



Thesauro multilingüe para la  
indización de contenidos en un  
sitio web:

El caso **icSemantic** en el gestor  
de contenidos (CMS) Plone

Gonzalo J. Auza  
Ana Sanllorenti  
Cristian S. Rocha



# Presentación de icSemantic / Platecom

- **icSemantic** es una de las aplicaciones de la plataforma **Platecom**.
- **Platecom** es una **plataforma** con diversas funcionalidades para la gestión de **sitios web** con contenidos en **diversos idiomas**.

# Presentación de icSemantic / Platecom

- Platecom es un proyecto de desarrollo en cooperación entre **Argentina y España**, bajo el programa **Iberoeka**.
- Tecnología de **código abierto**.
- CMS (Content Management System) **Plone**.

# Problemas de un sitio web multilingüe

- Un **único sitio** web multilingüe en lugar de un sitio con **varias** versiones (traducciones a varios idiomas).
- **Convivencia** de los idiomas en lugar de alternancia.

# Problemas de un sitio web multilingüe

- Recuperación de información en un entorno multilingüe.
- Necesidad de **indización** multilingüe.
  - **Metadatos** multilingües.
    - **Conceptos** → **Semántica**.

# Problemas de un sitio web multilingüe

- 2 vías para la web semántica
- ¿Indización profesional  
vs.  
etiquetado social?

# Los tesauros y el entorno Web multilingüe

- Mediados década 60
  - Tesauros para la **descripción** temática de documentos y **recuperación** de información.
- Actualidad
  - Tesauros, ontologías, clasificaciones para la **sistematización** de la información (falta de estructura y heterogeneidad de la información en la Web).
  - **Reutilización e interoperabilidad**

# Investigación en portales de información científica

- Science Direct, Wiley, Wilson
  - 97 documentos → 47  
(thesaurus or thesauri or ontology or ontologies or classification or classifications or taxonomy or taxonomies) and (multilingual or bilingual or translation or translating)

# Investigación en portales de información científica

<b>Líneas de investigación y desarrollo</b>	<b>#</b>	<b>%</b>
Construcción de tesauros multilingües (automática, asistida o manual) para recuperación de información o para traducción	17	36
Construcción de tesauros y ontologías en conjunto a partir de otras herramientas multilingües	1	2
Enriquecimiento de tesauros en ambientes multilingües	2	4
Uso de ontologías para recuperación multilingüe o para traducción	11	23
Uso de otras herramientas de representación del conocimiento en ambientes multilingües para recuperación de información o para traducción	2	4
Uso de tesauros multilingües para recuperación de información, clasificación o indización automática	14	30

# Investigación en portales de información científica

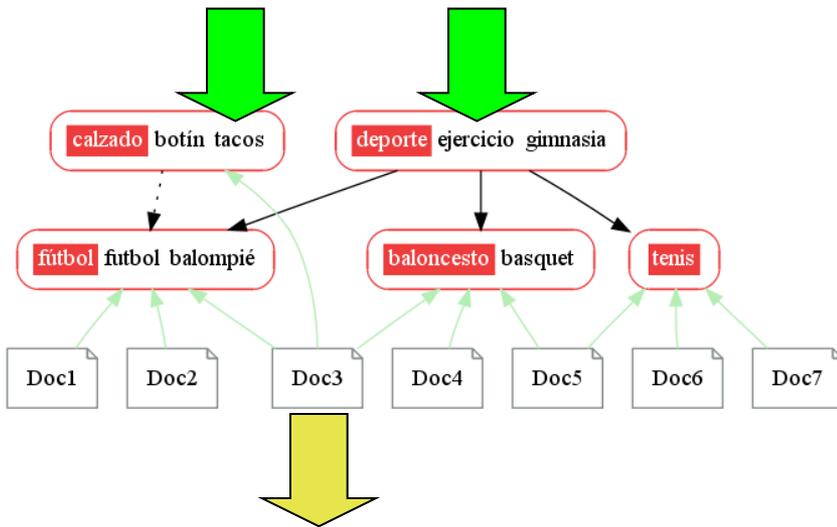
- Tendencias en el conjunto de documentos relevado:
  - Tesoros, clasificaciones, ontologías, etc. (aisladas o combinadas) son usadas para **crear o mejorar** otras herramientas.
  - 70 % → Tesoros
  - **Reutilización, interoperabilidad** entre tesoros, ontologías, clasificaciones, taxonomías, diccionarios.
  - **Finalidad:** **recuperación de información, indización asistida o automática, traducción asistida o automática.**

# Investigación en portales de información científica

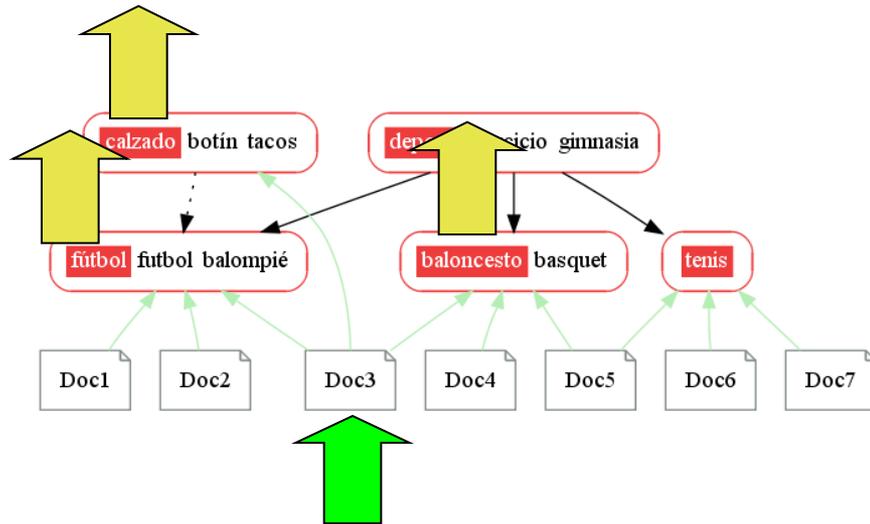
- Abordajes:
  - lingüística computacional / estadística
  - semántico / conceptual (Web semántica)
    - Experiencias de uso o construcción de herramientas de representación del conocimiento o de recuperación de información en particular o para entornos o sistemas particulares.
    - En este conjunto no se encontraron experiencias en el desarrollo de aplicaciones de software para la construcción de tesauros en ambientes multilingües.

# Uso tradicional

- **Búsqueda**  
¿Cuál documento... ?
- **Indizar**  
¿Qué términos...?
- **Vocabulario**  
¿Qué término...?



# Uso tradicional



- **Búsqueda**  
¿Cuál documento... ?
- **Indizar**  
¿Qué términos...?
- **Vocabulario**  
¿Qué término...?



# Uso no tradicional

- Reducir distancias  
Documentos, Usuarios...

- Asistencia a Formularios  
Inscripción, Consultas...

- Miroslav Klose es uno de los grandes delanteros del mundo y un símbolo en la historia reciente del **balompié** en Alemania.
- En lo que a **fútbol** respecta, Tevez se coronó campeón del torneo preolímpico disputado durante enero en Chile, para luego guiar a Boca Juniors hasta la final de la Copa Libertadores de América y la obtención de la Copa Sudamericana.
- El brasileño Luciano Emilio, delantero del DC United, se adjudicó este lunes el **Botín** de Oro 2007 del campeonato de Estados Unidos (MLS), concedido al goleador de la Liga durante la temporada regular.

# Uso no tradicional

- Reducir distancias  
Documentos, Usuarios...
- Asistencia a Formularios  
Inscripción, Consultas...

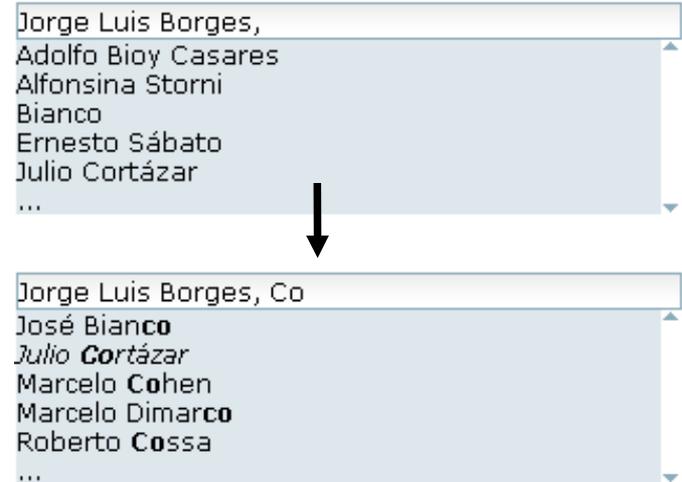
## Búsqueda Vertical



# Uso no tradicional

- Reducir distancias  
Documentos, Usuarios...
- Asistencia a Formularios  
Inscripción, Consultas...

## Búsqueda Horizontal



# Multilingüismo

- No-Simetría
- Equivalentes, Similares y Distintos
- Conceptos
- Construcción

<i>English</i>	<i>French</i>
Classroom environment USE LEARNING ENVIRONMENT	
LEARNING ENVIRONMENT UF Classroom environment UF Educational environment	MILIEU EDUCATIF EP Conditions d'apprentissage EP Milieu pédagogique EP Milieu d'apprentissage

# Multilingüismo

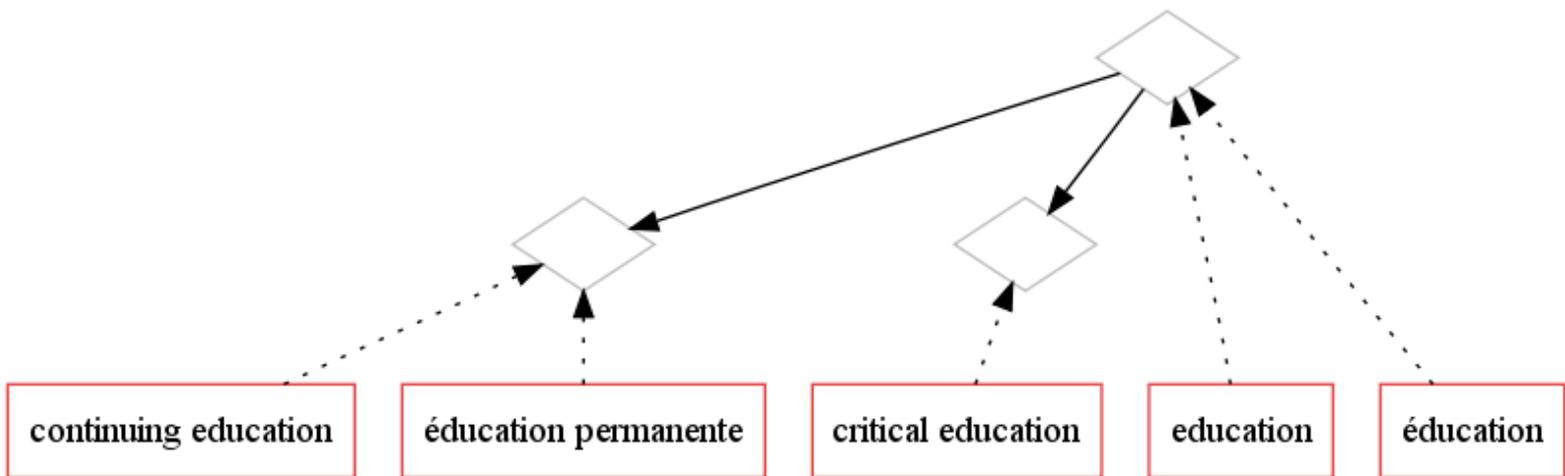
- No-Simetría
- Equivalentes, Similares y Distintos
- Conceptos
- Construcción

<i>English</i>	<i>Spanish</i>	<i>French</i>
historic settlements	= asentamientos históricos	≈ site de peuplement

<i>English (source)</i>	<i>French (target)</i>
EDUCATION	ÉDUCATION ENSEIGNEMENT FORMATION
TEACHING	ENSEIGNEMENT
INSTRUCTION	ENSEIGNEMENT INSTRUCTION

# Multilingüismo

- No-Simetría
- Equivalentes, Similares y Distintos
- **Conceptos**
- Construcción



# Multilingüismo

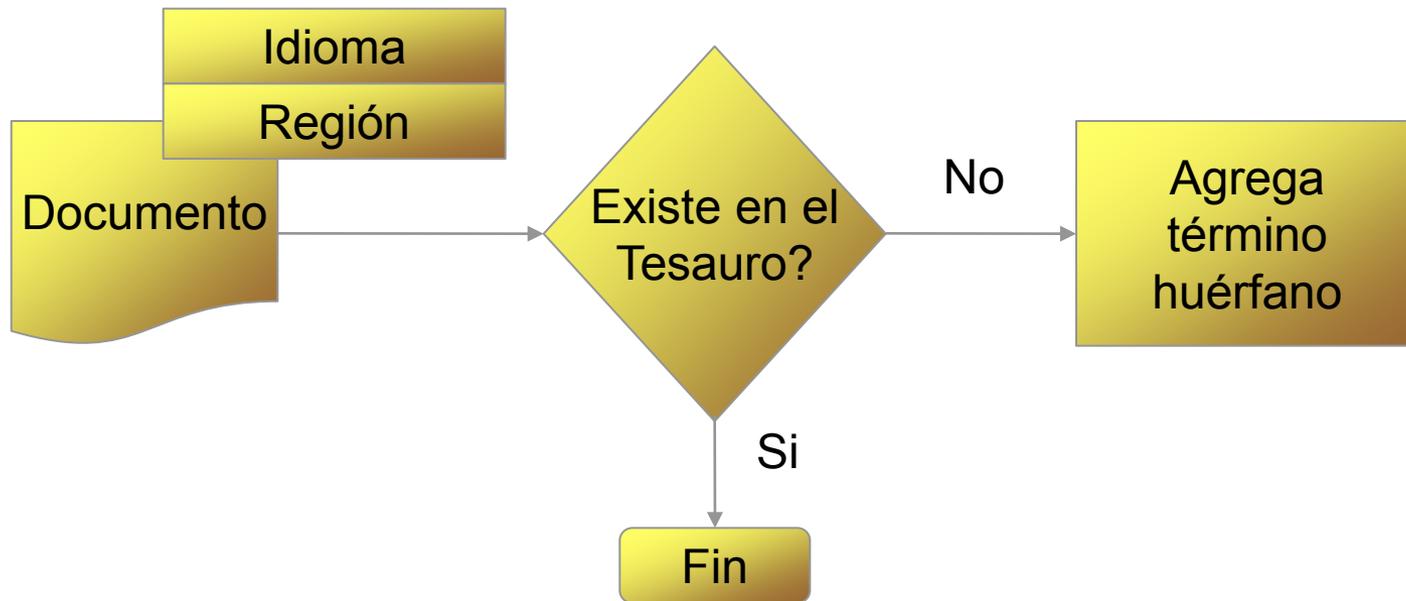
- Conceptos
- No-Simetría
- Equivalentes, Similares y Distintos
- **Construcción**
  - **Utilización de relaciones complejas para distinguir conceptos.**

# Implementación

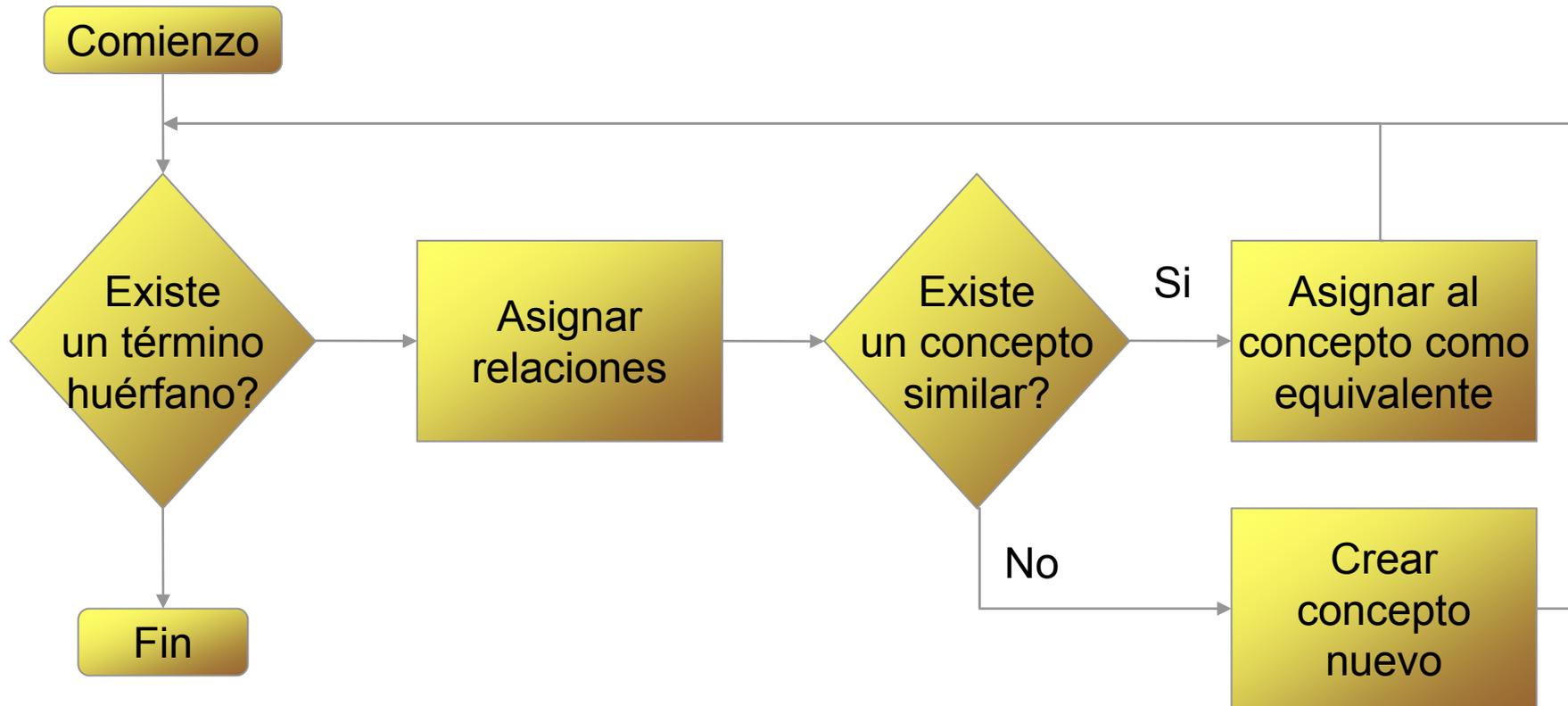
- Red de relaciones de conceptos.
- Conceptos: Términos + Relaciones.
- Término: Palabra ISO-639 ISO-3166 (ej: dulce de leche / es\_AR)
- Optimización: Búsqueda, Consultas.
- Lista de Términos.



# Vida de un término (Rol de Escritor)



# Vida de un término (Rol de Administrador)



## Relaciones básicas de un tesoro.

- Términos equivalentes (USE/UP),
- Términos relacionados (RT),
- Términos más amplios (BT),
- Términos más específicos (NT),

## Relaciones extendidas en platecom.

- Términos ocultos (HT) – SKOS.Core
- Términos distintos (DT) – Platecom
- Términos similares (ST)

# Trabajos a Futuro

- Implementar nuevas relaciones.
- Mejorar la interface de administración.  
(Notas, Comentarios, Historia, etc.)
- Asociar Términos con etiquetas gramaticales.
- ?

# Muchas Gracias



**[www.platecom.com](http://www.platecom.com)**

**[www.platecom.com](http://www.platecom.com)**

**Puede descargarlo y usarlo,  
aceptamos sugerencias**

## Relaciones propuestas

- Términos componente (COT)
- Términos compuestos (CUT)
- Términos elementos (PT)
- Términos componenetes Términos

