

The background of the slide features a close-up, artistic shot of a glass filled with orange juice. A white straw is visible on the right side, partially submerged in the liquid. The lighting is soft, creating a warm, golden glow on the surface of the juice and the rim of the glass. The overall aesthetic is clean and professional.

Explotación de repositorios OAI a través de la armonización de vocabularios controlados

Diego Ferreyra

2do Encuentro Nacional de Catalogadores

Biblioteca Nacional. Argentina - Noviembre / 2009

El modelo OAI-PMH

Reunión en Santa Fe (Nuevo México, 1999) orientada a aumentar el impacto y visibilidad de los eprints.

Se desarrolló y formalizó una arquitectura compuesta por un oferente de metadatos (servidor OAI-PMH) y un cliente

Conjunto simple y acotado de comandos, puede acceder masivamente a metadatos estructurados a través de XML utilizando Dublin Core no cualificado.

Modelo OAI

- Utilización de URI/URL
 - Recibir consultas
 - Identificar el servicio y recursos
- Utilización de XML
 - Representar respuestas del protocolo
 - Estructurar metadatos (Dublin Core)

Consolidación y adopción como estrategia para la exposición y difusión de recursos y modelos federados de explotación cooperativa, ya sea en comunidades especializadas, iniciativas gubernamentales, instituciones y como una capacidad estándar de un gran número de plataformas de gestión de recursos basadas en metadatos.

Tipos de repositorios:

- **Repositorios agregadores**
servicios basados en el cosechado de otros repositorios.
- **Repositorios disciplinares**
repositorios especializados según actividades, tópicos, disciplinas, etc.
- **Repositorios institucionales**
repositorios dedicados a garantizar la disponibilidad de recursos en base a producción o políticas de contenidos específicas de una institución.
- **Repositorios gubernamentales**
repositorios dedicados a garantizar la disponibilidad de recursos en base a lineamientos gubernamentales.

Algunos problemas

Variabilidad en la utilización del esquema Dublin Core:

En un esquema como el Dublin Core, *todo es optativo, todo es extensible, todo es modificable*, esto comporta un alto grado de variabilidad, aun en contextos de alta normalización local, ya que se trata en realidad de la integración de más de un repositorio provenientes de diferentes comunidades y lenguas que aún disponiendo de políticas de normalización para el llenado de metadatos, no tienen por qué ser homogéneas.

Reproducción de problemas de normalización de origen:

Cada repositorio implica asimismo diferentes grados de cumplimiento de las mencionadas políticas de llenado de metadatos. Más allá de que pudieran adoptar tal o cual pauta normativa o tal o cual vocabulario controlado, cada repositorio cumple y controla estas pautas de diferente forma. Los metadatos pueden ser llenados con diversos criterios, políticas y por distintas personas en diferentes momentos en un mismo repositorio.

Algunos problemas

Dificultad para implementar modelos locales:

Los metadatos obtenidos en los distintos repositorios rara vez responden a los modelos de dominio locales o a las prácticas lingüísticas del repositorio agregador.

En este sentido, el repositorio agregador representa un modelo de mediación emergente a partir de la agregación de los diferentes modelos de mediación implementados por cada uno de los repositorios agregados.

Un posible vía de solución

- **Procesos de armonización de vocabularios controlados**

Se consideró que si acaso es posible utilizar crosswalk para resolver la portabilidad entre esquemas de metadatos, la armonización de vocabularios podría considerarse y utilizarse un modelo de solución para lograr portabilidad de contenidos entre diferentes comunidades de interpretación.

Pasos a seguir

1. Selección y determinación de repositorios a cosechar.
2. Consolidación del vocabulario local del agregador
3. Establecer mapeos terminológicos entre los vocabularios de los repositorios y el vocabulario del repositorio de agregación.

Repositorios seleccionados

E-prints in Librarianship, Information Science (E-lis):

Surge en base a dos iniciativas RePEc y DoIS en el año 2003.

Dispone de una red de editores de 45 países de todo el mundo que participan voluntariamente.

- Utiliza una taxonomía en inglés denominada JITA Classification System of Library and Information Science.
- <http://eprints.rclis.org/>



Repositorios seleccionados

Digital Library of Information Science and Technology (DLIST):

inicia sus actividades en el año 2002 y se propone establecer un archivo abierto dedicado a las ciencias de la información. Es mantenido por la School of Information Resources and Library Science and Learning Technologies Center de la University of Arizona.

- Utiliza una lista de términos en inglés.
- <http://dlist.sir.arizona.edu/>



Repositorios seleccionados

Temaria:

Es sostenido por el Grup de recerca Organització i Recuperació de Continguts Digitals de la Facultat de Biblioteconomia i Documentació (Universitat de Barcelona) desde el año 2005 y se dedica a describir y ofrecer acceso a artículos de revistas científicas españolas de Información y Documentación.

- Utiliza el Tesauro de Biblioteconomía y Documentación elaborado por el CINDOC en español.
- <http://www.temaria.net/>



Temaria
revistas digitales de biblioteconomía y documentación

Tesauro

- ▶ **Biblioteconomía y documentación (cedido por el CINDOC)**
- ▶ Otras disciplinas
- ▶ Nombres de lugar
- ▶ Nombres de persona
- ▶ Nombres de entidad

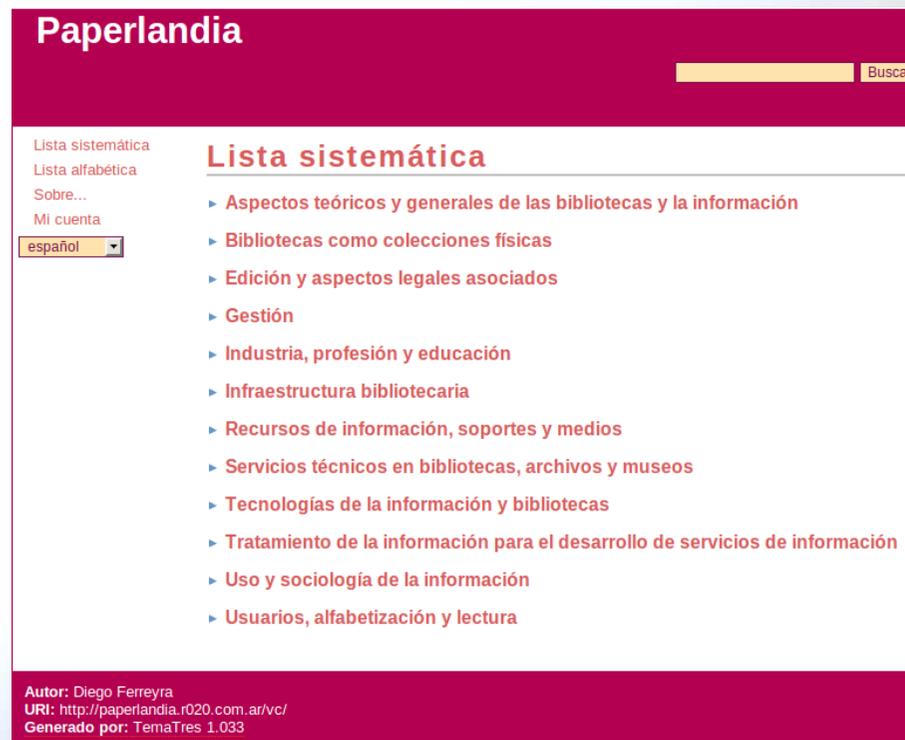
- 📁 Biblioteconomía y documentación
 - 📁 Archivística (158) TR
 - 📁 Biblioteconomía (609) TR
 - 📁 Ciencias y técnicas auxiliares (732)
 - 📁 Estudios métricos de la información (155) TR
 - 📁 Fuentes de información (864) TR
 - 📁 Lenguajes y lingüística (140)
 - 📁 Museología (37) TR
 - 📁 Proceso documental (322) TR
 - 📁 Profesionales de la información y usuarios (685)
 - 📁 Sociedad de la información (317) TR
 - 📁 Tecnologías de la información y las comunicaciones (999) TR
 - 📁 Unidades de información (881) TR

Vocabulario local

Se elaboró un vocabulario basado en una traducción al español del JITA Classification System of Library and Information Science.

Al vocabulario resultante se le agregaron relaciones asociativas y equivalencias.

<http://paperlandia.r020.com.ar/vocabularios/bibliotecologia/>



Paperlandia

Lista sistemática
Lista alfabética
Sobre...
Mi cuenta
español

Lista sistemática

- ▶ Aspectos teóricos y generales de las bibliotecas y la información
- ▶ Bibliotecas como colecciones físicas
- ▶ Edición y aspectos legales asociados
- ▶ Gestión
- ▶ Industria, profesión y educación
- ▶ Infraestructura bibliotecaria
- ▶ Recursos de información, soportes y medios
- ▶ Servicios técnicos en bibliotecas, archivos y museos
- ▶ Tecnologías de la información y bibliotecas
- ▶ Tratamiento de la información para el desarrollo de servicios de información
- ▶ Uso y sociología de la información
- ▶ Usuarios, alfabetización y lectura

Autor: Diego Ferreyra
URI: <http://paperlandia.r020.com.ar/vc/>
Generado por: TemaTres 1.033

Mapeo terminológico

La formalización de las relaciones de mapeo entre los vocabularios comprendió los siguientes niveles de armonización:

Armonización terminológica:

Eliminar inconsistencias terminológicas entre vocabularios cooperantes como ser variaciones en la puntuación, reglas sintácticas, nombres propios y regionales. Por ejemplo:

- Diseño, desarrollo, implementación y mantenimiento
 - IK. Design, development, implementation and maintenance
 - IK. Design, development, implementation and maintenance.

Mapeo terminológico

Referencias cruzadas:

formalización de relaciones de equivalencia entre vocabularios cooperantes. Se utilizaron las siguientes formas de equivalencia semántica entre conceptos:

- **Equivalencia exacta:** cuando dos conceptos tienen el mismo alcance y significado.
- **Equivalencia parcial:** cuando el alcance de un concepto en un vocabulario se corresponde completamente con el de un concepto de otro vocabulario, pero esto no ocurre a la inversa.
- **Equivalencia inexacta:** cuando el alcance de dos conceptos se solapan entre sí, sin llegar a corresponderse por completo.

Mapeo terminológico

Términos comunes:

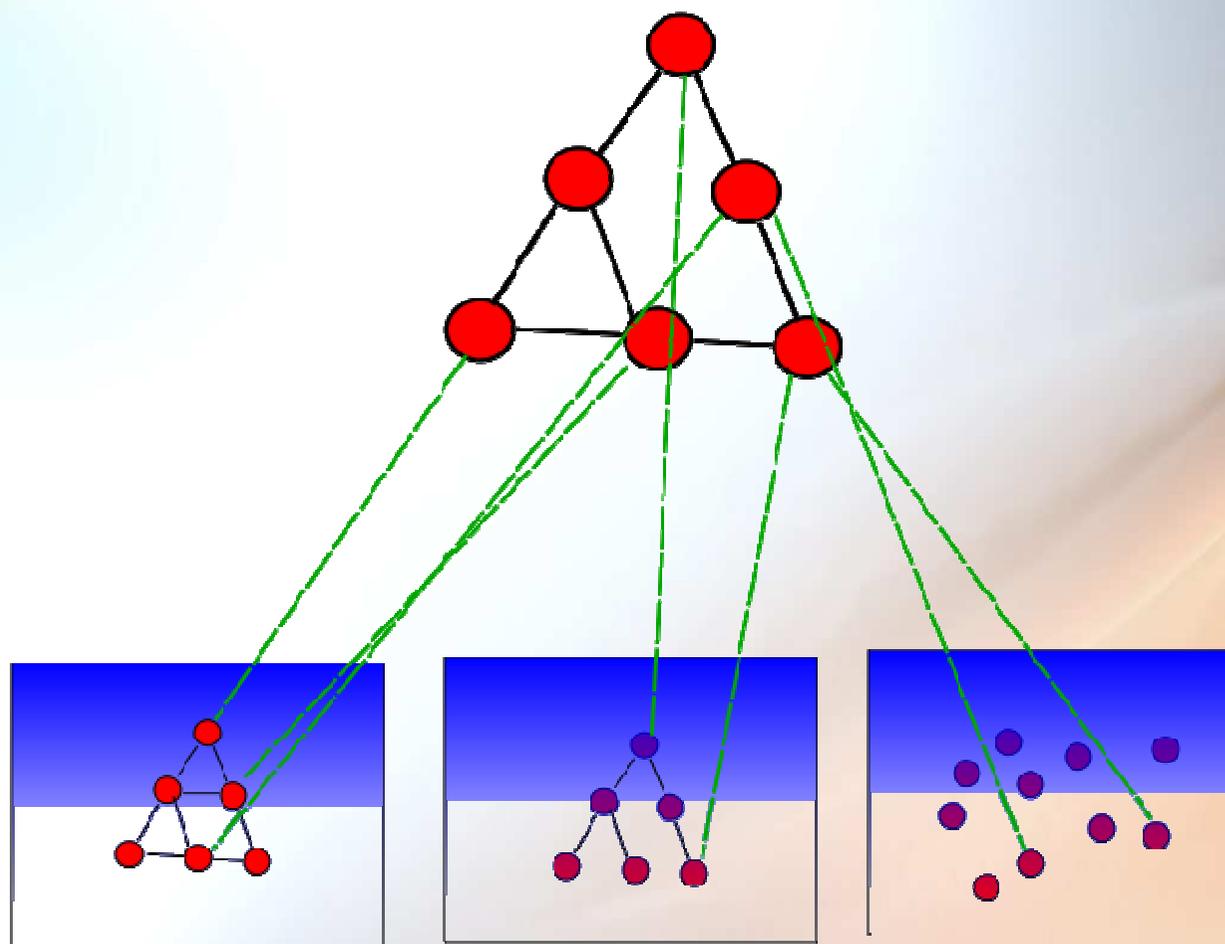
Utilización de términos idénticos para conceptos idénticos. Fue posible utilizar este tipo sólo con el *Tesaurus de Biblioteconomía y Documentación* elaborado por el CINDOC (utilizado por Temaria) ya que es el único en español.

Asociación:

Articulación consistente entre un vocabulario especializado en el contexto de la estructura jerárquica de uno vocabulario más general. Por ejemplo:

- Catalogación, control bibliográfico
 - Catalogación
 - Descripción bibliográfica
 - Directrices
 - Normas

Mapeo terminológico



Mapeo terminológico



- CONTENIDO**
- 1. Aspectos teóricos y generales de las bibliotecas y la información
 - 2. Edición y aspectos legales asociados
 - 3. Gestión
 - 4. Industria, profesión y educación
 - 5. Infraestructura bibliotecaria
 - 6. Procesos de información, prensa y medios
 - 7. Servicios asociados en bibliotecas, archivos y museos
 - 8. Tecnología de la información y la biblioteca
 - 9. Tratamiento de la información para el desarrollo de servicios de información
 - 10. Uso y socialización de la información
 - 11. Usuarios, alta tecnología y futuro

Paperlandia

Lista sistemática

- Aspectos teóricos y generales de las bibliotecas y la información
- Bibliotecas como colecciones físicas
- Edición y aspectos legales asociados
- Gestión
- Industria, profesión y educación
- Infraestructura bibliotecaria
- Procesos de información, prensa y medios
- Servicios asociados en bibliotecas, archivos y museos
- Tecnología de la información y la biblioteca
- Tratamiento de la información para el desarrollo de servicios de información
- Uso y socialización de la información
- Usuarios, alta tecnología y futuro

Juan Diego Sampedro
 URM: Biblioteca de la Universidad de Murcia
 Correo: jds@urmu.es

Buscar:

Aspectos teóricos y generales de las bibliotecas y la información

Bibliotecas y ciencias de la información como campo (30) - Historia de la bibliotecología y las ciencias de la información con afincamientos (371) - Teoría de información y teoría bibliotecológica (111) más...

Bibliotecas como colecciones físicas

Archivos (216) - Bibliotecas académicas (355) - Bibliotecas científicas (40) - Bibliotecas de alta tecnología, bibliotecas de música (113) - Bibliotecas de idiomas (15) - Bibliotecas digitales (174) - Bibliotecas escolares (54) más...

Edición y aspectos legales asociados

Centros (49) - Impresión, publicación electrónica (213) - Libertad intelectual (58) - Medios de comunicación (27) - Propiedad intelectual: derechos de autor, copyright y copyleft (248) - Venta de libros (16) más...

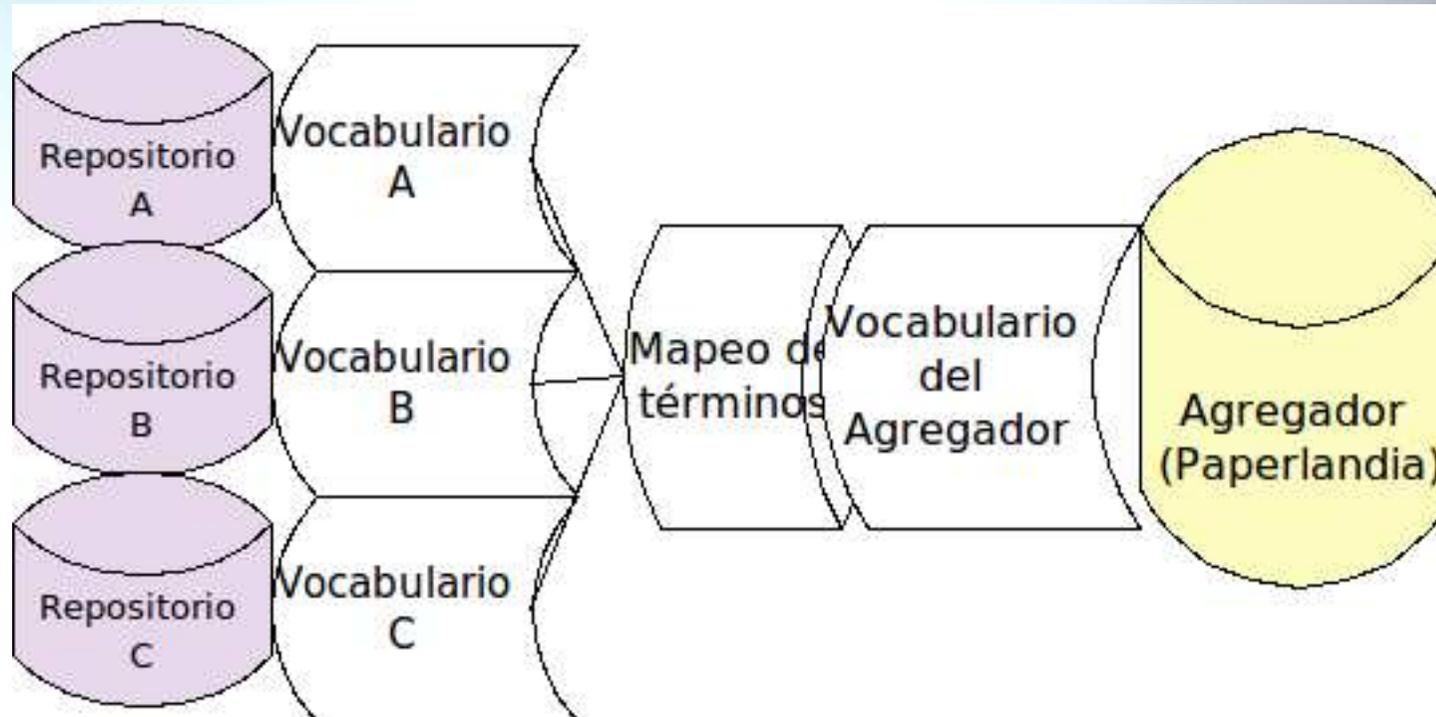
Gestión

Análisis de los libros (2) - Comercialización (11) - Cooperación (22) - Finanzamiento (158) - Recursos (24) - Gestión de personal (66) - Gestión social (16) más...

Industria, profesión y educación

Aspectos de los planes de estudios (160) - Bibliotecarios (64) - Educación (16) - Entrenamiento (12) - Industria de la información (41) - Industria del software (28) - Industria informática y de las telecomunicaciones (17) más...

Mapeo terminológico



Herramientas utilizadas

- PKP Open Archives Harvester :
 - Cosechado de metadatos
- TemaTres:
 - Gestión de vocabularios controlados

Desarrollos específicos:

- Herramienta para la detección de términos: orientada a detectar términos utilizados en un repositorio que no dispongan de un mapeo en el vocabulario del repositorio agregador
- Herramienta para presentación de contenidos: orientada a desarrollar las opciones de navegación, la presentación de metadatos y el esquema de búsqueda temático

Extensión y replicabilidad

Se extendió la misma metodología a otros campos previstos en el esquema Dublin Core e informados por los servicios OAI-PMH.

- *idioma* (language)

- <http://paperlandia.r020.com.ar/vocabularios/lengua/>

Español: es, spa, es-ES

Inglés: en, eng, en-US

- *tipo de recurso* (type)

- <http://paperlandia.r020.com.ar/vocabularios/tiposrecursos/>

- Artículo (impreso): Journal Article (Print/Paginated), Text.Article

Resultados funcionales

Implementación local basada en explotación de metadatos

Buscar:

Aspectos teóricos y generales de las bibliotecas y la información

[Bibliotecas y ciencias de la información como campo \(510\)](#) · [Relación de la bibliotecología y las ciencias de la información con otros campos \(271\)](#) · [Teoría de información y teoría bibliotecológica \(111\)](#) más ...

Bibliotecas como colecciones físicas

[Archivos \(216\)](#) · [Bibliotecas académicas \(555\)](#) · [Bibliotecas científicas \(46\)](#) · [Bibliotecas de la salud, bibliotecas médicas \(112\)](#) · [Bibliotecas del mundo \(15\)](#) · [Bibliotecas digitales \(174\)](#) · [Bibliotecas escolares \(34\)](#) más ...

Edición y aspectos legales asociados

[Censura \(45\)](#) · [Impresión, publicación electrónica \(211\)](#) · [Libertad intelectual \(59\)](#) · [Medios de comunicación \(27\)](#) · [Propiedad intelectual: derechos de autor, copyright y copyleft \(265\)](#) · [Venta de libros \(20\)](#) más ...

Gestión

[Autoridades unitarias \(2\)](#) · [Comercialización \(112\)](#) · [Cooperación \(225\)](#) · [Financiamiento \(18\)](#) · [Finanzas \(26\)](#) · [Gerencia de personal \(66\)](#) · [Gobierno local \(36\)](#) más ...

Industria, profesión y educación

[Aspectos de los planes de estudios \(164\)](#) · [Biografías \(49\)](#) · [Educación \(320\)](#) · [Entrenamiento \(133\)](#) · [Industria de la información \(61\)](#) · [Industria del software \(21\)](#) · [Industria informática y de las telecomunicaciones \(17\)](#) más ...

Resultados funcionales

Navegación facetada

Inicio · Edición y aspectos legales asociados

· idioma > Español : 48

1. [4º Foro de las evidencias electrónicas](#)
2. [Acerca del World Arts Resources](#)
3. [Acimed en el año 2003 : ¡Saltamos!](#)
4. [Acimed: frecuencia mensual + calidad. ¿Imposible?](#)
5. [Actualidad y perspectivas de las publicaciones electrónicas](#)
6. [Alfabetización digital: un peldaño hacia la sociedad de la información](#)
7. [Algunas consideraciones en torno al derecho de autor](#)
8. [Ante el reto actual de acelerar el proceso de publicación](#)
9. [Autoría y cultura posmoderna](#)
10. [Banco de artículos : una metodología para la edición racional de publicaciones periódicas científico-técnicas](#)
11. [Biblioteca electrónica de revistas científicas internacionales : proyecto de consorcio](#)
12. [Biblioteca Universitaria: una década de evolución editorial](#)
13. [Código Deontológico del Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya: ¿cómo? y ¿por qué?](#)
14. [Cómo auditar el cumplimiento de la LORTAD/LOPD/Reglamento de Seguridad = How to control the fulfillment of LORTAD/LOPD/Security regulation](#)
15. [Cómo realizar una comunicación científica. Aspectos lingüísticos \(II\)](#)
16. [Cómo realizar una comunicación científica. Aspectos tipográficos \(III\)](#)
17. [Confidencialidad y derechos de autor en un proyecto de intranet](#)
18. [Convergencia de jurisdicciones sobre propiedad intelectual](#)
19. [Derechos de autor en Internet](#)
20. [Desarrollo histórico del derecho de autor en la traducción](#)
21. [El acceso a la información pública: un acercamiento doctrinal](#)
22. [El control bibliográfico en la Comunidad Valenciana : producción editorial y Bibliografía Nacional](#)
23. [El dilema del editor de una revista biomédica : aceptar o no aceptar](#)
24. [El traductor y el derecho de autor](#)
25. [El tratamiento de los datos personales en Internet](#)
26. [Evaluación y aplicación de las normas de presentación de publicaciones periódicas: revisión bibliográfica](#)

Refinar según tema:

- [Censura](#) (45)
- [Impresión, publicación electrónica](#) (211)
- [Libertad intelectual](#) (59)
- [Medios de comunicación](#) (27)
- [Propiedad intelectual: derechos de autor, copyright y copyleft](#) (265)
- [Venta de libros](#) (20)

Refinar según tipo:

- [Artículo con revisión por pares](#) (39)
- [Artículo en línea \(sin paginar\)](#) (24)
- [Artículo \(impreso\)](#) (13)
- [Boletín/Artículo de revista](#) (13)
- [Artículo sin revisión por pares](#) (9)
- [Paper de conferencia](#) (6)
- [Capítulo de libro](#) (2)
- [Libro](#) (1)
- [Actas de conferencia](#) (1)
- [Otros](#) (1)

Resultados funcionales

Utilización de servicios semánticos del vocabularios controlado



The screenshot shows a search result page with a search bar at the top right containing the text 'busqueda solo por tema'. The main title of the article is 'The Role of Subjective Factors in the Information Search Process'. Below the title is a 'Descripción' section containing a paragraph of text. To the right of the description is a sidebar with sections: 'Temas:' listing 'Estudios de usuarios (275)', 'Interacción hombre-máquina (62)', 'Interfaces de usuario, usabilidad (219)', 'Presentación de información: hipertexto, hipermedia (110)', and 'Recuperación automática de información (187)'; 'Temas relacionados:' listing 'Uso y sociología de la información (611)' and 'Bibliotecas digitales(174)'; and 'Ampliar tema:' listing 'Usuarios, alfabetización y lectura(123)', 'Interfaces de usuario, usabilidad (219)', 'Tratamiento de la información para el desarrollo de servicios de información (450)', and 'Tecnologías de la información y bibliotecas(819)'. At the bottom of the article preview, there are fields for 'Autor' (Gwizdka, Jacek; Lopatovska, Irene), 'Fecha' (2009-01-01), 'Tipo' (Journal Article (Paginated)), and 'Identificador' (http://dlist.sir.arizona.edu/2656/).

The Role of Subjective Factors in the Information Search Process

Descripción

We investigated the role of subjective factors in the information search process. Forty eight participants each conducted six web searches in a controlled setting. We examined relationships between subjective factors (happiness levels, satisfaction with and confidence in the search results, feeling lost during search, familiarity with and interest in the search topic, estimation of task difficulty), and objective factors (search behavior, search outcomes and search task characteristics). Data analysis was conducted using a multivariate statistical test (Canonical Correlations Analysis). The findings confirmed existence of several relationships suggested by prior research, including relationships between objective search task difficulty and the perception of task difficulty; between subjective states and search behaviors and outcomes. One of the original findings suggests that higher happiness levels before the search and during the search correlate with better feelings after the search, but also correlates with worse search outcomes and lower satisfaction, suggesting that, perhaps, it pays off to feel some 'pain' during the search in order to 'gain' quality outcomes.

Autor
Gwizdka, Jacek
Lopatovska, Irene

Fecha
2009-01-01

Tipo
Journal Article (Paginated)

Identificador
<http://dlist.sir.arizona.edu/2656/>

Temas:

- Estudios de usuarios (275)
- Interacción hombre-máquina (62)
- Interfaces de usuario, usabilidad (219)
- Presentación de información: hipertexto, hipermedia (110)
- Recuperación automática de información (187)

Temas relacionados:

- Uso y sociología de la información (611)
- Bibliotecas digitales(174)

Ampliar tema:

- Usuarios, alfabetización y lectura(123)
- Interfaces de usuario, usabilidad (219)
- Tratamiento de la información para el desarrollo de servicios de información (450)
- Tecnologías de la información y bibliotecas(819)

Resultados

Autonomía local e interoperabilidad global:

El mapeo terminológico permitió construir una solución para la normalización de fuentes autónomas de datos en el contexto de redes federadas de cooperación, manteniendo condiciones de autonomía de cada uno de los nodos (repositorios) y el nodo agregador (Paperlandia)

Resultados

Metadatos y normalización:

Los desarrollos realizados para comparar los contenidos de un campo específico en el contexto de un esquema de metadatos y los términos disponibles en un vocabulario determinado resultaron una herramienta útil y fundamental a la hora de detectar de variaciones, frecuencias, regularidades y coincidencias en el contexto de cada repositorio en particular y en el contexto del repositorio agregador. Eventualmente, esta misma metodología podría utilizarse como vía de solución para la detección de problemáticas de normalización y control de autoridades.

Resultados

Interoperabilidad de contenidos:

Fue posible explorar una concepción amplia y compleja de la noción de interoperabilidad más allá de la portabilidad de datos y estructuras. Una noción capaz contemplar los modelos de mediación de cada comunidad de interpretación que ofrece un marco metodológico para la **apropiación de los contenidos según concepciones locales**.

Resultados

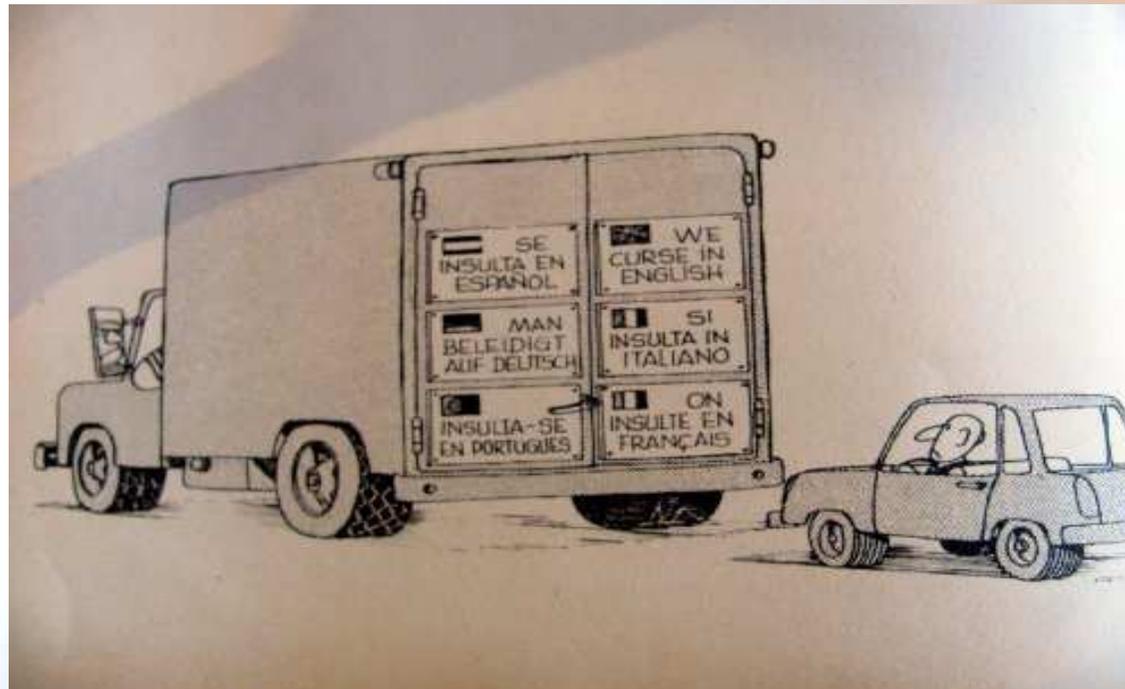
Muchos metatados, poca normalización:

A partir de la implementación real de Paperlandia, fue posible evaluar la viabilidad real del marco de armonización de vocabularios como esquema de solución para repositorios caracterizados por la **alta variabilidad** y el **bajo nivel de normalización** de sus metadatos en escenarios de cooperación abiertos pero delimitados.

Resultados

Universalidad vs. Apropiación:

El modelo de trabajo estuvo orientado a la experimentación con una metodología de apropiación basada en la utilización de los **vocabularios controlados como modelos de mediación** entre lenguas pero también entre comunidades de interpretación. La experiencia se basa en la formalización y visibilidad del modelo de mediación y no sobre su transparencia o invisibilidad.



Los vocabularios controlados como herramientas de representación y apropiación de recursos

