Collaborative indexing: towards the universal index

Glenda Browne
www.webindexing.biz
webindexing@optusnet.com.au

Collaborative indexing

- Indexing more important than ever, to deal with vast quantities of unstructured information.
- Collaborative indexing shares the workload.
- Collaborative indexing can be formal or informal.
- Web-based services enable the creation of 'de facto' indexes.
- Basedex has been written to enable the creation of small indexes which can be combined into a 'universal' index for the web



The information universe is too big to index...



and too big not to index!

One solution



Collaborative indexing – books

- Index Society (1877) created to construct a general index to all knowledge.
 - Many readers would readily jot down the name of any new subject they meet with in the book before them, and the page on which it occurs, and forward their notes to be sorted and arranged by any Society that would undertake the work (p.38).
 - I have heard two objections brought against the scheme of the Society:
 - 1. That it is needless to urge the compilation of indexes, because every worker worthy of the name makes his own...
 - 2. That the General Index is an impossibility, and that to attempt its preparation is a waste of time. Those who hold this opinion have not sufficient faith in the simplicity and usefulness of the alphabet....The object of the General Index is just this, that anything, however disconnected, can be placed there, and much that would otherwise be lost will there find a resting place (p.40).

Collaborative indexing – genealogy

- Genealogical indexes for cemetery inscriptions, school and hospital records, public shipping and BDM records.
 - consistency in fields and wording



Collaborative indexing – web directories

- Librarians' Internet Index lii.org
- Open Directory Project dmoz.org
- Mahalo alpha mahalo.com

Index structure – home page and related terms



Harry Potter and the Order of the Phoenix

Search

Categories: Entertainment | Film | Contemporary Films

Harry Potter and the Order of the Phoenix

Also try. Harry Potter Also try: Harry Potter Toys

Also try: Harry Potter and the Deathly Hallows

The Mahalo Top 7

- 1. Official WB Site: Harry Potter and the Order of the Phoenix
- 2. Official Film Site: Harry Potter and the Order of the Phoenix
- 3. Wikipedia: Harry Potter and the Order of the Phoenix Film and Novel
- 4. Harry Potter Wiki: Harry Potter and the Order of the Phoenix
- 5. MySpace Page: Harry Potter and the Order of the Phoenix
- IMDb: Harry Potter and the Order of the Phoenix





Login : Create Account



Guide Note:

Harry Potter and the Order of the Phoenix is the fifth fantasy novel by J.K. Rowling in the Harry Potter series. The film is being produced by Warner Bros...

Fast Facts:

- 1. US Release: July 11, 2007
- 2. Running time: 138 min
- 3. Budget: \$150-200 million
- 4. Director, David Yates
- 5. Cast: Daniel Radcliffe, Emma Watson, Rupert Grint

Antidepressants: Related Searches

<u>Depression</u> | <u>Seasonal Affective Disorder (SAD)</u> | <u>Bipolar Disorder</u> | <u>Anxiety Disorder</u>

Index structure – subheadings

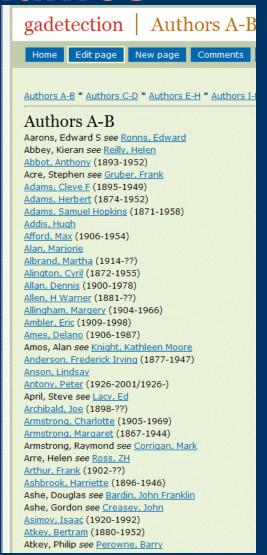


Wiki terms – disambiguation



Wiki terms – names

Golden Age of
 Detective Fiction
 wiki –
 gadetection.pbwiki.
 com/Authors+A-B



Wiki terms – names, topics

- WireFrames [redirected from WireFrame]
- CMS/ContentManageme ntSystems
- CardSort [software] / CardSorting [technique]
- CategoryBook [individual] / CategoryBooks [lists]
- CXO files before CafeTesting

- 43. AmbientFindability
- 44. AmitTungare
- 45. AnalysisTools
- 46. AndiLewis
- 47. AndreaTanzi
- 48. AndresSulleiro
- 49. AndrewBoyd
- 50. AndrewCates
- 51. AndrewChee
- 52. AndrewDillon
- 53. AndrewGent
- 54. AndrewHinton
- 55. / IARants
- 56. AndrewOtwell
- 57. AndrewsInIA
- 58. AndrésSulleiro
- AnnaGifford
- AntiThesaurus

Commercial indexing

- Periodical abstracting and indexing services (bibliographic databases)
 - long-term projects
 - professional staff
 - centralised control, including use of thesauruses
 - search techniques include proximity, fielded searching
 - expensive; subscription or pay-per-use
 - library service, eg,
 www.sl.nsw.gov.au/databases/athome.cfm
- Aggregating services/ePresses
 - www.informit.com.au
 - centralised searching interfaces being developed

Problems with indexing

- Professional indexing is:
 - expensive
 - time-consuming (delays in access)
 - formal and controlled (often)
- Users don't always speak the language of expert indexers...
 - ...so, social tagging lets users index (tag) sites for themselves and others.

Social tagging

common tags cloud | list

basedex collaboration collaborative index rails ruby rubyonrails

posting history

» first posted by Shanlalit to rails

Jun '07

by Igsicious to collaboration rails api list collaborative index ruby basedex

Mar '07

by pilarw to basedex collaborative index indexes

by wwwalker to rubyonrails index basedex collaboration

by Glendalndexer to basedex website_indexing

by racketjo to basedex web index useful ruby rails rubyonrails ror

by hsumh316 to rails

by kaizenfury7 to rails

Feb '07

by hajee to rails

by Shanlalit to rails

Social tagging

- But...users do not necessarily speak the same language as other users.
 - Tags such as 'odd', 'cool', 'must read', 'own'
 - Recall decreases through inconsistent wording:
 - 'information architecture',
 - 'information_architecture',
 - 'informationarchitecture' and 'IA'.
- Formal indexes have a set of basic principles which are (usually) understood.

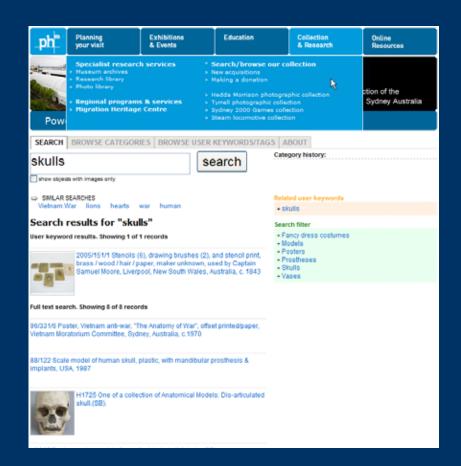
Institutional social tagging

- 'PictureAustralia'
 under the 'groups'
 tab in flickr
 (www.flickr.com/
 groups/pa_ourtown)
- Tagging of museum objects at steve.museum



Institutional social tagging

Powerhouse Museum
 (www.powerhouse
 museum.com/
 collection/database) –
 user-generated tags
 are combined with
 three controlled
 vocabularies



Page creators as metadata-based indexers



Structured collaborative indexing – Basedex

- Basedex by Nelson Hsu (blazingrails.basedex.com) is a new website indexing program, still in beta.
- It allows the creation of individual indexes which can be used alone or in combination with other indexes.

Basedex contributors and users

- Individuals choose topics or documents to create indexes on.
- These topics are grouped hierarchically.
- People 'subscribe' to indexes, combining those which are of interest to them.
- This could include indexes to a number of websites, or to a selection of ebooks, or both.
- The result is a large index to a selected portion of the web.
- The motivation for creating indexes will range from good will to self-promotion.

Basedex topic lists and duplicates

Basedex (v.noted	lone) basedex discussion blog join contact us help	<u>login</u>
Index: <u>Universe</u>	create sub-index view index info subscriptions	(
	all A B C D E F G H I J K L M N O P Q R S T everything current index	r u v w x y z s
top-level indexes Universe super topics sub topics Academia Apple Asian Culture Basedex Help Books Career Cars Cell Phones Computers Education Entertainment Fashion Finance Forums Howto Humor Indexing Internet Multimedia	Browsing Universe for entries beginning with S S.H.E english fansite & forum (in Asian Multimedia) >> S.H.E english fansite & forum (in Asian Multimedia) >> S.H.E english fansite & forum (in Asian Multimedia) >> S.H.E english fansite & forum (in Asian Multimedia) >> S.H.E english fansite & forum (in Asian Multimedia) >> saving compressed content >> susing with Ruby >> S.H.E english fansite & forum (in Asian Multimedia) >> saving compressed content >> safe (Finalizer) (in Ruby on Rails) >> safe (Finalizer) (in Ruby on Rails) >> safe (Finalizer) (in Ruby 1.8.6 Core API Methods) >> safe (Finalizer) (in Ruby 1.8.5 Core API Methods) >> safe (Finalizer) (in Ruby 1.8.5 Core API Methods) >> safe (Finalizer) (in Ruby 1.8.5 Core API Methods) >> safe (Finalizer) (in Ruby 1.8.5 Core API Methods) >> safe (Finalizer) (in Ruby 1.8.6 Core API Me	
News People Sample Index Shopping Software Engineering Sports Starcraft 2 Startups Torrent Sites Web Sites	safeconstname? (XSD::CodeGen::GenSupport) (in Ruby 1.8.5 Core API Methods) ≥> safemethodname (XSD::CodeGen::GenSupport) (in Ruby 1.8.5 Core API Methods) ≥> safemethodname (XSD::CodeGen::GenSupport) (in Ruby 1.8.6 Core API Methods) ≥> safemethodname (XSD::CodeGen::GenSupport) (in Ruby 1.9 Core API Methods) ≥> safemethodname? (XSD::CodeGen::GenSupport) (in Ruby 1.9 Core API Methods) ≥>	

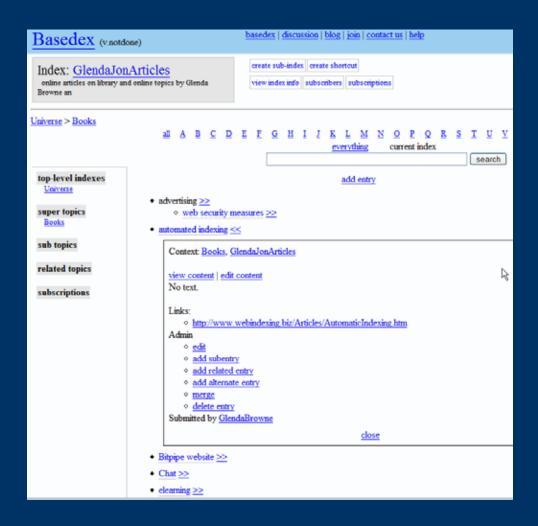
Basedex subscribing

- Any index can be subscribed to a number of other indexes, with the terms interfiling.
- The name in parentheses indicates the original source of the index term.
- When there is no name in parentheses the entry belongs to the home index.

Basedex granularity

- Basedex combines access at different levels of granularity: to parts of a book, to webpages and to websites.
- This may be frustrating for people who want a lot of information on a topic, and find themselves taken to one paragraph in a book
- The parenthetical qualifiers on entries which show the source will give some indication of the type of document being referred to.

Basedex edit window



Basedex subheadings

Dascucx (v.notdone)	
Index: Web Indexing SIG	create sub-index create shortcut
	view index info subscribers subscriptions
Create a subentry for members	
Entry type index entry 💌	
Entry	
Control	
Content	
Links (separate with new line)	
add cancel	

Basedex See references

Basedex (v.notdone)	basedex discussion blog join contact us help			
Index: Web Indexing SIG	create sub-index create shortcut view index info subscribers subscriptions			
Editing membership				
back to index properties links related entries alternate entries contexts merge delete entry				
Alternate Entries				
add alternate entry				
Alternate entry				
signing up				
update reset				

Basedex See Also references

Basedex (v.notdone)	basedex discussion blog join contact us help			
Index: Web Indexing SIG	create sub-index create shortcut view index info subscribers subscriptions			
Editing membership				
back to index properties links related entries alternate entries contexts merge delete entry				
Add a related entry Entry				
* Cross-references can only be created between entries in the same index.				
continue cancel				

Basedex practicalities

- Basedex indexes are created and stored centrally on the Basedex site. This means that the creator doesn't control the storage or display of the index.
- Indexes created independently in XML format can be uploaded to Basedex.

More information

- Australian and New Zealand Society of Indexers www.aussi.org, to become www.anzsi.org
- Browne, Glenda and Jermey, Jonathan 2004. Website indexing: enhancing access to information within websites, from www.webindexing.biz and www.lulu.com.
- Browne, Glenda and Jermey, Jon 2007. *The indexing companion*. (Cambridge University Press, 2007).
- Web Indexing Yahoo mailing list tech.groups.yahoo.com/group/web-indexing.
- Wheatley, Henry B. *What is an index?* (Society of Indexers (UK) facsimile 2002; original publication Longmans, Green & Co. 1879), pp. 38-40.