

En la gráfica de la Figura 33, se verifica lo expuesto en el caso de los resúmenes por ponderación:

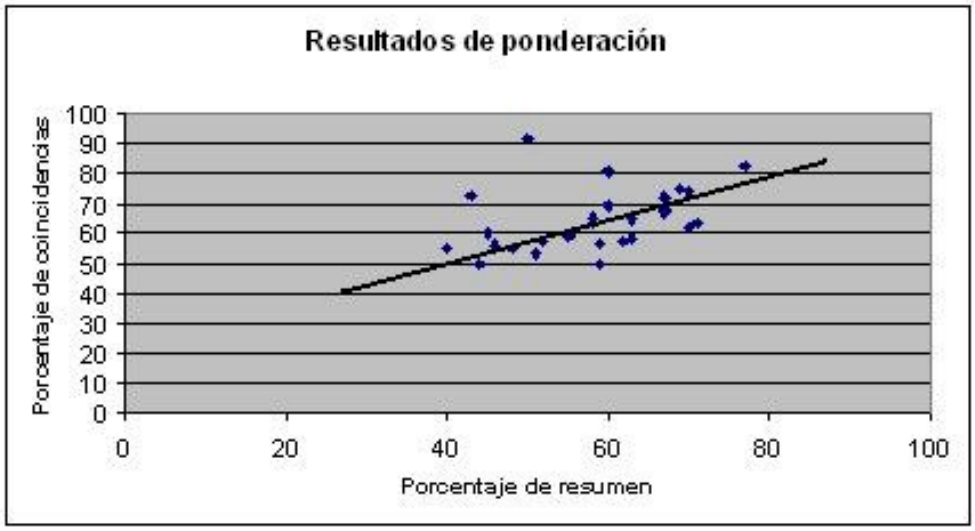

Figura 33 - Gráfico de Coincidencias vs. Porcentaje (Ponderación)

Para ambas gráficas, cada punto representa el porcentaje de resumen vs. porcentaje de coincidencias con el resumen manual para un caso descrito en la sección .

Se comprueba la tendencia al aumento del porcentaje de coincidencias cuando aumenta el porcentaje de resumen.

La media de los porcentajes de coincidencia obtenidos es de **68%** para ambos tipos de resúmenes, teniendo en cuenta que los resúmenes humanos pueden variar de una persona a otra (ver Figura 31), podemos afirmar que los resultados del resumidor son relativamente buenos.

Cabe destacar que en los casos donde los porcentajes de coincidencias entre el resumen manual y ambos tipos de resúmenes son iguales, no siempre se seleccionan las mismas oraciones, dado que la existencia de marcadores discursivos condiciona la selección de las oraciones.

Todos los casos de incoherencia debido a los marcadores discursivos en el resumen por ponderación, fueron solucionados en el resumen por coherencia.

6. Conclusiones y trabajo a futuro

En este capítulo se presentan las conclusiones obtenidas sobre el problema, producto y las tecnologías utilizadas. Se presentan además, posibles mejoras y extensiones a realizar sobre el proyecto.

6.1. Conclusiones

Como se puede apreciar en los dos casos descritos en la sección y en los casos de prueba (de coherencia y del resumidor) presentados en el capítulo , siempre se obtuvieron resultados positivos en la resolución de los problemas de cohesión y coherencia, que se presentaron por la ocurrencia de marcadores discursivos (binarios). Este es el principal aporte de la propuesta.

Se obtuvieron resultados satisfactorios de coincidencias entre resúmenes manuales y los propuestos por el sistema, 45% supera el 70% de coincidencias, ver .

En todos los casos de prueba se solucionaron los problemas de coherencia introducidos por las ocurrencias de los marcadores discursivos del tipo binario. Si bien el 7.5% (2 casos de 40, valor relativamente pequeño) de los casos de estudio de coherencia tuvieron un porcentaje menor de coincidencias con el resumen manual en comparación con las coincidencias del resumen ponderado, se observó que la calidad del resumen mejora en términos de coherencia.

En muchas de las pruebas (87.5%) se obtuvieron resultados idénticos de coincidencias entre el resumen manual y el ponderado, con el resumen manual y el resumen por coherencia. Esto sucede por tres principales motivos:

- Por la inexistencia de marcadores discursivos en los textos (resumen ponderado es idéntico al resumen por coherencia).
- En el estudio de coherencia, la oración que antecede a una oración que contiene un marcador (binario) es candidata, por lo tanto también conforma el resumen por ponderación.
- Misma cantidad de coincidencias pero las oraciones que componen el resumen ponderado y el resumen por coherencia difieren debido al estudio de coherencia.

Aunque no todos los problemas de coherencia del resumen son solucionados con la técnica propuesta –pueden existir problemas de coherencia por la ocurrencia de anáforas por ejemplo-, mediante la utilización de las funciones o propiedades que cumplen los marcadores discursivos presentes en los textos, se contribuye con el fin de mantener la coherencia del producto, y por lo tanto, mejorar la legibilidad del resumen propuesto.

Aunque el sistema no presenta una solución a las posibles incoherencias introducidas al resumen por los marcadores discursivos del tipo compuesto, se hizo un análisis y tratamiento de dichos marcadores para resolver el problema. La implementación de la solución a los problemas de incoherencias se enfocó hacia los marcadores discursivos del tipo binario, por ser los más frecuentes.

Se implementó un segmentador utilizando las herramientas JLex[14] y Java. Como ya se mencionó, resta solucionar algunos problemas existentes. Ejemplo: algunas ocurrencias

del carácter punto (".") que no indican la finalización de una oración (punto "Rearmado de oraciones" en la sección .). Una virtud de dicho reconocedor es que la lectura del texto no insume mayor costo, pero tiene la contra de que necesita una etapa de post – reconocimiento para corregir alguna "mala" interpretación del texto, como la recién mencionada.

La aplicación permite obtener el texto de entrada etiquetado (identificando párrafos, oraciones, etc.) presentado en formato XML con la idea de poder ser utilizado en otras aplicaciones.

Es posible realizar una fácil adaptación del sistema a otros lenguajes, solo bastaría con mantener la base de datos con los nombres, sustantivos y marcadores discursivos en el correspondiente idioma.

6.2. Trabajo a futuro

Los siguientes son algunos puntos que se podrían incorporar en un futuro para introducir mejoras al sistema resumidor:

- Implementación de una solución a la coherencia basada en marcadores discursivos compuestos (en este trabajo se presentó una propuesta o análisis del tratamiento de los marcadores compuestos).

- Incursión en otros métodos de ponderación de oraciones para tratar de obtener mejores resultados y así aumentar la cantidad de coincidencias con un resumen manual.

- Realizar pruebas del resumidor entregando mayor cantidad de textos a resumir a mayor cantidad de personas para obtener más datos y así realizar una evaluación del sistema más completa.

- Realizar mejoras al reconocedor del texto como por ejemplo: permitir seleccionar textos con diferentes formatos (no solo texto plano), distinguir que función esta cumpliendo una palabra que puede ser verbo o sustantivo (mediante métodos de exploración contextual), solucionar eficientemente alguno de los problemas planteados en el punto "Reconocimiento de nombres y sustantivos" de la sección .

- Crear un mecanismo de autorregulación de los parámetros o coeficientes de los ponderadores para que a partir de un conjunto de textos de prueba se obtenga la mejor combinación de los coeficientes (combinación donde la media o promedio del porcentaje de coincidencias entre el resumen del sistema y el resumen manual sea el mejor).

- Estudiar la posibilidad de adaptar al sistema, un tagger (etiquetador) de texto como Freeling[15] para evitar los problemas de interpretación del texto que tiene el reconocedor actual.

7. Glosario

Oración: es la menor unidad del habla que transmite un mensaje completo por sí misma, por lo que se dice que tiene sentido completo e independencia sintáctica. Las oraciones comienzan con mayúscula y terminan con un punto. El punto se utiliza para dar fin a una oración e indicar una pausa entre las distintas ideas que se expresan.

Score o ponderación: puntaje o valor que se le da a las oraciones en los diferentes métodos (palabras claves, palabras del título, etc.). Sinónimo de peso.

Marcadores discursivos: conjunto de términos que establecen relaciones entre segmentos textuales. Operadores que añaden estructura al texto.

Anáforas: función de ciertas palabras que asumen el significado de una parte del discurso ya emitida.

Tf.idf: método de ponderación de oraciones el cual recompensa la relativa frecuencia de una palabra en un documento (cuando se trata de un resumen multidocumento) ó párrafo (cuando se trata de un resumen de un solo documento).

Sustantivos: es la parte variable de la oración que sirve para nombrar y dar a conocer las personas, animales y cosas materiales e inmateriales. Son sustantivos por ejemplo, mujer, loro, lápiz, colombianos, escritor, etc. Los sustantivos abstractos (inmateriales) nombran cosas que no pueden ser percibidos en forma directa por los sentidos.

Nombres: se refiere a cualquier nombre propio, entre ellos, personas, regiones o países, siglas, cargos, etc.

JLex: es un generador de analizadores lexicográficos para Java, se utiliza para reconocer tokens a partir de una secuencia de caracteres, utilizando expresiones regulares.

XML: es el estándar de Extensible Markup Languaje. XML no es más que un conjunto de reglas para definir etiquetas semánticas que nos organizan un documento en diferentes partes. XML es un metalenguaje que define la sintaxis utilizada para definir otros lenguajes de etiquetas estructurados.

8. Referencias

[1]Prada, Juan José, "Marcadores del discurso en español, análisis y representación" – Tesis de maestría, Facultad de Ingeniería ROU- Octubre 2001.

[2]Zorraquino, M. A. y J. Portolés Lázaro (1999), "Los marcadores del discurso" en I. Bosque y V. Demonte (eds), "Gramática descriptiva de la lengua española", Madrid, España. http://kane.uab.es/cursos.lengua/Marcadores.pdf -
Último acceso: 18/05/2005

[3]Poblete B, María Teresa. "La cohesión de los marcadores discursivos en distintos tipos de discurso". http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0071-17131999003400012&lng=es&nrm=iso – Último acceso: 19/04/2005

[4]Hovy E., Chin-Yew Lin. Information Sciences Institute of the University of Southern California.

[5]Mateo P., González J, Villena J, Martínez J. – "Un sistema para resúmenes de texto en castellano". www.sepln.org/revistaSEPLN/revista/31/31-Pag29.pdf -
Último acceso: 19/04/2005.

[6]Acero I, Alcojor M, Díaz A, Gómez J. – "Generación automática de resúmenes personalizados".www.sepln.org/revistaSEPLN/revista/27/27-articulo33.pdf-24/06/2005

[7]Maña M – "Generación automática de resúmenes de texto para el acceso a la información".

[8]Luhn, H. P. 1958. "The Automatic Creation of Literature Abstracts". *IBM Journal of Research Development*.

[9]Edmundson, H. P. 1969. "New Methods in Automatic Extracting". *Journal of the Association for Computing Machinery*.

[10]Georgantopoulos B. MSc in Speech and Languaje Processing Dissertation: "Automatic summarizing based on sentence extraction: A statistical approach". http://apollo.u-gakugei.ac.jp/~jingjing/relation_work/Georgantopoulos96automatic.pdf
Último acceso: 23/06/2005

[11]Santos A. Larocca J., Kaestner C, Freitas A., Nievola J. "A trainable algorithm for summarizing news stories".http://citeseer.lcs.mit.edu/neto00trainable.html-
Último acceso: 23/06/05

[12]Kupiec J., Pedersen J., ChenF. "A Trainable Document Summarizer". www.dcs.shef.ac.uk/~mlap/teaching/kupiec95trainable.pdf -
Último acceso: 23/06/2005

[13]Marcu, D. 1997. "The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts". Ph.D. dissertation, University of Toronto.

[14]JLex: generador de analizadores lexicográficos para Java, se puede obtener en http://www.cs.princeton.edu/~appel/modern/java/JLex

[15]Freeling: segmentador o etiquetador de textos, se puede obtener en http://garraf.epsevg.upc.es/freeling/

[16]Desclés, J.P.; Cartier, E.; Jackiewicz, A.; Minel, J.L. (1997) "Textual Processing and Contextual Exploration Method" Context97, Río de Janeiro

9. Apéndice

En este apéndice se presentan los textos del caso 1 al caso 5 (descritos en la sección) junto a sus respectivos resúmenes.

Caso 1

Título del texto: "El pesquero argentino embistió al Eladia Isabel".

Texto:

"*El barco Eladia Isabel de la empresa Buquebus recibió en la madrugada de ayer un fuerte impacto por parte de un barco pesquero de bandera argentina denominado "Depemás 5". Los daños no fueron importantes, pero Buquebus realizará una investigación al respecto. Altas fuentes consultadas indicaron que el accidente se produjo por una "mala maniobra" del pesquero. Llama la atención de expertos que el pesquero argentino se pusiera delante de la proa del barco de pasajeros.*

El barco se sacudió y los pasajeros gritaron aterrados. Aunque era impensable que se tratara de un iceberg, muchos recordaron las escenas de "Titanic". Un pesquero habría rozado al Eladia Isabel, de Buquebus. El golpe no fue muy fuerte, pero alcanzó para que el barco cargado con 691 pasajeros quedara sin luz y varado frente a la costa porteña.

El pesquero de bandera argentina no sufrió muchos daños y contaba con una tripulación de 29 marinos. Dicho barco mide 45,05 metros de eslora y 8,60 de manga.

Había pasado poco más de media hora cuando los efectivos de Prefectura llegaron al lugar con la intención de atender a los heridos, que afortunadamente no existieron. La enorme embarcación, -Eladia Isabel es una de las embarcaciones más importantes de la flota de Buquebus-, debió ser remolcada y finalmente llegó a la dársena norte del puerto argentino cerca de las 3 de la mañana. Los medios de la vecina orilla captaron las imágenes de los pasajeros que no esperaban tanto "alboroto".

Mientras una versión narra que el accidente se produjo a raíz de una falla técnica en el instrumental de navegación, pero sin poder determinar de cuál de los dos barcos, otras fuentes consultadas indicaron que el accidente se produjo por una "mala maniobra" del pesquero". Con respecto a los daños sufridos, se supo que dos de los 55 autos que estaban en la bodega del Eladia Isabel fueron averiados por el impacto. "Los daños serán cubiertos por el seguro", confirmó el vocero de la empresa Buquebus que hoy atendía los reclamos de los pasajeros en las oficinas que la empresa tiene en Puerto Madero.

Por su parte, Buquebus emitió un comunicado en el que explica que el buque Eladia Isabel había partido desde el puerto de Colonia del Sacramento a las 19.30 horas. A las 22.20 horas, cuando se encontraba a sólo 6 kilómetros del puerto de Buenos Aires, y por motivos que se desconocen y "son objeto de investigación", un buque pesquero lo impactó en uno de sus laterales. La colisión fue leve, ocasionando solo algunos daños menores en la bodega y algunos sistemas eléctricos.

En consecuencia, se trabajó en forma conjunta con Prefectura Naval Argentina y dos remolcadores se encargaron de trasladar a ambas embarcaciones hacia el Puerto de Buenos Aires. Afortunadamente, los sistemas eléctricos fueron reparados durante el viaje, lo que posibilitó que el buque amarrara en puerto con sus sistemas propios.

En ese momento viajaban en el barco aproximadamente 600 pasajeros (la mitad de plazas disponibles), que fueron atendidos y contenidos por miembros de la tripulación, no habiéndose registrado heridos con motivo del accidente. Buquebus informó que los pasajeros que tenían programado su viaje a las 0.30 horas, viajaron en buques rápidos que salieron durante la madrugada desde el puerto de Buenos Aires, cumpliendo de esta manera con todos los servicios comprometidos. Lo mismo sucedió en la mañana de ayer, cuando se reubicó a los pasajeros que debían viajar en el Eladia Isabel, en otros buques rápidos. Por esto, se indicó que la situación se "encuentra normalizada" y las frecuencias previstas "se están respetando" mientras que los pasajeros que tenían pasaje para dicho buque fueron reubicados.

En estos momentos se están respetando las frecuencias previstas, por lo que no se "esperan mayores complicaciones". Buquebus expresó que las reparaciones "ya se iniciaron", aguardando ahora que el buque Eladia Isabel vuelva a su frecuencia habitual "en algunas horas".

Por su parte el jefe de la Prefectura del puerto de Buenos Aires, José Romero, declaró ayer a Clarín.com que "están investigando las causas del accidente. "Hubo un error de maniobra, pero hay que determinar de quién", afirmó. "Ambos barcos ya están en el Puerto para los peritajes correspondientes", dijo. Esta fuente coincide con una de las versiones manejadas hasta ahora. Aparentemente, se habría tratado de una falla técnica en el instrumental de navegación, pero la confirmación de la causa del choque aún se intenta determinar.

El capitán de ultramar de la Marina Mercante, Oscar Lebel, consultado por LA REPUBLICA explicó: "Desde Montevideo a Buenos Aires los barcos van por un canal dragado y balizado, o sea que tiene boyas de ambos lados. Esto hace que el barco "ande por el mar como anda un automóvil por una calle que tiene faroles de ambos lados". El canal hacia la ciudad vecina, en especial a partir de la ciudad de Mar del Plata, es más angosto y con más boyas. Además en el río hay corrientes y las corrientes por lo general no están en la misma línea que el canal.

En este caso no sé cómo fue el accidente, porque este barco hace la trayectoria todos los días, pero los errores en el mar existen, no sé por qué razón el pesquero se puso delante de la proa del barco, puedo

afirmar que los accidentes que existen nunca son porque se rompe un equipo de última generación, se producen por ese pequeño segundo, eso inesperado es lo que produce el accidente marítimo. Las cosas más tontas son las que producen desastres marítimos. No conozco a fondo el caso para opinar técnicamente. Aquí hubo un detalle y seguramente fue totalmente tonto e imprevisto, eso produjo el accidente"."

Origen: Diario La República - Uruguay
Total de oraciones: 40
Cantidad de uniones: 4
Oraciones propuestas en resumen manual: 18
Porcentaje de resumen: 46%

Cantidad de oraciones coincidentes: 8 – 44.5%

La poca coincidencia entre el resumen humano y el del sistema puede ser debida a una excesiva ponderación por posición. Dicha ponderación da un valor a la primera oración de 0.3 y resta para las siguientes oraciones del párrafo el valor 0.05 (según fórmula propuesta en el punto "Ponderación por posición" de la sección). Se cambiaron los valores a 0.4 y 0.1 respectivamente, obteniéndose el mismo resultado por lo cual se constató que el problema continuaba y se cambiaron nuevamente los valores a 0.15 para la primera oración del párrafo y restando de a 0.02 para el resto de las oraciones del párrafo. Los nuevos datos obtenidos fueron:
Cantidad de oraciones coincidentes: 10 – 55.6%

La coincidencia de oraciones fue mejor, superando el 50%. Los valores para la ponderación por posición se mantuvieron para el resto de las pruebas.

Resumen del sistema:

"*El barco Eladia Isabel de la empresa Buquebus recibió en la madrugada de ayer un fuerte impacto por parte de un barco pesquero de bandera argentina denominado "Depemás 5". Los daños no fueron importantes, pero Buquebus realizará una investigación al respecto. Altas fuentes consultadas indicaron que el accidente se produjo por una "mala maniobra" del pesquero. Llama la atención de expertos que el pesquero argentino se pusiera delante de la proa del barco de pasajeros.*

Un pesquero habría rozado al Eladia Isabel, de Buquebus.

La enorme embarcación, -Eladia Isabel es una de las embarcaciones más importantes de la flota de Buquebus-, debió ser remolcada y finalmente llegó a la dársena norte del puerto argentino cerca de las 3 de la mañana.

Mientras una versión narra que el accidente se produjo a raíz de una falla técnica en el instrumental de navegación, pero sin poder determinar de cuál de los dos barcos, otras fuentes consultadas indicaron que el accidente se produjo por una "mala maniobra" del pesquero". Con respecto a los daños sufridos, se supo que dos de los 55 autos que estaban en la bodega del Eladia Isabel fueron averiados por el impacto. "Los daños serán cubiertos por el seguro", confirmó el vocero de la empresa Buquebus que hoy atendía los reclamos de los pasajeros en las oficinas que la empresa tiene en Puerto Madero.

Por su parte, Buquebus emitió un comunicado en el que explica que el buque Eladia Isabel había partido desde el puerto de Colonia del Sacramento a las 19.30 horas. A las 22.20 horas, cuando se encontraba a sólo 6 kilómetros del puerto de Buenos Aires, y por motivos que se desconocen y "son objeto de investigación", un buque pesquero lo impactó en uno de sus laterales.

En consecuencia, se trabajó en forma conjunta con Prefectura Naval Argentina y dos remolcadores se encargaron de trasladar a ambas embarcaciones hacia el Puerto de Buenos Aires.

Buquebus informó que los pasajeros que tenían programado su viaje a las 0.30 horas, viajaron en buques rápidos que salieron durante la madrugada desde el puerto de Buenos Aires, cumpliendo de esta manera con todos los servicios comprometidos. Lo mismo sucedió en la mañana de ayer, cuando se reubicó a los pasajeros que debían viajar en el Eladia Isabel, en otros buques rápidos.

Buquebus expresó que las reparaciones "ya se iniciaron", aguardando ahora que el buque Eladia Isabel vuelva a su frecuencia habitual "en algunas horas".

Por su parte el jefe de la Prefectura del puerto de Buenos Aires, José Romero, declaró ayer a Clarín.com que "están investigando las causas del accidente.

El capitán de ultramar de la Marina Mercante, Oscar Lebel, consultado por LA REPUBLICA explicó: "Desde Montevideo a Buenos Aires los barcos van por un canal dragado y balizado, o sea que tiene boyas de ambos lados.

En este caso no sé cómo fue el accidente, porque este barco hace la trayectoria todos los días, pero los errores en el mar existen, no sé por qué razón el pesquero se puso delante de la proa del barco, puedo afirmar que los accidentes que existen nunca son porque se rompe un equipo de última generación, se producen por ese pequeño segundo, eso inesperado es lo que produce el accidente marítimo."

Caso propuesto por la segunda persona:

Oraciones propuestas en resumen manual: 23

Porcentaje de resumen: 58%

Cantidad de oraciones coincidentes: 15 – 65.3%

Resumen del sistema:

"*El barco Eladia Isabel de la empresa Buquebus recibió en la madrugada de ayer un fuerte impacto por parte de un barco pesquero de bandera argentina denominado "Depemás 5". Los daños no fueron importantes, pero Buquebus realizará una investigación al respecto. Altas fuentes consultadas indicaron que el accidente se produjo por una "mala maniobra" del pesquero. Llama la atención de expertos que el pesquero argentino se pusiera delante de la proa del barco de pasajeros.*

Aunque era impensable que se tratara de un iceberg, muchos recordaron las escenas de "Titanic". Un pesquero habría rozado al Eladia Isabel, de Buquebus.

El pesquero de bandera argentina no sufrió muchos daños y contaba con una tripulación de 29 marinos.

La enorme embarcación, -Eladia Isabel es una de las embarcaciones más importantes de la flota de Buquebus-, debió ser remolcada y finalmente llegó a la dársena norte del puerto argentino cerca de las 3 de la mañana.

Mientras una versión narra que el accidente se produjo a raíz de una falla técnica en el instrumental de navegación, pero sin poder determinar de cuál de los dos barcos, otras fuentes consultadas indicaron que el accidente se produjo por una "mala maniobra" del pesquero". Con respecto a los daños sufridos, se supo que dos de los 55 autos que estaban en la bodega del Eladia Isabel fueron averiados por el impacto. "Los daños serán cubiertos por el seguro", confirmó el vocero de la empresa Buquebus que hoy atendía los reclamos de los pasajeros en las oficinas que la empresa tiene en Puerto Madero.

Por su parte, Buquebus emitió un comunicado en el que explica que el buque Eladia Isabel había partido desde el puerto de Colonia del Sacramento a las 19.30 horas. A las 22.20 horas, cuando se encontraba a sólo 6 kilómetros del puerto de Buenos Aires, y por motivos que se desconocen y "son objeto de investigación", un buque pesquero lo impactó en uno de sus laterales.

En consecuencia, se trabajó en forma conjunta con Prefectura Naval Argentina y dos remolcadores se encargaron de trasladar a ambas embarcaciones hacia el Puerto de Buenos Aires.

En ese momento viajaban en el barco aproximadamente 600 pasajeros (la mitad de plazas disponibles), que fueron atendidos y contenidos por miembros de la tripulación, no habiéndose registrado heridos con motivo del accidente. Buquebus informó que los pasajeros que tenían programado su viaje a las 0.30 horas, viajaron en buques rápidos que salieron durante la madrugada desde el puerto de Buenos Aires, cumpliendo de esta manera con todos los servicios comprometidos. Lo mismo sucedió en la mañana de ayer, cuando se reubicó a los pasajeros que debían viajar en el Eladia Isabel, en otros buques rápidos.

Buquebus expresó que las reparaciones "ya se iniciaron", aguardando ahora que el buque Eladia Isabel vuelva a su frecuencia habitual "en algunas horas".

Por su parte el jefe de la Prefectura del puerto de Buenos Aires, José Romero, declaró ayer a Clarín.com que "están investigando las causas del accidente. "Ambos barcos ya están en el Puerto para los peritajes correspondientes", dijo.

El capitán de ultramar de la Marina Mercante, Oscar Lebel, consultado por LA REPUBLICA explicó: "Desde Montevideo a Buenos Aires los barcos van por un canal dragado y balizado, o sea que tiene boyas de ambos lados. El canal hacia la ciudad vecina, en especial a partir de la ciudad de Mar del Plata, es más angosto y con más boyas.

En este caso no sé cómo fue el accidente, porque este barco hace la trayectoria todos los días, pero los errores en el mar existen, no sé por qué razón el pesquero se puso delante de la proa del barco, puedo afirmar que los accidentes que existen nunca son porque se rompe un equipo de última generación, se producen por ese pequeño segundo, eso inesperado es lo que produce el accidente marítimo."

Caso 2

Título del texto: "Una marcha anti Bush desató el caos y violencia en Ciudad Vieja."

Texto:

"*Una turba de más de un centenar de personas, entonando cánticos antiimperialistas y bajo la consigna "Contra la cumbre del capital, América está en la calle", sembró el pánico -en la tarde de ayer- en las calles de la Ciudad Vieja, un par de horas antes que otra marcha repudiara la presencia del presidente de EEUU George Bush en Mar del Plata.*

Los agresores, muchos de elLos enmascarados, destrozaron todo a su paso, "a pedradas" y "patadas" parabrisas de autos, vidrieras de bancos, dependencias públicas, comercios, restoranes de comida rápida y de la Bolsa de Valores capitalina, además de agredir a cuanta persona se les cruzaba. Hubo quienes "grafitearon" consignas antiimperialistas, contrarias al presidente norteamericano George W. Bush e impactaron "bombas de alquitrán" y de pintura roja en cuanta pared o muro por donde pasaban, incluyendo a una dependencia de la Armada.

Varios propietarios de comercios y de vehículos reaccionaron ante Los desmanes que minutos después fueron controlados por la Policía. El ministro del Interior, José Díaz, repudió Los hechos y anunció que aplicará la ley "con todo rigor".

Sobre las 16:00 horas, comenzaron a marchar -al son del tamboril -desde la Matriz y por la peatonal Sarandí hacia el sur- con la intención de hacerlo por toda la Ciudad Vieja, para luego desplazarse hasta la sede de la embajada norteamericana, mientras que a su paso continuaban entonando las "agresivas" consignas.

Los hechos.

Apenas pasadas las 15:00 horas comenzaron a nuclearse en la Plaza Matriz Los manifestantes "que deseen oír su voz de denuncia y resistencia", Los cuales fueron convocadas por al menos una organización denominada "ALCArajo!", la cual en su sitio web (www.alcarajo.entodaspartes.org) promocionaba la concentración, que era en respuesta de la profunda disconformidad hacia la IV Cumbre de las Américas, que culmina hoy en la ciudad argentina de Mar del Plata, y a la presencia del primer mandatario estadounidense en la misma: "Los poderosos se encuentran para fortalecer la globalización neoliberal que beneficia a las potencias y sus corporaciones. No al ALCA y a la militarización de nuestros pueblos!".

Acto seguido, un pequeño grupo, que se reunió en el lugar, comenzó a entonar consignas agresivas hacia el "régimen imperialista y el presidente de EEUU, George W. Bush", y tras unos pocos minutos se les sumaron varias decenas de mujeres, niños, jóvenes, muchos de Los cuales traían sus rostros cubiertos.

Todo venía "bien" hasta que "voló" una piedra.

En su camino, la turba agredía a cuanta persona se les cruzaba, además de realizar pintadas en muros y paredes, como, por ejemplo, "muerte al capitalismo".

Pero, la explosión de mayor violencia, se dio recién cuando llegaron a la esquina de Sarandí y Misiones. Allí una piedra alcanzó Los vidrios de un banco de la zona, y de ahí en más Los manifestantes comenzaron a destrozar todo lo que encontraban a su paso mientras avanzaban por la calle Misiones.

Bancos, edificios públicos (como ser una dependencia de la Armada Nacional), la puerta de la Bolsa de Valores, la cual fue convertida -literalmente- en polvo debido a la lluvia de piedras, pilas, bombas de alquitrán y pintura roja, pero para tranquilidad de todos, no se registraron heridos en ese lugar, a pesar de que en ese momento había varias personas en el interior de la institución financiera.

Pero la cosa no paró ahí, sino que continuó, tanto por la calle Misiones, como por la calle Rincón, y en la intersección de ambas vías, Los revoltosos se encontraron con efectivos de la guardia Metropolitana y de la fuerza Puma, Los cuales intentaron contenerlos.

La intención de "pararlos" también pasó por la cabeza y la actitud de algunos comerciantes, personas que circulaban por el lugar, que se enfrentaron a Los agresores e intentaron aplicarles "su propia medicina", pero no pudieron cumplir su cometido, a pesar de que sí lograron golpear a más de uno de Los "muy jóvenes marchantes".

En dicho momento, Los manifestantes comenzaron a correr y a separarse en pequeños grupos de no más de una decena, aunque en la huida continuaron efectuando múltiples destrozos en comercios, como ser una regalería y una zapatería en Rincón y Juan Carlos Gómez.

Esa división en pequeños grupos dificultó que la Policía lograra detener al grueso de Los agresores, pero sobre la avenida 18 de Julio, fueron apresadas 15 personas -una de las cuales reconoció pertenecer a la Plenaria Memoria y Justicia, y fueron derivados por efectivos a la Seccional 1ª, lugar donde se encuentran recluidos en estos momentos.

En posesión de Los mismos, fueron encontradas bombas molotov, de alquitrán, de pintura; miguelitos; banderas con consignas antiimperialistas y con la cara del "Che", y diferentes colores de sprays.

Al declarar, varios de Los "muy" jóvenes detenidos no tenían idea de porqué participaron de este hecho.

Entre tanto, hoy a las 12.30 horas, Los detenidos concurrirán ante el juez de 1er. Turno, doctor Fernández Lecchini, quien se hizo presente en el lugar de Los hechos y constató Los destrozos ocasionados por Los revoltosos.

También fueron descubiertos, en Los lugares que recorrieron Los manifestantes, varias bombas molotov que habían sido "estratégicamente" dejadas, con el único objetivo de ser detonadas "más tarde", informaron a LA REPUBLICA fuentes policiales.

Cabe destacar que el saldo fue -además de Los destrozos materiales y de la enorme sensación de impotencia y rabia de la gente- de dos policías heridos, y de varias personas que no tenían "nada que ver" que sufrieron diversos rasguños, pero felizmente no tuvieron que lamentarse víctimas fatales, a pesar de la intensidad de la violencia de lo acontecido.

Por su parte, varios de Los comerciantes, testigos y víctimas -que prefirieron no ser identificados por miedo a una futura represalia- confesaron a LA REPUBLICA que vivieron momentos de extrema tensión, pánico y profundo miedo.

"Cuando vi que un montón de personas se vino sobre mi comercio y destrozó la vidriera, como si fueran unos salvajes, que se veía en sus rostros una gigantesca ira, pensé lo peor, pero menos mal que no paso más que eso", dijo una comerciante.

"Yo estaba dentro del banco y de repente escuché una explosión y salí a la calle y venía un grupo de personas enajenadas rompiendo todo y vi como un muchacho partía al medio al parabrisas de mi auto. Fue terrible", señaló otro de Los testigos.

*"Vino corriendo uno, me dio un empujón que me tiró al piso, pero no me hice nada, solo un raspón de nada, pero el miedo que me dio eso es muy fuerte, creo que nunca lo olvidaré, por la situación de no poder hacer nada para evitarlo", aseguró una joven señora. *"*

Origen: Diario La República - Uruguay

Total de oraciones: 28

Cantidad de uniones: 6

Oraciones propuestas en resumen manual: 11

Porcentaje de resumen: 40%

Cantidad de oraciones coincidentes: 6 – 54.6%.

Las coincidencias superan el 50%

Resumen del sistema:

"Una turba de más de un centenar de personas, entonando cánticos antiimperialistas y bajo la consigna "Contra la cumbre del capital, América está en la calle", sembró el pánico -en la tarde de ayer- en las calles de la Ciudad Vieja, un par de horas antes que otra marcha repudiara la presencia del presidente de EEUU George Bush en Mar del Plata.

Hubo quienes "grafitearon" consignas antiimperialistas, contrarias al presidente norteamericano George W. Bush e impactaron "bombas de alquitrán" y de pintura roja en cuanta pared o muro por donde pasaban, incluyendo a una dependencia de la Armada.

El ministro del Interior, José Díaz, repudió Los hechos y anunció que aplicará la ley "con todo rigor".

Apenas pasadas las 15.00 horas comenzaron a nuclearse en la Plaza Matriz Los manifestantes "que deseen oír su voz de denuncia y resistencia", Los cuales fueron convocadas por al menos una organización denominada "ALCArajo!", la cual en su sitio web (www.alcarajo.entodaspartes.org) promocionaba la concentración, que era en respuesta de la profunda disconformidad hacia la IV Cumbre de las Américas, que culmina hoy en la ciudad argentina de Mar del Plata, y a la presencia del primer mandatario estadounidense en la misma: "Los poderosos se encuentran para fortalecer la globalización neoliberal que beneficia a las potencias y sus corporaciones.

Acto seguido, un pequeño grupo, que se reunió en el lugar, comenzó a entonar consignas agresivas hacia el "régimen imperialista y el presidente de EEUU, George W. Bush", y tras unos pocos minutos se les sumaron varias decenas de mujeres, niños, jóvenes, muchos de Los cuales traían sus rostros cubiertos.

Sobre las 16.00 horas, comenzaron a marchar -al son del tamboril -desde la Matriz y por la peatonal Sarandí hacia el sur- con la intención de hacerlo por toda la Ciudad Vieja, para luego desplazarse hasta la sede de la embajada norteamericana, mientras que a su paso continuaban entonando las "agresivas" consignas.

Pero, la explosión de mayor violencia, se dio recién cuando llegaron a la esquina de Sarandí y Misiones.

Bancos, edificios públicos (como ser una dependencia de la Armada Nacional), la puerta de la Bolsa de Valores, la cual fue convertida -literalmente- en polvo debido a la lluvia de piedras, pilas, bombas de alquitrán y pintura roja, pero para tranquilidad de todos, no se registraron heridos en ese lugar, apesar de que en ese momento había varias personas en el interior de la institución financiera.

Pero la cosa no paró ahí, sino que continuó, tanto por la calle Misiones, como por la calle Rincón, y en la intersección de ambas vías, Los revoltosos se encontraron con efectivos de la guardia Metropolitana y de la fuerza Puma, Los cuales intentaron contenerlos.

Esa división en pequeños grupos dificultó que la Policía lograra detener al grueso de Los agresores, pero sobre la avenida 18 de Julio, fueron apresadas 15 personas -una de las cuales reconoció pertenecer a la Plenaria Memoria y Justicia, y fueron derivados por efectivos a la Seccional 1ª, lugar donde se encuentran recluidos en estos momentos.

"Vino corriendo uno, me dio un empujón que me tiró al piso, pero no me hice nada, solo un raspón de nada, pero el miedo que me dio eso es muy fuerte, creo que nunca lo olvidaré, por la situación de no poder hacer nada para evitarlo", aseguró una joven señora."

Caso 3

Título del texto: "Y el Corto agarró y se fue."

Texto:

"''El Corto'' nació el 23 de marzo de 1943, fue director teatral, actor, publicista, músico, además de ser uno de los fundadores de Canciones para no dormir la siesta, grupo que integró entre los años 1975 y 1990.

Durante su carrera participó en el Carnaval como letrista, actividad a la que volvió este año con Diablos Verdes.

Durante años fue columnista del diario LA REPUBLICA, de la columna amarilla de la contratapa.

Multifacético, el "Corto" Buscaglia le dio dura batalla a un cáncer de colon durante varios meses.

Cuando supo que estaba enfermo, lo enfrentó con el arma de su vida: el humor y lo escribió en la Columna Amarilla. No dejó de trabajar hasta el final, no sólo en su labor periodística o en su puesto como asesor en la Secretaría de Prensa y Difusión de la Presidencia, sino también componiendo junto a Rada canciones para niños o escribiendo obras de teatro y colaborando con los murguistas de los Diablos Verdes, con quienes estuvo hasta hace unos días.

El "Corto" Buscaglia fue un referente fundamental de la música y la cultura uruguaya, desde mediados de los años sesenta, compartiendo tareas de composición con Mateo, Pippo Spera, Urbano Moraes y Ruben Rada, entre otros, escribiendo para niños y adultos, comprometiéndose políticamente con su tiempo y su forma de encarar la cultura.

En una entrevista que fuera publicada en el portal Montevideo.com afirmó que "siempre estoy por escribir una obra sobre las cosas que me movilizan y me preocupan de esta sociedad pero siempre termino tirándola a la papelera". Escribía diariamente "La columna amarilla" en el diario LA REPUBLICA, donde "en unas 400 palabras hago pública mi opinión sobre diferentes hechos políticos, sociales y culturales que suelen suceder en el reino de este mundo".

Desde que nació, no paró de crear, incluso cuando la enfermedad lo quería postrar. Su batalla con el cáncer de colon culminó ayer en la madrugada.

"Y bueno... al final se hizo público y ya no puedo hacerme el gil. La cosa es así: en enero me operaron de un cáncer de colon, durante estos meses me he estado haciendo las duras y famosas quimioterapias. Las banqué bastante bien. Uno no sólo es un hueso duro de roer sino que además hace tiempo que tomé la decisión de que me voy a ir cuando yo lo decida y no cuando a cualquier vieja huesuda, venga en la forma en que venga, se le ocurra darme el guadañazo".

Buscaglia fue velado y sepultado ayer en el cementerio del Buceo. A su velatorio asistieron autoridades de gobierno, desde el presidente de la República, Tabaré Vázquez y el vice Rodolfo Nin Novoa, y otros dirigentes políticos y militantes sociales, hasta una pléyade de hombres y mujeres de la cultura uruguaya y amigos varios que tuvieron en el Corto un referente, pero más que eso, un ser humano excepcional.

La redacción de LA REPUBLICA, sus teléfonos, sus faxes y las casillas de correo electrónico no pararon de recibir muestras de dolor y saludos donde se expresa profunda consternación.

La Dirección Nacional de Fucvam, por ejemplo, envió nota que dice: "Podríamos definirlo de muchas maneras, su vida ha sido una permanente muestra de inacabable talento y batalla por causas dignas. Más de tres generaciones de uruguayos aplaudimos sus ocurrencias, sus poesías, sus músicas, sus personajes y sus muestras de compromiso inalterable con el cambio y la esperanza. Los cooperativistas nos sumamos al dolor y la impotencia que hoy los uruguayos sentimos al momento de perder un compañero de tantas jornadas. Hasta siempre "Corto", hasta el último aplauso."

Entre otras autoridades diplomáticas que se unieron al duelo del pueblo uruguayo por la muerte de Horacio Buscaglia, la embajadora de la República Bolivariana de Venezuela, María Urbaneja Durant, firma este testimonio y homenaje: "Animador fundamental de las últimas cuatro décadas en el movimiento cultural uruguayo, Buscaglia cumplió una brillante tarea, alinéandose siempre junto a las causas más justas y combatiendo el imperialismo y la desigualdad. Fue un militante por la vida que demostró con hechos en más de una ocasión su apoyo y solidaridad con procesos de transformación en el continente, y con el bolivariano especialmente, por lo que es considerado un amigo de nuestro pueblo (...), queremos darle el adiós a alguien que ha dejado una huella indeleble en nuestros corazones"."

Origen: Diario La República - Uruguay

Total de oraciones: 24

Cantidad de uniones: 7

Oraciones propuestas en resumen manual: 12

Porcentaje de resumen: 50%

Cantidad de oraciones coincidentes: 9 – 75%

En este caso podemos notar la gran coincidencia entre el resumen propuesto por una persona y el resumen que propone el sistema.

Luego de ingresar sustantivos y nombres que no se encontraban en la base de datos se obtuvieron nuevos resultados:

Cantidad de oraciones coincidentes: 11 – 91.7%

Se observa que luego de ingresar las palabras faltantes las coincidencias aumentaron hasta alcanzar casi el 100%.

Resumen del sistema:

"El Corto'' nació el 23 de marzo de 1943, fue director teatral, actor, publicista, músico, además de ser uno de los fundadores de Canciones para no dormir la siesta, grupo que integró entre los años 1975 y 1990.

Durante su carrera participó en el Carnaval como letrista, actividad a la que volvió este año con Diablos Verdes.

Durante años fue columnista del diario LA REPUBLICA, de la columna amarilla de la contratapa.

Multifacético, el "Corto" Buscaglia le dio dura batalla a un cáncer de colon durante varios meses.

No dejó de trabajar hasta el final, no sólo en su labor periodística o en su puesto como asesor en la Secretaría de Prensa y Difusión de la Presidencia, sino también componiendo junto a Rada canciones para niños o escribiendo obras de teatro y colaborando con los murguistas de los Diablos Verdes, con quienes estuvo hasta hace unos días.

El "Corto" Buscaglia fue un referente fundamental de la música y la cultura uruguaya, desde mediados de los años sesenta, compartiendo tareas de composición con Mateo, Pippo Spera, Urbano Moraes y Ruben Rada, entre otros, escribiendo para niños y adultos, comprometiéndose políticamente con su tiempo y su forma de encarar la cultura.

Escribía diariamente "La columna amarilla" en el diario LA REPUBLICA, donde "en unas 400 palabras hago pública mi opinión sobre diferentes hechos políticos, sociales y culturales que suelen suceder en el reino de este mundo".

Buscaglia fue velado y sepultado ayer en el cementerio del Buceo. A su velatorio asistieron autoridades de gobierno, desde el presidente de la República, Tabaré Vázquez y el vice Rodolfo Nin Novoa, y otros dirigentes políticos y militantes sociales, hasta una pléyade de hombres y mujeres de la cultura uruguaya y amigos varios que tuvieron en el Corto un referente, pero más que eso, un ser humano excepcional.

La Dirección Nacional de Fucvam, por ejemplo, envió nota que dice: "Podríamos definirlo de muchas maneras, su vida ha sido una permanente muestra de inacabable talento y batalla por causas dignas.

Entre otras autoridades diplomáticas que se unieron al duelo del pueblo uruguayo por la muerte de Horacio Buscaglia, la embajadora de la República Bolivariana de Venezuela, María Urbaneja Durant, firma este testimonio y homenaje: "Animador fundamental de las últimas cuatro décadas en el movimiento cultural uruguayo, Buscaglia cumplió una brillante tarea, alinéandose siempre junto a las causas más justas y combatiendo el imperialismo y la desigualdad. Fue un militante por la vida que demostró con hechos en más de una ocasión su apoyo y solidaridad con procesos de transformación en el continente, y con el bolivariano especialmente, por lo que es considerado un amigo de nuestro pueblo (...), queremos darle el adiós a alguien que ha dejado una huella indeleble en nuestros corazones".”

Caso 4

Título del texto: "La tiranía de Israel sobre Estados Unidos.”

Texto:

"*¿Qué país tiene en su territorio cientos de espías, topos y colaboradores trabajando, con total impunidad, para un gobierno extranjero desde hace más de 30 años como sucede en EEUU?. Según han informado antiguos y actuales periodistas que conocen bien el tema, algunos de los cuales han sido interrogados recientemente por el FBI, los agentes de la policía federal señalan a la policía secreta israelí Mossad como organizadora y promotora de esa red de espionaje.*

Durante el pasado año, en una de las más amplias investigaciones sobre el espionaje llevadas a cabo nunca, unos cien agentes del FBI estuvieron entrevistando, desde sus oficinas en ciudades por todo el país, a miles de testigos potenciales, informantes y sospechosos relacionados con el espionaje israelí en Estados Unidos.

Un antiguo reportero de un influyente semanario británico me contó que había sido interrogado en dos ocasiones, durante un total de unas doce horas, sobre la colaboración de los medios de comunicación con el Mossad a la hora de transmitir como "noticias" "información falsa" y propaganda a favor de Israel. De las conversaciones mantenidas con los periodistas entrevistados por el FBI surge un cuadro de penetración profunda y a gran escala de los espías israelíes y sus colaboradores en la sociedad y gobierno estadounidenses. Según mis fuentes, el FBI ha estado investigando durante treinta años las redes israelíes de espionaje, aunque la investigación se vio a menudo obstaculizada por políticos de ambos partidos en pago a los favores recibidos de lobbys israelíes y de ricos financieros para lograr que las campañas electorales acabaran favoreciendo a Israel. Según un escritor del británico Economist, hasta el FBI resultó infiltrado: el testimonio presentado por el escritor en los primeros años de la década de 1980 implicando a Richard Perle y Paul Wolfowitz en la entrega en mano de documentos a agentes del Mossad, "fue eliminado de los archivos del FBI y ha desaparecido".

Al pasar de los años, los servicios secretos israelíes se han ido haciendo más atrevidos y groseros en sus operaciones en EEUU. La red abarca a cientos de israelíes, a estadounidenses-israelíes (doble ciudadanía) y a sus colaboradores locales ("sayanin" o voluntarios seguidores judíos de los agentes israelíes fuera de Israel). Como secuelas del 11-S, cientos de agentes israelíes que estaban rondando por las oficinas gubernamentales,

fueron reunidos y deportados en silencio. En silencio, pero no porque no estuvieran cometiendo crímenes graves, sino para evitar que se incrementaran los ataques políticos desde las organizaciones pro-Israel más importantes y su clientela en el Congreso.

La expulsión masiva de espías israelíes fue una respuesta por el fallo de Israel cuando hubiera debido cooperar para impedir la masacre de miles de personas en Nueva York el 11 de septiembre de 2001. Parece que el FBI consiguió reunir pruebas de que la inteligencia israelí tenía detalladas evidencias del ataque terrorista del 11-S y no proporcionó la información a las autoridades estadounidenses. Sin embargo, siguieron afirmando que los israelíes les habían dado la información justo antes del ataque que sacó al FBI de la pista. Aunque el Mossad tiene la mayor red de espionaje y el sistema de apoyos más poderoso de cuantos países operan en EEUU, lo que resulta de especial interés es que, según los investigadores del FBI, esas operaciones están penetrando las más altas esferas del gobierno estadounidense, incluido el despacho del Vicepresidente Cheney. La prolongada investigación y la reciente y masiva asignación de recursos y agentes para investigar la conexión israelí se debe precisamente al espinoso asunto de tener que estar tratando con sospechosos en las esferas más altas de gobierno. Según un policía federal de Filadelfia, un paso en falso podría llevar a los peces gordos a cargarse la investigación. Por eso, los investigadores están extendiendo los interrogatorios para que alcancen a todas las fuentes posibles, acumulando miles de páginas con transcripciones, declaraciones juradas, intervención de conexiones telefónicas, videos de todos los posibles expertos o potencialmente implicados en las operaciones de espionaje de Israel desde hace mucho tiempo. A pesar de la intensificación de las investigaciones, montones de agentes israelíes y recientes reclutados continúan con las operaciones, muchos de ellos con la "cobertura protectora" de grupos cristianos evangélicos filo-sionistas así como de los "sayanin". Un objetivo clave de la investigación del FBI, pero uno muy difícil de forzar, es el AL - una unidad secreta de "katsas" experimentados (oficiales de caso del Mossad que reclutan agentes enemigos, como los describió Victor Ostrovsky, antiguo agente del Mossad, en "By Way of Deception").

Según las fuentes de mi periódico, el caso de Judith Miller pasando desinformación de origen israelí fue una práctica común durante los años de las décadas de 1980 y 1990. Muchos de los periodistas importantes y escritores de editoriales aceptaron y publicaron o divulgaron, a sabiendas, la información falsa israelí difundida por agentes del Mossad que actuaban como consejeros políticos desde la Embajada de Israel.

La investigación del FBI sobre las extensas operaciones de espionaje de Israel en EEUU es consecuencia de varios factores. Tras años de estrecha colaboración entre la inteligencia israelí y el FBI, éste (junto con la CIA) asumió la vergüenza por el "fracaso de los servicios de inteligencia en el 11-S" sin mencionar la falta de cooperación por parte de Israel al no haberles informado sobre lo que sabían. En segundo lugar, la descarada invasión a gran escala de los operativos israelíes sobre el área del FBI (en EEUU), ha socavado las actividades propias de las agencias, ha erosionado su posición como agencias de seguridad y ha desafiado de modo especial sus operaciones de contra-espionaje. En tercer lugar, el ascendente de Wolfowitz, Feith y Perle en los más altos escalones del Pentágono y de Elliot Abrams, Rubin y Libby en el Consejo Nacional de Seguridad, el Departamento de Estado y la Oficina del Vicepresidente, facilitó la transferencia rápida y masiva de documentación confidencial y decisiones delicadas al ejército de operativos del Mossad y a los altos funcionarios de la inteligencia militar tanto en EEUU como en Israel.

El flujo de información de EEUU a Israel se convirtió en un torrente incontrolado y, por lo que respecta al FBI, lo peor de todo fue que a nivel organizativo se convirtieron en actores marginales cuando no directamente despreciados. Lo que les resultó particularmente mortificante fue tener al menos cinco testigos deseando testificar contra Wolfowitz y Feith por un incidente de espionaje anterior y no poder ni tocarles a causa de sus altas puestos y del respaldo presidencial (especialmente tras el 11-S). El FBI estaba realmente preocupado por la profunda penetración en el Estado y por el papel clave que Israel jugaba asesorando, dirigiendo y transmitiendo propaganda y directrices a sus agentes, colaboradores y a las organizaciones sionistas más importantes en la carrera hacia la invasión estadounidense de Iraq. Dada la histeria de guerra y la propaganda "anti-terrorista" bombeada por todo el aparato ideológico pro-Israel, los agentes israelíes en el gobierno actuaron abiertamente y con total impunidad, desafiando tanto al FBI como a la CIA al establecer su propia Oficina de Planes Especiales como "operación clave de inteligencia" para transmitir información falsa directamente desde Israel hasta la Casa Blanca.

El inicio, y las inmediatas secuelas, de la guerra de Iraq y la subsiguiente ocupación supusieron el punto culminante de la tiranía israelí sobre Washington. "Asesores" pro Israel, miembros del gabinete, ideólogos, portavoces, miembros del Comité de Acción Política Israelo-Estadounidense (AIPAC, en sus siglas en inglés) y sus aliados en la Conferencia de Presidentes de las Organizaciones Judías más Importantes (CPMJO, en sus siglas en inglés) celebró su éxito presionando a EEUU a destruir completamente al principal adversario de Israel (Iraq), su ejército, su economía, sus sistemas administrativo y educativo y su infraestructura.

Sin embargo, la celebración de la victoria de Israel sobre el buen sentido e intereses nacionales de EEUU fue efímera. En cuanto la resistencia iraquí se fortaleció, en cuanto las bajas estadounidenses aumentaron y los costes de la guerra se dispararon, el pueblo estadounidense se volvió contra la guerra y el apoyo a la Administración Bush ha caído en picado. Con estos cambios políticos, los agentes israelíes y los colaboradores en el gobierno, autores y arquitectos de la guerra, debido a la investigación, perdieron parte de su inmunidad. Al detectar el FBI el cambio favorable en el clima político, procedió a ampliar enormemente su investigación; se sucedieron interrogatorios que incluyeron a Feith, Wolfowitz, Perle y otros neocon sionistas identificados con la inteligencia israelí. La siempre cautelosa agencia, temerosa de los ataques de los partidarios incondicionales de Israel en el Congreso de EEUU y en el Ejecutivo (Senadores Clinton y Lieberman, Secretaria de Estado Condi Rice y el Vicepresidente Cheney) se centró en los delitos de tres célebres elementos trabajando a favor de Israel - Irving "Scooter" Libby, de la oficina del Vicepresidente, por revelar la identidad de una agente secreta de la CIA; Larry Franklin, un funcionario del Pentágono de segundo rango unido a Feith y Wolfowitz, por espiar para Israel; y en dos dirigentes del AIPAC, el lobby pro Israel más importante, Rosen y Weissman, por pasar documentación confidencial a agentes del Mossad en la embajada israelí y por "complicidad" con periodistas de la corporación de prensa de Washington. Como la investigación del FBI sobre

la conexión israelí logró llegar hasta los niveles más altos en la jerarquía estatal, Wolfowitz, cuya ambición de toda la vida era ser el número uno en el Departamento de Defensa, dimitió de repente y fue nombrado para presidir el Banco Mundial; Feith también dimitió y se reincorporó a su firma legal israelo-estadounidense cuando la investigación llegó hasta uno de sus conductos más importantes (Franklin) por proporcionar inteligencia a los israelíes.

El FBI ha intensificado sus dragas en la muy extensa red de espionaje israelí y sus colaboradores en el AIPAC, la CPMJO y las organizaciones evangélicas cristiano-sionistas y muchas otras organizaciones comunales. Al mismo tiempo, los jerarcas israelíes, los operativos del Mossad y los funcionarios del gabinete israelí han intensificado su campaña para involucrar a EEUU en una nueva guerra contra Irán. Todas las organizaciones importantes pro Israel, los ideólogos y funcionarios de la Administración se han hecho eco de esa línea belicosa. Los Senadores Clinton y Lieberman declararon públicamente que, a la hora de "bombardear Irán", los intereses israelíes son el factor determinante de la política estadounidense hacia Oriente Próximo.

A pesar de las investigaciones del FBI, el AIPAC ha lanzado una de sus más virulentas y agresivas campañas de propaganda para satanizar a Irán, haciendo circular información falsa desde Israel sobre la amenaza de las (no existentes) armas nucleares de Irán y presionando con éxito al Congreso para que ladren ante la voz del Amo. A pesar del horrible desastre que para EEUU ha resultado ser la invasión de Iraq, en la cual los colaboradores israelíes jugaron un papel decisivo, están siguiendo el mismo guión a favor de la guerra con Irán - inventándose armas de destrucción masiva y amenazas para la seguridad de EEUU. El AIPAC está haciendo circular, entre todos los miembros del Congreso, fotos aéreas de bien conocidos e inspeccionados laboratorios experimentales iraníes como si fueran "lugares secretos de armas nucleares". Todos los ideólogos neocon sionistas importantes han producido como si fueran salchichas una serie de artículos en los que repetían como loros la compartida línea estatal israelí sobre la "amenaza iraní" y la necesidad urgente de imponerle o bien sanciones o bien llevar a cabo un ataque militar. En la actualidad, todo el aparato a favor de Israel supone la fuerza política más influyente presionando para la confrontación militar de EEUU con Irán, en contra de la opinión de todas las compañías petrolíferas importantes de dentro y fuera de EEUU.

Según un periodista que solía trabajar con el columnista Jack Anderson y al que el FBI pasó seis horas entrevistando, el FBI se ha asegurado la cooperación del ya condenado espía israelí y antiguo oficial del Pentágono, Lawrence Franklin, en el próximo juicio a los altos dirigentes del AIPAC Rosen y Weissman. Están ahora intentando alcanzar un acuerdo con el último para llegar hasta los escalones más altos de poder del AIPAC y del Gobierno Federal. Pero el proceso de investigación del espionaje israelí es lento y tedioso precisamente porque se introduce profundamente en las más altas instancias del gobierno y se irradia por una amplia red de organizaciones de la sociedad civil. Teniendo en cuenta la gran presión de los israelíes a favor de un inminente ataque militar contra Irán, no es probable que las investigaciones logren socavar su empeño en la guerra.

Sin embargo, puede suceder que las desastrosas consecuencias militares, políticas y económicas de la guerra contra Irán -añadidas a las pérdidas en Iraq y Afganistán- hagan aumentar más aún el rechazo hacia la Administración Bush y el aparato pro-Israel. Una decidida reacción popular podría impulsar que se llevaran a cabo más arrestos y más procesamientos de funcionarios públicos en altas instancias y entre los millonarios y operativos de las redes israelíes que están presionando a favor de la guerra.

Estas guerras desastrosas al servicio de Israel podrían lograr que los ciudadanos estadounidenses reflexionen y reaccionen frente al sometimiento de la política exterior estadounidense ante Israel. En última instancia, incluso podríamos ver la reinstauración de una República Americana "libre de enredos exteriores", por citar a George Washington, y de los "Benedict Arnold" [*], como alardean los Senadores estadounidenses."

Origen: Página Digital – Noticias y Artículos

http://www.paginadigital.com.ar/articulos/2006/2006prim/noticias/tirania-israeli-050206.asp

Total de oraciones: 55

Cantidad de uniones: 0

Oraciones propuestas en resumen manual: 25

Porcentaje de resumen: 46%

Cantidad de oraciones coincidentes: 14 – 56%

La cantidad de oraciones coincidentes supera el 50%

Luego de ingresado los nombres y sustantivos que faltaban en la base de datos se obtuvo el mismo resultado.

Resumen del sistema:

"Según han informado antiguos y actuales periodistas que conocen bien el tema, algunos de los cuales han sido interrogados recientemente por el FBI, los agentes de la policía federal señalan a la policía secreta israelí Mossad como organizadora y promotora de esa red de espionaje.

Según mis fuentes, el FBI ha estado investigando durante treinta años las redes israelíes de espionaje, aunque la investigación se vio a menudo obstaculizada por políticos de ambos partidos en pago a los favores recibidos de lobbys israelíes y de ricos financieros para lograr que las campañas electorales acabaran favoreciendo a Israel. Según un escritor del británico Economist, hasta el FBI resultó infiltrado: el testimonio presentado por el escritor en los primeros años de la década de 1980 implicando a Richard Perle y Paul Wolfowitz en la entrega en mano de documentos a agentes del Mossad, "fue eliminado de los archivos del FBI y ha desaparecido".

Un objetivo clave de la investigación del FBI, pero uno muy difícil de forzar, es el AL - una unidad secreta de "katsas" experimentados (oficiales de caso del Mossad que reclutan agentes enemigos, como los describió Victor Ostrovsky, antiguo agente del Mossad, en "By Way of Deception").

La investigación del FBI sobre las extensas operaciones de espionaje de Israel en EEUU es consecuencia de varios factores. Tras años de estrecha colaboración entre la inteligencia israelí y el FBI, éste (junto con la CIA) asumió la vergüenza por el "fracaso de los servicios de inteligencia en el 11-S" sin mencionar la falta de cooperación por parte de Israel al no haberles informado sobre lo que sabían. En tercer lugar, el ascendente de Wolfowitz, Feith y Perle en los más altos escalones del Pentágono y de Elliot Abrams, Rubin y Libby en el Consejo Nacional de Seguridad, el Departamento de Estado y la Oficina del Vicepresidente, facilitó la transferencia rápida y masiva de documentación confidencial y decisiones delicadas al ejército de operativos del Mossad y a los altos funcionarios de la inteligencia militar tanto en EEUU como en Israel.

El flujo de información de EEUU a Israel se convirtió en un torrente incontrolado y, por lo que respecta al FBI, lo peor de todo fue que a nivel organizativo se convirtieron en actores marginales cuando no directamente despreciados. El FBI estaba realmente preocupado por la profunda penetración en el Estado y por el papel clave que Israel jugaba asesorando, dirigiendo y transmitiendo propaganda y directrices a sus agentes, colaboradores y a las organizaciones sionistas más importantes en la carrera hacia la invasión estadounidense de Iraq. Dada la histeria de guerra y la propaganda "anti-terrorista" bombeada por todo el aparato ideológico pro-Israel, los agentes israelíes en el gobierno actuaron abiertamente y con total impunidad, desafiando tanto al FBI como a la CIA al establecer su propia Oficina de Planes Especiales como "operación clave de inteligencia" para transmitir información falsa directamente desde Israel hasta la Casa Blanca.

El inicio, y las inmediatas secuelas, de la guerra de Iraq y la subsiguiente ocupación supusieron el punto culminante de la tiranía israelí sobre Washington. "Asesores" pro Israel, miembros del gabinete, ideólogos, portavoces, miembros del Comité de Acción Política Israelo-Estadounidense (AIPAC, en sus siglas en inglés) y sus aliados en la Conferencia de Presidentes de las Organizaciones Judías más Importantes (CPMJO, en sus siglas en inglés) celebró su éxito presionando a EEUU a destruir completamente al principal adversario de Israel (Iraq), su ejército, su economía, sus sistemas administrativo y educativo y su infraestructura.

Sin embargo, la celebración de la victoria de Israel sobre el buen sentido e intereses nacionales de EEUU fue efímera. La siempre cautelosa agencia, temerosa de los ataques de los partidarios incondicionales de Israel en el Congreso de EEUU y en el Ejecutivo (Senadores Clinton y Lieberman, Secretaria de Estado Condi Rice y el Vicepresidente Cheney) se centró en los delitos de tres célebres elementos trabajando a favor de Israel - Irving "Scooter" Libby, de la oficina del Vicepresidente, por revelar la identidad de una agente secreta de la CIA; Larry Franklin, un funcionario del Pentágono de segundo rango unido a Feith y Wolfowitz, por espiar para Israel; y en dos dirigentes del AIPAC, el lobby pro Israel más importante, Rosen y Weissman, por pasar documentación confidencial a agentes del Mossad en la embajada israelí y por "complicidad" con periodistas de la corporación de prensa de Washington. Como la investigación del FBI sobre la conexión israelí logró llegar hasta los niveles más altos en la jerarquía estatal, Wolfowitz, cuya ambición de toda la vida era ser el número uno en el Departamento de Defensa, dimitió de repente y fue nombrado para presidir el Banco Mundial; Feith también dimitió y se reincorporó a su firma legal israelo-estadounidense cuando la investigación llegó hasta uno de sus conductos más importantes (Franklin) por proporcionar inteligencia a los israelíes.

El FBI ha intensificado sus dragas en la muy extensa red de espionaje israelí y sus colaboradores en el AIPAC, la CPMJO y las organizaciones evangélicas cristiano-sionistas y muchas otras organizaciones comunales. Al mismo tiempo, los jerarcas israelíes, los operativos del Mossad y los funcionarios del gabinete israelí han intensificado su campaña para involucrar a EEUU en una nueva guerra contra Irán. Todas las organizaciones importantes pro Israel, los ideólogos y funcionarios de la Administración se han hecho eco de esa línea belicosa. Los Senadores Clinton y Lieberman declararon públicamente que, a la hora de "bombardear Irán", los intereses israelíes son el factor determinante de la política estadounidense hacia Oriente Próximo.

A pesar de las investigaciones del FBI, el AIPAC ha lanzado una de sus más virulentas y agresivas campañas de propaganda para satanizar a Irán, haciendo circular información falsa desde Israel sobre la amenaza de las (no existentes) armas nucleares de Irán y presionando con éxito al Congreso para que ladren ante la voz del Amo. A pesar del horrible desastre que para EEUU ha resultado ser la invasión de Iraq, en la cual los colaboradores israelíes jugaron un papel decisivo, están siguiendo el mismo guión a favor de la guerra con Irán - inventándose armas de destrucción masiva y amenazas para la seguridad de EEUU. En la actualidad, todo el aparato a favor de Israel supone la fuerza política más influyente presionando para la confrontación militar de EEUU con Irán, en contra de la opinión de todas las compañías petrolíferas importantes de dentro y fuera de EEUU.

Según un periodista que solía trabajar con el columnista Jack Anderson y al que el FBI pasó seis horas entrevistando, el FBI se ha asegurado la cooperación del ya condenado espía israelí y antiguo oficial del Pentágono, Lawrence Franklin, en el próximo juicio a los altos dirigentes del AIPAC Rosen y Weissman.

Estas guerras desastrosas al servicio de Israel podrían lograr que los ciudadanos estadounidenses reflexionen y reaccionen frente al sometimiento de la política exterior estadounidense ante Israel. En última instancia, incluso podríamos ver la reinstauración de una República Americana "libre de enredos exteriores", por citar a George Washington, y de los "Benedict Arnold" [*], como alardean los Senadores estadounidenses."

Caso 5

Título del texto: "El último leño a la hoguera."

Texto:

"*Así que ahora se trata de cartones sobre el profeta Mahoma con un turbante en forma de bomba. Los embajadores son retirados de Dinamarca, los sauditas y los sirios se quejan, las naciones del Golfo Pérsico quitan de sus anaqueles todos los productos daneses y hombres armados en Gaza amenazan a la Unión Europea y a periodistas extranjeros. En Dinamarca, el editor de "cultura" del bobalicón diario en el que aparecieron esas tontas caricaturas -en septiembre pasado, por Dios- anuncia que "estamos siendo testigos de un choque de civilizaciones" entre las democracias laicas occidentales y las sociedades islámicas. Esto comprueba, supongo, que los periodistas daneses se mantienen fieles a la tradición de Hans Christian Andersen. ¡Ay, Dios, Dios! Lo que estamos presenciando es la puerilidad de las civilizaciones.*

Comencemos en el Departamento de Verdades Domésticas. Esto no es una cuestión de laicismo contra el Islam. Para los musulmanes, el profeta es el hombre que recibió las palabras divinas directamente de Dios. Nosotros vemos a nuestros santos y profetas, cuando mucho, como figuras históricas, que se contraponen a nuestros derechos humanos, a la alta tecnología y a nuestras libertades; los vemos casi como caricaturas. El hecho es que los musulmanes viven su religión, nosotros no.

Ellos han conservado su fe, pese a innumerables vicisitudes históricas. Nosotros hemos venido perdiendo nuestra fe desde que el poeta inglés Matthew Arnold escribió sobre "el largo y lejano rugido del mar". Hablamos de "occidente contra el Islam" en vez de "cristianos contra el Islam", porque tampoco quedan muchos cristianos en Europa que digamos. No hay forma de arreglar esto reuniendo a las religiones del mundo y preguntando por qué no se nos permite burlar de Mahoma.

Claro, siempre podemos ejercer nuestra propia hipocresía en torno de los sentimientos religiosos. Recuerdo que hace más de una década una película llamada La última tentación de Cristo mostraba a Jesús haciéndole el amor a una mujer. En París alguien le prendió fuego al cine que presentaba la cinta, y en el incendio murió un joven francés. También recuerdo que una de las principales universidades de Estados Unidos me invitó a dar una conferencia hace tres años. Lo hice. Mi conferencia se titulaba "Septiembre 11, 2001: pregunten quién lo hizo, pero por amor de Dios no pregunten por qué".

Cuando llegué a ofrecer la ponencia me encontré con que las autoridades habían eliminado la frase "por amor de Dios", alegando que "no querían ofender ciertas sensibilidades". Ajá, así que nosotros también tenemos "sensibilidades".

En otras palabras, a pesar de que exigimos que los musulmanes se comporten como buenos laicos cuando se trata de la libre expresión -o de caricaturas vulgares-, todavía tenemos que preocuparnos porque los adherentes a nuestra preciosa religión no se ofendan.

También disfruté enormemente las pomposas declaraciones de hombres de Estado europeos que afirman que no pueden controlar la libre expresión ni a los periódicos. Eso es una tontería. Si uno de los cartones hubiera mostrado a un rabino en vez de al profeta con un sombrero en forma de bomba nos hubieran vociferado al oído "antisemitas", y con toda razón. Esta es la queja que siempre hacen los israelíes de las caricaturas antisemitas que aparecen en los periódicos egipcios.

Más aún: en algunas naciones europeas -Francia es una, Alemania y Austria son otras- está prohibido en la ley negar genocidios. En Francia, por ejemplo, es ilegal decir que no existieron los holocaustos judío y armenio (nada más esperen a ver la reacción de Turquía ante este último punto, si es que este país llega a ingresar a la Unión Europea).

De modo que está prohibido hacer ciertas afirmaciones en Europa. No estoy seguro si esas leyes logran sus objetivos; no importa cuanto se prohíba la negación del holocausto, pues los antisemitas siempre encuentran forma de darle la vuelta a esas normas.

El punto, no obstante, es que a duras penas podemos hacer respetar nuestras prohibiciones políticas y leyes para evitar que haya caricaturas antisemitas o que se niegue el holocausto, y pese a ello nos ponemos a gritar en favor del laicismo cuando descubrimos que los musulmanes se ofenden por nuestras provocaciones e imágenes insultantes al profeta.

Para muchos musulmanes, la reacción "islámica" por todo ese escuálido asunto es una vergüenza. Es perfectamente razonable creer que a los musulmanes les gustaría ver que se introduzca algún elemento de reforma a su religión. Si los cartones hubieran promovido algún debate sobre el tema -si existiera la posibilidad de un diálogo serio-, nadie habría tenido objeciones.

Pero claramente hubo la intención de que las caricaturas fueran una provocación. Fueron tan absurdas, que lo que lo único que causaron fue una reacción.

Además, este no es el momento más adecuado para recalentar la vieja basura de Samuel Huntington sobre "el choque de civilizaciones". Irán tiene nuevamente un gobierno clerical. Lo mismo ocurre, para todo fin práctico, en Irak (donde supuestamente no iban a usar su democracia para elegir a un gobierno religioso, pero eso es lo que pasa cuando uno se pone a derrocar dictadores).

En Egipto, la Hermandad Musulmana ganó 20 por ciento de los escaños parlamentarios en las recientes elecciones legislativas. Ahora tenemos a Hamas a cargo de Palestina.

Aquí hay un mensaje, ¿no es cierto? Las políticas estadunidenses para el "cambio de régimen" y la "democracia" en Medio Oriente no están alcanzando sus objetivos. Estos millones de votantes prefieren el Islam a los gobiernos corruptos que les impusieron. El que los cartones sean arrojados a la situación para atizar el fuego es ciertamente peligroso.

En cualquier caso, no se trata de si el profeta debe o no ser retratado. El Corán prohíbe las imágenes del Profeta y aún así millones de musulmanes tienen y crean esas imágenes. El problema es que las caricaturas representan a Mahoma como imagen de violencia estilo Bin Laden. Muestran el Islam como religión violenta. Y no lo es. ¿O queremos que sí lo sea?."

Origen: Página Digital – Noticias y Artículos

Total de oraciones: 51

Cantidad de uniones: 0

Oraciones propuestas en resumen manual: 30

Porcentaje de resumen: 59

Cantidad de oraciones coincidentes: 15 – 50%

Resumen del sistema:

"Así que ahora se trata de cartones sobre el profeta Mahoma con un turbante en forma de bomba. Los embajadores son retirados de Dinamarca, los sauditas y los sirios se quejan, las naciones del Golfo Pérsico quitan de sus anaqueles todos los productos daneses y hombres armados en Gaza amenazan a la Unión Europea y a periodistas extranjeros. En Dinamarca, el editor de "cultura" del bobalicón diario en el que aparecieron esas tontas caricaturas -en septiembre pasado, por Dios- anuncia que "estamos siendo testigos de un choque de civilizaciones" entre las democracias laicas occidentales y las sociedades islámicas. Esto comprueba, supongo, que los periodistas daneses se mantienen fieles a la tradición de Hans Christian Andersen. ¡Ay, Dios, Dios! Lo que estamos presenciando es la puerilidad de las civilizaciones.

Comencemos en el Departamento de Verdades Domésticas. Esto no es una cuestión de laicismo contra el Islam. Para los musulmanes, el profeta es el hombre que recibió las palabras divinas directamente de Dios.

Hablamos de "occidente contra el Islam" en vez de "cristianos contra el Islam", porque tampoco quedan muchos cristianos en Europa que digamos.

Claro, siempre podemos ejercer nuestra propia hipocresía en torno de los sentimientos religiosos. Recuerdo que hace más de una década una película llamada La última tentación de Cristo mostraba a Jesús haciéndole el amor a una mujer. En París alguien le prendió fuego al cine que presentaba la cinta, y en el incendio murió un joven francés. También recuerdo que una de las principales universidades de Estados Unidos me invitó a dar una conferencia hace tres años. Mi conferencia se titulaba "Septiembre 11, 2001: pregunten quién lo hizo, pero por amor de Dios no pregunten por qué".

Cuando llegué a ofrecer la ponencia me encontré con que las autoridades habían eliminado la frase "por amor de Dios", alegando que "no querían ofender ciertas sensibilidades".

También disfruté enormemente las pomposas declaraciones de hombres de Estado europeos que afirman que no pueden controlar la libre expresión ni a los periódicos.

Más aún: en algunas naciones europeas -Francia es una, Alemania y Austria son otras- está prohibido en la ley negar genocidios. En Francia, por ejemplo, es ilegal decir que no existieron los holocaustos judío y armenio (nada más esperen a ver la reacción de Turquía ante este último punto, si es que este país llega a ingresar a la Unión Europea).

De modo que está prohibido hacer ciertas afirmaciones en Europa.

El punto, no obstante, es que a duras penas podemos hacer respetar nuestras prohibiciones políticas y leyes para evitar que haya caricaturas antisemitas o que se niegue el holocausto, y pese a ello nos ponemos a gritar en favor del laicismo cuando descubrimos que los musulmanes se ofenden por nuestras provocaciones e imágenes insultantes al profeta.

Pero claramente hubo la intención de que las caricaturas fueran una provocación.

Además, este no es el momento más adecuado para recalentar la vieja basura de Samuel Huntington sobre "el choque de civilizaciones". Irán tiene nuevamente un gobierno clerical.

En Egipto, la Hermandad Musulmana ganó 20 por ciento de los escaños parlamentarios en las recientes elecciones legislativas. Ahora tenemos a Hamas a cargo de Palestina.

Estos millones de votantes prefieren el Islam a los gobiernos corruptos que les impusieron.

En cualquier caso, no se trata de si el profeta debe o no ser retratado. El Corán prohíbe las imágenes del Profeta y aún así millones de musulmanes tienen y crean esas imágenes. El problema es que las caricaturas representan a Mahoma como imagen de violencia estilo Bin Laden. Muestran el Islam como religión violenta."

Caso propuesto por la segunda persona:

Oraciones propuestas en resumen manual: 39

Porcentaje de resumen: 77%

Cantidad de oraciones coincidentes por ponderación: 32 – 82.1%

Cantidad de oraciones coincidentes por coherencia: 33 – 84.6%

Aquí el resumen por ponderación y el resumen por coherencia dieron diferentes valores de coincidencia con el resumen propuesto por la persona. Ambos obtuvieron una coincidencia mayor al 80%

Resumen del sistema:

"Así que ahora se trata de cartones sobre el profeta Mahoma con un turbante en forma de bomba. Los embajadores son retirados de Dinamarca, los sauditas y los sirios se quejan, las naciones del Golfo Pérsico quitan de sus anaqueles todos los productos daneses y hombres armados en Gaza amenazan a la Unión Europea y a periodistas extranjeros. En Dinamarca, el editor de "cultura" del bobalicón diario en el que aparecieron esas tontas caricaturas -en septiembre pasado, por Dios- anuncia que "estamos siendo testigos de un choque de civilizaciones" entre las democracias laicas occidentales y las sociedades islámicas. Esto comprueba, supongo, que los periodistas daneses se mantienen fieles a la tradición de Hans Christian Andersen. ¡Ay, Dios, Dios! Lo que estamos presenciando es la puerilidad de las civilizaciones.

Comencemos en el Departamento de Verdades Domésticas. Esto no es una cuestión de laicismo contra el Islam. Para los musulmanes, el profeta es el hombre que recibió las palabras divinas directamente de Dios. Nosotros vemos a nuestros santos y profetas, cuando mucho, como figuras históricas, que se contraponen a nuestros derechos humanos, a la alta tecnología y a nuestras libertades; los vemos casi como caricaturas.

Nosotros hemos venido perdiendo nuestra fe desde que el poeta inglés Matthew Arnold escribió sobre "el largo y lejano rugido del mar". Hablamos de "occidente contra el Islam" en vez de "cristianos contra el Islam", porque tampoco quedan muchos cristianos en Europa que digamos. No hay forma de arreglar esto reuniendo a las religiones del mundo y preguntando por qué no se nos permite burlar de Mahoma.

Claro, siempre podemos ejercer nuestra propia hipocresía en torno de los sentimientos religiosos. Recuerdo que hace más de una década una película llamada La última tentación de Cristo mostraba a Jesús haciéndole el amor a una mujer. En París alguien le prendió fuego al cine que presentaba la cinta, y en el incendio murió un joven francés. También recuerdo que una de las principales universidades de Estados Unidos me invitó a dar una conferencia hace tres años. Mi conferencia se titulaba "Septiembre 11, 2001: pregunten quién lo hizo, pero por amor de Dios no pregunten por qué".

Cuando llegué a ofrecer la ponencia me encontré con que las autoridades habían eliminado la frase "por amor de Dios", alegando que "no querían ofender ciertas sensibilidades".

En otras palabras, a pesar de que exigimos que los musulmanes se comporten como buenos laicos cuando se trata de la libre expresión -o de caricaturas vulgares-, todavía tenemos que preocuparnos porque los adherentes a nuestra preciosa religión no se ofendan.

También disfruté enormemente las pomposas declaraciones de hombres de Estado europeos que afirman que no pueden controlar la libre expresión ni a los periódicos. Si uno de los cartones hubiera mostrado a un rabino en vez de al profeta con un sombrero en forma de bomba nos hubieran vociferado al oído "antisemitas", y con toda razón.

Más aún: en algunas naciones europeas -Francia es una, Alemania y Austria son otras- está prohibido en la ley negar genocidios. En Francia, por ejemplo, es ilegal decir que no existieron los holocaustos judío y armenio (nada más esperen a ver la reacción de Turquía ante este último punto, si es que este país llega a ingresar a la Unión Europea).

De modo que está prohibido hacer ciertas afirmaciones en Europa.

El punto, no obstante, es que a duras penas podemos hacer respetar nuestras prohibiciones políticas y leyes para evitar que haya caricaturas antisemitas o que se niegue el holocausto, y pese a ello nos ponemos a gritar en favor del laicismo cuando descubrimos que los musulmanes se ofenden por nuestras provocaciones e imágenes insultantes al profeta.

Es perfectamente razonable creer que a los musulmanes les gustaría ver que se introduzca algún elemento de reforma a su religión.

Pero claramente hubo la intención de que las caricaturas fueran una provocación.

Además, este no es el momento más adecuado para recalentar la vieja basura de Samuel Huntington sobre "el choque de civilizaciones". Irán tiene nuevamente un gobierno clerical. Lo mismo ocurre, para todo fin práctico, en Irak (donde supuestamente no iban a usar su democracia para elegir a un gobierno religioso, pero eso es lo que pasa cuando uno se pone a derrocar dictadores).

En Egipto, la Hermandad Musulmana ganó 20 por ciento de los escaños parlamentarios en las recientes elecciones legislativas. Ahora tenemos a Hamas a cargo de Palestina.

Aquí hay un mensaje, ¿no es cierto? Las políticas estadunidenses para el "cambio de régimen" y la "democracia" en Medio Oriente no están alcanzando sus objetivos. Estos millones de votantes prefieren el Islam a los gobiernos corruptos que les impusieron.

En cualquier caso, no se trata de si el profeta debe o no ser retratado. El Corán prohíbe las imágenes del Profeta y aún así millones de musulmanes tienen y crean esas imágenes. El problema es que las caricaturas

representan a Mahoma como imagen de violencia estilo Bin Laden. Muestran el Islam como religión violenta. Y no lo es."