

## Tratamiento automatizado de documentación judicial





## Serikat - Crecemos e innovamos contigo



- La Unidad de Tecnologías de la Información del **Grupo Erhardt**
- + de 20 años de experiencia
- + de **350 profesionales**
- 8 oficinas en España
- [www.serikat.es](http://www.serikat.es)





## La Gestión de la Justicia se encuentra ante el reto de incorporar de manera efectiva y eficiente datos que se generan en texto no estructurado

Cumpliendo con lo que a mi parecer debe ser una prioridad de común acuerdo con lo que dice Pablo, os paso el presupuesto por mi intervención profesional en el asunto contra MIRLO, en el mismo os detallo las bases para su calculo y las condiciones del mismo, mínimo posible, 2.802 € .

Por otro lado el Procurador me presupuesto 800 por su intervención profesional lo que haría con el Iva correspondiente 968 €.

Si hubiera recurso de apelación la cantidad se incrementaría en un 60%.

Si hubiera condena en costas, la cantidad se encarecería en el doble porque habría que abonar a la otra parte la mismas cantidades.

También si se ganase el procedimiento debería abonar los honorarios mis y del procurador la comunidad de propietarios de MIRLO.

Os pido que recordéis y es importante, que la Audiencia Previa esta prevista para abril de 2016 y el juicio será para un año después, razón, nos ha tocado el Juzgado N° x de LALALA, no es lento, es el mas retrasado de todos los de LALALA.

un saludo y a la espera de vuestras noticias

### ¿Qué queremos extraer para la gestión?

- Coste abogado:
- Coste Procurador
  
- Abogado X
- Objeto: Presupuesto
- Demandante
- Mirlo: Demandado

Asunto: No se sabe por el texto. Se podría deducir

“Voz: Presupuesto .... Condena.. Apelación....”

# La Gestión de la Justicia se encuentra ante el reto de incorporar de manera efectiva y eficiente datos que se generan en texto no estructurado



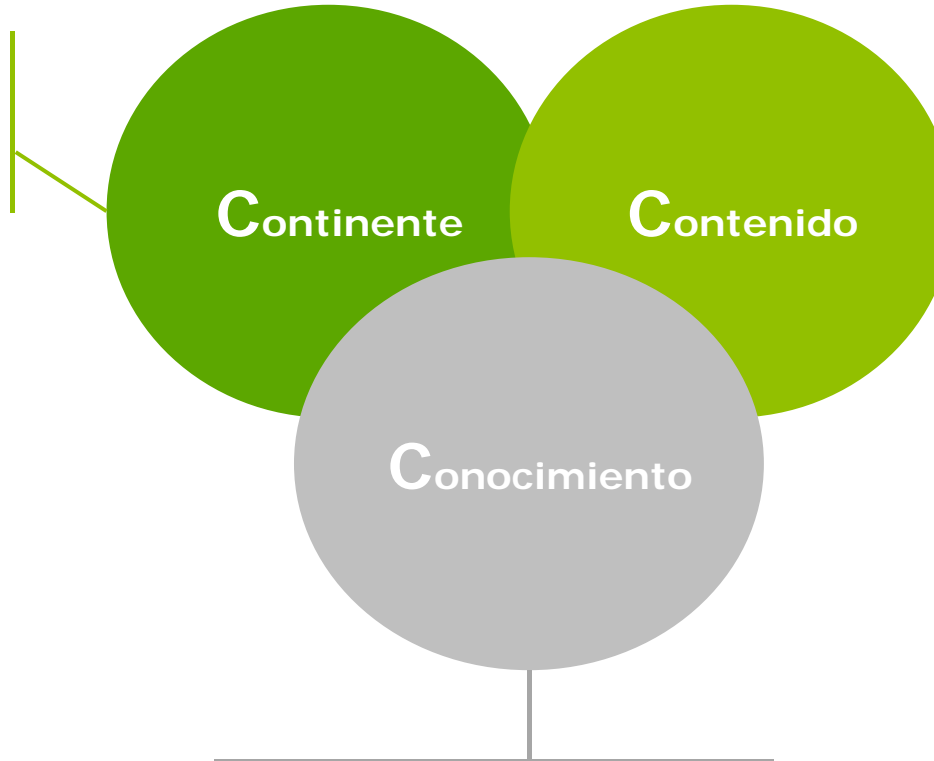
## En el mercado hay 3 aproximaciones para abordar el tratamiento automático de texto no estructurado

	Técnicas de localización posicional	Técnicas lingüísticas	Tratamiento holístico del documento
<b>¿En qué consiste?</b>	<ul style="list-style-type: none"> <li>Localización de contenido ubicado en una zona específica de un documento.</li> </ul>	<ul style="list-style-type: none"> <li>Aplicación de técnicas lingüísticas para intentar identificar el significado de un texto.</li> </ul>	<ul style="list-style-type: none"> <li>Localización de contenido independiente de su ubicación</li> <li>Análisis de la estructura de los documentos.</li> <li>Aplicación de técnicas semánticas, morfosintácticas y segmentación.</li> </ul>
<b>¿Para qué se aplica?</b>	<ul style="list-style-type: none"> <li>Extracción de información de documentos con estructura fija y conocida (formularios, facturas, albaranes, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>Análisis de sentimientos, identificación de patrones por agregación de información, etc, principalmente en entornos de Marketing.</li> </ul>	<ul style="list-style-type: none"> <li>Clasificación automática de documentos.</li> <li>Metadatación.</li> <li>Extracción estructurada de contenido.</li> <li>Grabación y explotación de datos.</li> </ul>
<b>Puntos fuertes</b>	<ul style="list-style-type: none"> <li>Precisión en la captura de datos.</li> <li>Velocidad de procesamiento.</li> </ul>	<ul style="list-style-type: none"> <li>Aplicaciones avanzadas basadas en el significado del texto.</li> </ul>	<ul style="list-style-type: none"> <li>Precisión en la captura de datos.</li> <li>Permite tratar documentos y textos con estructura variable.</li> <li>Flexibilidad de aplicación a distintos tipos documentales.</li> <li>ROI favorable.</li> </ul>
<b>Debilidades</b>	<ul style="list-style-type: none"> <li>Dificultad para tratar documentos con estructura variable.</li> <li>Requiere configuración específica para cada nuevo tipo documental.</li> <li>Dependiente de la precisión del OCR cuando trata imágenes de texto.</li> </ul>	<ul style="list-style-type: none"> <li>Muy baja precisión en la captura de datos concretos.</li> <li>ROI difícil de medir.</li> <li>Dependiente de la precisión del OCR cuando trata imágenes de texto.</li> </ul>	<ul style="list-style-type: none"> <li>Dependiente de la precisión del OCR cuando trata imágenes de texto.</li> </ul>

## Serikat apuesta por un enfoque holístico que explota las 3 C's en un documento

### ¿Cuál es el tipo documental?

Identifica el tipo de documento en base a su estructura y contenido.



### ¿Qué información hay en el documento?

Localiza, extrae y estructura los datos contenidos en el documento.

### ¿Sobre qué habla el documento?

Identifica voces en base a tesauros predefinidos.

- Maximizar **automatismos**.
  - Uso intensivo de reglas de negocio
  - Taxonomías
- Aplicar técnicas de **segmentación de texto y análisis morfosintáctico**.
  - Reconocimiento de elementos y estructuras de palabras relevantes para el objeto de análisis:
    - Metadatos, Voces,...
  - Estructura de documentos:
    - Publicaciones
    - Documentos aislados
  - Tipografía de letra:
    - Tamaños
    - Negritas
    - Cursivas
    - Subrayados,...
- Utilizar el máximo posible **componentes léxicos**:
  - Internos: Proporcionados por el Cliente: Tesoros, Ficheros de referencia...
  - Externos: Diccionarios públicos, normativas, NER, ...
  - Nuestra experiencia, Taxonomías, sinónimos, antónimos...





### 2. Análisis y tratamiento

Contenido , contenido y conocimiento

Explotación de la información recopilada ...

- Clasificación
- Catalogación
- Anonimización
- Identificación de voces
- Enriquecimiento

...



### 1. Entender, recopilar y preparar la información

De distintas fuentes de datos:

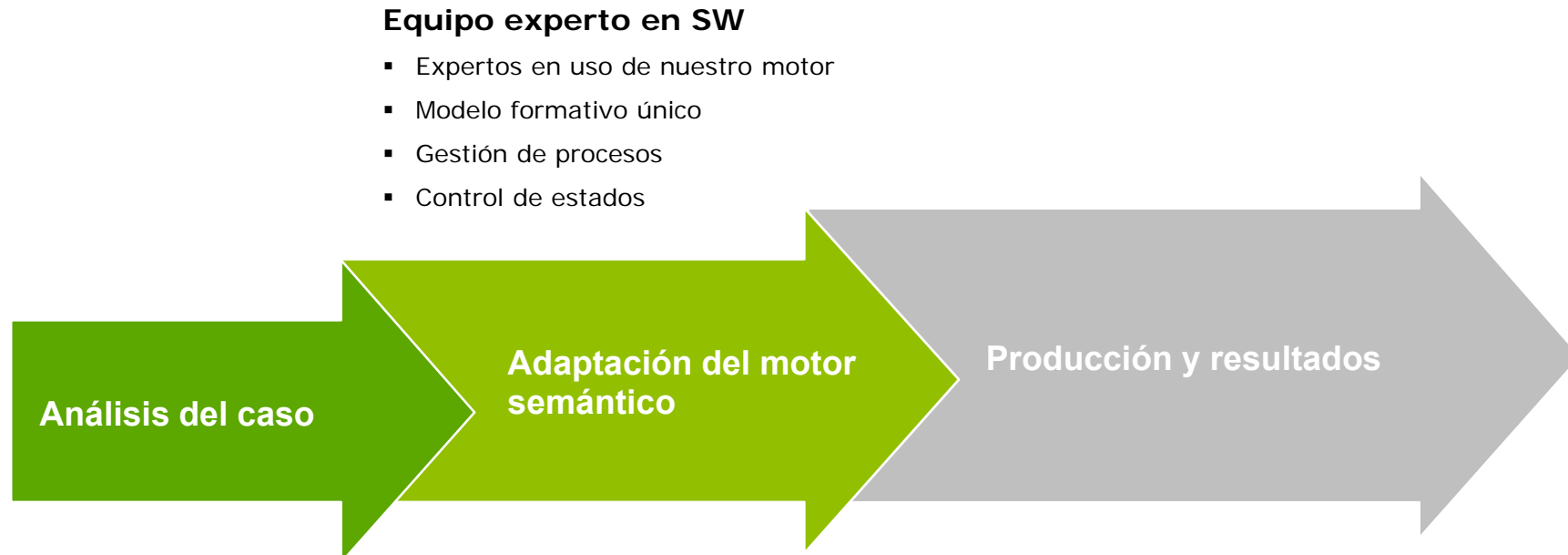
- Documentos físicos (digitalización y OCR).
- Documentos digitales.
- Publicaciones en la Web.
- Blogs.
- Redes sociales.
- Bases de datos con texto no estructurado.

### 3. Presentación e integración de la información

Generación de información de salida estructurada y presentación en múltiples formatos:

- XML
- Pdf
- Web
- ...

## Apojado en un equipo experto y multidisciplinar



### Equipo experto en SW

- Expertos en uso de nuestro motor
- Modelo formativo único
- Gestión de procesos
- Control de estados

### Equipo consultor

- Enfoque de la tecnología
- Estudio de la necesidad
- Estudio de las fuentes de datos
- Ponderación de atributos
- Estudio de los casos de uso
- Adaptación de procesos
- Proyecto definido

### Equipo de producción y mejora continua

- Gestión supervisada
- Proceso continuo de cambio
- Proceso continuo de mejora
- Gestión incidencias



CONSEJO GENERAL DEL PODER JUDICIAL  
CENTRO DE DOCUMENTACIÓN JUDICIAL

## CGPJ – CENTRO DE DOCUMENTACIÓN (CENDOJ)



Red.es / Ministerio de Justicia

### Tratamiento de jurisprudencia

- Digitalización y tratamiento de las sentencias judiciales en España:
  - Clasificación de sentencias.
  - Identificación automática de voces.
  - Despersonalización (eliminación de datos identificativos contenidos en las sentencias).
  - Enriquecimiento: Localización, normalización y vinculación de información relevante (jurisprudencia, legislación, ...).
  - + **de 5.000.000 de sentencias** tratadas a lo largo de los más de **10 años** de servicio.

### Clasificación de Expedientes Judiciales

- Proyecto perteneciente al Plan de Modernización de la Justicia.
- Digitalización y/o tratamiento de los expedientes judiciales.
  - Digitalización y/o corte automático de documentos.
  - Identificación de tipos documentales.
  - Catalogación de los expedientes.
  - Integración con el sistema de gestión procesal.

## Captura

LANTECEDENTES

1.- El Juzgado de Instrucción número 2 de Pamplona instruyó Sumario con el número 3/2007, contra **[REDACTED]**, por un delito continuado de abusos sexuales, y una vez concluso lo remitió a la Audiencia Provincial de Navarra que, con fecha 1 de abril de 2008 dictó sentencia que contiene los siguientes hechos probados:

«El procesado Sr. **[REDACTED]** mayor de edad, con antecedentes penales no computables a efectos de reincidencia cuyos anteriores datos de identidad se consigna, agredió sexualmente a su esposa la Srta. **[REDACTED]** las instalaciones destinadas a bar y restaurante anexo al santuario de San Miguel de Aralar ubicado en el Comarcado Foral de Navarra, desde mediados del mes de marzo del año 2006, hasta mediados del mes de enero de 2007.

El hermano de la Srta. **[REDACTED]** casado con D. **[REDACTED]** quienes habían adoptado en el mes de marzo del año 2002 a la Srta. **[REDACTED]** nacida el 4 de abril de 2001, había trabajado, en el año 2006, desde su profesión de abogado en la realización de actuaciones obras de construcción del Santuario antes al cual se encuentra el local destinado a bar y restaurante, cuya explotación, era llevada por su única hermana y cuñada.

Por razón de la relación familiar que los unía, el matrimonio formado por el Sr. **[REDACTED]** y la Srta. **[REDACTED]** junto a los **[REDACTED]** se desplazaron en diversas ocasiones, al menos cuatro o cinco, a comer algún día del fin de semana, normalmente los domingos, al bar y restaurante que regenta por el cuñado y la hermana del Sr. **[REDACTED]** produciéndose entre desplazamientos algún día del fin de semana para comer entre los meses de abril y septiembre del año 2006.

El procesado **[REDACTED]** movido por un intenso libidino o abstinencia de la relación su conducta que mantenía con su esposa **[REDACTED]** que varias de las veces a cinco ocasiones en las que **[REDACTED]** se había desplazado a comer al restaurante del santuario de San Miguel de Aralar entre los meses de abril y septiembre de 2006, aprovechó para abusar a la

#	Campo
01	NOMBRE TITULAR (1-4)
02	YFR APFI (1)O TITULAR (1-4)
03	YFR APFI (1)O TITULAR (1-4)
04	NIF/CIF TITULAR (1-4)
05	PARTICIPACIÓN TITULAR (1-5)
06	FORMA DE PROPIEDAD
07	TIPO INMUEBLE
08	PLZ/A RPT(A-EH)
09	NÚMERO REGISTRO
10	NÚMERO FINCA
11	LEUPO
12	LEBPO
13	FOLEO

## Anonimizador

LANTECEDENTES


1.- El Juzgado de Instrucción número 2 de Pamplona instruyó Sumario con el número 3/2007, contra **[REDACTED]**, por un delito continuado de abusos sexuales, y una vez concluso lo remitió a la Audiencia Provincial de Navarra que, con fecha 1 de abril de 2008 dictó sentencia que contiene los siguientes hechos probados:

«El procesado Sr. **[REDACTED]** mayor de edad, con antecedentes penales no computables a efectos de reincidencia cuyos anteriores datos de identidad se consigna, agredió sexualmente a su esposa la Srta. **[REDACTED]** las instalaciones destinadas a bar y restaurante anexo al santuario de San Miguel de Aralar ubicado en el Comarcado Foral de Navarra, desde mediados del mes de marzo del año 2006, hasta mediados del mes de enero de 2007.

El hermano de la Srta. **[REDACTED]** casado con D. **[REDACTED]** quienes habían adoptado en el mes de marzo del año 2002 a la Srta. **[REDACTED]** nacida el 4 de abril de 2001, había trabajado, en el año 2006, desde su profesión de abogado en la realización de actuaciones obras de construcción del Santuario antes al cual se encuentra el local destinado a bar y restaurante, cuya explotación, era llevada por su única hermana y cuñada.

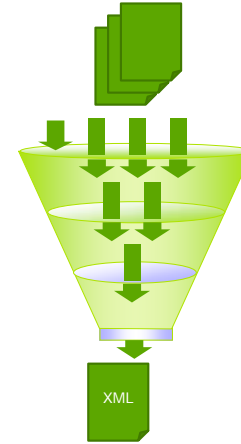
Por razón de la relación familiar que los unía, el matrimonio formado por el Sr. **[REDACTED]** y la Srta. **[REDACTED]** junto a los **[REDACTED]** se desplazaron en diversas ocasiones, al menos cuatro o cinco, a comer algún día del fin de semana, normalmente los domingos, al bar y restaurante que regenta por el cuñado y la hermana del Sr. **[REDACTED]** produciéndose entre desplazamientos algún día del fin de semana para comer entre los meses de abril y septiembre del año 2006.

El procesado **[REDACTED]** movido por un intenso libidino o abstinencia de la relación su conducta que mantenía con su esposa **[REDACTED]** que varias de las veces a cinco ocasiones en las que **[REDACTED]** se había desplazado a comer al restaurante del santuario de San Miguel de Aralar entre los meses de abril y septiembre de 2006, aprovechó para abusar a la



Juan Pérez López → Emilio  
 Josefa Andrés Fernández → Marina  
 Antonia Gómez Pérez → Apolonia  
 Luis Andrés Fernández → Pablo  
 Carla Andrés Gómez → Sofía

## Clasificador



### Qué hace

- Extrae información contenida en el documento no estructurado.

### Qué hace

- Identifica y elimina cualquier referencia a datos confidenciales o de carácter personal incluidos en un documento de texto no estructurado

### Qué hace

- Clasifica automáticamente por tipo de documento
- Clasifica automáticamente por contenido en base a "voces"

### Cómo lo hace

- Identificación de datos en base a diccionarios, taxonomías y análisis sintáctico.
- Utilización de tesauros personalizados.
- Reglas personalizadas.
- Integrable con el gestor documental.
- Combinado, cuando es necesario, con tratamiento "humano" para maximizar los niveles de calidad.

### Ámbitos de aplicación

- Documentación financiera y fiscal.
- Sentencias judiciales y documentos jurídicos.
- Historias clínicas.
- Notas registrales, escrituras y tasaciones.

### Ámbitos de aplicación

- Documentación financiera y fiscal.
- Sentencias judiciales y documentos jurídicos.
- Historias clínicas.
- LOPD

### Ámbitos de aplicación

- Cualquier proceso operativo en el que se procesen documentos
- Gestión del conocimiento

## Que descubren numerosos ámbitos de aplicación

### Persecución del fraude

- Sanitario
  - Extracción y análisis de datos de facturas / consumos derivados de informes médicos (informe de quirófano, ...)
- Fiscal
  - Facturas...
- Seguros
  - Reparaciones / peritaciones

### Reducción de costes AAPP

- Registro de entrada
  - Clasificación y ordenación de tipo documental
  - Catalogación automática de datos
  - Flujo directo al sistema
  - Detección de faltas de información
  - Reducción de duplicidades
- Disociación de datos
  - Cumplimiento de objetivos de transparencia

### Gestión del conocimiento

- Prensa
  - Clasificación automática de noticias de prensa
- Logística transporte
  - Detección de necesidades de carga en correos electrónicos

### Mejorar el negocio

- Precios
  - Extracción de información de precios por nichos
- Agilizar prestamos a empresas
  - Extracción de BS y PG



- Optimizar el **tiempo y coste** de tratamiento y explotación de la información.
- **Liberar recursos** para dedicarlos a tareas de mayor valor añadido.
- Incrementar la **calidad**, eliminando errores por intervención humana y homogeneizando los resultados.
- Incorporar **nuevas fuentes de información** para enriquecer los procesos de negocio.
- Mejorar la gestión, la calidad y la rapidez en el proceso administrativo judicial



## Otras colaboraciones

GARRIGUES



THOMSON REUTERS  
ARANZADI



Oficina Española  
de Patentes y Marcas

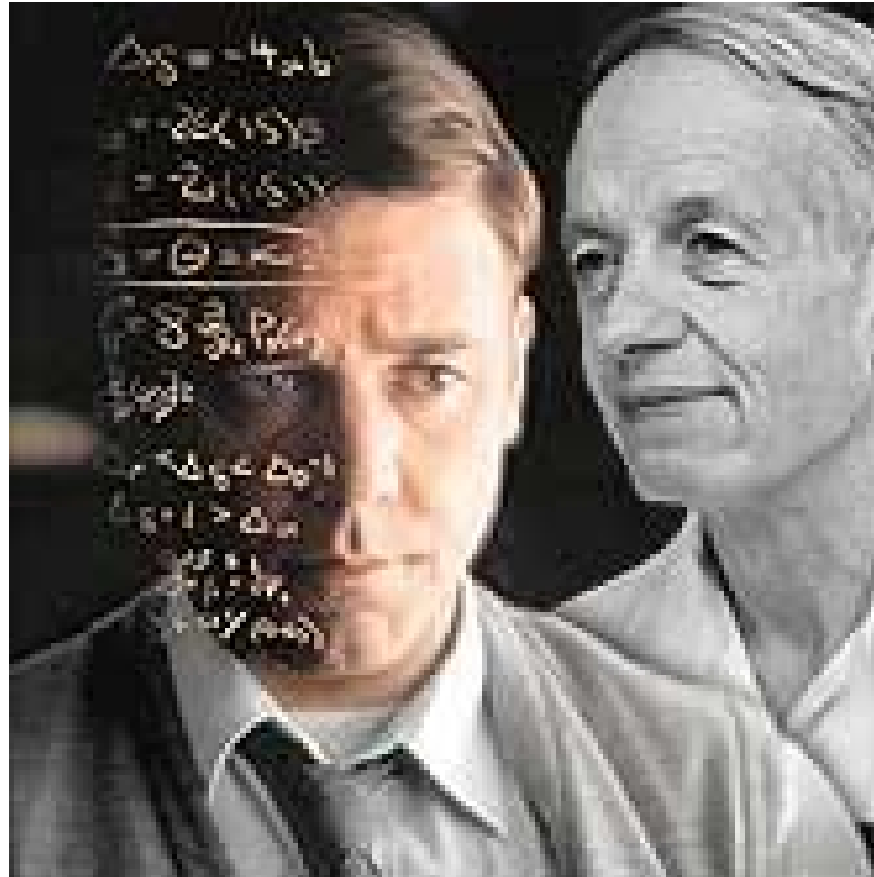


Gipuzkoako Foru Aldundia  
Diputación Foral de Gipuzkoa





## Resumen del taller





**serikat** >>