

# Plataforma para la generación dinámica de textos personalizados



Máster en Ingeniería Informática

## Trabajo Fin de Máster

Autor:

Antonio Guillén Espejo

Tutora:

Elena Lloret Pastor

Cotutor:

Yoan Gutiérrez Vázquez

Octubre 2015



Universitat d'Alacant  
Universidad de Alicante



# Plataforma para la generación dinámica de textos personalizados

---

## **Autor**

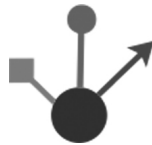
Antonio Guillén Espejo

## **Directores**

Elena Lloret Pastor

Yoan Gutiérrez Vázquez

*Departamento de Lenguajes y Sistemas Informáticos*



MÁSTER EN INGENIERÍA INFORMÁTICA



Escuela  
Politécnica  
Superior



Universitat d'Alacant  
Universidad de Alicante

ALICANTE, 16 de noviembre de 2015



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Contexto . . . . .	1
1.2. Objetivo y motivación . . . . .	1
1.3. Estructura de la memoria . . . . .	2
<b>2. Marco teórico</b>	<b>3</b>
2.1. Definición de Procesamiento del Lenguaje Natural (PLN) . . . . .	3
2.2. Niveles de análisis lingüístico y tareas del PLN . . . . .	3
2.3. Área de investigación . . . . .	4
2.4. Antecedente del sistema de integración . . . . .	5
<b>3. Objetivos</b>	<b>7</b>
3.1. Objetivos generales . . . . .	7
3.2. Objetivos específicos . . . . .	7
<b>4. Metodología</b>	<b>9</b>
4.1. Metodologías y tecnologías usadas en el proyecto . . . . .	9
<b>5. Desarrollo</b>	<b>11</b>
5.1. Arquitectura de TLH Suite . . . . .	11
5.2. Ficheros de configuración . . . . .	12
5.3. Servicios web seleccionados . . . . .	13
5.4. Wrappers de los servicios web . . . . .	14
5.5. Módulos y ejecutables . . . . .	14
5.5.1. Módulo Input . . . . .	15
5.5.2. Módulo Query . . . . .	16
5.5.3. Módulo Check Services . . . . .	17
5.5.4. Módulo Export/Import . . . . .	17
5.6. Entrada de datos, ejecución y almacenamiento de resultados . . . . .	18
5.6.1. Colección de datos “doc” . . . . .	19
5.6.2. Colección de datos “compendium” . . . . .	19
5.6.3. Colección de datos “semantic” . . . . .	20
5.6.4. Colección de datos “sentiment” . . . . .	20
5.6.5. Colección de datos “stanford” . . . . .	21
5.6.6. Colección de datos “tipsem” . . . . .	21
5.6.7. Colección de datos “general” . . . . .	21

---

<b>6. Evaluación y resultados</b>	<b>23</b>
6.1. Definición de escenarios de prueba . . . . .	23
6.2. Tabla de resultados . . . . .	24
6.3. Discusión de los resultados . . . . .	26
<b>7. Conclusiones</b>	<b>29</b>
7.1. Resumen del trabajo realizado . . . . .	29
7.2. Trabajo futuro . . . . .	29
<b>Bibliografía</b>	<b>31</b>
<b>A. Anexo I: Documentos de los escenarios</b>	<b>33</b>
<b>B. Anexo II: Resultados de las consultas</b>	<b>35</b>

# 1. Introducción

## 1.1. Contexto

Esta memoria describe el trabajo de fin de máster realizado por el alumno Antonio Guillén Espejo para el Grupo de Procesamiento del Lenguaje y Sistemas de Información (GPLSI), perteneciente al Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante. El grupo de investigación está centrado en el Procesamiento del Lenguaje Natural (PLN) por lo tanto este trabajo se define en dicho contexto. Este proyecto de final de máster se ha realizado en base a un proyecto de investigación emergente (GRE13-15) financiado por la Universidad de Alicante [Lloret et al., 2015]. En este proyecto emergente participan otros miembros del grupo de investigación como doctores y doctorandos.

## 1.2. Objetivo y motivación

El objetivo de este trabajo es anotar automáticamente un documento con información de distinto tipo para luego almacenarla y explotarla. Esto se conseguirá a partir de una selección e integración de herramientas de Tecnologías del Lenguaje Humano (TLH) que existen actualmente. Estas se pueden usar remotamente mediante servicios web, pero se generará un wrapper o encapsulado que permita ejecutarlas adecuadamente adaptando sus entradas y salidas, para finalmente combinar toda la información resultante de su ejecución. Esta combinación de datos se llamará “paquete semántico”.

Se diseñará una plataforma que permita cumplir este objetivo, la cual se probará mediante una serie de documentos textuales obtenidos de diversas fuentes. El paquete semántico generado a partir de estos documentos servirá posteriormente para consultar información de interés con la que evaluar la eficacia de la plataforma.

Uno de los fines del estudio del PLN es conseguir desarrollar herramientas que lleguen a la comprensión del lenguaje y por tanto, a la comunicación de ideas y al razonamiento. La llamada Sociedad de la Información no sólo exige acceso a todo tipo de materiales, principalmente textos, sino que genera una cantidad de nuevos documentos que crece exponencialmente cada día. Actualmente, ya es imposible manejar tal cantidad de información sin la ayuda de un ordenador y de las herramientas lingüísticas adecuadas. Nadie, por ejemplo, se plantea su trabajo, hoy en día, sin recurrir a un corrector ortográfico, un traductor o un buscador.

### 1.3. Estructura de la memoria

Este trabajo tiene la siguiente estructura:

**Introducción:** donde se define el contexto, el objetivo principal y la motivación del trabajo.

**Marco teórico:** definición de las áreas de investigación vinculadas al trabajo y sus tareas más comunes.

**Objetivos:** se establecen los objetivos general y específicos del trabajo.

**Metodología:** se indicará los métodos y tecnologías que serán utilizados para llevarlo a cabo.

**Desarrollo:** es el cuerpo del trabajo, donde se explica con detalle el diseño del sistema, su implementación, su funcionamiento, etc.

**Evaluación y resultados:** incluirá los resultados de evaluación del sistema, así como el análisis y la discusión de los mismos.

**Conclusiones:** conclusiones sobre el trabajo realizado y el trabajo futuro.



## 2. Marco teórico

### 2.1. Definición de Procesamiento del Lenguaje Natural (PLN)

El Procesamiento del Lenguaje Natural consiste principalmente en gestionar de forma automática cualquier fuente escrita o hablada generalmente no estructurada, como textos y conversaciones siendo de cualquier origen y dominio de aplicación [GPLSI, 2011]. Dado que el propio lenguaje humano es complejo a nivel estructural y cognitivo, no es un área de investigación trivial. Esto es debido a que se debe tener en cuenta aspectos del lenguaje como la morfología, sintaxis, semántica y pragmática. Así mismo, la multilingüidad es otro aspecto que incluso adquiere mayor relevancia dado el abanico de diferentes lenguas que abundan en Internet.

### 2.2. Niveles de análisis lingüístico y tareas del PLN

Un sistema de PLN debe usar abundante conocimiento de las estructuras del lenguaje (oraciones, palabras) pero también sobre el significado de las palabras u oración completa. De la misma manera, es necesario comprender la clasificación tradicional que los lingüistas han hecho de las formas de conocimiento de la lengua [Moreno et al., 1999]:

- *Fonética*: ciencia que se ocupa del estudio de los sonidos para formar palabras.
- *Morfología*: ciencia que estudia cómo las palabras son construidas a partir de unidades más pequeñas, son los monemas (morfemas y lexemas).
- *Sintaxis*: ciencia que estudia las combinaciones de las palabras para formar secuencias correctas y las relaciones de las mismas unas con otras.
- *Semántica*: ciencia que se ocupa del significado de las palabras y cómo éstas se combinan para formar el significado completo de una oración.
- *Pragmática/Gramática de contexto*: ciencia que estudia cómo el contexto afecta a la interpretación de las oraciones.

También se debe tener en cuenta el conocimiento general de las estructuras del mundo que todo usuario del lenguaje debe conocer, además de cómo este conocimiento afecta a la interpretación de las estructuras del lenguaje. Con respecto a la anterior clasificación, un sistema computacional divide las fases o niveles de análisis de una oración o texto en:

- *Análisis morfológico-léxico*: transforma la secuencia de caracteres de entrada en una secuencia de unidades significativas haciendo uso del diccionario y reglas morfológicas.

- *Análisis sintáctico*: analiza la secuencia de unidades léxicas y produce una representación de su estructura (árbol, red, ...).
- *Análisis semántico*: a partir de la estructura generada por el proceso sintáctico genera otra estructura o forma lógica asociada que representa el significado o sentido de la sentencia.
- *Análisis contextual o función pragmática*: utiliza la forma lógica o estructura semántica de la fase anterior para desarrollar la interpretación final de la oración, en función de las circunstancias de contexto.

## 2.3. Área de investigación

Debido a que es un área de investigación que representa un gran reto, necesariamente se ha de fraccionar en subproblemas más o menos complejos. Algunos de los más importantes y vinculados a este trabajo son:

**Recuperación y extracción de información:** aborda la búsqueda de textos, documentos e, incluso, archivos multimedia sobre grandes volúmenes de información para filtrar la información relevante para el usuario de la que no. La extracción de información es un tipo de recuperación de la información cuyo objetivo es detectar información muy específica y almacenarla de forma estructurada [Moreno et al., 1999].

**Análisis semántico:** el análisis semántico consiste en la obtención del significado oracional a partir de modelos semánticos probabilísticos [Moreno et al., 1999]. Por ejemplo, tenemos ISR-Wornet [Gutiérrez et al., 2010] que es una herramienta que es capaz concebir una red semántica con multitud de recursos lingüísticos.

**Generación de resúmenes:** se trata de obtener una versión reducida del documento o documentos fuente, disminuyendo su contenido de tal forma que se seleccionen y queden presentes en el resumen los conceptos más importantes de dichos documentos. Es el ejemplo de Compendium [Lloret and Palomar, 2013], un sistema de generación de resúmenes para varios propósitos y dominios.

**Minería de opiniones o análisis de sentimientos:** nos permite detectar si una persona está satisfecha con un producto, descubrir la opinión de usuarios del todo el mundo con respecto a un político o empresa incluso prever actos delictivos o posibles casos de suicidios a partir de comentarios en blogs, foros, redes sociales, etc. Esto es muy interesante para por ejemplo generar aplicaciones de seguimiento que analizan en tiempo real entidades de política o turismo entre otras, a partir de las redes sociales [Fernández et al., 2015].

**Reconocimiento de expresiones temporales:** permite identificar expresiones temporales desde una perspectiva multilingüe, utilizando roles semánticos y redes semánticas. Tenemos el caso de TIPSem [Llorens Martínez et al., 2009] un sistema que detecta y extrae expresiones temporales en un texto.

**Desarrollo de corpus y recursos:** es de especial importancia disponer de adecuados recursos lingüísticos y corpus anotados por especialistas de diferentes disciplinas. Como ejemplo, disponemos de la base de datos WordNet [Miller, 1995] utilizada para varios propósitos. Así mismo es interesante disponer de estructuras adecuadas donde guardar y consultar los datos y recursos disponibles para su tratamiento, como pueden ser las ontologías [Dávila et al., 2012].

## 2.4. Antecedente del sistema de integración

La idea de este trabajo es realizar un sistema que integre varias de las herramientas mencionadas en el apartado anterior. Como antecedente a esta plataforma tenemos InTiMe [Gómez, 2008] un proyecto del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante. Esta plataforma trata de integrar una gran parte de los recursos TLH utilizados actualmente. De este modo muchos investigadores tendrán acceso a todas las herramientas y corpus integrados en el sistema de forma remota e independientemente del sistema operativo. Así mismo, permite a desarrolladores e investigadores incorporar sus propios recursos a la plataforma de forma sencilla.

Sin embargo, con el sistema que se acaba de describir solo es posible ejecutar de forma independiente las herramientas. En este trabajo se pretende diseñar un nuevo sistema que aporte la mejora de vincular los resultados de los servicios mediante el paquete semántico, además otras nuevas características que se irán viendo durante la memoria.



## **3. Objetivos**

### **3.1. Objetivos generales**

Se propone crear y desarrollar un marco inteligente que combine e integre las distintas tecnologías del lenguaje humano para procesar textos en base a distintos niveles de análisis lingüístico. Esto dará como salida un paquete de información semántica que almacene de forma estructurada los datos para que mediante determinadas consultas proporcione al usuario la información de interés.

### **3.2. Objetivos específicos**

Por otra parte pretende también cumplir algunos objetivos más concretos:

- Estudiar nociones básicas sobre las tecnologías del lenguaje humano.
- Desarrollar una plataforma que integre diferentes tecnologías de procesamiento del lenguaje humano en un único ecosistema informático.
- Analizar e implementar las técnicas o aplicaciones más apropiadas para relacionar los distintos resultados de las tecnologías implicadas en el ecosistema.
- Establecer un escenario de pruebas del sistema para comprobar que la plataforma responde a las necesidades de información que pueda tener el usuario.



## 4. Metodología

En este apartado se describe las principales metodologías con las que se aborda el proyecto. Por una parte se expone la metodología seguida para este trabajo. Teniendo en cuenta que el objetivo es desarrollar una aplicación informática que integre una serie de herramientas de TLH, también se abordará las tecnologías usadas para el desarrollo de la aplicación.

### 4.1. Metodologías y tecnologías usadas en el proyecto

Los tutores y el alumno han seguido una metodología de trabajo Scrum<sup>1</sup>. Se ha llevado a cabo mediante reuniones periódicas de cada dos semanas aproximadamente, y se han tratado temas referentes a la gestión de tareas, diseño o refinamiento del proyecto, su arquitectura, las tecnologías que utilizará, etc. Así mismo se han planificado hitos para posteriores reuniones.

Durante las reuniones se han ido proponiendo ideas sobre las tecnologías con las que abordar el proyecto. Finalmente en el proyecto se usan las siguientes tecnologías:

- *Java*: el proyecto se ha realizado íntegramente en lenguaje Java sobre el IDE NetBeans. Este dispone de un amplio conjunto de librerías para implementar recursos TLH y son adecuadas para por ejemplo el consumo de los servicios web. Además tiene la ventaja de poder desplegar y ejecutar el proyecto en prácticamente cualquier plataforma.
- *XML*: se ha utilizado el lenguaje XML para la creación de ficheros de configuración del proyecto Java. Se ha decidido que es el mejor formato para definir la configuración ya que es estándar, organizado e intuitivo para usuarios que tienen que editarlo.
- *JSON*: se ha utilizado para el formato de entrada y salida de datos del sistema. Principalmente por ser un formato muy extendido y bastante flexible a la hora de procesarlo.
- *MongoDB*: hemos visto adecuado utilizar una base de datos NoSQL para almacenar de forma rápida todos los resultados de la invocación de servicios TLH (lo que llamamos paquete semántico). A su vez permite diseñar consultas de datos muy similares a las de SQL.

---

<sup>1</sup>Scrum es una metodología ágil y flexible para gestionar el desarrollo de software.

- *Servicios web (SOAP y REST)*: los servicios web son útiles para poner a disposición de todos (remotamente) muchas de las herramientas TLH. SOAP (siglas de Simple Object Access Protocol) es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos XML. Mientras que los servicios REST (Representational State Transfer) utilizan una comunicación más simple y ligera. Ambas tecnologías son usadas en los servicios que se deben integrar en el sistema y son perfectamente incorporables con el proyecto Java del mismo. En el siguiente apartado se describirá concretamente qué servicios de herramientas TLH se han utilizado.



# 5. Desarrollo

## 5.1. Arquitectura de TLH Suite

TLH Suite es un proyecto desarrollado en Java para línea de comandos Linux. Dispone de una serie de ejecutables BASH Script correspondientes a los diferentes módulos que componen el sistema, los cuales se explican en las próximas secciones. La salida de estos se muestra por consola o se guarda en una base de datos MongoDB con una estructura JSON, siendo este resultado lo que se conoce como paquete semántico.

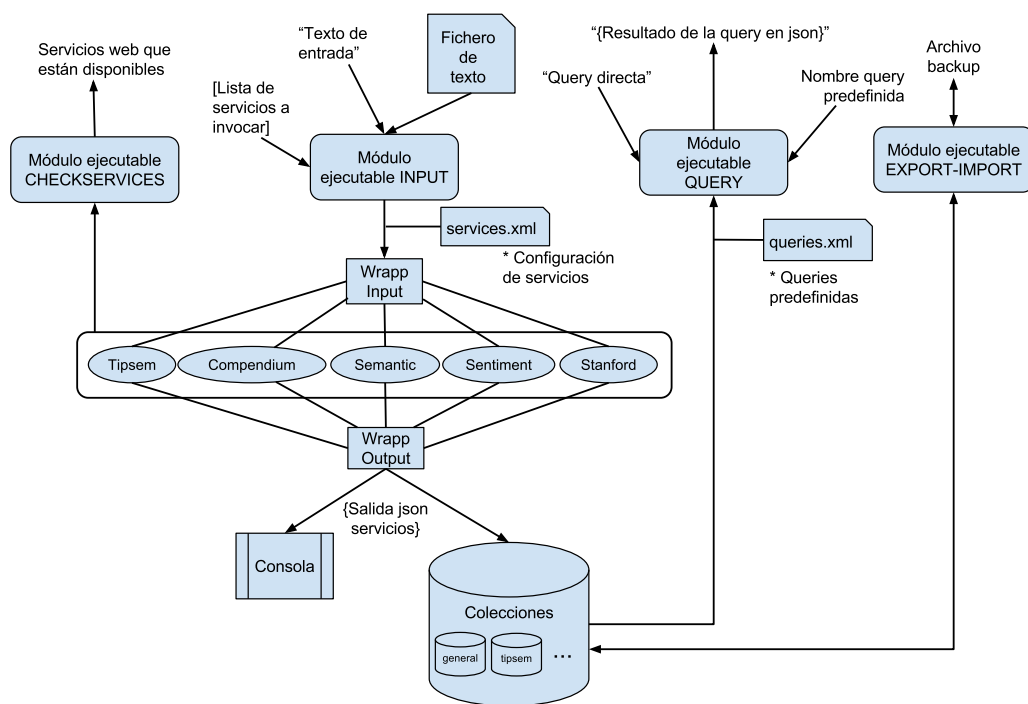


Figura 5.1.: Arquitectura TLH Suite.

La arquitectura mostrada en la figura 5.1 trata de describir la composición y funcionamiento de los módulos, así como la interacción de estos con los elementos de entrada, salida, procesado y almacenamiento.

## 5.2. Ficheros de configuración

Los ficheros de configuración del proyecto son:

- `services.xml`: fichero donde se definen los servicios web asociados al sistema (lo usa el módulo Input). Se puede observar un ejemplo de este fichero en la figura 5.2.

```
<?xml version="1.0" encoding="UTF-8"?>
<services>
  <service>
    <!-- Activar o desactivar servicio (Valores para activar: yes, true, on, 1) -->
    <active>yes</active>
    <!-- Nombre del servicio -->
    <name>sentiment</name>
    <!-- Clase donde se ha implementado los wrappers y el uso del servicio -->
    <class>es.ua.tlhsuite.wrapper.SentimentWrapper</class>
  </service>
  <service>
    ...
  </service>
  ...
</services>
```

Figura 5.2.: Ejemplo de fichero `services.xml`.

- `queries.xml`: fichero para predefinir consultas y usarlas indicando su nombre (lo usa el módulo Query). Se puede observar un ejemplo de este fichero en la figura 5.3.

```
<?xml version="1.0" encoding="UTF-8"?>
<queries>
  <query>
    <!-- Nombre de la query -->
    <name>basic</name>
    <!-- Colección a usar (FROM) -->
    <collection>compendium</collection>
    <!-- Campos que mostrar (SELECT) -->
    <fields>{ "_id" : 0 , "summary" : 1 }</fields>
    <!-- Filtro de la query (WHERE) -->
    <filter>{ "summary" : "Este resumen" }</filter>
    <!-- Limite de resultados a obtener (LIMIT) -->
    <limit>10</limit>
    <!-- Saltar resultados (LIMIT) -->
    <skip>5</skip>
  </query>
  <query>
    ...
  </query>
  ...
</queries>
```

Figura 5.3.: Ejemplo de fichero `queries.xml`.

En la figura 5.4 se muestra una equivalencia entre el formato de query nativa de TLH Suite y una query MySQL.

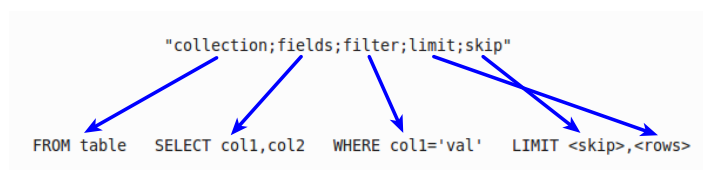


Figura 5.4.: Equivalencia entre query TLH Suite y MySQL.

### 5.3. Servicios web seleccionados

Se ha hecho una selección de servicios web de herramientas TLH que serán incorporadas en el sistema. Uno de los criterios para la selección es que sean herramientas capaces de caracterizar la información de un documento en base a distintos niveles. Debido a que algunos servicios solo funcionan para el idioma inglés, se ha decidido que el proyecto funcione solo para este idioma, aunque algunos de los servicios también funcionan para español. En la tabla 5.1 se muestran las herramientas seleccionadas e información sobre ellas.

Servicio	Tarea TLH	Nombre técnico	Tipo	Idioma	Público
Tipsem	Reconocedor de expresiones temporales	TIPSem	REST	Inglés	Si
Compendium	Generación de resúmenes	COMPENDIUM	SOAP	Inglés	Si
Semantic	Análisis semántico	ISR-WN	SOAP	Inglés	No
Sentiment	Análisis de sentimientos	SA	REST	Inglés Español	Si
Stanford	Reconocedor entidades nombradas	Stanford NER	REST	Inglés Español	Si

Tabla 5.1.: Selección de servicios web para incluirlos en TLH Suite.

Todos los servicios han sido desarrollados y publicados en el seno del GPLSI (salvo Stanford<sup>1</sup>) como resultado de métodos y tesis realizadas en el grupo de investigación.

<sup>1</sup>La herramienta StanfordNER ha sido creada por investigadores de la universidad de Stanford, en el grupo solo la hemos incorporado en un servicio web.

## 5.4. Wrappers de los servicios web

Para cada uno de los servicios se ha implementado un wrapper con Java. Un wrapper es una implementación que encapsula la invocación de un servicio web, de manera que se puedan adaptar sus entradas y salidas a otras tecnologías o formatos de representación. En el caso particular de TLH Suite los wrappers son clases Java que hacen esto, como podemos ver en la figura 5.5.

En la implementación del wrapper se pueden ver 3 partes diferenciadas:

- *Wrapp input*: para adaptación de las entradas al servicio.
- *Call service*: para la invocación del servicio web.
- *Wrapp output*: para la adaptación de la salida del servicio.

En particular para los servicios seleccionados de TLH Suite, en *Wrapp input* se suele adaptar el texto plano a un formato aceptado por el servicio web. Mientras que en *Wrapp output* se transforma la salida a formato JSON. Muchos de los servicios web trabajan con JSON directamente, así que en ocasiones no es necesario reformatear la entrada o salida.

```

@Override
public String run(String input) throws ServiceWrapperException {
    try {
        // Wrapp input
        input = "{\"ugcs\" : [\"" + input + "\",\"contexts\" : [],\"subjects\" : []]";

        // Call service
        String output = serviceRest.call(input);

        // Wrapp output
        JsonParser parser = new JsonParser();
        JsonElement o = (JsonElement) parser.parse(output);
        JsonObject resultObj = new JsonObject();
        resultObj.add("category", o.getAsJsonObject().get("semanticAttributes").getAsJs
        resultObj.add("intensity", o.getAsJsonObject().get("semanticAttributes").getAsJs

        return resultObj.toString();
    } catch (CallServiceException ex) {
        throw new ServiceWrapperException(ex);
    }
}

```

Figura 5.5.: Ejemplo de implementación de un wrapper.

## 5.5. Módulos y ejecutables

En este apartado se describirá los módulos que componen el sistema encargados de realizar las tareas principales de la plataforma. Así mismo, cada uno de los módulos dispone de su correspondiente programa BASH Script para ser ejecutado en línea de comandos, ya que en esta versión de TLH Suite solo se dispone de esta interfaz para su uso. En los apartados siguientes se hará una breve descripción de cada módulo seguido de la explicación del uso de su correspondiente ejecutable.

### 5.5.1. Módulo Input

El módulo Input se ocupa del procesamiento de documentos. Se utiliza para invocar los servicios web vinculados al sistema, procesar la entrada con estos y almacenar los resultados. La entrada es texto plano o un fichero de texto. La salida es por consola (formato JSON) o almacenamiento en MongoDB. El módulo realiza invocación de los servicios de forma secuencial.

El ejecutable que se encarga de hacer funcionar el módulo es `input.sh`. Su sintaxis es la siguiente:

```
# ./input [-h] [--text TEXT] [--file FILE] [--metadata METADATA] --services SERVICES
[SERVICES ...] [--dbname DBNAME] [--services-file SERVICESFILE] [--storedb] [--no-storedb]
[--storepartial] [--no-storepartial] [--splitphrases] [--no-splitphrases] [--overwritedata]
[--no-overwritedata] [--escid ESCID] [--cleantext] [--no-cleantext]
```

A continuación se describe cada parámetro:

- `--text`: entrada de texto plano del documento.
- `--file`: entrada del documento (fichero de texto plano).
- `--metadata`: entrada de metadatos asociados al documento de entrada. Deben estar en formato JSON.
- `--services`: hay que indicar obligatoriamente una lista de 1 a N valores separados por espacios. Los valores corresponden a los servicios que se van a invocar en la ejecución: `compendium`, `sentiment`, `semantic`, `stanford` o `tipsem`.
- `--dbname`: nombre de la base de datos MongoDB donde se van a guardar los resultados de la ejecución (si no existe, se crea).
- `--services-file`: fichero de configuración XML de los servicios web definidos para TLH Suite.
- `--storedb`, `--no-storedb`: activa o desactiva el flag de almacenamiento en base de datos de los resultados. Si no se activa solo se muestran por consola. Por defecto está activado.
- `--storepartial`, `--no-storepartial`: activa o desactiva el flag de guardar parcialmente datos a medida que se invocan los servicios. Si no se activa se mantienen los resultados en memoria y no se guardan en base de datos hasta que finaliza la ejecución. Si se activa se guardan los resultados parciales de cada servicio que se ha ido invocando. Por defecto está desactivado.
- `--splitphrases`, `--no-splitphrases`: activa o desactiva el flag de separar el documento en frases para procesarlas individualmente. Por defecto está activado.
- `--overwritedata`, `--no-overwritedata`: activa o desactiva el flag de sobrescribir los resultados de la ejecución. Por defecto está desactivado.

- `--escid`: indicar el identificador de escenario manualmente por parte del usuario. Si no se indica se generará un identificador genérico.
- `--cleantext`, `--no-cleantext`: activa o desactiva el flag de limpieza de texto del documento de entrada. Si se activa eliminará del texto los caracteres que no sean alfanuméricos, punto, coma o espacio. Por defecto está activado.

### 5.5.2. Módulo Query

Se ocupa de las consultas de los resultados generados y procesados por el módulo Input y guardados en la base de datos MongoDB. Las consultas se pueden indicar directamente por línea de comandos o usando un fichero XML de queries. La salida puede ser por consola y/o en un fichero, siempre en formato JSON. Hay dos formas de realizar las consultas, mediante una query nativa indicada como parámetro, o invocando queries predefinidas a partir de un fichero de configuración XML. En el caso de utilizar el fichero XML de queries también se puede indicar el fichero de salida donde queremos guardar el resultado de la consulta.

El ejecutable que se encarga de hacer funcionar el módulo es `query.sh`. Su sintaxis es la siguiente:

```
# ./query [-h] [--query QUERY] [--qname QNAME] [--allqexec] [--no-allqexec] [--qfile QFILE] [--escid ESCID] [--printout] [--no-printout] [--prettyout] [--no-prettyout] [--dbname DBNAME]
```

Los parámetros que recibe son:

- `--query`: para ejecutar una query nativa, esta se indica con una cadena con el formato que describe la figura 5.4.
- `--qname`: en el caso de que se indique un fichero de queries, con este parámetro se indica el nombre de la query a ejecutar del fichero.
- `--qfile`: para indicar la ruta de un fichero XML de queries predefinidas. Un ejemplo de formato para el fichero es el descrito en la figura 5.3.
- `--allqexec`, `--no-allqexec`: activar o desactivar el flag de ejecutar todas las queries del fichero. Por defecto está desactivado.
- `--dbname`: nombre de la base de datos MongoDB donde realizar las consultas.
- `--escid`: para filtrar el ámbito de la consulta en un escenario concreto.
- `--prettyout`, `--no-prettyout`: activar o desactivar el flag de embellecer el resultado JSON de la consulta. Por defecto está activado.
- `--printout`, `--no-printout`: activar o desactivar el flag de mostrar el resultado de la consulta por la consola. Por defecto está activado.

Para el caso de las queries predefinidas se debe crear un fichero XML con el formato que describe la figura 5.3. El formato de las queries nativas se expone en la figura 5.4, se trata de indicar valores entre “;”. Estos valores corresponden ciertos campos propios de las queries de MongoDB.

Los valores de las queries nativas se describen a continuación:

- **collection:** se trata de la colección de datos MongoDB a seleccionar. Es equivalente a la sentencia FROM de MySQL.
- **fields:** se trata de de los campos a seleccionar de la colección. Es equivalente a la sentencia SELECT de MySQL.
- **filter:** sirve para definir ciertas reglas y condiciones con las que filtrar los datos. Es equivalente a la sentencia WHERE de MySQL.
- **limit:** para limitar el número de resultados. Es equivalente al segundo campo de la sentencia LIMIT de MySQL.
- **skip:** para saltar los primeros resultados. Es equivalente al primer campo de la sentencia LIMIT de MySQL.

### 5.5.3. Módulo Check Services

Se ocupa de comprobar la disponibilidad de los servicios. Comprueba secuencialmente cada uno los servicios web vinculados al sistema y se indica por consola si está disponible y funcional. Cada servicio se invoca con un ejemplo de entrada para comprobar el correcto funcionamiento del mismo.

El ejecutable que se encarga de hacer funcionar el módulo es `checkservices.sh`. Su sintaxis es la siguiente:

```
# ./checkservices.sh
```

No recibe ningún parámetro.

### 5.5.4. Módulo Export/Import

Este módulo se encarga de exportación e importación de datos. La exportación consiste en extraer y guardar en un fichero backup las colecciones de datos MongoDB donde se almacenan los resultados de las ejecuciones de TLH Suite. Por otra parte la importación consiste en lo contrario, generar estas colecciones de datos a partir del fichero backup. El fichero backup para la exportación e importación es un comprimido zip que contiene ficheros de texto JSON correspondientes a las colecciones.

A diferencia del resto de módulos este dispone de dos ejecutables distintos llamados `export.sh` e `import.sh`. El ejecutable `export.sh` exporta todos los datos de las colecciones de un escenario guardado en la base de datos MongoDB cuyo nombre es indicado como parámetro. Los datos de cada colección son guardados en un fichero y después se crea el archivo zip de backup con todos ellos. El ejecutable `import.sh` por el contrario,

extrae del zip los ficheros de las colecciones y las importa una a una a la base de datos MongoDB indicada como parámetro. A continuación se describe la sintaxis de ambos ejecutables, los parámetros se deben indicar en el orden que aparecen.

La sintaxis de `export.sh` es la siguiente:

```
# ./export.sh <dbname_export> <filename_export>
```

Los parámetros que recibe son:

- `<dbname_export>`: es el nombre de la base de datos MongoDB que se exportará.
- `<filename_export>`: es el nombre para el exportado, se creará un fichero zip cuyo nombre se compone del valor indicado en este parámetro junto con la fecha de generación.

La sintaxis de `import.sh` es la siguiente:

```
# ./import.sh <dbname_import> <file_import>
```

Los parámetros que recibe son:

- `<dbname_import>`: es el nombre de la base de datos MongoDB donde se importará. Si no existe una base de datos con este nombre se crea.
- `<file_import>`: el nombre del fichero zip a importar.

## 5.6. Entrada de datos, ejecución y almacenamiento de resultados

La ejecución de TLH Suite se hace a nivel de documento, siendo este un texto de una temática concreta. Sin embargo, TLH Suite divide este documento en frases para invocar cada servicio web con cada una de ellas (exceptuando el servicio Compendium que se invoca con el texto completo del documento para generar adecuadamente el resumen). Por lo tanto los resultados que se almacenan son a nivel de frase (donde la frase es la sentencia que hay hasta un punto). Por otra parte se puede anotar que el documento pertenece a un escenario particular, esto quiere decir que los documentos del mismo escenario pertenecen a la misma temática, y se tendrá en cuenta a la hora de almacenar y consultar resultados. En la figura 5.6 se describe de forma gráfica cada uno de estos contextos.



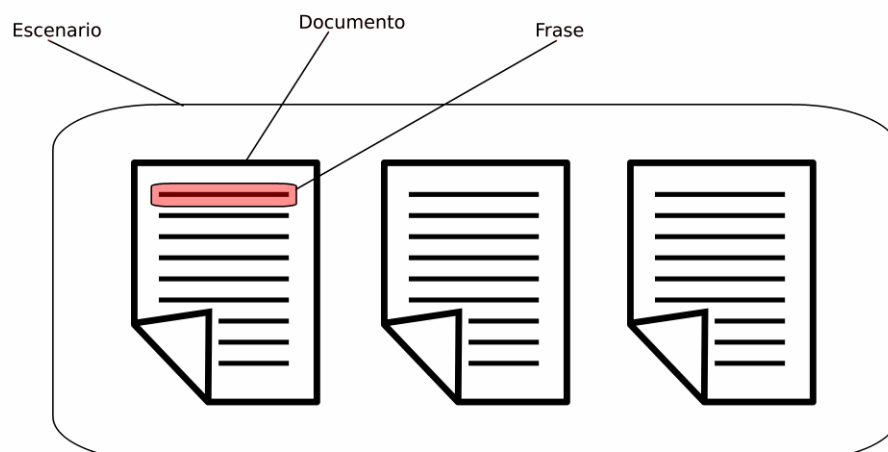


Figura 5.6.: Partes de la entrada de datos.

El almacenamiento de los resultados se hace en la base de datos MongoDB. Esta contiene una serie de colecciones (equivalente a las tablas en una base de datos relacional). En cada colección se guarda la información en formato JSON resultante de la transformación realizada por el wrapper de cada uno de los servicios web de TLH Suite. Cada dato de la información se guarda a su vez en campos definidos en cada colección. A continuación se describe con más detalle la información que almacena cada colección y sus campos definidos.

### 5.6.1. Colección de datos “doc”

Esta colección se utiliza para guardar el documento original con el que se ejecuta TLH Suite.

Los campos definidos en esta colección son:

- **doc\_id**: identificación del documento, si el usuario no la indica MongoDB crea una genérica.
- **esc\_id**: identificación del escenario, si el usuario no la indica MongoDB crea una genérica.
- **phrases**: array de las frases del documento separadas. Una frase es una sentencia del texto original acabada en punto.
- **text**: texto original del documento de entrada.

### 5.6.2. Colección de datos “compendium”

Esta colección es donde se guarda la generación del resumen resultado de la invocación del servicio Compendium.

Los campos definidos en esta colección son:

- **doc\_id**: identificación del documento, si el usuario no la indica, MongoDB crea una genérica.
- **esc\_id**: identificación del escenario, si el usuario no la indica, MongoDB crea una genérica.
- **text**: texto original del documento con el que se ha invocado el servicio Compendium.
- **compendium**: salida del servicio Compendium en formato JSON. En este caso solo se guarda un sub-campo “full-summary” con el resumen completo.

### 5.6.3. Colección de datos “semantic”

En esta colección se guarda la información semántica resultado de la invocación del servicio Semantic.

Los campos definidos en esta colección son:

- **doc\_id**: identificación del documento, si el usuario no la indica, MongoDB crea una genérica.
- **esc\_id**: identificación del escenario, si el usuario no la indica, MongoDB crea una genérica.
- **phrase**: frase original del documento con la que se ha invocado el servicio Semantic.
- **semantic**: salida del servicio Semantic en formato JSON.

### 5.6.4. Colección de datos “sentiment”

En esta colección guarda la evaluación de sentimientos resultado de la invocación del servicio Sentiment.

Los campos definidos en esta colección son:

- **doc\_id**: identificación del documento, si el usuario no la indica, MongoDB crea una genérica.
- **esc\_id**: identificación del escenario, si el usuario no la indica, MongoDB crea una genérica.
- **phrase**: frase original del documento con la que se ha invocado el servicio Sentiment.
- **sentiment**: salida del servicio Sentiment en formato JSON.

### 5.6.5. Colección de datos “stanford”

Esta colección es donde se guarda las entidades nombradas resultado de la invocación del servicio Stanford.

Los campos definidos en esta colección son:

- **doc\_id**: identificación del documento, si el usuario no la indica, MongoDB crea una genérica.
- **esc\_id**: identificación del escenario, si el usuario no la indica, MongoDB crea una genérica.
- **phrase**: frase original del documento con la que se ha invocado el servicio Stanford.
- **stanford**: salida del servicio Stanford en formato JSON.

### 5.6.6. Colección de datos “tipsem”

Esta colección se utiliza para guardar las expresiones temporales resultado de la invocación del servicio Tipsem.

Los campos definidos en esta colección son:

- **doc\_id**: identificación del documento, si el usuario no la indica, MongoDB crea una genérica.
- **esc\_id**: identificación del escenario, si el usuario no la indica, MongoDB crea una genérica.
- **phrase**: frase original del documento con la que se ha invocado el servicio Stanford.
- **tipsem**: salida del servicio Tipsem en formato JSON.

### 5.6.7. Colección de datos “general”

Esta colección reúne toda la información guardada de las colecciones anteriormente descritas. Con esto se pretende conseguir consultas que relacionen los datos de las diferentes colecciones.

Los campos definidos en esta colección son:

- **doc\_id**: identificación del documento, si el usuario no la indica, MongoDB crea una genérica.
- **esc\_id**: identificación del escenario, si el usuario no la indica, MongoDB crea una genérica.
- **phrase**: frase original del documento con la que se ha invocado los servicios.
- **text**: texto original del documento con el que se ha invocado el servicio Compendium.

- `compendium`: salida del servicio Compendium en formato JSON.
- `semantic`: salida del servicio Semantic en formato JSON.
- `sentiment`: salida del servicio Sentiment en formato JSON.
- `stanford`: salida del servicio Stanford en formato JSON.
- `tipsem`: salida del servicio Tipsem en formato JSON.

## 6. Evaluación y resultados

### 6.1. Definición de escenarios de prueba

Se han propuesto 3 escenarios para las pruebas del sistema. Cada escenario consiste en determinar una serie de documentos de una temática con el fin de evaluar la plataforma ante diversas áreas de interés. Así mismo el tamaño y la cantidad de documentos seleccionados para cada escenario variará para comprobar el rendimiento y correcto funcionamiento del sistema. Los documentos han sido extraídos de páginas web de noticias y cine. Estos documentos completos se pueden consultar vía web en el Anexo I: Documentos de los escenarios.

En concreto estos son los escenarios seleccionados:

**Escenario 1:** Artículo sobre la apertura de McDonalds en la Unión Soviética

**Temática:** gastronomía

**Documentos:** 1 documento grande

**Número de frases:** 47

**Número de palabras:** 1052

**Idioma:** inglés

**Escenario 2:** Noticia sobre un partido de tenis de Rafa Nadal

**Temática:** deportes

**Documentos:** 2 documentos pequeños

**Número de frases:** 15 (Primer documento: 9, Segundo documento: 6)

**Número de palabras:** 281 (Primer documento: 157, Segundo documento: 124)

**Idioma:** inglés

**Escenario 3:** Biografía del director de cine Stanley Kubrick

**Temática:** cine

**Documentos:** 2 documentos grandes

**Número de frases:** 119 (Primer documento: 52, Segundo documento: 67)

**Número de palabras:** 2053 (Primer documento: 1115, Segundo documento: 938)

**Idioma:** inglés

## 6.2. Tabla de resultados

Se ha diseñado un paquete de 50 consultas de distintos tipos y complejidad para evaluar el sistema, y así compararlo con respecto a la utilización individual de las herramientas de TLH. Esto supondrá una mejora significativa con respecto a la forma actual de utilizar las diferentes herramientas.

En la tabla 6.1 se proporciona la siguiente información: el número de la consulta, el enunciado de la consulta, si se ha podido resolver con TLH Suite, si se puede resolver sin TLH Suite, es decir, utilizando los servicios web individualmente como se hace actualmente. Los resultados detallados de cada consulta se pueden encontrar en el Anexo II: Resultados de las consultas.

#	Consulta	Con TLH Suite	Sin TLH Suite
01	Entidades lugar	✓	✓
02	Entidades persona	✓	✓
03	Expresiones temporales que aparecen	✓	✓
04	Número de frases positivas	✓	✓
05	Número de frases negativas	✓	✓
06	Frases que están en los resúmenes	✓	✓
07	Dominios generales del escenario 1	✓	✓
08	Dominios generales del escenario 2	✓	✓
09	Dominios generales del escenario 3	✓	✓
10	Frases negativas que contengan la entidad LOCATION	✓	X
11	Frases del resumen que contengan una referencia temporal que sea de los años 2014 o 2015	✓	X
12	Resumen que contenga alguna frase positiva	✓	X
13	Frases que contengan la entidad LOCATION y al menos una referencia temporal	✓	X
14	Frases que contengan la entidad PERSON y al menos una referencia temporal	✓	X
15	Dominios de las frases del resumen	✓	X
16	Frases del resumen que contienen alguna entidad nombrada	✓	X
17	Entidades nombradas de frases negativas	✓	X
18	Dominios de las frases positivas	✓	X
19	Frases del resumen negativas	✓	X
20	Frases que no están en el resumen	✓	X
21	Frases que hagan referencia al año 2015 y sean neutrales o positivas	✓	X

Sigue en la página siguiente.

#	Consulta	Con TLH Suite	Sin TLH Suite
22	Dominios de frase que contengan alguna referencia temporal	✓	X
23	Frases positivas que contengan alguna entidad NUMBER	✓	X
24	Frases negativas que tengan alguna referencia temporal y la entidad LOCATION	✓	X
25	Frases que contengan la entidad LOCATION y no contengan alguna referencia temporal	✓	X
26	Entidades nombradas que están implicadas con el dominio Gastronomy (Escenario 1)	✓	X
27	Entidades nombradas que están implicadas con el dominio Sport (Escenario 2)	✓	X
28	Entidades nombradas que están implicadas con el dominio Cinema (Escenario 3)	✓	X
29	Frases que hablen de la entidad McDonalds (Escenario 1)	✓	X
30	Frases que hablen de la entidad Nadal (Escenario 2)	✓	X
31	Frases que hablen de la entidad Kubrick (Escenario 3)	✓	X
32	Frases positivas y negativas con el dominio Gastronomy (Escenario 1)	✓	X
33	Frases positivas y negativas con el dominio Sport (Escenario 2)	✓	X
34	Frases positivas y negativas con el dominio Cinema (Escenario 3)	✓	X
35	Frases de resumen que contengan una referencia temporal y el dominio Gastronomy (Escenario 1)	✓	X
36	Frases de resumen que contengan una referencia temporal y el dominio Sport (Escenario 2)	✓	X
37	Frases de resumen que contengan una referencia temporal y el dominio Cinema (Escenario 3)	✓	X
38	Frases positivas/negativas y que hablen de la entidad McDonalds (Escenario 1)	✓	X
39	Frases positivas/negativas y que hablen de la entidad Nadal (Escenario 2)	✓	X

Sigue en la página siguiente.

#	Consulta	Con TLH Suite	Sin TLH Suite
40	Frases positivas/negativas y que hablen de la entidad Kubrick (Escenario 3)	✓	X
41	Frases negativas que contengan la entidad McDonalds (Escenario 1)	✓	X
42	Frases negativas que contengan la entidad Nadal (Escenario 2)	✓	X
43	Frases negativas que contengan la entidad Kubrick (Escenario 3)	✓	X
44	Obtener todas las entidades de las frases que contienen McDonalds (Escenario 1)	✓	X
45	Obtener todas las entidades de las frases que contienen Nadal (Escenario 2)	✓	X
46	Obtener todas las entidades de las frases que contienen Kubrick (Escenario 3)	✓	X
47	Todas las expresiones temporales distintas	X	X
48	Entidades nombradas NUMBER coincidentes en el escenario 2 y 3	X	X
49	Fechas asociadas a una persona	X	X
50	Polaridad del escenario completo	X	X

Tabla 6.1.: Consultas de evaluación de TLH Suite.

Es importante añadir que las consultas de prueba se realizan sobre la colección general que contiene los datos de todas las colecciones. Las consultas siempre se realizan sobre los 3 escenarios juntos, pero en algunos casos se aplican ciertos filtros o condiciones que limitan los resultados de un escenario concreto. Para más información consultar el Anexo II: Resultados de las consultas.

### 6.3. Discusión de los resultados

Con la evaluación que se ha hecho del sistema se pretende demostrar que TLH Suite alcanza un nivel más de refinamiento a la hora de consultar y vincular toda la información almacenada. Como se puede observar, TLH Suite puede resolver consultas donde se realizan filtros relacionando la información obtenida de diferentes servicios. Anteriormente, usando individualmente estos servicios no se podían hacer conexiones de la información y por lo tanto no se puede consultar con criterios que abarcan más de uno de los servicios. De ahí el motivo de que las consultas que hay a partir de la número 10 no se puedan resolver sin TLH Suite.

Por otra parte, todavía falta mayor conexión entre datos, mayor nivel de análisis y dar soporte a consultas más complejas. De ahí que las consultas número 47, 48, 49 y 50 no se puedan resolver. A continuación se describe con más detalle los motivos de por qué no ha sido posible resolverlas con TLH Suite:



- La consulta 47 no se puede resolver por que las consultas TLH Suite no tienen la funcionalidad para omitir los resultados repetidos (equivalente a DISTINCT de MySQL).
- La consulta 48 no se puede resolver por que las consultas TLH Suite no soportan las combinaciones de conjuntos de resultados.
- La consulta 49 no se puede resolver con TLH Suite porque no existe una relación entre una expresión temporal de Tipsem y una entidad PERSON del servicio Stanford. Sin esta relación por tanto, no se puede obtener una asociación de estos conceptos.
- La consulta 50 no se puede resolver con TLH Suite porque no se evalúa ni se guarda el análisis de sentimientos de todo el texto del escenario.

Por lo tanto, es precioso contar con una forma de almacenamiento que pueda hacer conexiones de datos en los casos similares a los de la consulta 49, un tipo de almacenamiento que se aproxima a esta propuesta puede ser una ontología. Por otra parte, en preguntas del estilo 50 es necesario contar con un mayor nivel de análisis para poder disponer de más información semántica. En los casos de las consultas 47 y 48 simplemente sería implementar estas funcionalidades en futuras versiones de TLH Suite.



## 7. Conclusiones

### 7.1. Resumen del trabajo realizado

La propuesta inicial era desarrollar un marco inteligente que combine e integre las distintas tecnologías del lenguaje humano para procesar textos en base a distintos niveles de análisis lingüístico. El sistema TLH Suite cumple este objetivo con las herramientas que se han seleccionado y está capacitado para añadir nuevas herramientas en el futuro. El procesamiento en los servicios integrados y el diseño de las consultas basadas en MongoDB permiten análisis lingüísticos a diferentes niveles. Por otra parte, la arquitectura y el diseño del sistema permite flexibilizar la incorporación de nuevas características para aportar nuevas funcionalidades y afrontar nuevos objetivos futuros.

### 7.2. Trabajo futuro

Se proponen las siguientes ideas como trabajo futuro:

**Incorporar una ontología:** Se haría de la misma forma que se ha incorporado MongoDB, tendría módulos Input y Query propios, además hay que definir un nuevo fichero de configuración `queries.xml` para las consultas tipo SPARQL propias de ontologías. A parte, una ontología mejoraría la conexiones entre toda la información almacenada y aportaría un mayor nivel de análisis lingüístico.

Sobre la forma de poblar la ontología, se consideran 2 propuestas:

- *A partir de consultas MongoDB.* Esta propuesta sería más fácil, ya que se puede plantear añadiendo un nuevo módulo a la arquitectura manteniendo el diseño actual de la aplicación, aunque es necesario almacenar previamente los resultados en MongoDB.
- *A partir de la salida de los servicios.* Esta propuesta es más complicada en términos de implementación, ya que implementar una gestión de ontologías complicaría el diseño actual de la aplicación.

**Realizar interfaz web:** Para el uso de los módulos desde web. Sería fácil de implementar porque los módulos están preparados para ser instanciados desde una interfaz proporcionando al módulo los parámetros necesarios. La interfaz web daría mayores facilidades al usuario a la hora de utilizar el sistema y de evaluar los resultados.

**Más flexibilidad en la incorporación de nuevos servicios:** mediante wrappers y el fichero de configuración `services.xml` se ha intentado conseguir una forma fácil

e intuitiva de incorporar un servicio web a la plataforma. Sin embargo la idea, si es factible llevarla a cabo, es poder incorporar nuevos servicios únicamente cambiando el fichero de configuración o buscando un nuevo método que evite la programación al añadir nuevos servicios a la plataforma.

## Bibliografía

- [Dávila et al., 2012] Dávila, H., Fernández, A., Gutiérrez, Y., Muñoz, R., Montoyo, A., and Vázquez, S. (2012). Método de extracción de información semántica en ontologías semantic information extraction method on ontologies.
- [Fernández et al., 2015] Fernández, J., Gutiérrez, Y., Gómez, J. M., and Martínez-Barco, P. (2015). Social rankings: análisis visual de sentimientos en redes sociales. *Procesamiento del Lenguaje Natural*, 55:199–202.
- [GPLSI, 2011] GPLSI (2011). Líneas de investigación. <http://gplsi.dlsi.ua.es/gplsi11/researchlines>. [Online; accessed 2015-11-09].
- [Gutiérrez et al., 2010] Gutiérrez, Y., Fernández, A., Montoyo, A., and Vázquez, S. (2010). Integration of semantic resources based on wordnet.
- [Gómez, 2008] Gómez, J. M. (2008). InTiMe: plataforma de integración de recursos de PLN. <http://hdl.handle.net/10045/5039>. [Online; Apr-2008].
- [Llorens Martínez et al., 2009] Llorens Martínez, H., Navarro Colorado, F. d. B., Saquete Boró, E., et al. (2009). Detección de expresiones temporales timeml en catalán mediante roles semánticos y redes semánticas.
- [Lloret et al., 2015] Lloret, E., Gutiérrez, Y., Peregrino, F. S., Soriano, J. M. G., Guillén, A., and Llopis, F. (2015). Explotación y tratamiento de la información disponible en internet para la anotación y generación de textos adaptados al usuario. *Procesamiento del Lenguaje Natural*, 55:177–180.
- [Lloret and Palomar, 2013] Lloret, E. and Palomar, M. (2013). Compendium: a text summarisation tool for generating summaries of multiple purposes, domains, and genres. *Natural Language Engineering*, 19(02):147–186.
- [Miller, 1995] Miller, G. A. (1995). Wordnet: A lexical database for english. *COMMUNICATIONS OF THE ACM*, 38:39–41.
- [Moreno et al., 1999] Moreno, L., Palomar, M., Molina, A., and Ferrández, A. (1999). Introducción al procesamiento del lenguaje natural. *Servicio de Publicaciones Universidad de Alicante*. *Universidad de Alicante*.



## A. Anexo I: Documentos de los escenarios

En este anexo se pueden consultar los documentos completos de cada escenario. Debido a que son muy extensos, no se han podido exponer en esta memoria, con lo que se han publicado en la red. Mediante los siguientes enlaces se puede acceder al contenido de cada uno:

**Escenario 1:** Artículo sobre la apertura de McDonalds en la Unión Soviética.

- [\[Documento 1\]](#)

**Escenario 2:** Noticia sobre un partido de tenis de Rafa Nadal.

- [\[Documento 1\]](#)
- [\[Documento 2\]](#)

**Escenario 3:** Biografía del director de cine Stanley Kubrick.

- [\[Documento 1\]](#)
- [\[Documento 2\]](#)

Por otra parte, en este anexo también se exponen las frases sueltas que componen cada escenario. Estas se muestran en los siguientes enlaces entre saltos de línea:

- [\[Frases del escenario 1\]](#)
- [\[Frases del escenario 2\]](#)
- [\[Frases del escenario 3\]](#)





## B. Anexo II: Resultados de las consultas

En este anexo se exponen los resultados de cada una de las consultas. Estos han sido reformateados y reducidos por motivos de legibilidad y espacio. Las consultas siempre se realizan sobre los 3 escenarios juntos. Hay consultas en las que se aplica un filtro o condición propia de un escenario concreto, en estos casos se indica entre paréntesis en el enunciado de la consulta a que escenario pertenece dicho valor de filtro. Por ejemplo en una consulta tipo “cuya entidad nombrada sea Kubrick (Escenario 3)”, el valor Kubrick es un filtro para el Escenario 3.

### C01: Entidades lugar

Resultado:

```
1 { "EntityLemma": ["America", "Atlanta", "Barcelona", "Bronx", "Budapest", "California", "England", "Kubricks", "Manhattan", "Montreal", "Moscow", "Nashua", "New", "oscar nominated", "Sao", "Soviet", "UK", "United"] }
```

### C02: Entidades persona

Resultado:

```
1 { "EntityLemma": ["Arthur", "Barry", "Beeg", "Christiane", "Clarke", "Federico", "Fognini", "Fogninis", "Jack", "Kei", "Kirk", "Kubrick", "Larissa", "Lebedeva", "Lubov", "Marc", "Metty", "Mikhail", "Nadal", "Rafael", "Rumour", "Russell", "Ruth", "Sam", "Stanley", "Stephen", "Steven", "Strangelove", "Victor", "Walter", "Yahel", "Zvetlana"] }
```

### C03: Expresiones temporales que aparecen

Resultado:

```
1 { "times": [{ "value": "2015-09-23", "term": "Wednesday"
  }, { "term": "14 years", "value": "P14Y" }, { "term": "
  nine months", "value": "P9M" }, { "value":
  "2015-09-23", "term": "Wednesday" }, { "term": "recent
  years", "value": "PAST_REF" }, { "term": "three months
  ", "value": "P3M" }, { "term": "now", "value": "
  PRESENT_REF" }, { "term": "the 23rd of April 2015", "
  value": "2015-04-23" }, { "term": "12 years", "value":
  "P12Y" }, { "term": "2003", "value": "2003" }, { "term
  ": "earlier this year", "value": "2015" }, { "term": "
  last years", "value": "2014" }, { "value": "1940", "
  term": "1940" }, { "term": "1941", "value": "1941" }, {
  "value": "1946", "term": "1946" }, { "term": "1951", "
  value": "1951" }, { "term": "23 years", "value": "P23Y"
  }, { "term": "1962", "value": "1962" }, { "term":
  "1957", "value": "1957" }, { "term": "1960", "value":
  "1960" }, { "term": "now", "value": "PRESENT_REF" }, {
  "term": "1964", "value": "1964" }, { "term": "now", "
  value": "PRESENT_REF" }, { "term": "now", "value": "
  PRESENT_REF" }, { "term": "the early 1970s", "value":
  "197" }, { "term": "1987", "value": "1987" }, { "term":
  "early 1990s", "value": "FUTURE_REF" }, { "term": "the
  late 1990s", "value": "199" }, { "value": "FUTURE_REF
  ", "term": "7 March, 1999" }, { "value": "1990", "term
  ": "1990" }, { "term": "the years", "value":
  "2015-09-27" }, { "term": "Padre 1951", "value": "
  FUTURE_REF" }] }
```

### C04: Número de frases positivas

Resultado:

```
1 { 23 }
```

### C05: Número de frases negativas

Resultado:

```
1 { 25 }
```

## C06: Frases que están en los resúmenes

Resultado:

```
1 {
2   "phrase": "4 in Empire Magazines list of Th...",
3   "phrase": "AI was finally released in 2001 ...",
4   "phrase": "Although cynics might complain o...",
5   "phrase": "Another customer, Victor Kunyase...",
6   "phrase": "As a child, Stanley was consider...",
7   "phrase": "At the age of 17, he landed a jo...",
8   "phrase": "At training sessions before open...",
9   "phrase": "But for all his wide, engrossing...",
10  "phrase": "But it was nice, a good place to...",
11  "phrase": "But my forehand was vulgar, it w...",
12  "phrase": "But those strawberry milk cockta...",
13  "phrase": "But when Full Metal Jacket was e...",
14  "phrase": "By 1951, when he was just 23 yea...",
15  "phrase": "By now, Stanley had become renow...",
16  "phrase": "By the early 1970s, he had also ...",
17  "phrase": "Canadian and American officials ...",
18  "phrase": "Clarke. This was a groundbreakin...",
19  "phrase": "Clockwork Orange, 2001, Full Met...",
20  "phrase": "Coffee is not on the menu, for e...",
21  "phrase": "Despite all of the planning, som...",
22  "phrase": "Finally, thanks to the money hed...",
23  "phrase": "Fognini levelled again at 55 tha...",
24  "phrase": "For example, we had to teach our...",
25  "phrase": "For instance, Ive read that I we...",
26  "phrase": "From now on, he would always hav...",
27  "phrase": "He also allowed them free run of...",
28  "phrase": "He developed a project called Th...",
29  "phrase": "He moved to the UK in 1962 and r...",
30  "phrase": "He turned down an invitation to ...",
31  "phrase": "He used his own savings to finan...",
32  "phrase": "He was very interested in music,...",
33  "phrase": "His father, Jack, who worked as ...",
34  "phrase": "His father then hit upon the ide...",
35  "phrase": "His next two movies, Killers Kis...",
36  "phrase": "I am convinced that this situati...",
37  "phrase": "I certainly want my family to tr...",
38  "phrase": "if you saw no other films, you w...",
39  "phrase": "In desperation, his father taugh...",
40  "phrase": "In honor of the grand opening, a...",
41  "phrase": "In the early 1990s, he began an ...",
42  "phrase": "I played poorly, I didnt play li...",
43  "phrase": "Is there any fear that when the ...",
44  "phrase": "It tasted very, I would say, unu...",
45  "phrase": "It was Fogninis second straight ...",
```

46 "phrase": "It was very interesting, that ga...",  
47 "phrase": "Ive been able to avoid flying fo...",  
48 "phrase": "Kubrick changed pace yet again w...",  
49 "phrase": "Kubrick took command of the proj...",  
50 "phrase": "Lebedeva, 25, who formerly worke...",  
51 "phrase": "McDonalds employees are paid 1,5...",  
52 "phrase": "Metty apparently did as he was t...",  
53 "phrase": "Most customers seemed thrilled t...",  
54 "phrase": "My forehand has been my biggest ...",  
55 "phrase": "Nadal battled back to level at 6...",  
56 "phrase": "Next came one of Kubricks bigges...",  
57 "phrase": "Next, Kubrick turned his attentio...",  
58 "phrase": "On Wednesday, though, it stayed ...",  
59 "phrase": "Preproduction work on AI was sti...",  
60 "phrase": "Rumour has it that during the lo...",  
61 "phrase": "Sam Yahel, from McDonalds in Atl...",  
62 "phrase": "Several other short films follow...",  
63 "phrase": "Since Stanley had frequently dis...",  
64 "phrase": "Some reported waiting up to 1 ho...",  
65 "phrase": "Soviet farmers, who are supplyin...",  
66 "phrase": "Stanley acted as producer, direc...",  
67 "phrase": "Stanley originally wrote the scr...",  
68 "phrase": "Stanley soon developed a passion...",  
69 "phrase": "Starring Peter Sellers, Dr.",  
70 "phrase": "Strangelove or How I Learned To ...",  
71 "phrase": "Strangelove was a huge success, ...",  
72 "phrase": "The Big Mac costs 3,75 rubles, o...",  
73 "phrase": "The film explored the intricacie...",  
74 "phrase": "The film told the controversial ...",  
75 "phrase": "The Italian, seeded 13th, won 64...",  
76 "phrase": "The joint venture agreement betw...",  
77 "phrase": "The McDonalds prices are high co...",  
78 "phrase": "The Moscow McDonalds will be ope...",  
79 "phrase": "The next film he actually comple...",  
80 "phrase": "The next major film Kubrick rele...",  
81 "phrase": "The previous record for openingd...",  
82 "phrase": "There is a brass plaque near the...",  
83 "phrase": "There werent many ideological di...",  
84 "phrase": "The Shining, which starred Jack ...",  
85 "phrase": "The Soviet Union is definitely a...",  
86 "phrase": "The staff was chosen from 25,000...",  
87 "phrase": "The theme, whats behind the emti...",  
88 "phrase": "The worlds largest McDonalds, wi...",  
89 "phrase": "The young Kubrick knew nothing a...",  
90 "phrase": "They practiced cooking using yel...",  
91 "phrase": "This film caused a furore when i...",  
92 "phrase": "This is a blow for me, but I acc...",  
93 "phrase": "This is very different from anyt...",  
94 "phrase": "To give one example, the cinemat...

```

95 "phrase": "Top seed Kei Nishikori earlier b...",
96 "phrase": "Virtually every other foreign fi...",
97 "phrase": "When the doors swung open Wednes...",
98 "phrase": "While most of the Soviet workers...",
99 "phrase": "Whilst he was still a teenager, ...",
100 "phrase": "Whilst the technology labs strug...",
101 "phrase": "With his second marriage, to Rut..."
102 }

```

## C07: Dominios generales del escenario 1

Resultado:

```

1 { "DomainValue": ["Administration", "Agriculture", "
  Anatomy", "Animal_Husbandry", "Animals", "Anthropology
  ", "Applied_Science", "Architecture", "Art", "
  Artisanship", "Astrology", "Astronomy", "Banking", "
  Baseball", "Betting", "Biology", "Body_Care", "
  Book_Keeping", "Bowling", "Boxing", "Buildings", "Card
  ", "Chemistry", "Chess", "Cinema", "Color", "Commerce",
  "Computer_Science", "Drawing", "Earth", "Economy", "
  Electricity", "Enterprise", "Entomology", "Ethnology",
  "Exchange", "Fashion", "Finance", "Food", "Furniture",
  "Gastronomy", "Geography", "Geology", "Geometry", "Golf
  ", "Grammar", "History", "Hydraulics", "Industry", "
  Jewellery", "Law", "Linguistics", "Literature", "
  Mathematics", "Mechanics", "Medicine", "Metrology", "
  Military", "Money", "Music", "Nautical", "Number", "
  Painting", "Pedagogy", "Person", "Philology", "
  Philosophy", "Photography", "Physics", "Physiology", "
  Plants", "Play", "Politics", "Post", "
  Psychological_Features", "Psychology", "Publishing", "
  Quality", "Racing", "Radio+Tv", "Religion", "School", "
  Sexuality", "Sociology", "Sport", "Table_Tennis", "
  Telecommunication", "Telephony", "Tennis", "Theatre", "
  Time_Period", "Tourism", "Town_Planning", "Transport",
  "University", "Vehicles", "Wrestling"] }

```

## C08: Dominios generales del escenario 2

Resultado:

```

1 { "DomainValue": ["Administration", "Agriculture", "
  Anatomy", "Animals", "Anthropology", "Applied_Science",
  "Art", "Astronomy", "Baseball", "Betting", "Biology",
  "Boxing", "Buildings", "Card", "Chemistry", "Dance", "
  Economy", "Electricity", "Ethnology", "Fashion", "
  Fencing", "Finance", "Fishing", "Food", "Football", "
  Free_Time", "Furniture", "Gastronomy", "Geography", "
  Geology", "Golf", "Grammar", "Health", "History", "
  Industry", "Law", "Linguistics", "Literature", "
  Mathematics", "Mechanics", "Medicine", "Meteorology", "
  Metrology", "Military", "Money", "Music", "Nautical", "
  Number", "Person", "Pharmacy", "Physics", "Physiology",
  "Plants", "Play", "Psychological_Features", "
  Psychology", "Publishing", "Quality", "Racing", "
  Religion", "RootDomain", "School", "Sexuality", "
  Sociology", "Sport", "Table_Tennis", "Telecommunication
  ", "Tennis", "Theatre", "Time_Period", "Tourism", "
  Transport"] }

```

### C09: Dominios generales del escenario 3

Resultado:

```

1 { "DomainValue": ["Acoustics", "Administration", "
  Agriculture", "Anatomy", "Animals", "Applied_Science",
  "Art", "Astrology", "Astronomy", "Aviation", "Banking",
  "Baseball", "Betting", "Biochemistry", "Biology", "
  Body_Care", "Book_Keeping", "Boxing", "Buildings", "
  Card", "Chemistry", "Chess", "Cinema", "Color", "
  Commerce", "Computer_Science", "Dance", "Drawing", "
  Economy", "Electricity", "Electronics", "
  Electrotechnology", "Engineering", "Enterprise", "
  Ethnology", "Exchange", "Fashion", "Finance", "Fishing
  ", "Food", "Football", "Free_Time", "Furniture", "
  Gastronomy", "Geography", "Geology", "Geometry", "Golf
  ", "Grammar", "History", "Industry", "Law", "
  Linguistics", "Literature", "Mathematics", "Mechanics",
  "Medicine", "Metrology", "Military", "Money", "Music",
  "Mythology", "Nautical", "Number", "Numismatics", "
  Painting", "Pedagogy", "Person", "Pharmacy", "Philology
  ", "Philosophy", "Photography", "Physics", "Physiology
  ", "Plants", "Play", "Politics", "Post", "
  Psychological_Features", "Psychology", "Publishing", "
  Pure_Science", "Quality", "Racing", "Radiology", "Radio
  +Tv", "Railway", "Religion", "School", "Sexuality", "
  Social_Science", "Sociology", "Sport", "Surgery", "Tax
  ", "Telecommunication", "Telephony", "Tennis", "Theatre

```

```
", "Time_Period", "Tourism", "Town_Planning", "
Transport", "University", "Vehicles"] }
```

### **C10: Frases negativas que contengan la entidad LOCATION**

Resultado:

```
1 {
2 "phrase": "The Italian, seeded 13th, won 64 7...",
3 "phrase": "His father, Jack, who worked as a ...",
4 "phrase": "He even played chess for money at ...",
5 "phrase": "Finally, thanks to the money hed e..."
6 }
```

### **C11: Frases del resumen que contengan una referencia temporal que sea de los años 2014 o 2015**

Resultado:

```
1 {
2 "phrase": "When the doors swung open Wednesda...",
3 "phrase": "On Wednesday, though, it stayed op...",
4 "phrase": "Rafael Nadal beaten by Fabio Fogni...",
5 "phrase": "It was Fogninis second straight wi...",
6 "phrase": "Top seed Kei Nishikori earlier bea...",
7 "phrase": "He also allowed them free run of h..."
8 }
```

### **C12: Resumen que contenga alguna frase positiva**

Resultado:

```
1 {
2   "compendium": {
3     "full-summary": "When the doors swung ope...",
4   },
5   "compendium": {
6     "full-summary": "Rafael Nadal beaten by F...",
7   },
}
```

```

8     "compendium": {
9         "full-summary": "My forehand has been my ...",
10    },
11    "compendium": {
12        "full-summary": "As a child, Stanley was ...",
13    },
14    "compendium": {
15        "full-summary": "Next, Kubrick turned his..."
16    }
17 }

```

### C13: Frases que contengan la entidad LOCATION y al menos una referencia temporal

Resultado:

```

1  {
2  "phrase": "When the doors swung open Wednesda...",
3  "phrase": "Although cynics might complain of ...",
4  "phrase": "Virtually every other foreign firm...",
5  "phrase": "The Italian, seeded 13th, won 64 7...",
6  "phrase": "It was Fogninis second straight wi...",
7  "phrase": "His father, Jack, who worked as a ...",
8  "phrase": "He even played chess for money at ...",
9  "phrase": "He moved to the UK in 1962 and rem...",
10 "phrase": "With his second marriage, to Ruth ...",
11 "phrase": "He also allowed them free run of h..."
12 }

```

### C14: Frases que contengan la entidad PERSON y al menos una referencia temporal

Resultado:

```

1  {
2  "phrase": "When the doors swung open Wednesda...",
3  "phrase": "Rafael Nadal beaten by Fabio Fogni...",
4  "phrase": "Rafael Nadal says he played poorly...",
5  "phrase": "It was Fogninis second straight wi...",
6  "phrase": "Top seed Kei Nishikori earlier bea...",
7  "phrase": "His father, Jack, who worked as a ...",
8  "phrase": "He moved to the UK in 1962 and rem...",
9  "phrase": "His next two movies, Killers Kiss ...",

```



```

10 "phrase": "With his second marriage, to Ruth ...",
11 "phrase": "Strangelove or How I Learned To St...",
12 "phrase": "By now, Stanley had become renowne...",
13 "phrase": "The next major film Kubrick releas...",
14 "phrase": "In the early 1990s, he began an in...",
15 "phrase": "The film explored the intricacies ...",
16 "phrase": "He also allowed them free run of h...",
17 "phrase": "Strangelove or How I Learned to St..."
18 }

```

## C15: Dominios de las frases del resumen

Resultado:

```

1 { "DomainValue": ["Acoustics", "Administration", "
  Agriculture", "Anatomy", "Animal_Husbandry", "Animals",
  "Anthropology", "Applied_Science", "Architecture", "
  Art", "Artisanship", "Astrology", "Astronomy", "
  Aviation", "Banking", "Baseball", "Betting", "
  Biochemistry", "Biology", "Body_Care", "Book_Keeping",
  "Bowling", "Boxing", "Buildings", "Card", "Chemistry",
  "Chess", "Cinema", "Color", "Commerce", "
  Computer_Science", "Dance", "Drawing", "Earth", "
  Economy", "Electricity", "Electronics", "
  Electrotechnology", "Engineering", "Enterprise", "
  Entomology", "Ethnology", "Exchange", "Fashion", "
  Fencing", "Finance", "Fishing", "Food", "Football", "
  Free_Time", "Furniture", "Gastronomy", "Geography", "
  Geology", "Geometry", "Golf", "Grammar", "Health", "
  History", "Hydraulics", "Industry", "Jewellery", "Law",
  "Linguistics", "Literature", "Mathematics", "Mechanics
  ", "Medicine", "Meteorology", "Metrology", "Military",
  "Money", "Music", "Mythology", "Nautical", "Number", "
  Numismatics", "Painting", "Pedagogy", "Person", "
  Pharmacy", "Philology", "Philosophy", "Photography", "
  Physics", "Physiology", "Plants", "Play", "Politics", "
  Post", "Psychological_Features", "Psychology", "
  Publishing", "Pure_Science", "Quality", "Racing", "
  Radiology", "Radio+Tv", "Railway", "Religion", "
  RootDomain", "School", "Sexuality", "Social_Science", "
  Sociology", "Sport", "Surgery", "Table_Tennis", "Tax",
  "Telecommunication", "Telephony", "Tennis", "Theatre",
  "Time_Period", "Tourism", "Town_Planning", "Transport",
  "University", "Vehicles", "Wrestling"] }

```

## C16: Frases del resumen que contienen alguna entidad nombrada

Resultado:

```
1 {
2   "phrase": "4 in Empire Magazines list of Th...",
3   "phrase": "AI was finally released in 2001 ...",
4   "phrase": "Although cynics might complain o...",
5   "phrase": "Another customer, Victor Kunyase...",
6   "phrase": "As a child, Stanley was consider...",
7   "phrase": "At the age of 17, he landed a jo...",
8   "phrase": "At training sessions before open...",
9   "phrase": "But for all his wide, engrossing...",
10  "phrase": "But it was nice, a good place to...",
11  "phrase": "But my forehand was vulgar, it w...",
12  "phrase": "But those strawberry milk cockta...",
13  "phrase": "But when Full Metal Jacket was e...",
14  "phrase": "By 1951, when he was just 23 yea...",
15  "phrase": "By now, Stanley had become renow...",
16  "phrase": "By the early 1970s, he had also ...",
17  "phrase": "Canadian and American officials ...",
18  "phrase": "Clarke. This was a groundbreakin...",
19  "phrase": "Clockwork Orange, 2001, Full Met...",
20  "phrase": "Coffee is not on the menu, for e...",
21  "phrase": "Despite all of the planning, som...",
22  "phrase": "Everything was very soft and mil...",
23  "phrase": "Finally, thanks to the money hed...",
24  "phrase": "Fognini levelled again at 55 tha...",
25  "phrase": "For example, we had to teach our...",
26  "phrase": "For instance, Ive read that I we...",
27  "phrase": "From now on, he would always hav...",
28  "phrase": "He also allowed them free run of...",
29  "phrase": "He developed a project called Th...",
30  "phrase": "He even played chess for money a...",
31  "phrase": "He had married again for the thi...",
32  "phrase": "He moved to the UK in 1962 and r...",
33  "phrase": "He turned down an invitation to ...",
34  "phrase": "He used his own savings to finan...",
35  "phrase": "He was very interested in music,...",
36  "phrase": "His dad was right Stanley took t...",
37  "phrase": "His father, Jack, who worked as ...",
38  "phrase": "His father then hit upon the ide...",
39  "phrase": "His next two movies, Killers Kis...",
40  "phrase": "I am convinced that this situati...",
41  "phrase": "I certainly want my family to tr...",
42  "phrase": "if you saw no other films, you w...",
43  "phrase": "In desperation, his father taugh..."
```

44 "phrase": "In honor of the grand opening, a...",  
45 "phrase": "In the early 1990s, he began an ...",  
46 "phrase": "I played poorly, I didnt play li...",  
47 "phrase": "Is there any fear that when the ...",  
48 "phrase": "It has such power that it can re...",  
49 "phrase": "It should be a progression of mo...",  
50 "phrase": "It tasted very, I would say, unu...",  
51 "phrase": "It was Fogninis second straight ...",  
52 "phrase": "It was very interesting, that ga...",  
53 "phrase": "Ive been able to avoid flying fo...",  
54 "phrase": "I will keep working.",  
55 "phrase": "Kubrick changed pace yet again w...",  
56 "phrase": "Kubrick received mixed reviews f...",  
57 "phrase": "Kubrick took command of the proj...",  
58 "phrase": "Lebedeva, 25, who formerly worke...",  
59 "phrase": "McDonalds employees are paid 1,5...",  
60 "phrase": "McDonalds invested 50 million in...",  
61 "phrase": "McDonalds officials say it is no...",  
62 "phrase": "Metty apparently did as he was t...",  
63 "phrase": "Most customers seemed thrilled t...",  
64 "phrase": "Most of the Soviet workers had n...",  
65 "phrase": "My forehand has been my biggest ...",  
66 "phrase": "Nadal battled back to level at 6...",  
67 "phrase": "Next came one of Kubricks bigges...",  
68 "phrase": "Next, Kubrick turned his attentio...",  
69 "phrase": "On Wednesday, though, it stayed ...",  
70 "phrase": "Our food and our standards its a...",  
71 "phrase": "Preproduction work on AI was sti...",  
72 "phrase": "Rafael Nadal beaten by Fabio Fog...",  
73 "phrase": "Rafael Nadal says he played poor...",  
74 "phrase": "Rumour has it that during the lo...",  
75 "phrase": "Russians say gamburger because t...",  
76 "phrase": "Sam Yahel, from McDonalds in Atl...",  
77 "phrase": "Several other short films follow...",  
78 "phrase": "Since Stanley had frequently dis...",  
79 "phrase": "Some expressed wonder at the spe...",  
80 "phrase": "Some reported waiting up to 1 ho...",  
81 "phrase": "Soviet farmers, who are supplyin...",  
82 "phrase": "Stanley acted as producer, direc...",  
83 "phrase": "Stanley originally wrote the scr...",  
84 "phrase": "Stanley soon developed a passion...",  
85 "phrase": "Starring Peter Sellers, Dr. ...",  
86 "phrase": "Strangelove or How I Learned to ...",  
87 "phrase": "Strangelove or How I Learned To ...",  
88 "phrase": "Strangelove was a huge success, ...",  
89 "phrase": "That is more important to them r...",  
90 "phrase": "The Big Mac costs 3,75 rubles, o...",  
91 "phrase": "The extra effort was necessary t...",  
92 "phrase": "The film explored the intricacie..."

```

93 "phrase": "The film told the controversial ...",
94 "phrase": "The Italian, seeded 13th, won 64...",
95 "phrase": "The joint venture agreement betw...",
96 "phrase": "The McDonalds prices are high co...",
97 "phrase": "The Moscow McDonalds will be ope...",
98 "phrase": "The next film he actually comple...",
99 "phrase": "The next major film Kubrick rele...",
100 "phrase": "The previous record for openingd...",
101 "phrase": "There is a brass plaque near the...",
102 "phrase": "There is no other way forward ot...",
103 "phrase": "There werent many ideological di...",
104 "phrase": "The Shining, which starred Jack ...",
105 "phrase": "The Soviets understood that the ...",
106 "phrase": "The Soviet Union is definitely a...",
107 "phrase": "The Spaniard was making an uncha...",
108 "phrase": "The staff was chosen from 25,000...",
109 "phrase": "The theme, whats behind the emti...",
110 "phrase": "The worlds largest McDonalds, wi...",
111 "phrase": "They graduated from the Canadian...",
112 "phrase": "The young Kubrick knew nothing a...",
113 "phrase": "They practiced cooking using yel...",
114 "phrase": "This film caused a furore when i...",
115 "phrase": "This frequently involved actors ...",
116 "phrase": "This is a blow for me, but I acc...",
117 "phrase": "This is very different from anyt...",
118 "phrase": "To give one example, the cinemat...",
119 "phrase": "Top seed Kei Nishikori earlier b...",
120 "phrase": "Virtually every other foreign fi...",
121 "phrase": "When the doors swung open Wednes...",
122 "phrase": "While most of the Soviet workers...",
123 "phrase": "Whilst he was still a teenager, ...",
124 "phrase": "Whilst the technology labs strug...",
125 "phrase": "With his second marriage, to Rut..."
126 }

```

## C17: Entidades nombradas de frases negativas

Resultado:

```

1 { "NER": ["DATE", "DURATION", "LOCATION", "MISC", "NUMBER",
  "ORDINAL", "ORGANIZATION", "PERSON"] }

```

## C18: Dominios de las frases positivas

Resultado:

```

1 { "DomainValue": ["Acoustics", "Administration", "
  Agriculture", "Anatomy", "Animal_Husbandry", "Animals",
  "Anthropology", "Applied_Science", "Architecture", "
  Art", "Artisanship", "Astrology", "Astronomy", "
  Aviation", "Banking", "Baseball", "Betting", "
  Biochemistry", "Biology", "Body_Care", "Book_Keeping",
  "Bowling", "Boxing", "Buildings", "Card", "Chemistry",
  "Chess", "Cinema", "Color", "Commerce", "
  Computer_Science", "Dance", "Drawing", "Earth", "
  Economy", "Electricity", "Electronics", "
  Electrotechnology", "Engineering", "Enterprise", "
  Entomology", "Ethnology", "Exchange", "Fashion", "
  Finance", "Fishing", "Food", "Football", "Free_Time", "
  Furniture", "Gastronomy", "Geography", "Geology", "
  Geometry", "Golf", "Grammar", "History", "Hydraulics",
  "Industry", "Jewellery", "Law", "Linguistics", "
  Literature", "Mathematics", "Mechanics", "Medicine", "
  Meteorology", "Metrology", "Military", "Money", "Music
  ", "Mythology", "Nautical", "Number", "Numismatics", "
  Painting", "Pedagogy", "Person", "Pharmacy", "Philology
  ", "Philosophy", "Photography", "Physics", "Physiology
  ", "Plants", "Play", "Politics", "Post", "
  Psychological_Features", "Psychology", "Publishing", "
  Pure_Science", "Quality", "Racing", "Radiology", "Radio
  +Tv", "Railway", "Religion", "School", "Sexuality", "
  Social_Science", "Sociology", "Sport", "Surgery", "
  Table_Tennis", "Tax", "Telecommunication", "Telephony",
  "Tennis", "Theatre", "Time_Period", "Tourism", "
  Town_Planning", "Transport", "University", "Vehicles",
  "Wrestling"] }

```

## C19: Frases del resumen negativas

Resultado:

```

1 {
2   "phrase": "Some reported waiting up to 1 hour...",
3   "phrase": "They graduated from the Canadian I...",
4   "phrase": "For example, we had to teach our e...",
5   "phrase": "Coffee is not on the menu, for exa...",
6   "phrase": "Is there any fear that when the go...",
7   "phrase": "That is more important to them rig..."

```

```
8 "phrase": "The Italian, seeded 13th, won 64 7...",
9 "phrase": "I played poorly, I didnt play like...",
10 "phrase": "This is a blow for me, but I accep...",
11 "phrase": "There is no other way forward othe...",
12 "phrase": "The Spaniard was making an unchara...",
13 "phrase": "Nadal battled back to level at 66 ...",
14 "phrase": "As a child, Stanley was considered...",
15 "phrase": "His father, Jack, who worked as a ...",
16 "phrase": "He even played chess for money at ...",
17 "phrase": "Finally, thanks to the money hed e...",
18 "phrase": "To give one example, the cinematog...",
19 "phrase": "Kubrick received mixed reviews for...",
20 "phrase": "By the early 1970s, he had also ac...",
21 "phrase": "Next, Kubrick turned his attention...",
22 "phrase": "He developed a project called The ...",
23 "phrase": "The film explored the intricacies ...",
24 "phrase": "For instance, Ive read that I wear...",
25 "phrase": "Ive been able to avoid flying for ...",
26 "phrase": "Strangelove or How I Learned to St..."
27 }
```

## C20: Frases que no están en el resumen

Resultado:

```
1 { }
2
3 (No hay resultados.)
```

## C21: Frases que hagan referencia al año 2015 y sean neutrales o positivas

Resultado:

```
1 {
2 "phrase": "When the doors swung open Wednesda...",
3 "phrase": "On Wednesday, though, it stayed op...",
4 "phrase": "Rafael Nadal beaten by Fabio Fogni...",
5 "phrase": "It was Fogninis second straight wi...",
6 "phrase": "He also allowed them free run of h..."
7 }
```

## C22: Dominios de frase que contengan alguna referencia temporal

Resultado:

```

1 { "DomainValue": ["Acoustics", "Administration", "
  Agriculture", "Anatomy", "Animal_Husbandry", "Animals",
  "Anthropology", "Applied_Science", "Architecture", "
  Art", "Artisanship", "Astrology", "Astronomy", "
  Aviation", "Banking", "Baseball", "Betting", "
  Biochemistry", "Biology", "Body_Care", "Book_Keeping",
  "Bowling", "Boxing", "Buildings", "Card", "Chemistry",
  "Chess", "Cinema", "Color", "Commerce", "
  Computer_Science", "Dance", "Drawing", "Earth", "
  Economy", "Electricity", "Electronics", "
  Electrotechnology", "Engineering", "Enterprise", "
  Entomology", "Ethnology", "Exchange", "Fashion", "
  Fencing", "Finance", "Fishing", "Food", "Football", "
  Free_Time", "Furniture", "Gastronomy", "Geography", "
  Geology", "Geometry", "Golf", "Grammar", "Health", "
  History", "Hydraulics", "Industry", "Jewellery", "Law",
  "Linguistics", "Literature", "Mathematics", "Mechanics
  ", "Medicine", "Meteorology", "Metrology", "Military",
  "Money", "Music", "Mythology", "Nautical", "Number", "
  Numismatics", "Painting", "Pedagogy", "Person", "
  Pharmacy", "Philology", "Philosophy", "Photography", "
  Physics", "Physiology", "Plants", "Play", "Politics", "
  Post", "Psychological_Features", "Psychology", "
  Publishing", "Pure_Science", "Quality", "Racing", "
  Radiology", "Radio+Tv", "Railway", "Religion", "
  RootDomain", "School", "Sexuality", "Social_Science", "
  Sociology", "Sport", "Surgery", "Table_Tennis", "Tax",
  "Telecommunication", "Telephony", "Tennis", "Theatre",
  "Time_Period", "Tourism", "Town_Planning", "Transport",
  "University", "Vehicles", "Wrestling"] }

```

## C23: Frases positivas que contengan alguna entidad NUMBER

Resultado:

```

1 {
2   "phrase": "It tasted very, I would say, unusu...",
3   "phrase": "It was very interesting, that gamb...",
4   "phrase": "McDonalds invested 50 million in s..."

```

```

5 "phrase": "Fognini levelled again at 55 thank...",
6 "phrase": "Metty apparently did as he was tol...",
7 "phrase": "Clarke. This was a groundbreaking ...",
8 "phrase": "AI was finally released in 2001 bu...",
9 "phrase": "4 in Empire Magazines list of The ...",
10 "phrase": "He also allowed them free run of h..."
11 }

```

## C24: Frases negativas que tengan alguna referencia temporal y la entidad LOCATION

Resultado:

```

1 {
2 "phrase": "The Italian, seeded 13th, won 64 7...",
3 "phrase": "His father, Jack, who worked as a ...",
4 "phrase": "He even played chess for money at ..."
5 }

```

## C25: Frases que contengan la entidad LOCATION y no contengan alguna referencia temporal

Resultado:

```

1 {
2 "phrase": "The worlds largest McDonalds, with...",
3 "phrase": "Sam Yahel, from McDonalds in Atlan...",
4 "phrase": "The Soviet Union is definitely a d...",
5 "phrase": "The Moscow McDonalds will be open ...",
6 "phrase": "The previous record for openingday...",
7 "phrase": "Canadian and American officials wh...",
8 "phrase": "McDonalds invested 50 million in s...",
9 "phrase": "There werent many ideological disc...",
10 "phrase": "But for all his wide, engrossing w...",
11 "phrase": "Stanley soon developed a passion f...",
12 "phrase": "At the age of 17, he landed a job ...",
13 "phrase": "The young Kubrick knew nothing abo...",
14 "phrase": "Finally, thanks to the money hed e...",
15 "phrase": "But when Full Metal Jacket was eve...",
16 "phrase": "AI was finally released in 2001 bu..."
17 }

```



### C26: Entidades nombradas que están implicadas con el dominio Gastronomy (Escenario 1)

Resultado:

```
1 { "NER": ["DATE", "DURATION", "LOCATION", "MISC", "NUMBER", "ORDINAL", "ORGANIZATION", "PERSON", "TIME"] }
```

### C27: Entidades nombradas que están implicadas con el dominio Sport (Escenario 2)

Resultado:

```
1 { "NER": ["DATE", "DURATION", "LOCATION", "MISC", "NUMBER", "ORDINAL", "ORGANIZATION", "PERSON"] }
```

### C28: Entidades nombradas que están implicadas con el dominio Cinema (Escenario 3)

Resultado:

```
1 { "NER": ["DATE", "DURATION", "LOCATION", "MISC", "NUMBER", "ORDINAL", "ORGANIZATION", "PERSON"] }
```

### C29: Frases que hablen de la entidad McDonalds (Escenario 1)

Resultado:

```
1 {  
2 "phrase": "When the doors swung open Wednesda...",  
3 "phrase": "Although cynics might complain of ...",  
4 "phrase": "In honor of the grand opening, a b...",  
5 "phrase": "The worlds largest McDonalds, with...",  
6 "phrase": "Sam Yahel, from McDonalds in Atlan...",  
7 "phrase": "The Moscow McDonalds will be open ...",  
8 "phrase": "On Wednesday, though, it stayed op...",  
9 "phrase": "The McDonalds prices are high comp..."
```

```
10 "phrase": "McDonalds employees are paid 1,50 ...",
11 "phrase": "McDonalds invested 50 million in s...",
12 "phrase": "The joint venture agreement betwee...",
13 "phrase": "McDonalds officials say it is not ..."
14 }
```

### C30: Frases que hablen de la entidad Nadal (Escenario 2)

Resultado:

```
1 {
2 "phrase": "Rafael Nadal beaten by Fabio Fogni...",
3 "phrase": "Rafael Nadal says he played poorly...",
4 "phrase": "I played poorly, I didnt play like...",
5 "phrase": "My forehand has been my biggest vi...",
6 "phrase": "Fognini levelled again at 55 thank...",
7 "phrase": "Nadal battled back to level at 66 ..."
8 }
```

### C31: Frases que hablen de la entidad Kubrick (Escenario 3)

Resultado:

```
1 {
2 "phrase": "But for all his wide, engrossing w...",
3 "phrase": "The young Kubrick knew nothing abo...",
4 "phrase": "He moved to the UK in 1962 and rem...",
5 "phrase": "Kubrick took command of the projec...",
6 "phrase": "To give one example, the cinematog...",
7 "phrase": "Kubrick received mixed reviews for...",
8 "phrase": "Kubrick changed pace yet again wit...",
9 "phrase": "Clarke. This was a groundbreaking ...",
10 "phrase": "Next, Kubrick turned his attention...",
11 "phrase": "The next major film Kubrick releas...",
12 "phrase": "Rumour has it that during the long...",
13 "phrase": "In the early 1990s, he began an in...",
14 "phrase": "Preproduction work on AI was still...",
15 "phrase": "The film explored the intricacies ...",
16 "phrase": "AI was finally released in 2001 bu..."
17 }
```

### C32: Frases positivas y negativas con el dominio Gastronomy (Escenario 1)

Resultado:

```
1 {
2 "phrase": "It tasted very, I would say, unusu...",
3 "phrase": "Most customers seemed thrilled to ...",
4 "phrase": "It was very interesting, that gamb...",
5 "phrase": "Another customer, Victor Kunyasev,...",
6 "phrase": "But it was nice, a good place to t...",
7 "phrase": "Some reported waiting up to 1 hour...",
8 "phrase": "At training sessions before openin...",
9 "phrase": "They graduated from the Canadian I...",
10 "phrase": "For example, we had to teach our e...",
11 "phrase": "McDonalds invested 50 million in s...",
12 "phrase": "Coffee is not on the menu, for exa...",
13 "phrase": "Is there any fear that when the go...",
14 "phrase": "That is more important to them rig..."
15 }
```

### C33: Frases positivas y negativas con el dominio Sport (Escenario 2)

Resultado:

```
1 {
2 "phrase": "The Italian, seeded 13th, won 64 7...",
3 "phrase": "It was Fogninis second straight wi...",
4 "phrase": "I played poorly, I didnt play like...",
5 "phrase": "This is a blow for me, but I accep...",
6 "phrase": "There is no other way forward othe...",
7 "phrase": "The Spaniard was making an unchara...",
8 "phrase": "Fognini levelled again at 55 thank...",
9 "phrase": "Nadal battled back to level at 66 ..."
10 }
```

### C34: Frases positivas y negativas con el dominio Cinema (Escenario 3)

Resultado:

```
1 {
2 "phrase": "if you saw no other films, you wou...",
3 "phrase": "As a child, Stanley was considered...",
4 "phrase": "His father, Jack, who worked as a ...",
5 "phrase": "He even played chess for money at ...",
6 "phrase": "The young Kubrick knew nothing abo...",
7 "phrase": "Finally, thanks to the money hed e...",
8 "phrase": "To give one example, the cinematog...",
9 "phrase": "Metty apparently did as he was tol...",
10 "phrase": "Kubrick received mixed reviews for...",
11 "phrase": "The film told the controversial st...",
12 "phrase": "Strangelove or How I Learned To St...",
13 "phrase": "Strangelove was a huge success, bo...",
14 "phrase": "Clarke. This was a groundbreaking ...",
15 "phrase": "Stanley originally wrote the scrip...",
16 "phrase": "By the early 1970s, he had also ac...",
17 "phrase": "Next, Kubrick turned his attention...",
18 "phrase": "Rumour has it that during the long...",
19 "phrase": "But when Full Metal Jacket was eve...",
20 "phrase": "In the early 1990s, he began an in...",
21 "phrase": "He developed a project called The ...",
22 "phrase": "The film explored the intricacies ...",
23 "phrase": "AI was finally released in 2001 bu...",
24 "phrase": "4 in Empire Magazines list of The ...",
25 "phrase": "He also allowed them free run of h...",
26 "phrase": "For instance, Ive read that I wear...",
27 "phrase": "Ive been able to avoid flying for ...",
28 "phrase": "Strangelove or How I Learned to St..."
29 }
```

### **C35: Frases de resumen que contengan una referencia temporal y el dominio Gastronomy (Escenario 1)**

Resultado:

```
1 {
2 "phrase": "When the doors swung open Wednesda...",
3 "phrase": "Although cynics might complain of ...",
4 "phrase": "While most of the Soviet workers t...",
5 "phrase": "On Wednesday, though, it stayed op...",
6 "phrase": "Virtually every other foreign firm...",
7 "phrase": "McDonalds employees are paid 1,50 ...",
8 "phrase": "That is more important to them rig..."
9 }
```

### C36: Frases de resumen que contengan una referencia temporal y el dominio Sport (Escenario 2)

Resultado:

```
1 {  
2 "phrase": "Rafael Nadal beaten by Fabio Fogni...",  
3 "phrase": "Rafael Nadal says he played poorly...",  
4 "phrase": "The Italian, seeded 13th, won 64 7...",  
5 "phrase": "It was Fogninis second straight wi...",  
6 "phrase": "Top seed Kei Nishikori earlier bea..."  
7 }
```

### C37: Frases de resumen que contengan una referencia temporal y el dominio Cinema (Escenario 3)

Resultado:

```
1 {  
2 "phrase": "His father, Jack, who worked as a ...",  
3 "phrase": "He even played chess for money at ...",  
4 "phrase": "By 1951, when he was just 23 years...",  
5 "phrase": "He moved to the UK in 1962 and rem...",  
6 "phrase": "His next two movies, Killers Kiss ...",  
7 "phrase": "With his second marriage, to Ruth ...",  
8 "phrase": "Strangelove or How I Learned To St...",  
9 "phrase": "From now on, he would always have ...",  
10 "phrase": "By now, Stanley had become renowne...",  
11 "phrase": "By the early 1970s, he had also ac...",  
12 "phrase": "The next major film Kubrick releas...",  
13 "phrase": "In the early 1990s, he began an in...",  
14 "phrase": "The film explored the intricacies ...",  
15 "phrase": "He also allowed them free run of h...",  
16 "phrase": "Strangelove or How I Learned to St..."  
17 }
```

### C38: Frases positivas/negativas y que hablen de la entidad McDonalds (Escenario 1)

Resultado:

```
1 {
2 "phrase": "McDonalds invested 50 million in setting up the
   restaurant and a food processing plant in a Moscow
   suburb that turns out everything from meat patties to
   sesameseed buns."
3 }
```

### **C39: Frases positivas/negativas y que hablen de la entidad Nadal (Escenario 2)**

Resultado:

```
1 {
2 "phrase": "I played poorly, I didnt play like...",
3 "phrase": "Fognini levelled again at 55 thank...",
4 "phrase": "Nadal battled back to level at 66 ..."
5 }
```

### **C40: Frases positivas/negativas y que hablen de la entidad Kubrick (Escenario 3)**

Resultado:

```
1 {
2 "phrase": "The young Kubrick knew nothing abo...",
3 "phrase": "To give one example, the cinematog...",
4 "phrase": "Kubrick received mixed reviews for...",
5 "phrase": "Clarke. This was a groundbreaking ...",
6 "phrase": "Next, Kubrick turned his attention...",
7 "phrase": "Rumour has it that during the long...",
8 "phrase": "In the early 1990s, he began an in...",
9 "phrase": "The film explored the intricacies ...",
10 "phrase": "AI was finally released in 2001 bu..."
11 }
```

### **C41: Frases negativas que contengan la entidad McDonalds (Escenario 1)**

Resultado:

```
1 { }
2
3 (No hay resultados.)
```

### **C42: Frases negativas que contengan la entidad Nadal (Escenario 2)**

Resultado:

```
1 {
2 "phrase": "I played poorly, I didnt play like...",
3 "phrase": "Nadal battled back to level at 66 ..."
4 }
```

### **C43: Frases negativas que contengan la entidad Kubrick (Escenario 3)**

Resultado:

```
1 {
2 "phrase": "To give one example, the cinematog...",
3 "phrase": "Kubrick received mixed reviews for...",
4 "phrase": "Next, Kubrick turned his attention...",
5 "phrase": "The film explored the intricacies ..."
6 }
```

### **C44: Obtener todas las entidades de las frases que contienen McDonalds (Escenario 1)**

Resultado:

```
1 { "NER": ["DATE", "ORDINAL", "ORGANIZATION", "LOCATION", "MISC", "PERSON", "DURATION", "NUMBER", "TIME"] }
```

### **C45: Obtener todas las entidades de las frases que contienen Nadal (Escenario 2)**

Resultado:

```
1 { "NER": ["PERSON", "ORGANIZATION", "DATE", "DURATION", "MISC", "ORDINAL", "NUMBER", "LOCATION"] }
```

### **C46: Obtener todas las entidades de las frases que contienen Kubrick (Escenario 3)**

Resultado:

```
1 { "NER": ["NUMBER", "PERSON", "LOCATION", "MISC", "ORGANIZATION", "DATE", "ORDINAL"] }
```

### **C47: Todas las expresiones temporales distintas**

No se puede realizar.

### **C48: Entidades nombradas NUMBER coincidentes en el escenario 2 y 3**

No se puede realizar.

### **C49: Fechas asociadas a una persona**

No se puede realizar.

### **C50: Polaridad del escenario completo**

No se puede realizar.