

Metadatos y ontologías para el tratamiento de textos legislativos: la experiencia del grupo DELi-EnergiaTech

Joseba Abaitua



I Simposio de XML Legislativo
Universidad de Valladolid, 15-16 de noviembre de 2007

Metadatos y ontologías (i)

- 1995-1999 Boletines oficiales bilingües
 - DELi se constituye en 1997
 - Traducción asistida por ordenador (CAT)
 - Tratamiento de corpus bilingües (TEI/SGML)
- 2000-2005 Metadatos y ontologías
 - Protocolo OAI-PMH
 - RDF, OWL
- 2005-2007 KnowBoard: buscador semántico
 - Se crea un spinoff: EmergiaTech
 - Alianza con THERA - Centre de Llenguatges i Computació

- Proyecto Legebiduna (1995-99)
 - Creación de corpus (manual)
 - Etiquetado TEI/SGML
 - Estudio estructural
 - 900 documentos, 22 géneros
 - Memoria de traducción TMX/XML
 - Tesis doctorales:
 - Martínez 2000
 - Casillas 2000

- Raquel Martínez (2000)
 - *Alineación automática de corpus paralelos: una propuesta metodológica y su aplicación a un dominio de especialidad*
- Arantza Casillas (2000)
 - *Explotación de corpus paralelos alineados para el desarrollo de entornos de composición de documentos estructurados bilingües*

Metadatos y ontologías: proyectos DELi

- 2000-01 Xtra-Bi: Extracción automática de unidades bitextuales para memorias de traducción
- 2001-02 XML-Bi: Procedimientos para la gestión del flujo documental multilingüe sobre XML/TEI-P3
- 2003-04 X-Flow: Gestión de flujo de contenidos multilingües sobre XLIFF y TMX
- 2003-05 OAC-onto: Open Archives Cataloger, metadatos y ontologías (OAC-onto)
- 2005-06 SemB-UDDI: Semantic UDDI register of Basque high-tech companies

Tipos de metadatos

- Text Exchange Initiative (**TEI**)
- Translation Memory eXchange (**TMX**)
- XML Localization Interchange Format (**XLIF**)
- Darwin Information Typing Architecture (**DITA**)
- Open Archive Initiative-Protocol for Metadata Harvesting (**OAI-PMH**)
- Dublin Core Metadata Initiative (**DCMI**)
- Resource Description Format (**RDF**)
- Web Ontology Language (**OWL**)

Douglas Biber. 1989. A Typology of English Texts. *Linguistics* 27: 3-43.

```
<bibtex:entry bibtex:id="Biber89">  
  <bibtex:article>  
    <bibtex:author>Biber, Douglas </bibtex:author>  
    <bibtex:title> Typology of English Texts </bibtex:title>  
    <bibtex:journal> Linguistics </bibtex:journal>  
    <bibtex:year> 1989 </bibtex:year>  
    <bibtex:number> 27 </bibtex:volume>  
    <bibtex:pages> 3-43 </bibtex:pages>  
    <bibtex:url>http://www.nau.edu/english/ling/faculty/biber  
    </bibtex:url>  
  </bibtex:article>  
</bibtex:entry>
```

- Metadatos y metacontenidos
- Procesamiento lingüístico
- Identificación de palabras clave (frecuencia y relevancia)
- Correspondencia con hispaWordNet
- Red conceptual > Ontología
- Consulta, acceso, selección semántica
- **DEMO** (Iberlex, Boletines)

- DELi <http://www.deli.deusto.es>
- EmergiaTech <http://www.emergiatech.com>
- Thera <http://www.thera-clic.com>