

---

Cooperativa CubaSWL

# Arquitectura de información

29 de junio de 2014

**Proyecto:** ToNgueLP, editor de corpus de textos para tareas NLP

**Producto:** ToNgueLP Corpus Tools

**Versión:** 1.0

Autores: Abel Meneses Abad  
Leonel Salazar Videaux

Control de versiones

Fecha	Versión	Iteración	Descripción	Autor
29/6/2014	0.1	2	Digitalizada el prototipo de la vista principal con Pencil.	Abel Meneses Abad

Reglas de Confidencialidad

Clasificación:

Software particular para la Universidad de Camagüey, factible de exportar por las partes desarrolladoras, de carácter nacional a evaluar en la convocatoria del CITMA en el 2015.

Este documento contiene información propietaria de [Abel Meneses Abad](#), y es emitido confidencialmente para orientar al equipo de desarrollo de EScorpus.

El que recibe el documento asume la custodia y control, comprometiéndose a no reproducir, divulgar, difundir o de cualquier manera hacer de conocimientos público su contenido, excepto para cumplir el propósito para el cual se ha generado.

Estas reglas son aplicables a las **6** páginas de este documento.

Índice

<b>1. Introducción</b>	<b>5</b>
<b>1.1. Alcance</b>	5
<b>1.2. Definiciones, Acrónimos y Abreviaturas</b>	5
<b>1.3. Referencias</b>	5
<b>2. Esbozo de la estructura o taxonomía</b>	<b>5</b>
<b>2.1. Descripción de los elementos de la arquitectura</b>	5
<b>2.2. Definición de los elementos de usabilidad y accesibilidad</b>	6
<b>3. Definición de la estructura</b>	<b>6</b>
<b>3.1. Diagrama de arquitectura</b>	6
<b>3.2. Elementos del sistema de navegación</b>	6
<b>3.3. Diseño de las pantallas tipo</b>	7
<b>3.3.1. Vista Pricipal</b>	7
<b>3.3.2. Vista de Comparación (Borrador)</b>	10
<b>3.3.3. Vista de Diccionarios (Borrador)</b>	11
<b>3.4. Descripción de la interacción de los usuarios</b>	11

<b>4. Componentes relevantes del diseño</b>	<b>12</b>
<b>5. Anexos</b>	<b>12</b>

# 1. Introducción

Este documento define la arquitectura de información de ToNgueLP.

## 1.1. Alcance

Este documento refiere en su versión 1 los elementos de la arquitectura de información para las 3 vistas fundamentales de ToNgueLP:

- Vista Principal.
- Vista de Comparación.
- Vista de Diccionarios.

## 1.2. Definiciones, Acrónimos y Abreviaturas

## 1.3. Referencias

1. Estudio de Homólogos.
2. Ayuda generada con Sphinx.
3. Documentos manuscritos de Abel.
4. Lista de chequeo del estándar ISO 9241-210:2010

# 2. Esbozo de la estructura o taxonomía

<<Representación simple de la estructura del portal en cuanto a etiquetas y jerarquía de los contenidos.>>

<<Ejemplo (los niveles deben verse de esta forma):

Inicio

1.1 Quiénes somos

1.1.1 Nuestra historia

1.1.2. Nuestro Personal

1.1.2.1 Biografías del personal

1.2 Qué hacemos

1.2.1 Productos

1.2.2 Servicios

## 2.1. Descripción de los elementos de la arquitectura

<<Descripción textual de cada uno de los elementos de la estructura, características, comportamiento. etc.>>

### 2.2. Definición de los elementos de usabilidad y accesibilidad

<<Descripción textual de las características de los elementos de la aplicación que pueden influir en la usabilidad y la accesibilidad en el sistema, utilizar como referencia los documentos "Principios de Nielsen" y "Pautas de Accesibilidad".>>

## 3. Definición de la estructura

### 3.1. Diagrama de arquitectura

### 3.2. Elementos del sistema de navegación

<<Listado de los elementos del Sistema de Navegación>>

<<Ejemplo:

- Menú General: siempre presente en todo el sitio, permite el acceso a cada una de las áreas del sitio.
- Pié de Página: usualmente ubicado en la parte inferior de cada página, indica el nombre de la institución, teléfonos, dirección física y de correo electrónico.
- Barra Corporativa: ofrece diversas opciones de información respecto del sitio y tal como el anterior, se muestra en todas las páginas.
- Ruta de Acceso: listado que aparece en la parte superior de cada página y que muestra el trazado de páginas que hay entre la Portada del sitio hasta la página actual que se esté revisando; cada una de ellas debe tener un enlace, para acceder al área de la cual depende la página. Cada uno de los elementos que conforman este «camino» debe tener un enlace que permita el acceso a esas áreas. En la literatura internacional en inglés sobre este tema, se llama a este elemento como «breadcrumbs».
- Fecha de publicación: para saber la vigencia de publicación del contenido desplegado. • Botón Inicio: para ir a la portada.
- Botón Mapa del sitio: para ver el mapa del Sitio Web
- Botón Contacto: para enviar un mensaje al encargado del sitio
- Buscador: presente en cada página si la funcionalidad existe en el sitio.
- Botón Ayuda: para recibir ayuda sobre qué hacer en cada pantalla del sitio.
- Botón Imprimir: para imprimir el contenido de la página; se espera que el formato de impresión del documento que se muestra en pantalla sea más simple que la página normal del Sitio Web, para dar la impresión al usuario de que hay una preocupación por ayudarlo en la tarea de llevar impreso el contenido.>>

### 3.3. Diseño de las pantallas tipo

Se describen 3 pantallas tipos:

- Vista Principal.
- Vista de Comparación.
- Vista de Diccionarios.

#### 3.3.1. Vista Pricipal

Esta pantalla es la pantalla inicial de ToNgueLP, su objetivo principal es visualizar los casos contenidos en los XML de los corpus contenidos por la aplicación.

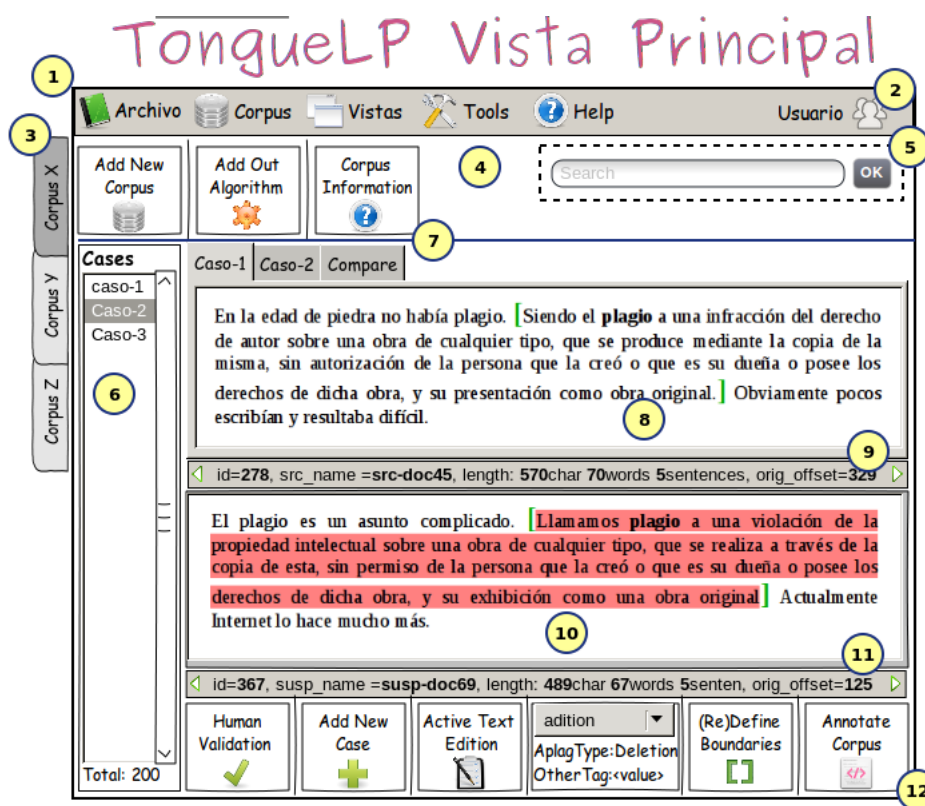


Figura 1: Vista Principal ToNgueLP

- Breve descripción de los elementos que componen la pantalla:

1. Barras de Menú
2. Opción de Loguin, para expertos autorizados a confirmar manualmente los casos generados automáticamente.

3. Tabs para diferentes corpus cargados.
4. Barra de Plugins
5. Componente de búsqueda.
6. Lista de casos
7. Tabs de, casos o vistas de la aplicación.
8. Fragmento del texto del documento fuente u original.
9. Barra de datos del fragmento fuente.
10. Fragmento del texto del documento sospechoso.
11. Barra de datos del fragmento sospechoso.
12. Elementos para editar fragmentos.

■ **Descripción detallada de los elementos que componen la pantalla:**

1. Barras de Menú (definición beta)
  - a) Archivo: permite cargar un archivo de corpus/ salvar el corpus en edición/ o todos los corpus.
  - b) Corpus: Información del Corpus/Seleccionar el Parser/Sugerir automáticamente el parser compatible.
  - c) Vistas: Vista Principal/Vista Comparación/Vista de Diccionario. (esta puede ser sustituida por la barra de plugins y entonces cambiar a Configuración: Idioma/Parser/Datos del Editor/...)
  - d) Tools: Acceso directo a los procesos de EScorpus para ejecutarlos manualmente:
    - 1) Convertir documentos TXT o PDF a ZZZdoc.XML
    - 2) Generar XML estandar a partir de la carpeta OUT de la salida de un algoritmo de detección.
    - 3) Comparación del XML estandar con XML del corpus cargado.
    - 4) Actualizar ciertos tags del XML de datos del algoritmo y actualizar ciertos tags del XML reporte de casos detectados.
  - e) About: enlaces a Copyright/Ayuda de Usuarios/Manual de Desarrollo
2. Opción de Loguin, para expertos autorizados a confirmar manualmente los casos generados automáticamente.
3. Tabs para diferentes corpus cargados: Tener acceso a más de un corpus simultáneamente.
4. Barra de Plugins: Similar a Doclux, permite accesar directamente las funciones principales de la aplicación.
5. Componente de búsqueda: buscar casos que contienen determinada secuencia de palabras (aquí se debe agregar un selector con: id, corpus, case, query)
6. Lista de casos: Mostrará en una lista todos los casos existentes en el corpus para accederlos rápidamente(en caso de que sean muchos este componente deberá tener también una búsqueda para filtrarlos y accesarlos rápidamente)



7. Tabs de, casos o vistas de la aplicación: casos en edición, o Vistas abiertas en las que se trabaja simultáneamente.
8. Fragmento del texto del documento fuente u original: visualiza el fragmento original del caso (este componente no permite edición al menos que se active el ELEMENTO “Active Text Edition” después de autenticado)
9. Barra de datos del fragmento fuente: datos que contiene el corpus sobre el fragmento original.(no se edita manualmente, solo automáticamente)
10. Fragmento del texto del documento sospechoso: visualiza el fragmento sospechoso del caso (este componente no permite edición al menos que se active el ELEMENTO “Active Text Edition” después de autenticado)
11. Barra de datos del fragmento sospechoso: datos que contiene el corpus sobre el fragmento sospechoso.(no se edita manualmente, solo automáticamente)
12. Elementos para editar fragmentos: barra de ELEMENTOS.
  - a) Human Validation: para que se confirme manualmente que el caso generado por un algoritmo automático realmente es consistente.
  - b) Add New Case: adicionar un nuevo caso. Se colocan en blanco 8 y 10. (un algoritmo(basado en análisis temático) sugiere 3 documentos donde serán insertados, se abre un popup para seleccionar y estos son insertados automáticamente)
  - c) Active Text Edition: permite editar 8 y 10 en el caso de la pestaña en el foco.
  - d) Datos generados automáticamente y un selector para que manualmente se categorice a este caso.
  - e) Define or Re-Define los límites(boundaries) del texto que compone el caso. Siempre podrá verse texto en los límites superior e inferior que no pertenecen al caso.
  - f) Annotate Corpus: permitirá adicionar etiquetas o visualizar aquellas que existen en el corpus, como: POS, SRL, Name Entity, etc.

## 3.3.2. Vista de Comparación (Borrador)

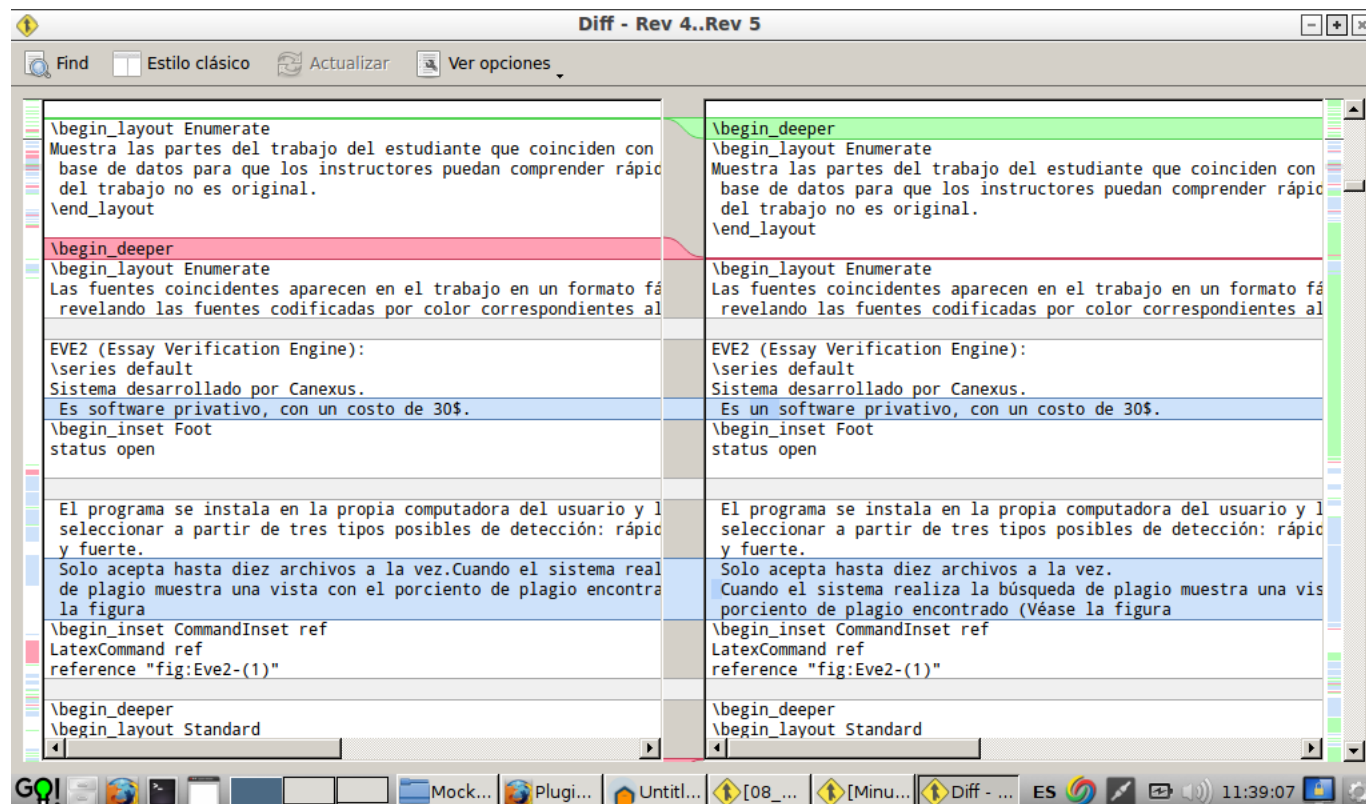


Figura 2: Vista de Diff de Bazaar Explorer, prototipo de como quedará la Vista de Comparación.

3.3.3. Vista de Diccionarios (Borrador)

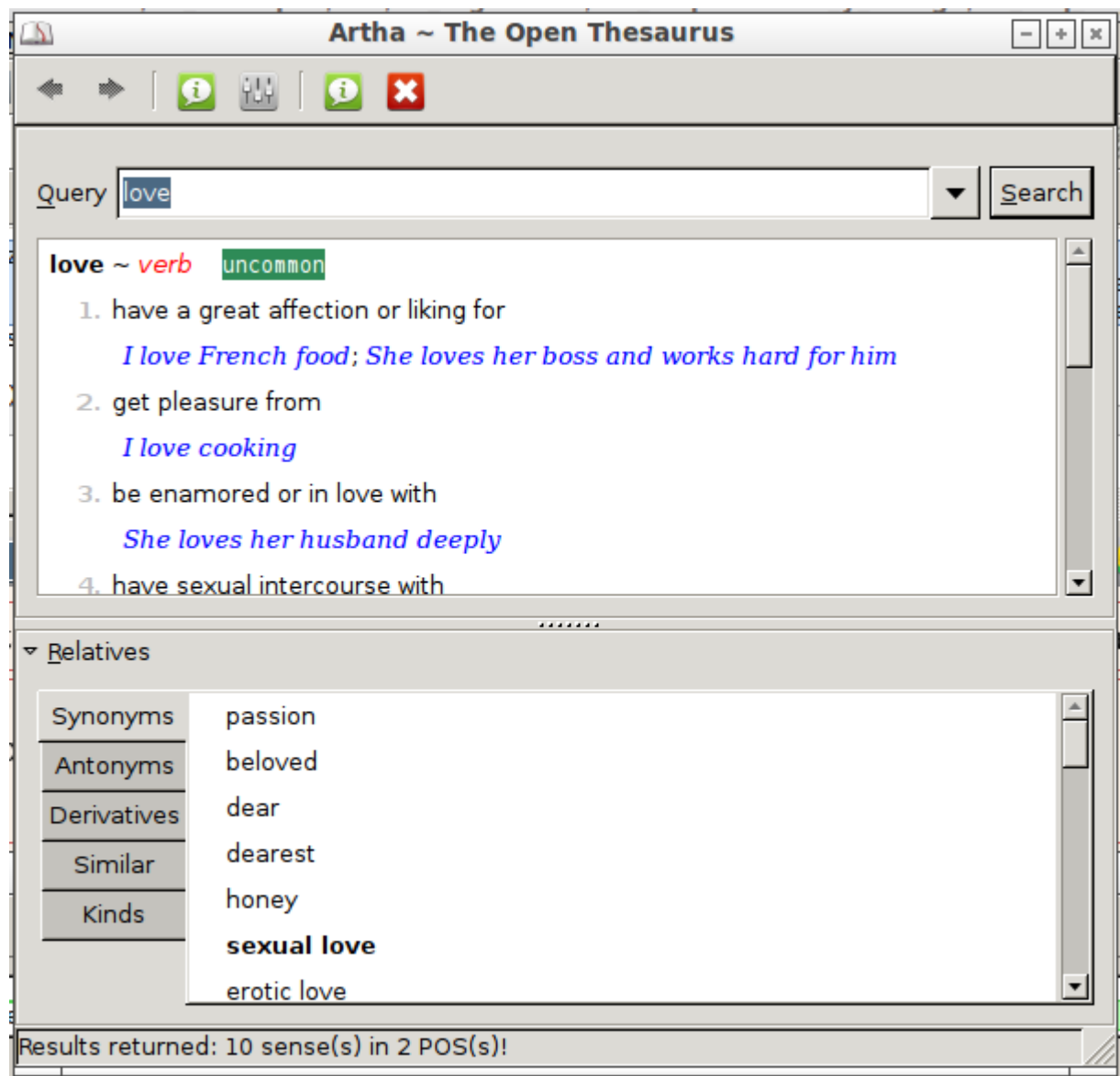


Figura 3: Vista del Front-End de Wordnet: Artha.

3.4. Descripción de la interacción de los usuarios

<<Debe explicarse la interacción de los elementos y los resultados en sus pantallas correspondientes.>>

## 4. Componentes relevantes del diseño

<<Deben listarse y explicar los componentes relevantes que se incluirán en el diseño con su nombre técnico y su correspondiente referencia internacional.>>

<<La idea de este punto es ir aumentando el acervo del equipo de desarrollo en cuanto a diseño de funcionalidades de acuerdo a lo más novedoso usado actualmente.>>

## 5. Anexos

<<Poner aquí imágenes para un mayor entendimiento de los componentes.>>