Especificación de Requerimientos de Software (ERS)

*WS para descubrimiento, desambiguación y enlace en*

*Linked Data*

Versión [1.0]

Información de Documento

|  |  |
| --- | --- |
| Título: | Especificación de casos de Uso |
| Subtítulo: | Servicio Web de Extracción de Entidades |
| Versión: | [1.0] |
| Archivo: | ECS\_SW\_ExtracciónEntidades |
| Autor: | Fabricio Montaño |
| Estado: | Borrador |

Lista de Cambios

|  |  |  |  |
| --- | --- | --- | --- |
| Versión | Fecha | Autor | Descripción |
| 1.0 | 17/07/2014 | Fabricio Montaño | Emisión inicial |
|  |  |  |  |

Firmas y Aprobaciones

|  |  |  |  |
| --- | --- | --- | --- |
| Elaborado por: | Fabricio Montaño | |  |
| Fecha: | 17/07/2014 | Firma: |
|  | | | |
| Revisado por: | Ing. Nelson Piedra | |  |
| Fecha: |  | Firma: |

Contenido

[1 Introducción 4](#_Toc399771639)

[1.1 Descripción 4](#_Toc399771640)

[1.2 Problemas Conocidos 4](#_Toc399771641)

[1.3 Referencias 4](#_Toc399771642)

[2 Descripción General 5](#_Toc399771643)

[2.1 Perspectiva del Proyecto 5](#_Toc399771647)

[2.2 Características del Producto 5](#_Toc399771649)

[2.3 Características del Usuario 6](#_Toc399771650)

[2.4 Limitaciones Generales 7](#_Toc399771651)

[2.5 Asunciones y Dependencias 7](#_Toc399771652)

[3 Requerimientos Funcionales. 7](#_Toc399771653)

[3.1 REQ001 Configuraciones 8](#_Toc399771654)

[3.2 REQ002 Tutorial Interactivo 8](#_Toc399771655)

[3.3 REQ003 Juego modo manual 8](#_Toc399771656)

[3.4 REQ004 Juego modo automático 9](#_Toc399771657)

[3.5 REQ005 Selección de operación 9](#_Toc399771658)

[3.6 REQ006 Colocar triángulos en la Pizza Matemática 10](#_Toc399771659)

[4 Otros Requerimientos. 10](#_Toc399771660)

[5 Firmas de aprobación 11](#_Toc399771661)

**Especificación de Requerimientos de Software (ERS)**

# Introducción

## Descripción

El presente documente tiene como finalidad redactar las funcionalidades con las que debe contar el sistema de descubrimiento, desambiguación y enlace en datos enlazados.

## Problemas Conocidos

Después de un análisis inicial se detectan los siguientes problemas:

* Los *abstract* de las publicaciones universitarias que contienen datos relevantes a los que se desea acceder se encuentran en texto plano entendible solo para humanos.
* Los datos en texto no son referenciados hacia fuentes externas.
* Los datos en texto plano pueden ser ambiguos y tener más de un enlace posible en LOD Cloud específicamente DBpedia.
* No existe un proceso estándar para procesos de enlace y desambiguación lingüística.
* Las conexiones y/o consultas hacia el DataSet de DBpedia pueden demorar o fallar.

## Referencias

|  |  |
| --- | --- |
|  | ANSI/IEEE Std. 830-1984 Guía del IEEE para la Especificación de Requerimientos Software.[[1]](#footnote-1) |

# Descripción General

El fin del sistema es el descubrimiento de datos relevantes dentro del texto plano en los *abstracts* en las publicaciones universitarias, y si en caso un término extraído es ambiguo determinar el significado usado, para luego ser enlazado a DBpedia (LOD Cloud) si en caso existiera un recurso al cual referencie.

Para acceder al sistema se levantara servicios web diferentes para cada proceso relevante dentro del sistema que interactuaran entre sí. Se desarrollara un interfaz web para usuarios que permitirá visualizar los resultados individuales e integrales de los servicios.

## Perspectiva del Proyecto

Esto se pretende desarrollar en base a la relevancia que toman los datos en la web semántica, buscando enlazar las publicaciones científicas a las fuentes de Datos Enlazados, donde se ubican los recursos a los cuales hacen referencia y permitiendo de esta forma ampliar la información y descubrir nuevos enlaces.

El sistema que permita extraer datos relevantes dentro del texto de las publicaciones científicas, desambiguar estos términos de ser necesario y enlazarlos a LOD Cloud. Se construirá separando e integrando los procesos relevantes mediante servicio web.

## Características del Producto

El producto a desarrollar constara de las siguientes características:

* **Servidor**
  + Servicios web REST
  + Separados en los procesos importantes:
    - Procesamiento de lenguaje natural,
    - Desambiguación y
    - Enlace
  + Integrados entre si
  + Servidor local con DataSet de DBpedia (SPARQL EndPoint)
* **Cliente**
  + REST web
  + Visualizar resultados individuales de los servicios
  + Visualizar resultado integrado de los servicios
  + Permitir ver el JSON resultante del servicio consumido

## Características del Usuario­­

**Usuarios anónimos:** a través de la construcción del cliente Rest web cualquier usuario podrá interactuar con el sistema.

**Usuarios clientes Rest:** los servicios web implementaos podrán responder al cualquier servicio que se pueda construir a partir de esos.

## Limitaciones Generales

A continuación se detallan limitaciones en cuanto al software:

* Todos los enlaces que se puedan realizar se los hará con recursos disponibles en el DataSet de DBpedia, esto significa que pueden existir recursos en otro u otros repositorios a los cuales no se los enlazara directamente.
* De no existir el término extraído en DBpedia, no podrá ser enlazado.
* La desambiguación de un “término” extraído de una publicación se realizara en base a los recursos disponibles en la DBpedia que son nombrado mediante este “termino”.

## Asunciones y Dependencias

**Asunciones**

* Posibles errores en extracción de términos del texto de las publicaciones debido a faltar tipeado o error humano en la escritura del texto.
* Posibles errores el en enlace de a DBpedia producto de no existir recurso o error de desambiguación.

**Dependencias**

* Se desarrollara en lenguaje de programación de alto nivel Python 2.7 y algunas de sus librerías especializadas en procesamiento de lenguaje natural, levantamiento de servicios, consultas SPARQL, etc.
* Cliente desarrollara en base a HTML, CSS, JavaScript, etc.
* De navegadores web que soporte tecnologías en las que se construirá en cliente para poder acceder a este.

# Requerimientos Funcionales.

## REQ001 Extraer entidades y palabra relevantes

**Descripción**

Descubrir datos relevantes en el texto, a quien se describe y las palabra relevantes que lo acompañan

**Entrada**

* Texto

**Proceso**

1. Tokenización del texto en sentencias (oraciones), separa todas las sentencias.
2. Tokenización de las sentencias en palabras.
3. Etiquetar (Part of Speech).
4. Extracción en base a etiquetas.

**Salida**

* JSON con entidades y palabras claves extraídas

## REQ002 Enlazar entidades y palabra relevantes con LOD Cloud

**Descripción**

Se enlazara los términos encontrados en caso de que sea posible con la LOD Cloud

**Entrada**

* JSON estructurado por procesos anteriores con entidades a enazar

**Proceso**

1. Consultar a DBpedia por recursos que sean nombrados con las entidades y palabras relevantes extraídas del texto de entrada.

**Salida**

* JSON estructurado con los enlaces de los recursos de DBpedia.

## REQ002 Desambiguar entidades y palabra relevantes

**Descripción**

Se determinara el sentido con que las palabras estas siendo usadas en caso de que estas sean ambiguas

**Entrada**

* JSON procesos anteriores

**Proceso**

1. Aplicar algoritmo s de desambiguación en base a contexto.
2. Determinar sentido utilizado en base a mejor resultado de coincidencia.

**Salida**

* JSON con términos desambiguados.

## REQ004 Levantar servicios REST separados para los procesos relevantes.

**Descripción**

Para que los procesos relevantes dentro del sistema puedan ser reutilizados se levantaran servicios individuales.

**Entrada**

* JSON.
* Texto.

**Proceso**

1. Se invoca las funciones necesarias del sistema para resolver la petición del servicio llamado.
2. Se estructurará la data en JSON.
3. Se devuelve el JSON ya sea al servicio que lo invocó o al cliente si este fuese el origen de la invocación del servicio.

**Salida**

* JSON estructurado de acuerdo al servicio invocado.

## REQ005 Frontal UI Web

**Descripción**

Construir una interfaz web que permita visualizar el comportamiento del sistema, es decir, la integración de los servicios y su funcionamiento individual

**Entrada**

* Texto

**Proceso**

1. Introducir texto a ser procesado
2. Seleccionar servicios a ser invocados
3. Esperar resultado
4. Procesar resultado
5. Presentar resultado procesado

**Salida**

* Resultado gráfico de servicios invocados.

|  |  |
| --- | --- |
|  |  |



1. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=2228 [↑](#footnote-ref-1)