Fusion of Libraries and the Web: Subject-based Information Retrieval in the Web 2.0 Era

Digital Library Research Division
Information Technology Center
University of Tokyo
Yoji Kiyota (Assistant Professor)

- 1. What is happening on the information retrieval world in *the Web 2.0 era?*
- 2. Toward *the Lib 2.0:* what is needed to enhance the values of libraries?
- 3. A solution: fusion of folksonomy and taxonomy
- 4. Application for reference services: Littel Navigator

- 1. What is happening on the information retrieval world in *the Web 2.0 era?*
- 2. Toward *the Lib 2.0:* what is needed to enhance the values of libraries?
- 3. A solution: fusion of folksonomy and taxonomy
- 4. Application for reference services: Littel Navigator

History of information access methods

- Until the 20th century
 - To research in libraries
 - To go to bookstores
 - To ask someone
 - To ask consultant services
 (librarians, government offices, consultants, ...)
 - To look up Web directories, BBS, databases, ...
- From the 21st century: in addition,
 - To google (Web search engines)
 - To access QA sites, SNS, folksonomy, ...

So-called the Web 2.0 paradigm

What is Web 2.0?

Definition by Tim O'Reilly (2005)

- The Web as platform
- Harnessing collective intelligence
- Data is the next Intel Inside
- End of the software release cycle
- Lightweight programming models
- and more...

Just a buzzword? --- No, if we choose one of the definitions

What is Web 2.0?

Definition by Tim O'Reilly (2005)

- The Web as platform
- Harnessing collective intelligence
 - Improvements on web search engines
 - Folksonomy: organization by the wisdom of crowds
- Data is the next Intel Inside
- End of the software release cycle
- Lightweight programming models

Possibility of collective intelligence as an Information access method

- The broadest gate for information
 - Any keyword hits using Google
 - The best way to find cues?
- Diversity of information
 - Exceeds diversity of mass-media
- Self-organization of information
 - Improvements of web search engines
 - PageRank: democracy on the Web
 - Folksonomy: meta-tagging based on the wisdom of crowds

Folksonomy

= "folks" + "taxonomy"

Meta-tagging by ordinary people

- Web services diffusing from 2005 Flickr, del.icio.us, YouTube, ...
- Every participant assigns tags to contents, based on each viewpoint
- As a result, diverse tags are assigned to each content
- Example approaches of libraries
 - Ann Arbor District Library "Social OPAC"

ル(F)

表示(V)

履歴(S)

ブックマーク(B) ツール(T) ヘルプ(H)







We found 2,703 photos matching B747.

View: Most relevant • Most recent • Most interesting

Photos searched by "B747" Using Flicker

Sponsored Results



New Treo™ 680

The Treo™ 680 smartphone has arrived. With new phone features, fast web connectivity, camera, and more. Visit www.onTreo.com for details.

B747

and Games. Related tags co-occurrence)

Shop and save on Toys and Games. Compare products, prices & stores.

www.Shopping.com/toysandgames



B747

Search the Web

Also try: boeing, airport, airplane, ircraft or plane







B747 Classic

Uploaded on 13 January 2007



By Down in the...

See more photos, or visit his profile.

🖺 blue, atlanta, winter, sky

JAL B747-146B & B747-246B @ OSA/RJOO

Uploaded on 3 April 2006

Tags assigned to each content

By Hyougushi

See more photos, or visit

airplane airport boeing

JAL B747-146B/SUD (JA8170 or JA8176) @ HND/RJTT

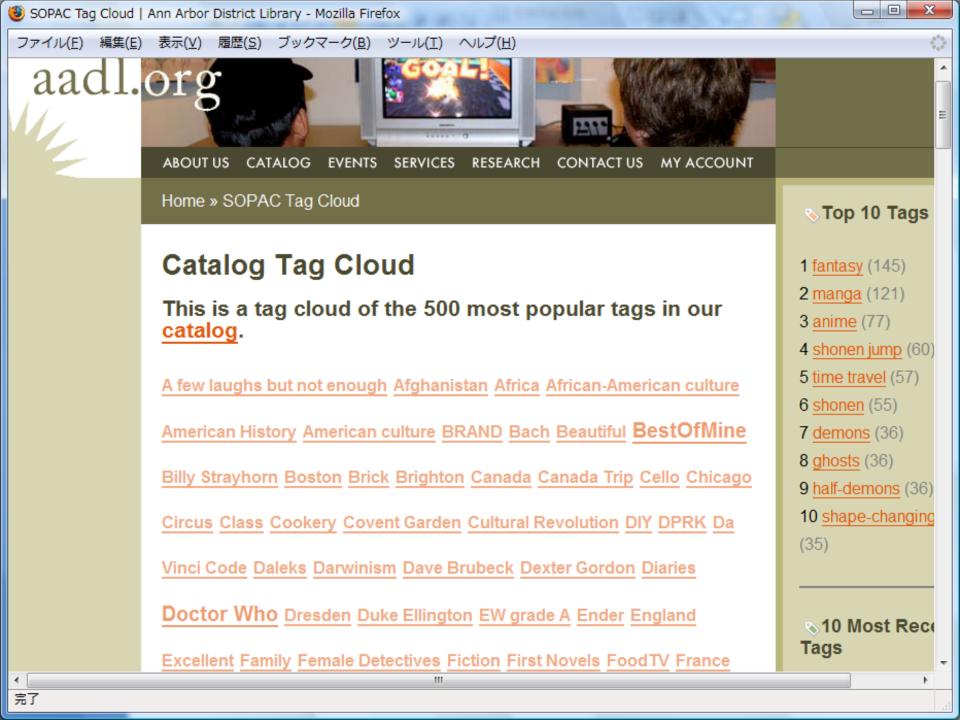
Uploaded on 13 November 2005

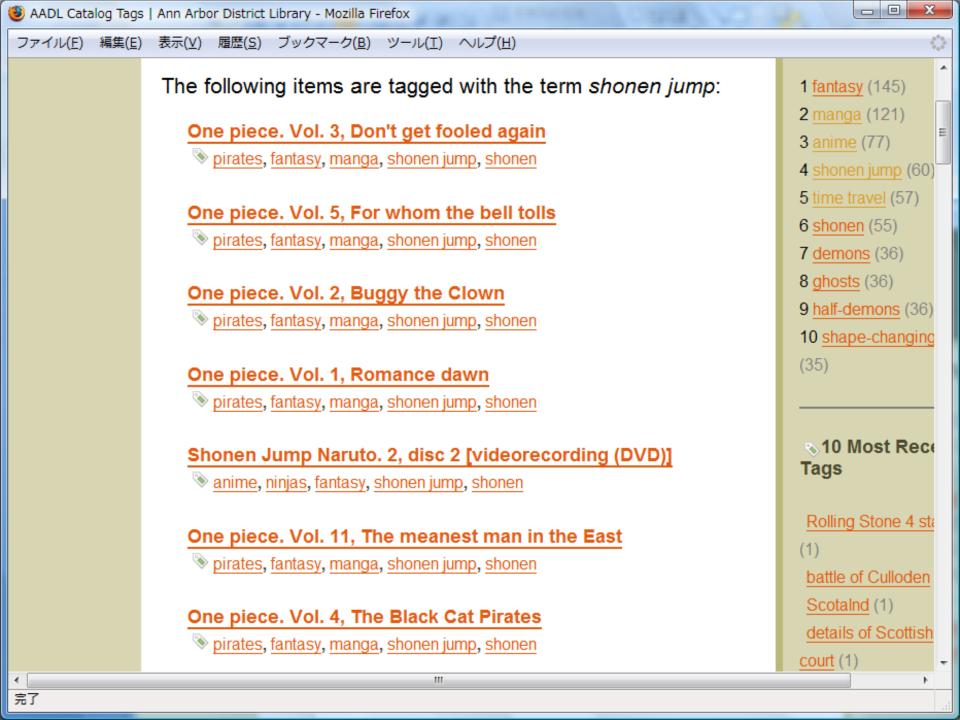


By Hyougushi

See more photos, or visit has profile.

🖺 japan, airplane, tokyo, airpor





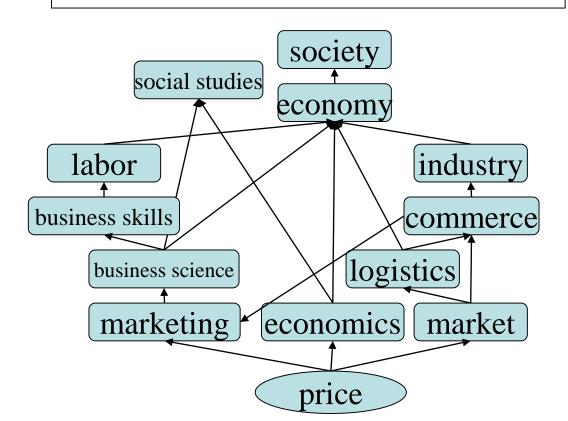
Wikipedia categories

- A multilingual encyclopedia edited by ordinary people
- Categories are assigned to each article
 - categories are regarded as folksonomy tags
 - multi-level folksonomy
 - multiple hyponyms

Price

Price in <u>economics</u> and <u>business</u> is the assigned numerical <u>monetary value</u> of a <u>good</u>, <u>service</u> or <u>asset</u>. The concept of price is central to <u>microeconomics</u> where it is one of the most important variables in <u>resource</u> <u>allocation</u> theory (also called price theory).....

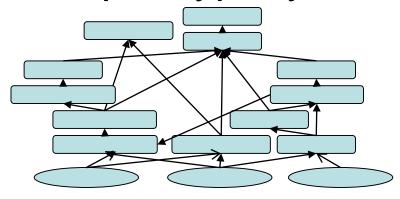
Categories: [marketing][economics][market]



Two classification paradigms

Folksonomy

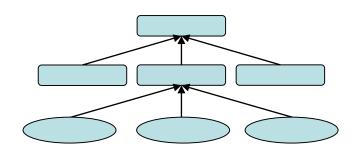
- Bottom-up approach
- Multiple hyponyms



Classification suitable for Web resources

Taxonomy

- Top-down approach
- Only one hyponym



Classification suitable for library materials

Limitation of the folksonomy

Browsing-oriented

- Floods of information without validation
- Prefer new information to old information
 - How many web pages survive 10 years?
 cf. web.archive.org
- Shallow organization
 - Meta noise of folksonomy
- Lack of methods for evaluation and validation

Can libraries complement the limitation of folksonomy?

- 1. What is happening on the information retrieval world in *the Web 2.0 era?*
- 2. Toward the Lib 2.0: what is needed to enhance the values of libraries?
- 3. A solution: fusion of folksonomy and taxonomy
- 4. Application for reference services: Littel Navigator

Toward the Lib 2.0

Fusion of libraries and the Web

- To provide accesses to stocks in libraries
 - Induction from the Web to libraries
 - Enables validations of information on the Web
- To provide viewpoints of information
 - Various methods, including reference books, dictionaries, ...

The role of library classification systems is very important!

Library classification systems (gaps between ideal and reality)

- Classification (UDC, LC, DDC, ...)
 - ideal: browsing of organized bookshelves is useful for information retrieval
 - reality: lack of browsing methods online (cf. OPAC)
- Subject headings (LCSH, ...)
 - ideal: concept-based search of catalogs
 - reality: lack of flexibility

a high hurdle for ordinary people?

A hint: pathfinders

- Firstly developed at MIT library (in 1970s)
- The most demanded information resources at the beginning of information retrieval?
- Manually created by librarians
 - low coverage

Can we create pathfinders automatically?

Possibilities of automation of pathfinder creation

- Templates are available
- Subject headings and library classifications play the major role for pathfinder generation

Possible if subject headings are estimated for query keywords

 databases of reference books, elementary books, ...

A Problem

too few subject headings!

- 1. What is happening on the information retrieval world in *the Web 2.0 era?*
- 2. Toward *the Lib 2.0:* what is needed to enhance the values of libraries?
- 3. A solution: fusion of folksonomy and taxonomy
- 4. Application for reference services: Littel Navigator

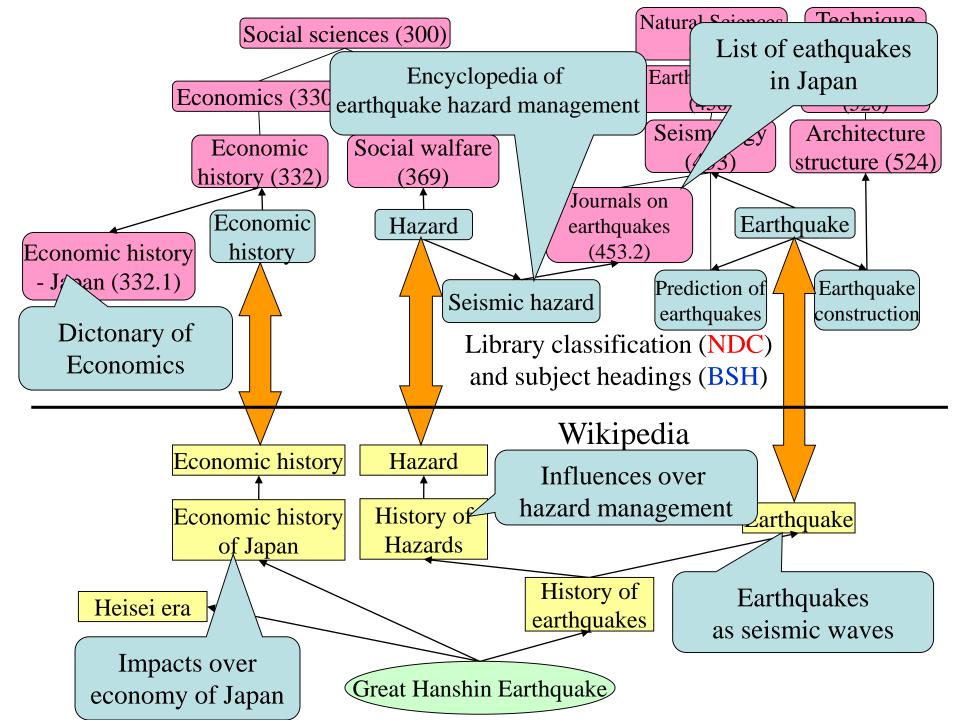
A solution: fusion of Wikipedia categories and library classifications

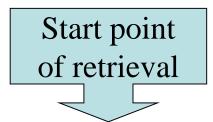
Expansion of classifications using Wikipedia

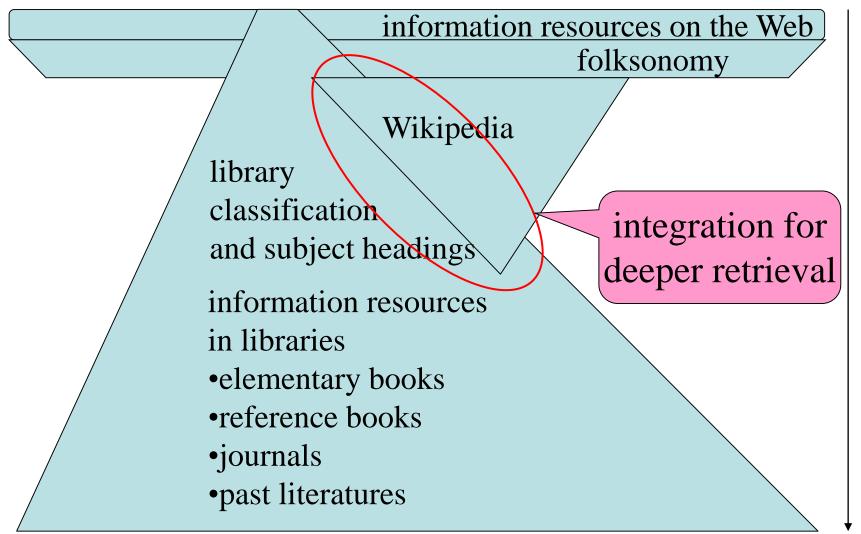
- Wikipedia: folksonomy-based organization
 - The most organized information resource on the Web
 - Compatibility of other Web resources
- Library classifications: taxonomy-based organization
 - Reflect the knowledge structure of human history
 - Powerful tools for demanded information resources

Automated Subject Induction from Query Keywords through Wikipedia Categories and Subject Headings

-> recommendation of useful information





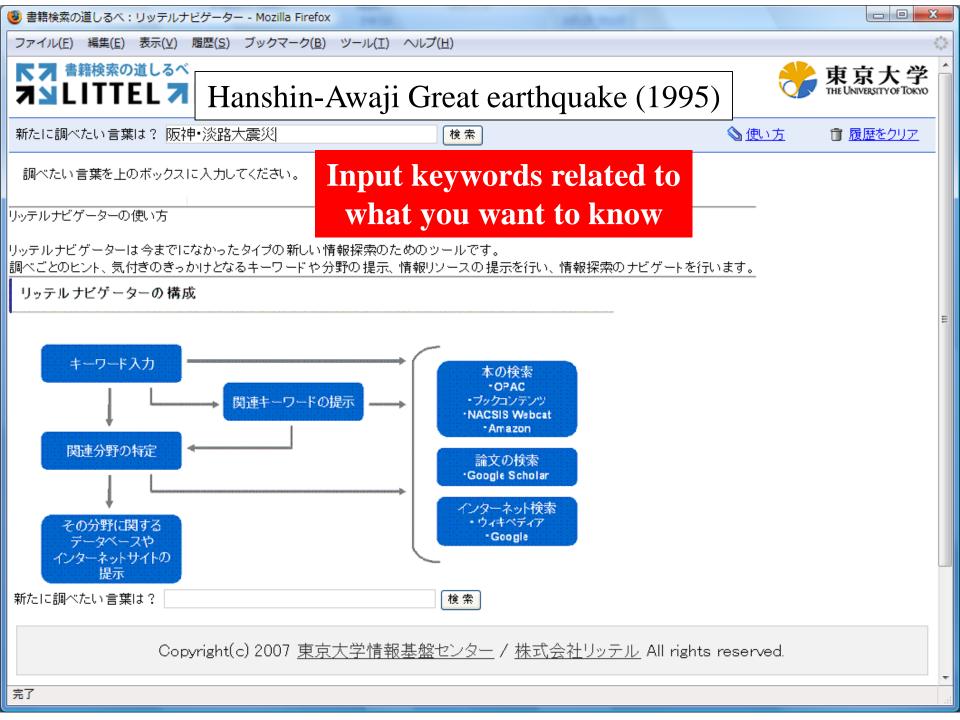


- 1. What is happening on the information retrieval world in *the Web 2.0 era?*
- 2. Toward *the Lib 2.0:* what is needed to enhance the values of libraries?
- 3. A solution: fusion of folksonomy and taxonomy
- 4. <u>Application for reference services:</u>
 <u>Littel Navigator</u>

What is Littel Navigator?

A search engine for hints of information retrieval

- = A pathfinder generator
- Fusion of various information sources
 - Subject headings, Library classifications, reference books
 - Web sites, Wikipedia...
- Automated theme induction from any keyword
 - Traversal of the Wikipedia network
- Navigation from vague keywords to specific materials
 - Stored history of user inputs











ファイル(E) 編集(E) 表示(Y) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

🗾 書籍検索の道しるべ:リッテルナ... 🖂 📗 🕟 阪神・淡路大震災 - Google Sch... 🖸



阪神·淡路大震災

Scholar ヘルブ

◎ ウェブ全体から検索 ◎ 日本語のページを検索

Scholar すべての記事 - 最近の記事

阪神・淡路大震災 の検索結果 約 2,190 件中 1 - 10 件目 (0.08 秒)

すべての検索結果

林春男 木村玲欧

立木茂雄

田村圭子 青野文江 gm 阪神・淡路大震災後の被災者の移動とすまいの決定に関する研究,地域安全学会論文集

木村玲欧, 林春男, 立木茂雄, 浦田康幸 - No. 1999 引用元 8 - 関連記事 - ウェブ検索

gm 阪神・淡路大震災における被災者の対応行動に関する研究~西宮市を事例として~ 地域安全学会論文報告集 青野文江,田中聡,林春男,重川希志依,宮野道雄 - No, 1998 引用元 6 - 関連記事 - ウェブ検索

gm 阪神・淡路大震災後のすまい再建バターンの再現-2001 年京大防災研復興調査報告- 地域安全学会論文集 木村玲欧, 林春男, 立木茂雄, 田村圭子 - No. 2001 引用元 5 - 関連記事 - ウェブ検索

⑤Ⅰ雨 阪神・淡路大震災被災者の生活再建課題とその基本構造の外的妥当性に関する研究、地域安全学会論文集 田村圭子,立木茂雄,林春男 - No. 2000 引用元 4 - 関連記事 - ウェブ検索

阪神・淡路大震災からの生活再建 7 要素モデルの検証 - 2001 年京大防災研復興調査報告 - - 3 グルーブ »

田村圭子, 林春男, 立木茂雄, 木村玲欧 - No, 2001 - seis.nagoya-u.ac.jp Page 1. 地域安全学会論文集 No.2, 2000.11 1 阪神・淡路大震災からの生活再建7 要素モデルの検証 -2001年京大防災研復興調査報告- A Quantitative Verification of

the Seven Elements Model of Socio-Econmic Recovery from the Kobe Earthquake ...

引用元 3 - 関連記事 - HTML バージョン - ウェブ検索

gim 阪神・淡路大震災-神戸市の記録 199 神戸市 - 1996年 引用元 3 - 関連記事 - ウェブ検索

A search result of Google Scholar (articles for Hanshin-Awaji Earthquake)

阪神・淡路大震災による産業被害の推定







Conclusion

- Fusion of library classification systems and the Web 2.0 paradigm enables "A search engine of hints for information retrieval
 - Various viewpoints for information retrieval
 - Derivation from Web resources to reliable resources
- Complement each other
 - Web: any keywords can be uses as cues
 - Library: reliability and organization