Metadata and XML
Improving the Findability of Information

Peter J. Bogaards (BogieLand.com)
Information Designer & Information Architect
“Sharing knowledge is better than having it.”

EIDC 2004 - Wiesbaden
10 november 2004
Introduction

- Background in instructional design. (1987)
- Design of (tech) facilities to enhance human learning processes.
- Interface, document and information designer. (@Informaat ‘90-'97).
- W3: Electronic documentation and user interface design merger.
Agenda

• Purpose: To paint the landscape
• Findability of information
• XML and metadata
• Subject-based classification: Controlled vocabularies, Thesaurus, Taxonomy, and Ontology
• Faceted classification (XFML)
• Technologies: Topic maps and RDF
• A vision for the future
• ?&!
Findability of Information
Finding anyone or anything from anywhere at anytime
Findability of information

- A wealth of information = a poverty of attention
- Structure versus chaos
- Information architecture: How to organize information in order to let people find things?
- Applying concepts, methods and techniques from Library and Information Science
- How to improve information retrieval?
- Documents are for humans, data is for machines
XML

eXtensible Markup Language
XML: eXtensible Markup Language

- SGML -> HTML/XHTML -> XML
- A language for making <tag> sets.
- Meaningful tags for search and information retrieval.
- Machine understandable information.
- Document structure: XML schema
- Document content: XML name spaces
- Document presentation: XSL(T), SVG et al.
Metadata
Data about data
Metadata: Data about Data, not Code

- Information about objects on subjects - metadata describes objects.
- Purposes: Information management and discovery.
- Metadata enables content to be retrieved, tracked, and assembled automatically.
- Metadata is machine understandable information about (web) resources and is the foundation of all information retrieval.
- Metadata is any statement about an information resource.
- Metadata is a writing skill.
Email document: Attribute value pairs

From: Peter J. Bogaards (pjb@bogieland.com)
To: Michael Fritz (michael@tekom.de)
Date: Nov. 10, 2004

Hi Michael,
How are you?
Best,
Peter
Dublin Core Metadata Initiative: 15 Elements

- Title, Subject/Keywords*, Description
- Creator, Publisher, Contributor
- Date, Type, Format
- Relation, Coverage, Rights
- Source, Language, Identifier

<meta name="DC.Identifier" scheme="URI" content="http://www.informationdesign.org" />

*Meaning in the SUBJECT/KEYWORDS tag, other tags are for document management.

See also: dublincore.org
Controlled vocabularies
Organizing words and phrases
CVs: Organized Words and Phrases

• “… organized lists of words and phrases (…) that are used to initially tag content, and than to find it through navigation or search.” (Amy Warner)

• No CV: multiple terms for identical concepts -> chaos
• Closed list of named subjects, which can be used for classification.
• Creating a common language between user and system.
• A type of metadata that functions as a subset of natural language.
Taxonomy

Carl Linnaeus Goes Digital
Taxonomy: Carl Linnaeus (1700’s) Goes Digital

- A taxonomy is a complex CV
- One type of relation between terms: broader/narrower term in the hierarchy.
- A subject-based classification that arranges the terms in the CV into a hierarchy.
Thesaurus
BT/NT, RT, SN, and USE/UF
Thesaurus: BT/NT, USE/UF, SN and RT

• Extend taxonomies to describe the world better.
• ISO standard 2788 - Properties:
  – BT: Broader term - one level up in the hierarchy
  – NT: Narrow term / Inversed with BT
  – SN: Scope note (Explanation of meaning of the term)
  – RT: Related term (No synonym or BT/NT: ‘See also’)
  – USE: Other term preferred/synonym /Inversed with UF
• To provide a much richer vocabulary for describing the terms than taxonomies do.
Thesaurus: Example (Karl Fast et al.)

Jeans
• BT Pants
• NT Levis
• NT Wranglers
• UF Dungarees
• UF Waist Overalls
• RT Denim
• RT Overalls

Denim
• BT Fabrics
• NT Ring Spun
• NT Dark Indigo
• NT Stonewash
• RT Jeans
Ontology
A Specification of a Conceptualization
Ontology: A Specification of Conceptualization

• Derivate of artificial intelligence (Logical inferencing)
• “… a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)).”
• There is no one correct way to describe a domain.
• A model for describing the world that consists of a set of topics, properties, and relationship types.
• Fixed versus open vocabularies.
Faceted Classification
Analysis and Synthesis
Faceted Classification: The Elephant

Post-Web Information Systems Design

Knowledge Management
Interaction Design
Information Architecture
Experience Design
Information Design
Customer Relationship Management
Faceted Classification: Analysis and Synthesis

- S.R. Ranganathan (1892-1972)
- Facet: ‘a clearly defined, mutually exclusive, and collective exhaustive aspects, properties or characteristics of a class or specific subject.’
- Describing documents from various perspectives.
- A special purpose controlled vocabulary.
eXchangable Faceted Metadata Language

- A language to exchange metadata between websites.
- XFML Core aka XFML 1.0 (Peter van Dijck et al. 2002)
- Categories, subcategories, and faceted metadata.
- Open XML format for publishing and connecting faceted metadata of websites.
- An XFML file contains TOPICS, organized in FACETS.
- Effectively separating navigation from content.

See also: xfml.org
**XFML: Example**

```xml
<?xml version="1.0"?>
<xfml version="1.0" url="http://xfml.org/spec/example.xml" language="en-us">
  <mapInfo>
    <managingEditor>
      <name>Peter Van Dijck</name>
      <email>peter@poorbuthappy.com</email>
      <url>http://petervandijck.net</url>
    </managingEditor>
  </mapInfo>
  <facet id="place_to_go">places to go</facet>
  <!-- TOPICS -->
  <topic id="bogota" facetid="place_to_go" parentTopicid="colombia">
    <name>Bogota</name>
  </topic>
  <page url="http://poorbuthappy.com/colombia/topics.php">
    <title>Guide to Colombia topics page</title>
    <occurrence topicid="diving"/>
    <occurrence topicid="bogota"/>
  </page>
</xfml>
```
Topic Maps

The GPS of the Information Universe
Topic Maps: The GPS of the Info Universe

• “The purpose of a topic map is to convey knowledge about resources through a superimposed layer, or map, of the resources. A topic map captures the subjects of which resources speak, and the relationships between subjects, in a way that is implementation-independent.”

• A model to describe knowledge structures.
• A topic map is a data structure.
• Key concepts: (typed) Topics, Associations, and Occurences.
• Topic Maps can represents controlled vocabularies, taxonomies, thesauri, and faceted classification.
• XML Topic Maps 1.0 (valid XML)

See also: topicmaps.org
Topic Map: Example (Garshol)
RDF

Resource Description Framework
RDF: Resource Description Framework

- Alternative to Topic Maps.
- R.V. Guha @ Apple Meta Content Framework.
- A framework for representing information on the Web.
- XML app -> W3C Recommendation (1999) for the expression of any kind of target.
- Key concepts: Resources, Properties, and Statements.
The Semantic Web
A Vision of the Future
SemWeb: A Vision of the Future

• “The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” - TBL

• Vision: Well-defined data on the Web that can be used by machines for automation, integration and re-use.

• The Web can reach its full potential: data to be shared and processed by automatic tools.

• Based upon RDF and the Web Ontology Language

See also: semanticweb.org
Discussion

? & !
BogieLand
information design & information architecture
http://www.bogieland.com

Peter J. Bogaards
pjb@bogieland.com