

---

# MAPA76 | Casos de uso: (documento de trabajo)

## A) Subir documento nuevo

El usuario sube un documento nuevo, mediante un WIZARD.

- > Selecciona que tipo de documento es
  - ? puede agregar una categoría nueva?
- > Le agrega descripción
- > Le pone fecha de PUBLICACION (cuando el documento se dio a conocer)

## B) Consultar base de documentos

- > Comparación de múltiples documentos
- > Bajar/exportar documentos (punto **D**)
- > Búsqueda de documentos
- > Filtro/orden de documentos
- > **Herramientas de la toolbox**
- > **Datos Contextuales del documento actual**

## C) Analisis y Edición

- > ALTA/BAJA/MODIFICACION de Documentos
- > Verificación de datos (un nombre es realmente un nombre? etc)
  - ? distribuida, wikipedia-style o votada o likes?
- > Creación de **HITOS** (persona-> lugar -> tiempo)

### >>> Creación de relaciones

#### Ver aparte **RELACIONES**

- Persona → acción → persona
- Persona → comparte lugar con → persona
- etc

## D) Dump/export

- > Cola de edición (poder tener dos o hitos en edición a la vez)
- > Citas bibliográficas  
Selecciono un párrafo o texto, y me lo prepara con sus respectivos datos bibliográficos
- > Trackback: Cualquier cita o copia o export que se haga de mapa76, pide un trackback para saber en que se utiliza esa información, para poder agregarla a los datos del documento y el usuario

## E) Visualización

- > Línea de tiempo/plotline
- > Mapa/geolocalización
- > Ejemplo utilizando herramienta TreeMap (para visualizar las relaciones entre HITOS)  
sitio web: <http://www.cs.umd.edu/hcil/treemap/>  
Teniendo los datos estructurados, podemos utilizar la herramienta de treemap para agrupar los casos de las diferentes personas en distintas jerarquías. Por ejemplo: centro de detención, provincia, fecha de detención, etc. De esta forma podemos descubrir la relación de una persona con otra en función de su pertenencia a un mismo grupo. Al ser las jerarquías configurables por el usuario, se podría investigar en función de la variable de interés del investigador.
- > Ejemplo de relaciones y timeline: EDGEMAPS

sitioweb: <http://mariandoerk.de/edgemaps/>

paper: <http://mariandoerk.de/edgemaps/ivj2012.pdf>

Puntos en común de los ejemplos con el caso de Mapa76:

La herramienta permite visualizar relaciones explícitas e implícitas en dos gráficos diferentes:

*Mapa de Similitud:*

Analiza la similitud entre músicos, pintores y filósofos en función de sus palabras claves. Las palabras claves son parte de la base de datos que los autores utilizaron. En nuestro caso, deberíamos definir qué utilizar como palabras claves para adaptar la visualización. Por ejemplo, definiendo como palabra clave el centro de detención, las personas que estuvieron en un mismo centro van a estar cerca en el mapa, mientras que las que no estuvieron estarán alejadas.

*Timeline:*

En la herramienta utilizan la fecha de nacimiento de la persona para situarlo en la línea del tiempo. En nuestro caso podríamos definir fecha de detención, fecha de desaparición, etc.

*Relaciones explícitas:*

Son modeladas como las influencias que recibe una persona y las personas en las que influyen. Son explícitas porque son un dato que surge de la base. En nuestro caso a partir de la definición de la influencia podremos definir las relaciones. Por ejemplo represor-víctima.

*Relaciones implícitas:*

Se definen en función del tiempo y las palabras claves.

*Elementos de diseño:*

El tamaño de los nodos representa la cantidad de influencias sobre otros que tiene una persona. El color está armado en función de las palabras claves (ver paper para mayor referencia). En nuestro caso en función de lo que definamos como influencia podremos determinar la relevancia de estos elementos en el diseño.

---

## MAPA76 | DEFINICIONES:

**FRONT END:** visualizaciones, mapa, consulta de datos

**BACK END:** carga de documentos, administracion de documentos

**HECHOS:** Conexión contextual de datos de nombre, lugar y tiempo. Se puede llegar a lograr una detección cuasi automática de los mismos. *EJ: Pedro estuvo en la perla en enero del '77*

**RELACIONES:** Conexión conceptual, contextual, gramatical de HITOS o personas o lugares. Pueden incluir HECHOS. Pueden ser multi personales, o multi locaciones, o multi temporales. *EJ: Pedro vio a Laura al ser trasladados desde La Perla a Mansion Sere.*

**DOCUMENTOS:** Son los archivos (pdf, doc, txt) que contienen la información (fallos, testimonios, etc)

**CONSULTAR:** Leer textos, hacer búsquedas, ver relaciones, seguir vínculos

**EDITAR:** Realizar nuevas relaciones, hitos, cruces. Editar datos obtenidos automáticamente

**NOMBRE DE ENTIDAD (Named Entity o NE):** Es un elemento atómico en un texto que representa el nombre de una persona, un lugar, una fecha, una organización, o cualquier otro concepto que merezca ser destacado<sup>1</sup>.

**PERFIL:** ...

### Procesos

**NORMALIZACIÓN:** Si aún no lo es, se convierte el documento a PDF para facilitar la extracción de texto. Luego se extraen las líneas de texto del documento, manteniendo la ubicación de estas y su tipografía. Se normaliza la codificación (a UTF8). Se remueven caracteres inválidos o innecesarios<sup>2</sup>.

**ANÁLISIS DE COMPOSICIÓN:** Se realiza un análisis de composición geométrico y lógico para extraer el texto plano en el orden de lectura óptimo. Se eliminan los grupos de texto repetidos o que no tienen relevancia para la detección de NEs.

**EXTRACCIÓN:** Es el proceso de detección y clasificación de NEs (nombres de entidad, o *named entities*), como nombres de personas, lugares, organizaciones, fechas, direcciones, etc.

**RESOLUCIÓN DE CORREFERENCIA:** Luego de la extracción de NEs, este proceso analiza si dos NEs *correfieren*, es decir, si ambos NEs refieren al mismo objeto. Por ej. si *Juan Pérez* y *Juan* correfieren, son la misma persona. El contexto en donde aparece la entidad influye en la decisión tomada.

**GEODECODIFICACIÓN:** ...

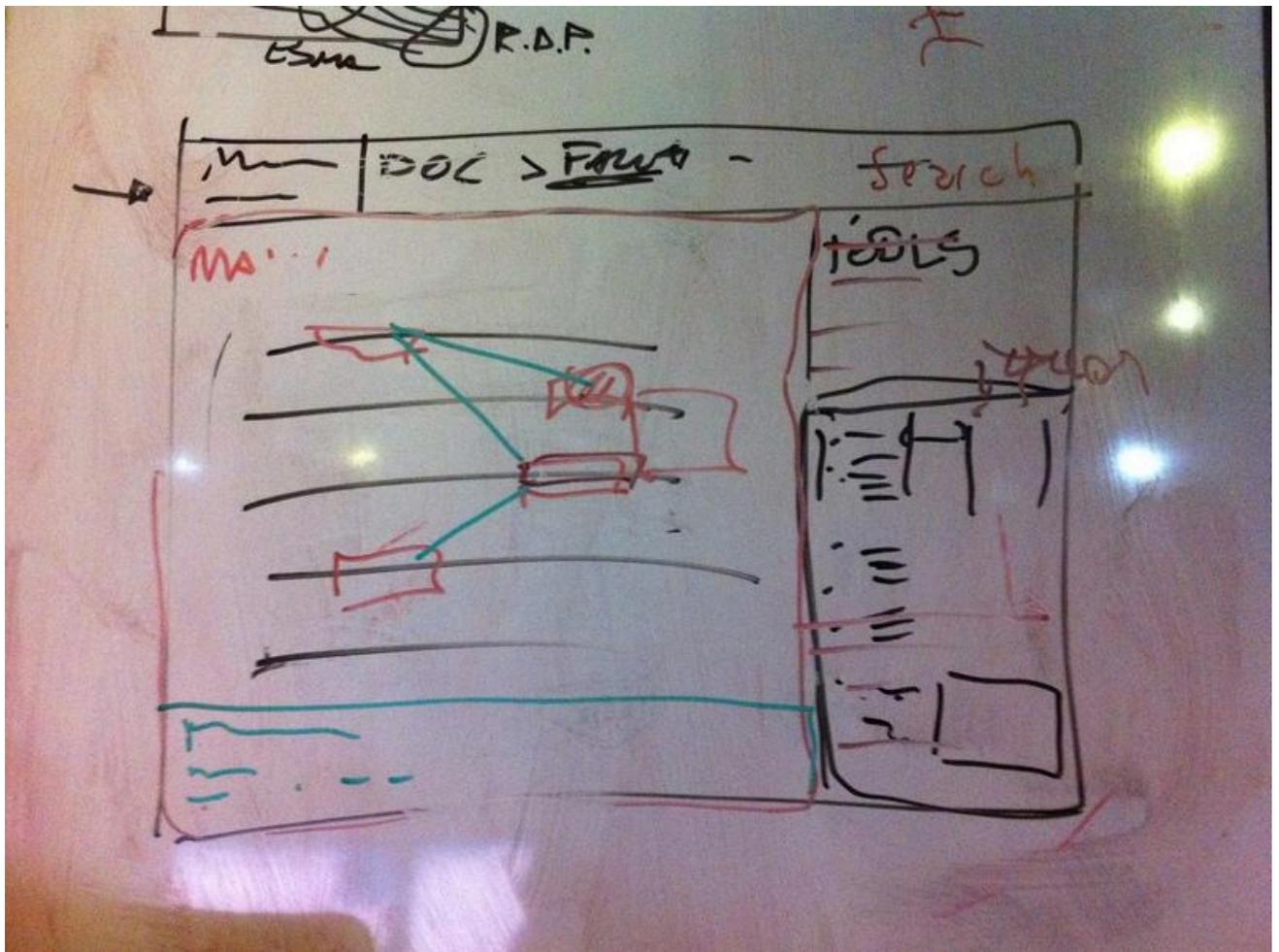
---

<sup>1</sup> **NOMBRE DE ENTIDAD != PERSONA:** Los nombres de entidad, como los nombres de una persona, son sólo *referencias* a una persona. Un registro identificado como nombre en el proceso de **extracción**, tiene un nivel de confianza bajo, y será asociado a una persona cuando su nivel de confianza supere el *threshold* asignado.

<sup>2</sup> Sean caracteres inválidos que ya estaban en el documento original, o fueron producidos luego de la extracción por OCR.

DIVISION CONCEPTUAL DE LA INTERFACE.

UBICACION: DONDE ESTOY?	TOOLBOX: HERRAMIENTAS DE TRABAJO
MAIN: LO QUE ESTOY VIENDO/ ANALIZANDO.	DATOS CONTEXTUALES DE LO QUE ESTOY VIENDO



## MAPA76 | TAREAS:

Blog de Mapa76.info

Apuntar mapa76.info a hhba

Instalar Wordpress

Buscar Theme x

Reunion con Gonzalo Archivo de la memoria

# MAPA76 | SISTEMA DE CONFIABILIDAD DE REGISTROS:

EL PROBLEMA DE LOS NOMBRES REPETIDOS.

El sistema de scrapping levanta los registros asumiendolos como datos unicos (todos los nombres encontrados pertenecen a ala misma persona), y les asigna un nivel de confiabilidad 0

La confiabilidad ira subiendo dependiendo del resultado de aplicacion de diferentes modulos.

Ej: Se encuentra un nombre XX, se busca en ANM. Si en conadep hay un solo registro para ese nombre, entonces la confiabilidad de ese registro nombre, es elevada. Se busca tambien un rango de registros que rodean ese nombre en el pdf fuente (el subido a mapa76). se comparan dichos registros con los que estan en la ficha de ANM. si coincidcen esto le da mas confiabilidad al nombre y tambien a los registros scrapeados.

En el caso de que no figure en ANM o que figuren varios con el mismo nombre, la confiabilidad sera menor, y aumentara solo si el resto de registros conicida, o mediante otros modulos futuros.

## MEJORA DE CONFIABILIDAD EN LA CARGA DEL DOCUMENTO

Aprovechar el input humano en el momento de subir el documento a mapa76

Utilizando plantillas, al preguntarle al humano que tipo de documento es, puede el wizard guiarlo y ofrecerle los testigos, acusados, victimas, etc, sacados de la portada del documento si este fuera un legajo judicial.

Esto permitiria que en ese documento los registros conicidentes con los datos suministrados en la subida del documento, ya tengan un nivel de confiabilidad mayor.

# MAPA 76

