Quick Semantic Collaborative Authoring and Classifying on the Web

David Aumüller
Motive: Organizing Information

• Example: Scientific Papers, Biblio.
  – Author(s), title, year, venue
  – Keywords
  – Abstract/full text
  – Citation count

• Annotations
  – Comments, reviews
  – Categories, see-also’s
Collaboration – What’s possible

• Dirty data/metadata – Community to ...
  ... help clean the data (Collab. Curation)
  ... classify documents
  ... annotate, rate, review, relate
  ... update

• Interface/platform
  – Wiki-like – quick, that is.
  – Easy to fix errors, not hard to make them
Agenda

• Accessible Classification Systems
  – Searching
  – Browsing

• Quick Authoring/Classifying
  – Semantic Wiki
    • Example/Demo
    • Challenges

• What’s ahead?
Making Information Accessible

• Full Text Search
• Authority List
  – Simple controlled vocabulary
• Taxonomy
  – Hierarchical controlled vocabulary
• Faceted Classification
  – Multiple taxonomies
  – Each item can live in multiple slots

Cf. [Otwell02]
Full Text Search

Pros:
- Easy to implement, simple form accessing database
- Most words are indexed, great search range, high search specificity

Cons:
- Data is not organized
- Requires content authors to include all relevant terms
- Word count frequency not a good determinant of relevance/quality
- Easy to get too many or no results – patience needed
- Any effort, time, and cost, saved during database production are passed on to the searcher
Authority List (Simple C. V.)

Pros:
• Controlled access searching supports efficient retrieval
• Spelling errors, acronyms, other common variants mapped to preferred terms
• Input data quality improved by including known keywords
• Addresses some usability concerns: error tolerance, showing context for search results

Cons:
• Relationships between concepts/content items hard to manage
• Keyword list must be manually maintained
• May not scale easily
**Taxonomy (hierarchical C. V.)**

**Pros:**
- Relationships between concepts, made explicit
- Domain knowledge clarified for all users
- Similar to known reference tools: encyclopaedias, phone books, library catalogs

**Cons:**
- Increased maintenance complexity, editor needed
- Strict hierarchy of concepts may not be flexible or clear enough for all uses
- Taxonomy cannot be automatically generated

Adding hierarchical structure to the c.v. list improves site navigation.
Faceted Classification

• Facets = multiple attributes
• Facet values = attribute values
  – Can be hierarchical, e.g.:
    • Location: Europe > Germany > Leipzig
• Faceted navigation/browsing/searching
  – Find items that match multiple criteria:
    • Location: England
    • Era: 19th century
    • Profession: Literary author
<table>
<thead>
<tr>
<th>Concepts</th>
<th>Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Access (1)</strong></td>
<td><strong>17 &amp; 18th c (1400)</strong></td>
</tr>
<tr>
<td><strong>Barrier free design (4)</strong></td>
<td><strong>19th century (2901)</strong></td>
</tr>
<tr>
<td><strong>Circulation (2)</strong></td>
<td><strong>20th century (7557)</strong></td>
</tr>
<tr>
<td><strong>Cultural (145)</strong></td>
<td><strong>Islamic-hegira, 622 ce (1391)</strong></td>
</tr>
<tr>
<td><strong>Economic (17)</strong></td>
<td><strong>Modern (6054)</strong></td>
</tr>
<tr>
<td><strong>Environmental (471)</strong></td>
<td><strong>People</strong></td>
</tr>
<tr>
<td><strong>Housing (1188)</strong></td>
<td><strong>Agency (245)</strong></td>
</tr>
<tr>
<td><strong>Locations</strong></td>
<td><strong>Architect (16258)</strong></td>
</tr>
<tr>
<td><strong>Africa (140)</strong></td>
<td><strong>Artist (1773)</strong></td>
</tr>
<tr>
<td><strong>Antarctica (47)</strong></td>
<td><strong>Author (259)</strong></td>
</tr>
<tr>
<td><strong>Asia (2854)</strong></td>
<td><strong>Culture (690)</strong></td>
</tr>
<tr>
<td><strong>Australia &amp; pacific islands (188)</strong></td>
<td><strong>Source</strong></td>
</tr>
<tr>
<td><strong>Central america (70)</strong></td>
<td><strong>A &amp; y (49)</strong></td>
</tr>
<tr>
<td><strong>Eastern eupe (1165)</strong></td>
<td><strong>A + u (151)</strong></td>
</tr>
<tr>
<td><strong>Materials</strong></td>
<td><strong>A. jaffe pc (1)</strong></td>
</tr>
<tr>
<td><strong>Animal material (38)</strong></td>
<td><strong>Aalto7 (1)</strong></td>
</tr>
<tr>
<td><strong>Artistic materials (31)</strong></td>
<td><strong>Aama alam (6)</strong></td>
</tr>
<tr>
<td><strong>Asphalt (2)</strong></td>
<td><strong>Ahmed ahmed (3)</strong></td>
</tr>
<tr>
<td><strong>Brick (138)</strong></td>
<td><strong>Ahite (4)</strong></td>
</tr>
<tr>
<td><strong>Building materials (727)</strong></td>
<td><strong>more...</strong></td>
</tr>
<tr>
<td><strong>Cement (3)</strong></td>
<td><strong>more...</strong></td>
</tr>
<tr>
<td><strong>Ceramic (5)</strong></td>
<td><strong>more...</strong></td>
</tr>
<tr>
<td><strong>Building Names</strong></td>
<td><strong>Styles</strong></td>
</tr>
<tr>
<td><strong>1001 pennsylvania avenue (3)</strong></td>
<td><strong>Africa (599)</strong></td>
</tr>
<tr>
<td><strong>133 s 22nd st (2)</strong></td>
<td><strong>Ancient (72)</strong></td>
</tr>
<tr>
<td><strong>1515 9th avenue (3)</strong></td>
<td><strong>Asian (3783)</strong></td>
</tr>
<tr>
<td><strong>16th street baptist church (2)</strong></td>
<td><strong>Australian &amp; pacific island (157)</strong></td>
</tr>
<tr>
<td><strong>18thkansas lofts (3)</strong></td>
<td><strong>Bronze age (20)</strong></td>
</tr>
<tr>
<td><strong>234 kansas bownhouses (4)</strong></td>
<td><strong>Central american (2)</strong></td>
</tr>
<tr>
<td><strong>1927 stones: monument against racism (6)</strong></td>
<td><strong>Early mediterranean (2112)</strong></td>
</tr>
<tr>
<td><strong>more...</strong></td>
<td><strong>Early near eastern (777)</strong></td>
</tr>
<tr>
<td><strong>more...</strong></td>
<td><strong>European (13699)</strong></td>
</tr>
<tr>
<td><strong>more...</strong></td>
<td><strong>Iron age (26)</strong></td>
</tr>
<tr>
<td><strong>more...</strong></td>
<td><strong>Islamic (663)</strong></td>
</tr>
<tr>
<td><strong>more...</strong></td>
<td><strong>North american (14232)</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>View Types</th>
<th><strong>View Types</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Axiornetricals (170)</strong></td>
<td><strong>Design drawings (318)</strong></td>
</tr>
<tr>
<td><strong>City aerial views (702)</strong></td>
<td><strong>Drawings (178)</strong></td>
</tr>
</tbody>
</table>
Faceted Classification

Pros
• Multiple categories more flexible, scalable
• Maintenance easier: concept structures small
• Accommodates new concepts and relationships
• Local changes do not disturb other parts
• Easier for searchers to understand context
• Supports complex searches and browsing

Cons
• Increased maintenance complexity — domain expert required
• Facets cannot be automatically generated
Social/Community Approach

• Tags/keywords as so-called *folksonomy*
  – Keywords given *freely* to ...
    • Bookmarks – del.icio.us
    • Photos – flickr
    • Papers/books – cite-u-like
    • ...

• Out of control?
  – Community aspect: feedback
    • Effort gets rewarded – enticement to tag
Results for google base

showing 1 - 10 of about 1358

« previous | next »

Google Base
to google search tools database googlebase by 258 people

Google Base
to google search database web2.0 googlebase by 206 people

Google Base: All your base are, in fact, belong to us
to google search base database internet by 97 people

Google Base
to google search database googlebase base by 102 people

Official Google Blog: First Base
to google googlebase base search web by 87 people

Burnham’s Beat: RSS and Google Base: Google Feeds Off The Web
to google rss xml googlebase web2.0 by 85 people

Google Base Was Sort of Live
to google googlebase database web tools by 56 people

More Google Base Screenshots
to google database base googlebase screenshot by 47 people

seweso’s blog: Google base
to google googlebase screenshot database folksonomy by 35 people
17

Comments

helinars pro says:

mich wundert, dass das noch keiner zerstört hat, am besten abbauen und mit nach Hause nehmen!

Posted 2 weeks ago. (permalink)

hauphasse pro says:

und dort gibt's dann doppelt konzentrierte suppe!

Posted 3 days ago. (permalink)

Would you like to comment?

Sign up for a free account, or sign in (if you're already a member).

Tags

- Leipzig
- GDR
- DDR
- Reklame
- Schild
- advertisement
- Konserven
- tin
- can
- Graffiti
Sky over Leipzig 2

Taken on the same day as the previous photo. A few minutes later the wind had blown the clouds away. I wish there was an ensemble of skyscrapers and not merely one famous landmark.

Comments

bakpacker pro says:
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>See on Amazon</th>
<th>Average Rating</th>
<th>First Shelved</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Things That Make Us Smart: Defending Human Attributes</td>
<td>Donald A. Norman</td>
<td></td>
<td>★★★★★</td>
<td>November 18th, 2005</td>
<td></td>
</tr>
<tr>
<td>The Design of Everyday Things</td>
<td>Donald A. Norman</td>
<td></td>
<td>★★★★★</td>
<td>October 8th, 2005</td>
<td></td>
</tr>
<tr>
<td>The European Dream</td>
<td>Jeremy Rifkin</td>
<td></td>
<td>★★★★☆</td>
<td>November 18th, 2005</td>
<td></td>
</tr>
<tr>
<td>Code and Other Laws of Cyberspace</td>
<td>Lawrence Lessig</td>
<td></td>
<td>★★★★☆</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Which tablet would you want more?

EliteGroup ECS EZ30  vs.  Lenovo ThinkPad X41 Tablet

Choose your tablet and place your vote!
Classification Schemes /Scope

• **Folksonomy**: propose other *related* tags to users based on frequency, forming clusters

• **Taxonomy**: more query and navigation possibilities, e.g. research topic taxonomy
  – Not only by title/author/year, tags/keywords
  – But also via related categories/concepts
    • either by way of **transitive relationships**, i.e. super/sub-classes of research-topics
      e.g. ‘Model Management’ as super-concept to ‘Schema Matching’, ‘Ontology Matching’, ‘Evolution’
    • or other relationships, e.g. “see-also” within an ontology
  – Giving context; consider a concept ‘Evolution’
Agenda

• Accessible Classification Systems
  – Searching
  – Browsing

• Quick Authoring/Classifying
  – Semantic Wiki
    • Example/Demo
    • Challenges

• Are we there yet?
What is a Wiki? – Two things.

» A Wiki enables documents to be written collectively in a simple markup using a web browser « [wikipedia].

» A collective website where a large number of participants are allowed to modify any page or create a new page using their Web browser « [Desilets05].
What is a Wiki? – Keypoints

• Quick
• Collaborative
• Authoring
• On the Web

» Making it easy to correct mistakes, rather than making it hard to make them « [WikiWay]
What is a Semantic Wiki? /1

• Intersection of Wiki and Semantic Web
• Semantics/semantikos: significant meaning, derived from sema/sign
• Semantic Web
  – Giving meaning processable by machines
    • Conceptual graphs [John F. Sowa] (as triples)
    • URIs as identifier of concepts/relationships
    • Common vocabulary, e.g. Dublin Core, FOAF
The cat sat on the mat

- Conceptual Graph representation

- Semantic Web triple/statement
  - Subject, predicate, object/literal
  - Each identified via unique URIs
Vocabulary, Generalization

- **Thing**
  - **Person**
    - Lecturer
    - Author
      - Literary Author
  - **Potable Liquid**
    - Beer
      - Stout
    - Wine
Vocabulary, Namespaces/URIs

- **OWL:** Thing
  - **FOAF:** Person
    - **MyNS:** Lecturer
    - **DC:** Author
      - **MyNS:** LiteraryAuthor
  - **W3C:** PotableLiquid
    - **David:** Beer
      - **David:** Stout
    - **W3C:** Wine
More triples/statements

or...

http://purl.org/dc/elements/1.1/title

urn:isbn:0596002637

Practical RDF

machine processable

[Miles05] 30
Semantic Web – Triple space

• Spans labelled directed acyclic graphs

• Query possibility using W3C’s specs.
  – RDQL
  – SPARQL
What is a Wiki? – Revisited

• Quick
• Collaborative
• Authoring
• On the Web
What is a Semantic Wiki? /2

• Quick
• Collaborative
• Authoring
• On the Web
• Of concepts and their relationships
• ... offering querying, navigating, ...
  ... a lot.
How? – A semantic wiki demo

• Editing relationships between concepts
• *Quick?* – Keep it simple!
  – Simple syntax (extended CamelCase)
  – Single form to enter everything
• Entering/authoring
  – Relationships create navigational means
  – Querying
  – Interactivity
Demo

» make it easy to correct mistakes, rather than making it hard to make them.

[WikiWay]
Query Interface/Translation

• Translating user-entered query into SPARQL/RDQL-query, e.g.:
  – InstanceOf=LiteraryAuthor
    DateOfBirth between 1800 and 1900

  ... becomes by way of a regular expression:
  – SELECT ?resource ?tempDateOfBirth
    ?tempDateOfBirth FROM
    <http://localhost/triples.n3> WHERE
    ( ?resource <#InstanceOf> <#LiteraryAuthor> )
    ( ?resource <#DateOfBirth> ?tempDateOfBirth )
    AND ?tempDateOfBirth >= 1800 ,
    ?tempDateOfBirth <= 1900
$query =~ s/$wikipattern$op$literal/
  my $pred = $1; my $xop = $2; my $lit = $3;
  if ($xop ne ") {  
    $selects.=" ?temp$pred";
    if ($xop eq "=") {
      $andclause.=" ?temp$pred $xop \"/$lit\" ",";
    }
    else {
      $andclause.=" ?temp$pred $xop $lit ",";
    }
  }
  "( $var1 <$wuri$pred> ?temp$pred ) "
} else {
  if ($lit=~$wikipattern) {
    "( $var1 <$wuri$pred> <$wuri$lit> ) "
  } else {
    "( $var1 <$wuri$pred> $lit ) "
  }
}
xge;
A Semantic Wiki Hype?

• Related Work
  – Platypus
  – SnipSnap
  – pOWL/OntoWiki
  – MediaWiki
  – SemperWiki
  – DiamondWiki
  – …
Submit Publication

Title:
RevOWLution - Rapid Ontology Evolution

Research area:
-DB evolution
--Ontology evolution
--Web service evolution
--Workflow evolution
--Software evolution
--Schema language / data model
--relational
--object-oriented
--XML
--ER / UML
--OWL / RDF
--Versioning

Publication venue:
Unpublished / Tech report

Year:
2006-2007

Paper type:

Authors:
Potter, Harry; Squarpants, Spongebob
Title

Details (optional)

Details are attributes that define the item and help searchers find it. For example, "Author: Ernest Hemingway" or "Area: 400 square km".

Author: Vannevar Bush
Publication Name: As We May Think
Publish Date: 07/1945
Pages:
Publication Volume: The Atlantic Monthly
System: Memex

Labels (optional)

Keywords or phrases that describe your item. Maximum of 10. Separate with commas.
Authoring Information w. Class

- Free tags/keywords
  - No relationships except via clustering
- Free relationships/categories
  - Too many relationship types/categories?
- Providing selection of categories
  - Inflexible?
  - Cumbersome?
Challenges

• High effort populating initial database
  – Community does it
  – Web scraping per item, e.g. Bookmarklets
  – Bulk import, online data provider

• Query interface too cumbersome
  – Auto-completion may help
  – Combination of browsing and querying

• Usable but scaleable system wanted
Wanted

• Bulk import
  – E.g. biblio. domain: Selection of articles thru Web Services by iFuice/biblioFuice

• Issues
  – Categories – enter how, retrieve from where?
  – Static vs. dynamic content – updateability
    • Author, title, venue, date
    • Citation count
Last slide

Thank you