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

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

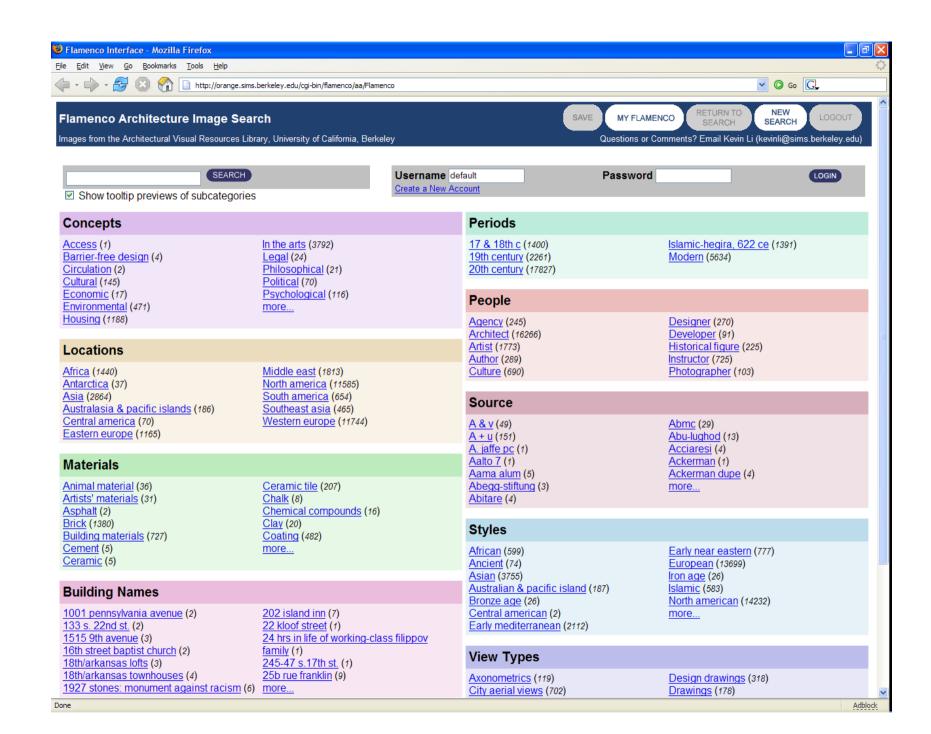
Cons:

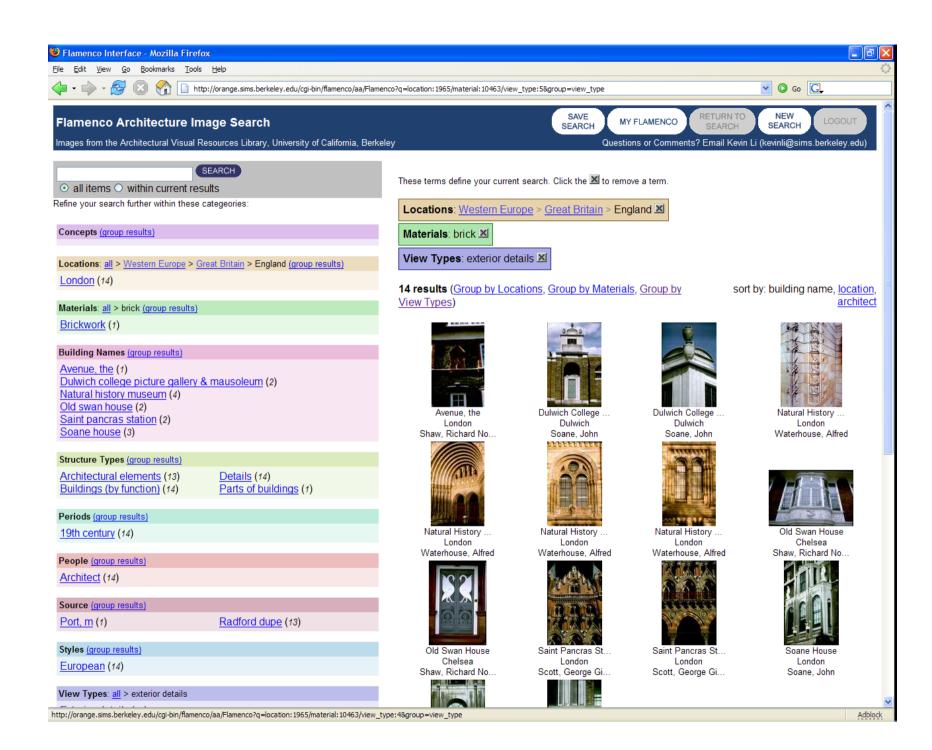
- Increased maintenance complexity, editor needed
- Strict hierarchy of concepts may not be flexible or clear enough for all uses
- Taxonomy cannot be automatically generated

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





Faceted Classification

Pros

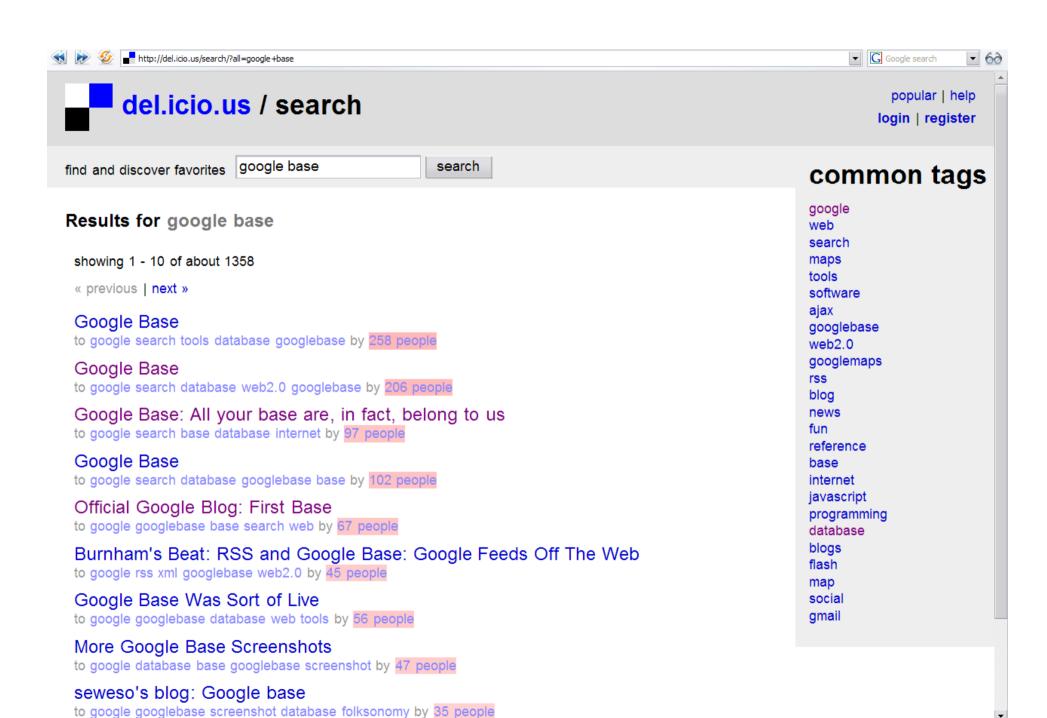
- Multiple categories more flexible, scalable
- Maintenance easier: concept structures small
- Accommodates new concepts and relationships
- Local changes do not disturb other parts
- Easier for searchers to understand context
- Supports complex searches and browsing

Cons

- Increased maintenance complexity—domain expert required
- Facets cannot be automatically generated

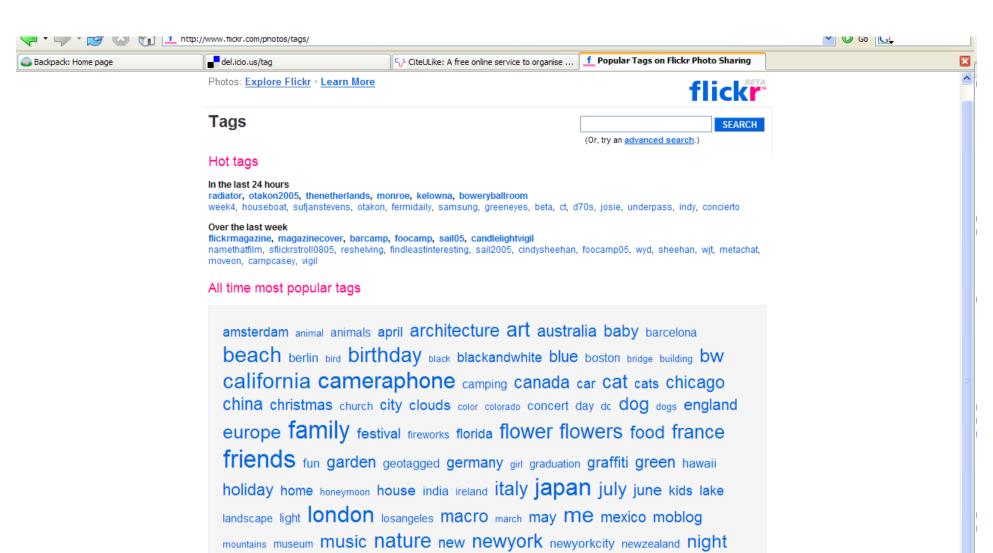
Social/Community Approach

- 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





software web design blog reference tools programming music css linux howto art news mac blogs webdesign fun ajax photography windows internet games javascript humor tutorial cool osx technology free development science security politics business tips flash java search shopping freeware toread google rss opensource funny webdev php tech ruby books food apple productivity gtd video flickr graphics education travel xml html media history computer python culture diy del.icio.us daily blogging maps tutorials research article hacks hardware inspiration firefox gaming audio photos religion writing wiki photo community resources social rails utilities work download comics code language geek lifehacks mp3 useful mobile wordpress web2.0 poker tool images email cooking cards usability movies health gambling learning cms links japan interesting photoshop todo onlinepoker sportsbook casino jobs podcast radio personal gambler gamble cash network online architecture recipes humour game tv articles magazine computers career betting casinos film electronics information book hosting database forum job marketing microsoft misc fonts hack unix dev entertainment resource .net stud bet management literature organization shop movie nyc guide macosx folksonomy tagging hacking china environment perl money philosophy bittorrent tags gmail library animation accessibility home economics finance life p2p illustration visualization website people sex xhtml ipod standards voip color fashion networking gallery bookmarks podcasting info uk gui interface 3d gadgets psychology society japanese apps sysadmin partypoker weird porn documentation ideas clothing evolution wishlist palm furniture math mapping ui portfolio extensions googlemaps weather utility weblog law reading temp os friends dvd sports advertising pc comic webservices torrent usa read wireless physics server ebooks api government english future service map privacy plugin macintosh teaching it canada coding house theory rubyonrails image framework reviews dhtml torrents school oracle digital kids delicious



nyc ocean orange oregon paris park party people phone photo pink portrait red reflection river roadtrip rock rome sanfrancisco school scotland sea seattle sign sky snow spain spring street summer sun sunset taiwan texas thailand tokyo toronto travel tree trees trip uk unfound urban usa vacation vancouver washington water wedding white winter yellow zoo



Comments



helimars pro says:

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

Posted 2 weeks ago. (permalink)



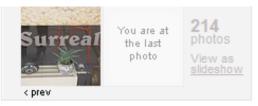
haupthase pro says:

und dort gibts dann dopppelt 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).



This photo also belongs to:



- + Street Corners (Pool)
- + FlickrCentral (Pool)
- + S!gns (Pool)
- + Murals & Graffiti (Pool)
- + Facades (Pool)
- + Schilder, Schmilder [signs] (Pool)
- + emptyphoto (Pool)

Tags

- S Leipzig
- GDR
- ODR
- Reklame
- Schild
- advertisment
- Konserven
- tin
- can
- Graffiti

Sky over Leipzig 2



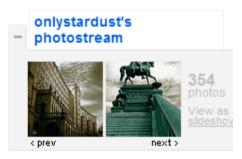
Taken on the same day as the previous photo. A few minutes later the wind had blown the clouds away. I wish there was an ensemble of skyscrapers and not merely one famous landmark ...

Comments



bakpacker pro says:





This photo also belongs to:



Tags

S Leipzig

Additional Information

- © All rights reserved
- o Taken on June 9, 2005
 - o 7 people call this photo a favorite.
- Viewed 93 times.

Ning Pivot ->

Ning

For the latest news or updates on the Ning Playground, visit the Ning

Example Apps

Restaurant Reviews Restaurant Reviews w/Maps Bookshelf Rate My Teachers This or That How Much Do I Rock? Bookmarks Photo Sharing Dating Questions & Answers Simple Social Network Anytown Listings Anytown Marketplace Find-A-Developer Wishlist Discussion Board Blog Confess! Storvteller Bay Area Hiking Trails

Welcome to the Ning Playground

With Ning, all you need is a good idea to run your own social application. **Social apps let you match, transact, and communicate with other people.** A few clicks, no coding (unless you want to), and you can be up and running with your own social app in minutes.

For any City

- Create your own Classified Listings
- For any Interest

 Create your own Online Reviews
- For any Event

 Create your own Photo Sharing App
- For any School

 Create your own Social Network

Sign up and build your own social apps on Ning »









Top Social Apps on the Ning Playground

More Top Apps »

all things stupid anytown listings bay area hiking trails bead store reviews with maps bookmarks bookshelf bulldogster choose your beer confess! dating developer discussion douchetracker find-a-developer girl on girl got milf? how much do i rock? ning developer documentation photo sharing restaurant reviews with maps roshambo storyteller superhero dating tablet pc duel! technical archaics Wishlist

Top Tags on the Ning Playground

More Top Tags »

art battle blog classics css delphi design email fantasy fiction food fun funny geek history milf music ning nudes php scifi test tutoriales to web

Bookshelf

All Books Top Tags

Search

All Books

There are 521 readers and 968 books on this bookshelf. Showing 1 - 15.

99
100
-

The Wind-Up Bird Chronicle: A Novel (Vintage International)

By Haruki Murakami | See this on Amazon

Avg. Rating: ★★★☆☆

First shelved by sakanasato on November 20th, 2005 0 people currently reading this book.



Things That Make Us Smart: Defending Human Attributes in the Age of the Machine

Sort by most recently added

☐ 1 Comment

□ 1 Comment

O Comments

1 Comment

☐ 1 Comment

By Donald A. Norman | See this on Amazon

Avg. Rating: ★★★☆☆

Most recently shelved by huerlisi on November 18th, 2005 First shelved by crosenblath on October 4th, 2005, and 2 others since 0 people currently reading this book.



The Design of Everyday Things

By Donald A. Norman | See this on Amazon

Avg. Rating: ***

Most recently shelved by huerlisi on November 18th, 2005 First shelved by sibisoro on October 8th, 2005, and 2 others since 0 people currently reading this book.



The European Dream

By Jeremy Rifkin | See this on Amazon

Avg. Rating: ★★★☆☆

First shelved by huerlisi on November 18th, 2005 0 people currently reading this book.



Code and Other Laws of Cyberspace

By Lawrence Lessig | See this on Amazon

Avg. Rating: ***

TABLET PC DUEL!

NEW DUEL

WINNERS

LOSERS

NEWEST

ADD A TABLET

GRIPES

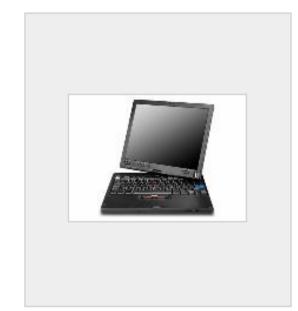
WISHES

Which tablet would you want more?

VS.



EliteGroup ECS EZ30



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/subclasses 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 'xsdsadadghdafasf' or... 'xsdsadadghdafasf' 'xsdsadadghdafasf' 'Practical RDF' http://purl.org/dc/elements/1.1/title urn:isbn:0596002637 machine processable

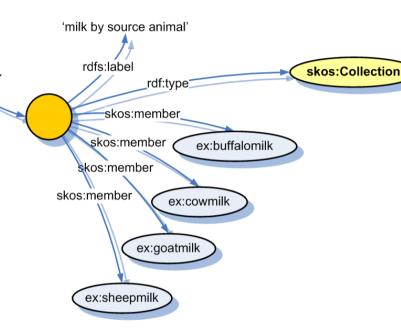
[Miles05]

30

Semantic Web - Triple space

ex:milk

• 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</pre>
```

Query Translation: Reg. Exp.

```
$query =~ s/$wikipattern$op$literal/
   my pred = 1; my per = 2; my lit = 3;
   if ($xop ne "=") {
         $selects.=" ?temp$pred";
         if (xop eq == ) {
                   $andclause.=" ?temp$pred $xop \"V$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

– ...



database group

search



Publication Categorizer & Browser

Publications

- Research Area
- Publication venue
- Year
- Paper type

Features

- Publication Index
- category browser

Recent comments

cat

- Forums
- create content
 - Publication
 - forum topic
- my account
- recent posts
- blocks
- categories
- menus
- □ log out

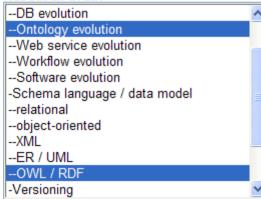
Home » create content

Submit Publication

Title:*

RevOWLution - Rapid Ontology Evolution

Research area:



Publication venue:

Unpublished / Tech report >

Year:

2006-2007

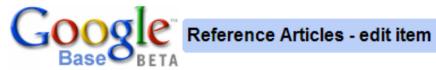
Paper type:

<none> Implementation -Product -Prototype Theory

Authors:

Potter, Harry; Squarepants, Spongebob

list of authors as in Aumueller, David; Simpson, Homer



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".

Autnor: Text	Vannevar Bush	ĪXĪ
Publication Name:	As We May Think	[x]
Publish Date: Date range	07/1945 Time: (opt	ional) hh:mm am [x]
Pages: Number		[x]
Publication Volume:	The Atlantic Monthly	[x]
Webpage Web URL ▼	nttp://www.theatlantic.co	om/doc/194507/bush
System	Memex	[x]
Text ▼. Add another attribute.		

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