

Estructura, semántica, extracción de información y XML legislativo: experiencias en la U. de Valladolid



Universidad de Valladolid

**M. Mercedes Martínez, Dámaso-Javier Vicente, Pablo de la Fuente,
Joaquín Adiego, Alma M. Pisabarro, Jose Manuel Sánchez**
(U. de Valladolid, España)



15-16 de noviembre de 2007

El equipo

Informática

Mercedes Martínez González
Pablo de la Fuente Redondo
Joaquín Adiego Rodríguez
Alma M. Pisabarro Marrón

*GRINBD (Grupo de Recuperación de Información y Bibliotecas Digitales).
Departamento de Informática. U. de Valladolid
(España).*

- Ingeniería de contenidos
- Recuperación de información
- XML
- Extracción de información
- Textos legislativos

Derecho

Dámaso-Javier Vicente Blanco
Jose-Manuel Sánchez Felipe

*Área de Derecho Internacional Privado. Facultad
de Derecho. Universidad de Valladolid (España).*

- Inversiones internacionales
- Contratos electrónicos y comercio electrónico de ámbito internacional
- Protección de intereses generales en el comercio internacional
- Legislación Comunitaria

Ideas principales

- Textos normativos
- Dominio de usuario
- Conocimiento del entorno jurídico
- Código abierto
- Esfuerzos de implementación razonables

Council Regulation (EC) No 44/2001 of 22 December 2000 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters

...

Chapter II

JURISDICTION

Section 1

General provisions

Article 2.

1. Subject to this Regulation, persons domiciled in a Member State shall, whatever their nationality, be sued in the courts of that Member State.

2. Persons who are not nationals of the Member State ...

Article 3

1. Persons domiciled in a Member State may be sued in the courts of another Member State only by virtue of the rules set out in Sections 2 to 7 of this chapter.

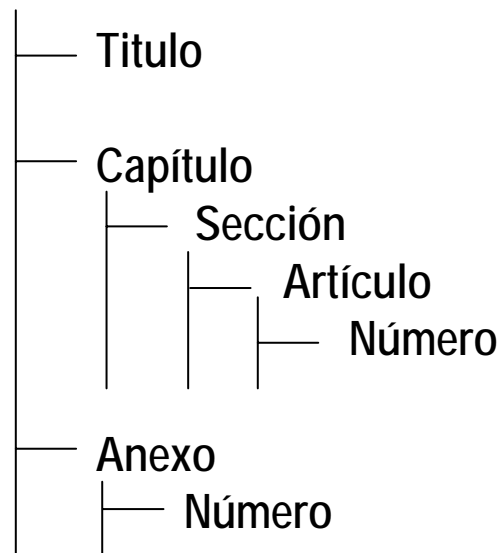
2. In particular the rules of national jurisdiction set out in Annex I shall not be applicable as against them.

...

Aproximación (I): Estructura

- Documentos cuya estructura está implícita en su contenido

LEY / DECRETO /...



Council Regulation (EC) No 44/2001 of 22 December 2000 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters

...

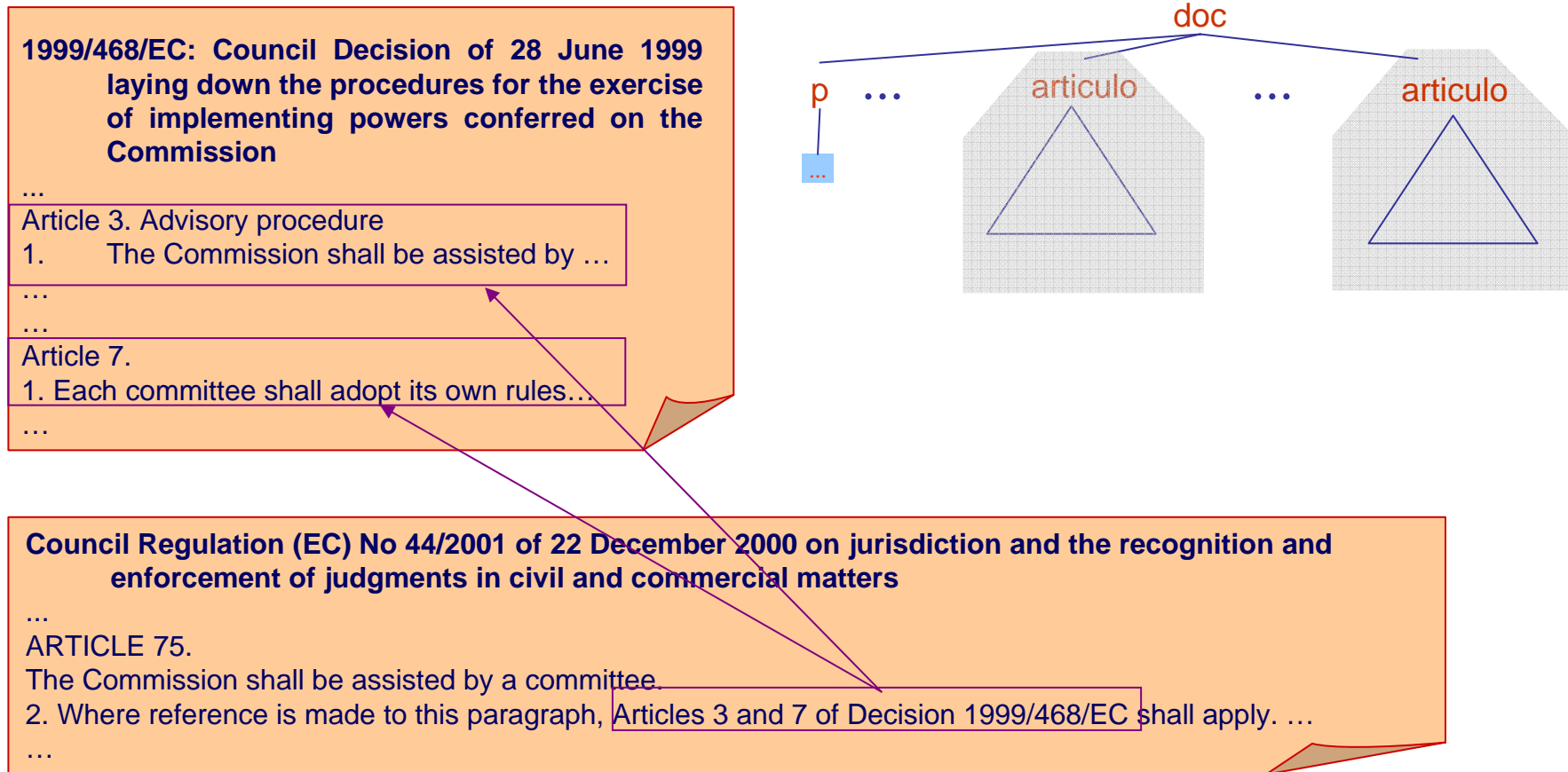
Chapter II
JURISDICTION
Section 1
General provisions
Article 2.
1. Subject to this Regulation, persons domiciled in a Member State shall, whatever their nationality, be sued in the courts of that Member State.
2. Persons who are not nationals of the Member State ...

Article 3
1. Persons domiciled in a Member State may be sued in the courts of another Member State only by virtue of the rules set out in Sections 2 to 7 of this chapter.
2. In particular the rules of national jurisdiction set out in Annex I shall not be applicable as against them.

...

Aproximación (II): Relaciones y estructura

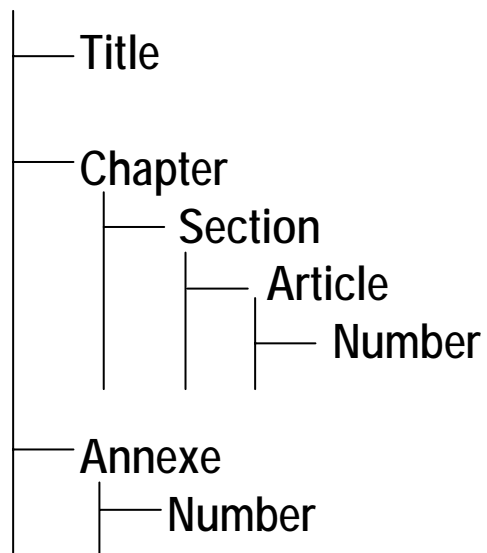
■ Relaciones entre elementos de estructura



Aproximación (III): Metadatos y Estructura

- Localización de patrones de información

RULE/DECREE/REGULATION/...



Convention on Jurisdiction and the enforcement of judgements in civil and commercial matters, done at Lugano on 16 September 1988

...

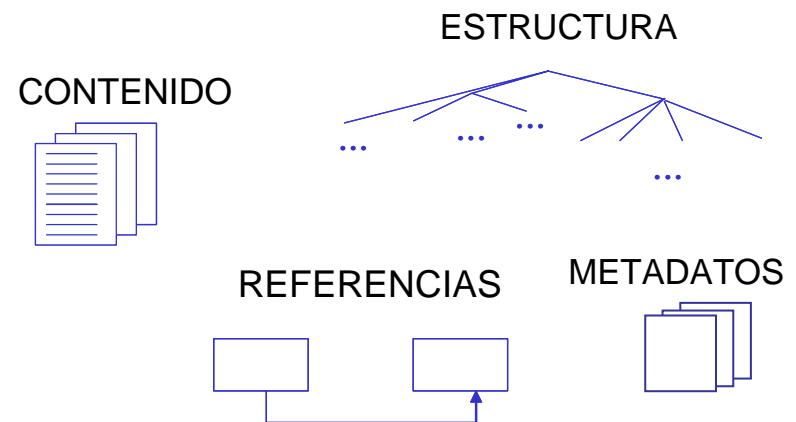
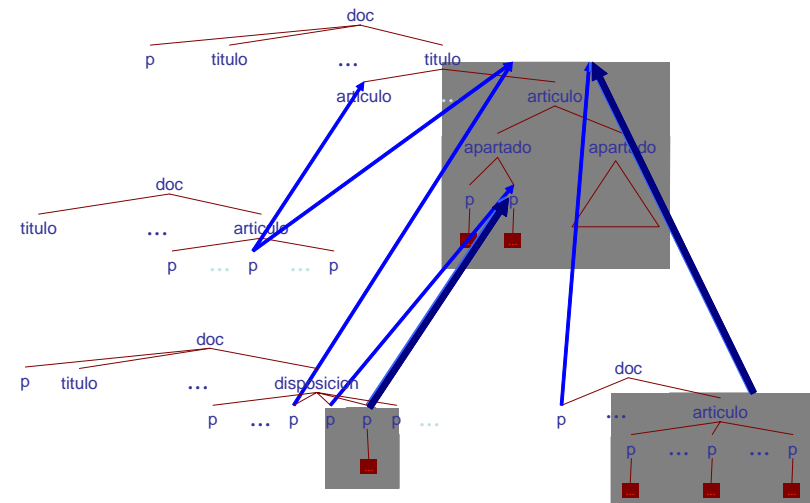
Article 68
This Convention, drawn up in a single original in the Danish, Dutch, Finland,
...

Por su Majestad el Rey de España
Pour le président de la République française

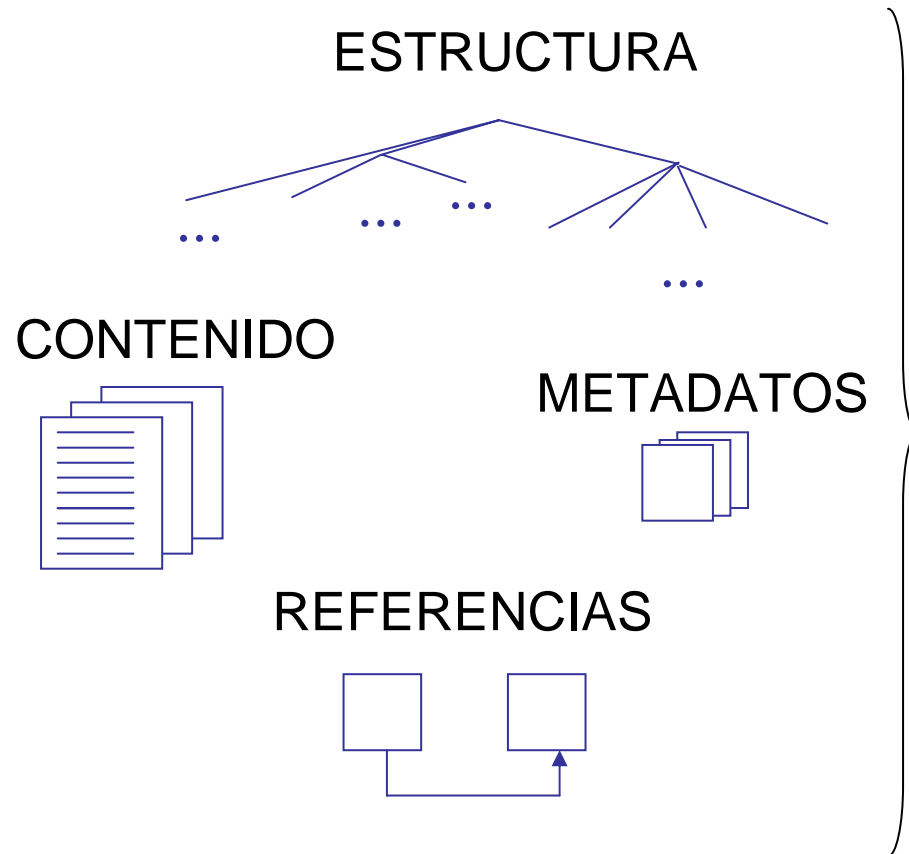
...

Nuestra propuesta investigadora

- **Modelo**
 - Grafo de relaciones
- **Procesos relacionados con la estructura**
 - Extracción de estructura
 - Extracción de referencias, basada en la estructura
 - Recuperación de información estructurada
 - Obtención de estructuras virtuales (consolidación, enlaces hipertexto)
 - Anotación a elementos de estructura
- **Arquitectura**
 - Servicios. Facilita la integración
- **Implementación de la propuesta**
 - XML



XML en la implementación



- Textos legislativos: documentos XML
 - norma.dtd, norma.xsd
- Metadatos, referencias, anotaciones: documentos XML adicionales
 - notas.dtd
 - enlaces.dtd
 - ...
- XPath, XSLT, DTD / XML Schema, SAX, XLink, XPointer

Extracción de estructura (normalización)

- Extracción de la estructura **implícita** de un documento mediante análisis de su contenido

*Ley Orgánica 2/1980, de 18 de enero, sobre ...
Don Juan Carlos I, Rey de España*

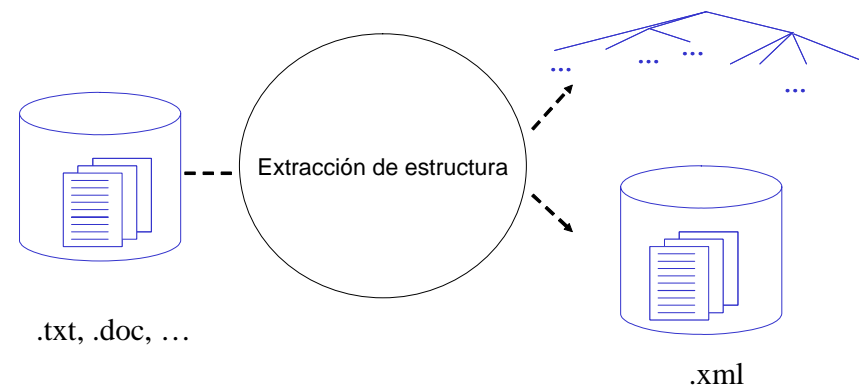
CAPÍTULO I. DEL REFERENDUM Y SUS DISTINTAS MODALIDADES

SECCIÓN PRIMERA. DISPOSICIONES GENERALES

Artículo Primero.
El referendun en sus distintas modalidades, se celebrará de acuerdo con las condiciones y procedimientos regulados en la presente Ley Orgánica.

Artículo Segundo.
...

- Proporciona
 - nodos del grafo de relaciones
 - Colección de documentos estructurados (XML)



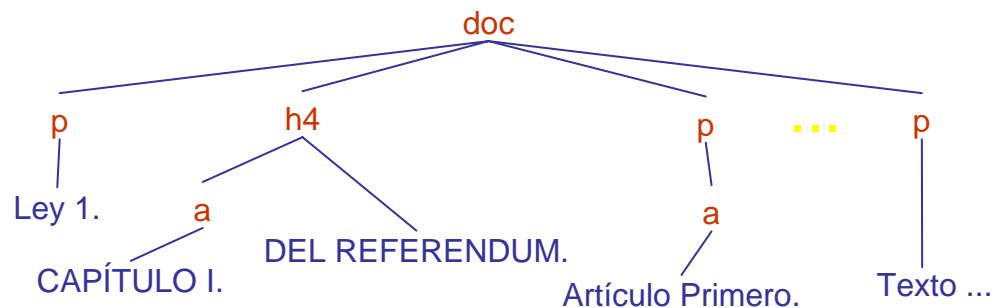
Extracción de estructura: Normalización a XML

■ Análisis del contenido

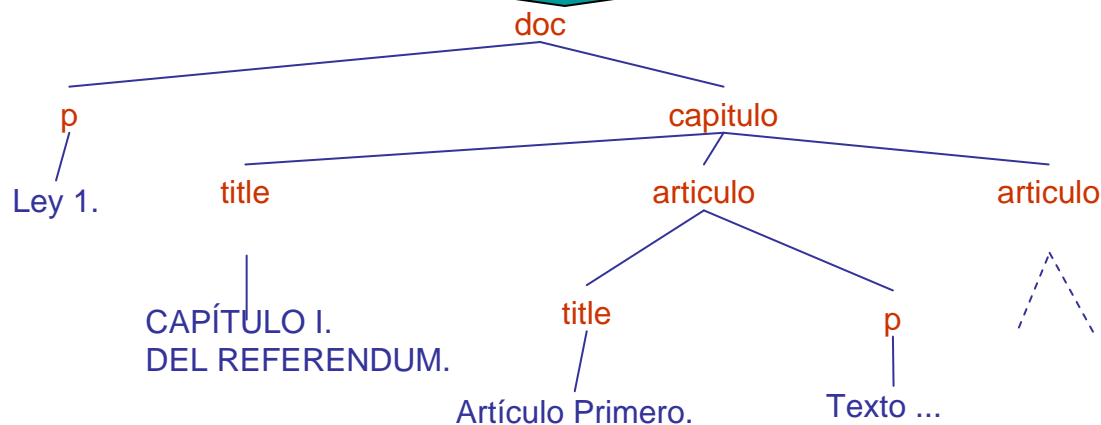
```
<doc>
  <p>Ley 1.</p>
  <h4><a>CAPÍTULO I.</a>
    DEL REFERENDUM.</h4>
  <p><a>Artículo Primero.</a></p>
  <p>Texto del artículo primero.</p>
  <p><a>Artículo Segundo.</a></p>
  <p>Texto del artículo segundo.</p>
</doc>
```

Extracción de estructura

```
<doc>
  <p>Ley 1.</p>
  <capitulo>
    <title>CAPÍTULO I. DELREFERENDUM.</title>
    <articulo><title>Artículo Primero.</title>
      <p>Texto del artículo primero.</p>
    </articulo>
    <articulo><title>Artículo Segundo.</title>
      <p>Texto del artículo segundo.</p>
    </articulo>
  </capitulo>
</doc>
```



Extracción de estructura



Extracción de información con estructura

- **Información que se extrae:**
 - Referencias
 - Otros metadatos
- **Problemas:**
 - Heterogeneidad de esquemas: múltiples DTDs / XML Schema

Heterogeneidad de esquemas

Reglamento (CEE) número 1970 de la Comisión, de 29 de junio, relativo al derecho de los trabajadores a permanecer ...

Según "lex.dtd"

```
<apartado>
...
<p id="art1">
Artículo Primero.
</p>
<p id="art1p1">
Las disposiciones del presente Reglamento serán
aplicables [...] tal como se define en el
<em>
artículo 10 del Reglamento (CEE) número 1612/1968 del
Consejo relativo a la libre circulación de los trabajadores
dentro de la Comunidad.
</em>
</p>
</apartado>
```

Según "norma.dtd"

```
...
<articulo>
<head> Artículo Primero </head>
<p>
Las disposiciones del presente Reglamento serán
aplicables a los nacionales de un Estado miembro que
hayan ejercido una actividad como trabajadores
asalariados en el territorio de otro Estado miembro, así
como a los miembros de su familia, tal como se define en
el artículo 10 del Reglamento (CEE) número 1612/1968
del Consejo relativo a la libre circulación de los
trabajadores dentro de la Comunidad.
</p>
</articulo>
```

Extracción de información con estructura

- **Información que se extrae:**
 - Referencias
 - Metadatos
- **Problemas:**
 - Heterogeneidad de esquemas: múltiples DTDs
- **Distinguimos dos tipos de manipulación:**
 - De la estructura
 - Análisis del documento XML
 - Del contenido
 - Procesos de extracción de información

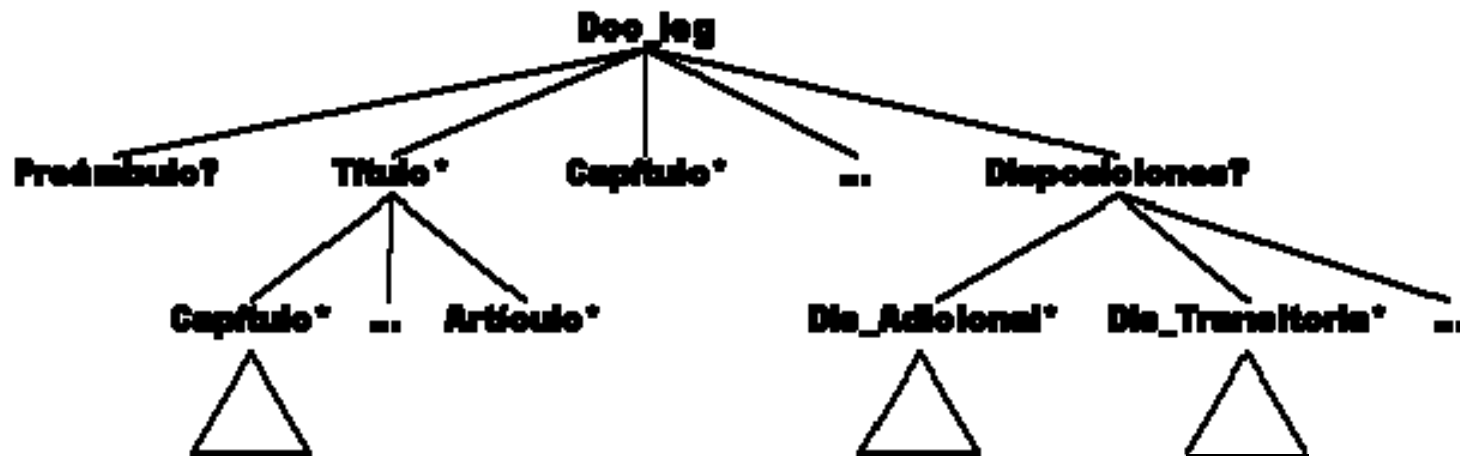
Extracción de información con estructura

- **Distinguimos dos tipos de manipulación:**
 - De la estructura
 - Análisis del documento XML
 - Del contenido
 - Procesos de extracción de información

- **Etapas:**
 1. Transformación a un ESQUEMA PIVOTE
 2. Procesamiento

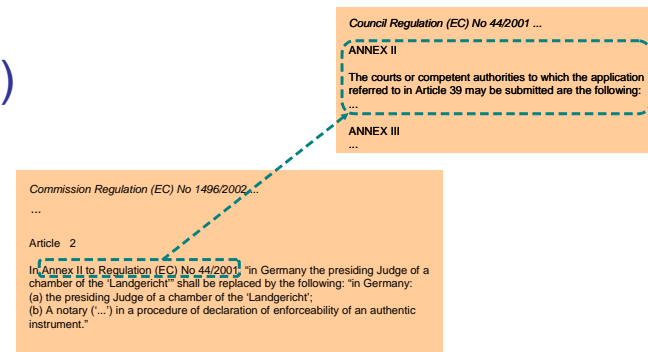
Esquema pivote: reglas de estructura

ESQUEMA PIVOTE = Divisiones + Reglas de inclusión

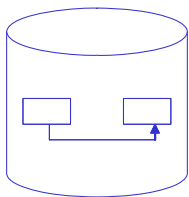


Extracción de referencias

- Extracción de referencias a otros documentos
 - Análisis del contenido
 - Reconocimiento de expresiones (patrones)
 - Guiado por una taxonomía
 - Resolución, a elementos de estructura
 - Particularizado para el entorno jurídico



- Resultado:
 - Colección enlaces (xlink)



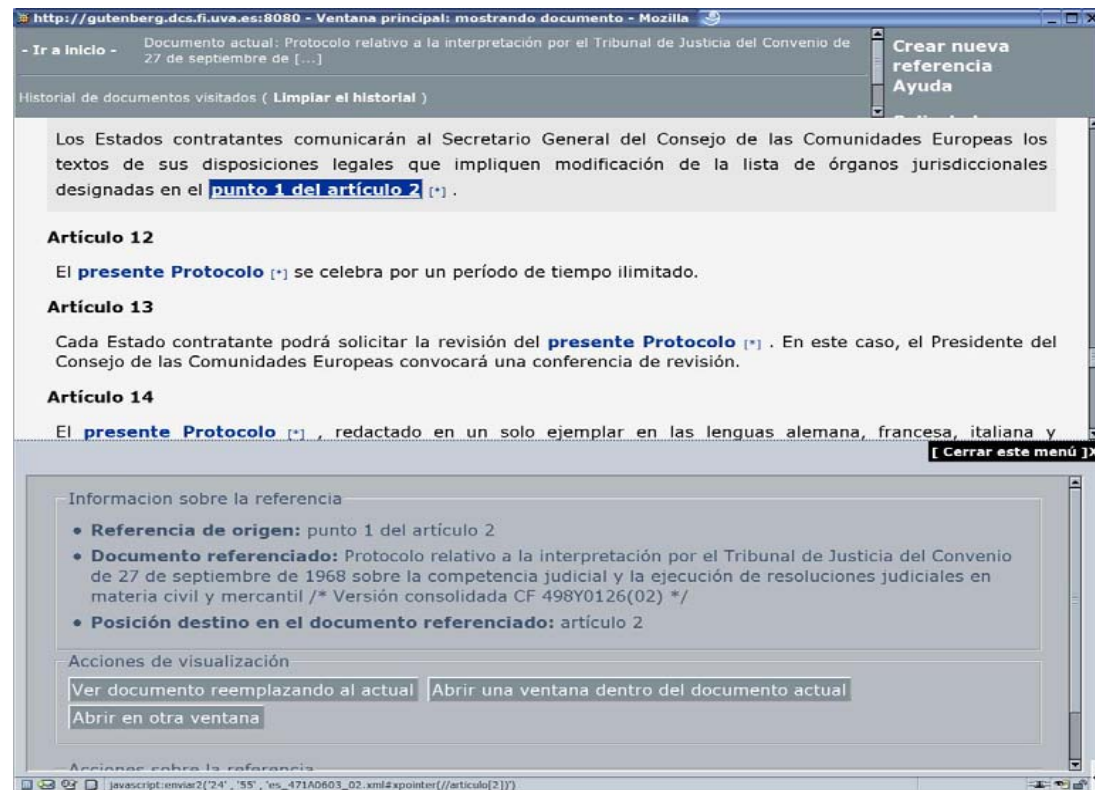
```
<ENLACE>
  <ORIGEN xlink:href="rd775-1997.xml#xpointer(child:disposicion[1]/p[3])"
    xlink:role="substitution"
    date="1997" doctype="norma" />
  <DESTINO xlink:href="rd685-1982.xml#xpointer(child:articulo[72]/apartado[1]/p[2])"
    xlink:role="target"
    date="1982" doctype="norma" />
  <ARCO
    xlink:from="substitution" xlink:to="target"
    xlink:show="undefined" xlink:actuate="undefined" />
</ENLACE>
```

Prototipos: Colecciones

- **Textos normativos españoles**
 - 50 documentos normalizados, que se utilizan en las pruebas y experimentos
- ***Cooperación judicial en materia civil (Legislación Comunitaria)***
 - Aprox. 30 documentos normalizados, usados en el primer prototipo de usuario que manipula esta colección
 - Semilla del proyecto de investigación *Creación de una aplicación de Web Semántica para el tratamiento y manipulación electrónicos de los textos jurídicos de la Unión Europea y el Espacio Económico Europeo en materia de Conflictos Internacionales de Jurisdicción*, VA010B06

Trabajo actual: Herramientas (I)

Manipulación de referencias



The screenshot shows a Mozilla browser window displaying a document from the Gutenberg project. The document is a protocol relative to the interpretation of the Convention of 27 September 1968. The text includes several articles, with Article 12, 13, and 14 visible. A reference manipulation tool is overlaid on the document, showing the following information:

- **Referencia de origen:** punto 1 del artículo 2
- **Documento referenciado:** Protocolo relativo a la interpretación por el Tribunal de Justicia del Convenio de 27 de septiembre de 1968 sobre la competencia judicial y la ejecución de resoluciones judiciales en materia civil y mercantil /* Versión consolidada CF 498Y0126(02) */
- **Posición destino en el documento referenciado:** artículo 2

Below the information, there are actions for visualization:

- Ver documento reemplazando al actual
- Abrir una ventana dentro del documento actual
- Abrir en otra ventana

Trabajo actual: Herramientas (II)

Cooperación judicial en materia civil: 1er prototipo

Reglamento de Bruselas I de la Unión Europea

Archivo Documentos Notas Ayuda

Convenio de Bruselas de 1968 relativo a la competencia judicial y a la ejecución de resoluciones judiciales en materia civil y mercantil. Versión consolidada 1978

documento

- PREÁMBULO [...ÉMBULO]
- TÍTULO I [...ICACIÓN]
- TÍTULO II [...ETENCIA]
- TÍTULO III [...ECUCION]
- TÍTULO IV [...ICIALES]
- TÍTULO V [...NERALES]
- TÍTULO VI [...ITORIAS]
- TÍTULO VII [...NVENIOS]
- TÍTULO VIII [...FINALES]

Convenio de Bruselas de 1968 relativo a la competencia judicial y a la ejecución de resoluciones judiciales en materia civil y mercantil (Versión consolidada 1978).

[ver nota](#)

PREÁMBULO

LAS ALTAS PARTES CONTRATANTES DEL TRATADO CONSTITUTIVO DE LA COMUNIDAD ECONOMICA EUROPEA ,

Deseando aplicar las disposiciones del artículo 220 de dicho Tratado , en virtud del cual se comprometían a la simplificación de las formalidades a que están sometidos el reconocimiento y la ejecución recíprocos de las decisiones judiciales ,

Preocupadas por reforzar en la Comunidad la protección jurídica de las personas establecidas en la misma ,

Considerando que , a este fin , es importante determinar la competencia de sus jurisdicciones en el orden internacional , facilitar el reconocimiento y establecer un procedimiento rápido con el fin de asegurar la ejecución de las resoluciones , de los documentos auténticos y de las transacciones judiciales ,

Han decidido celebrar el presente Convenio y han designado con tal fin como plenipotenciarios :

SU MAJESTAD EL REY DE LOS BELGAS :

al señor Pierre HARMEL ,

Ministro de Asuntos Exteriores ;

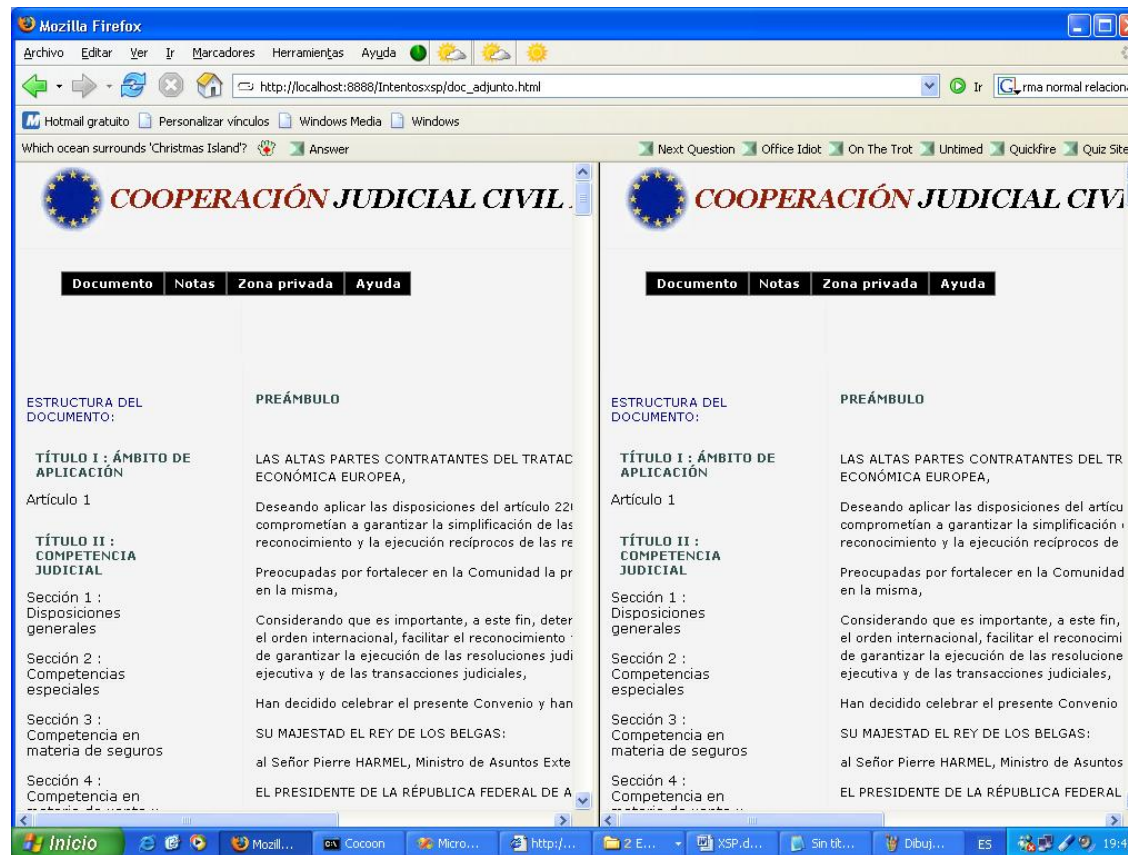
EL PRESIDENTE DE LA REPUBLICA FEDERAL DE ALEMANIA :

al señor Willy BRANDT ,

Abierto: Convenio de Bruselas de 1968 relativo a la competencia judicial y a la ejecución de resoluciones judiciales en materia civil y mercantil. Versión consolidada 1978

Trabajo actual: Herramientas (III)

Cooperación judicial en materia civil: 2o prototipo



Trabajo en curso

- Ampliación de la colección sobre Cooperación Internacional en material civil
- Mejoras en la extracción de referencias
- Recuperación de información estructurada
- Búsquedas semánticas
- Nuevos prototipos

Estructura, semántica, extracción de información y XML legislativo: experiencias en la U. de Valladolid



Universidad de Valladolid

**M. Mercedes Martínez, Dámaso-Javier Vicente, Pablo de la Fuente, Joaquín Adiego,
Alma M. Pisabarro, Jose Manuel Sánchez**
(U. de Valladolid, España)



15-16 de noviembre de 2007