

# Anexo 1: Entrevista de requerimientos

## 1. ¿Cuál es el objetivo primordial del sistema?

El objetivo principal del sistema es permitir la ejecución de las tareas de preprocesamiento de minería de texto en las tecnologías de Big Data para aumentar la escalabilidad y el desempeño. Las tareas mínimas a ejecutar son: POS, Splitter, Tokenizer, EntityRecognition (sea con listas precargadas como con listas propias), Búsqueda de reglas basado en expresiones regulares (NE Transducer de GATE), identificación de coreferencias.

## 2. ¿Cuáles serían los usuarios del sistema?

Existen dos tipos de usuarios:

Usuario de negocio: Este usuario podrá ingresar un texto o un conjunto de texto a través de una GUI y le dirá al sistema qué tareas de preprocesamiento quiere ejecutar y en qué orden y el sistema le debe entregar al finalizar las anotaciones de cada texto (similar al Gate Application)

Usuario programador: El usuario debe poder incluir las librerías necesaria en su programa para hacer llamados al API del sistema de tal manera que pueda llamar las diferentes tareas de preprocesamiento.

## 3. ¿Cuáles son las entradas del sistema?

Depende de la tarea de preprocesamiento. Cada Tarea exige cosas diferentes. En términos generales se exigirán las mismas entradas que un procesador de texto como GATE. Por ejemplo, para el reconocimiento de entidades debe recibir los documentos y adicionalmente las listas con lo que se quiere reconocer y el sistema de entregar la anotaciones de cada documento correspondiente a cada uno de los elementos identificados en la lista..

## 4. ¿Cuáles son las salidas del sistema?

Te respondí en el anterior.

## 5. ¿Cómo se imagina el sistema?

Me imagino un gate reducido en donde sólo están disponibles un conjunto de recursos de procesamiento, pero por debajo usando tecnologías de big data.

## 6. ¿Cuáles son las principales operaciones de preprocesamiento en minería de texto que debe implementar el sistema?

POS, Splitter, Tokenizer, EntityRecognition (sea con listas precargadas como con listas propias), Búsqueda de reglas basado en expresiones regulares (NE Transducer de GATE), identificación de coreferencias.

7. ¿Existe alguna prioridad entre esas operaciones de preprocesamiento en minería de texto?

Splitter, Tokenizer, POS, EntityRecognition (sea con listas precargadas como con listas propias), Búsqueda de reglas basado en expresiones regulares (NE Transducer de GATE), identificación de coreferencias.

8. ¿Qué funcionalidades adicionales le gustaría que tuviera el sistema? Facilidad de instalación
9. ¿De los siguientes requerimientos no funcionales, cuáles le parecen importantes y en qué prioridad de mayor a menor?
- a. Usabilidad - 5
  - b. Operabilidad Qué es esto?
  - c. Confiabilidad - 4
  - d. Tolerancia a fallos
  - e. Mantenibilidad - 1
  - f. Capacidad de cambio - 2
  - g. Funcionalidad - 3
  - h. Interoperabilidad - 6
10. ¿De qué manera debe interactuar el sistema con otras aplicaciones? Debe poder ser importado como librería y debe proveer un API.
11. ¿De qué manera debe interactuar el sistema con un usuario final? Interfaz gráfica para usuario de negocio.
12. ¿Se requiere que el sistema almacene datos?, de ser así, cuáles? Las listas para hacer entity recognition. Los resultados deben estar temporalmente.
13. ¿Se requiere que el sistema implemente algún tipo de mecanismo de seguridad, como por ejemplo autenticación y autorización o manejo de certificados digitales? No.
14. Si con una palabra pudiera definir lo que es más importante para el sistema, ¿cuál sería? Desempeño.
15. ¿Qué tipo de aplicación debe ser, una de escritorio o una Web? Escritorio.