





Search 2.0 & Expert Content

Special Libraries Association
Strategic Technology Alliance Series
Washington, D.C., June 16, 2009

*Carol Knoblauch,
Product Manager
Open Text, Inc.*

- About Search
 - About Content
 - About User Behavior
 - Search Technology Review
 - Social Search vs Semantic Search
 - Next Generation Search in Libraries
-
- 

About Open Text -- Booth #1545



16 years of experience
Over 45,000 customers
Hundreds of content management
deployments tailored for
libraries and archives
ONE enterprise-class
integrated library system

**What makes Open Text
The Content Experts?
We believe the numbers
speak for themselves.**

Trust Open Text to help you manage
your library and digital collections
and get the most value from
your content.

OPEN TEXT
The Content Experts™
www.opentext.com/library

SLA Booth #1025
basis_sales@opentext.com
800-328-2648 or 614-658-3836

Enterprise Content Management

Document management

Web content management

Digital asset management

Communities of Practice

Content lifecycle management

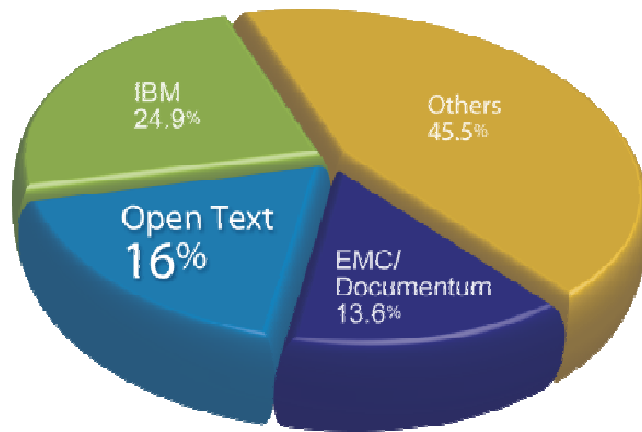
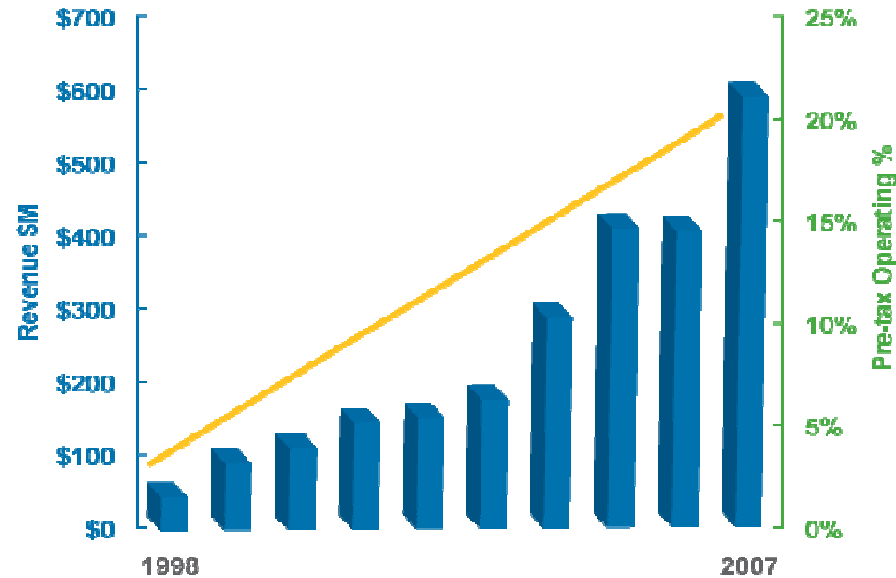
Enterprise archives

Library automation

More...

Where Are We Today?

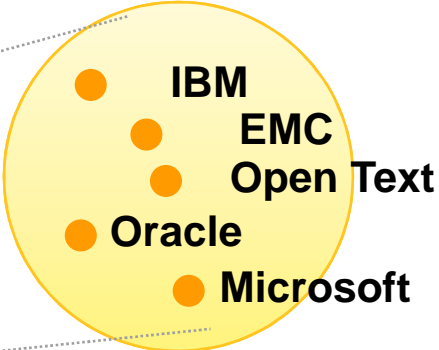
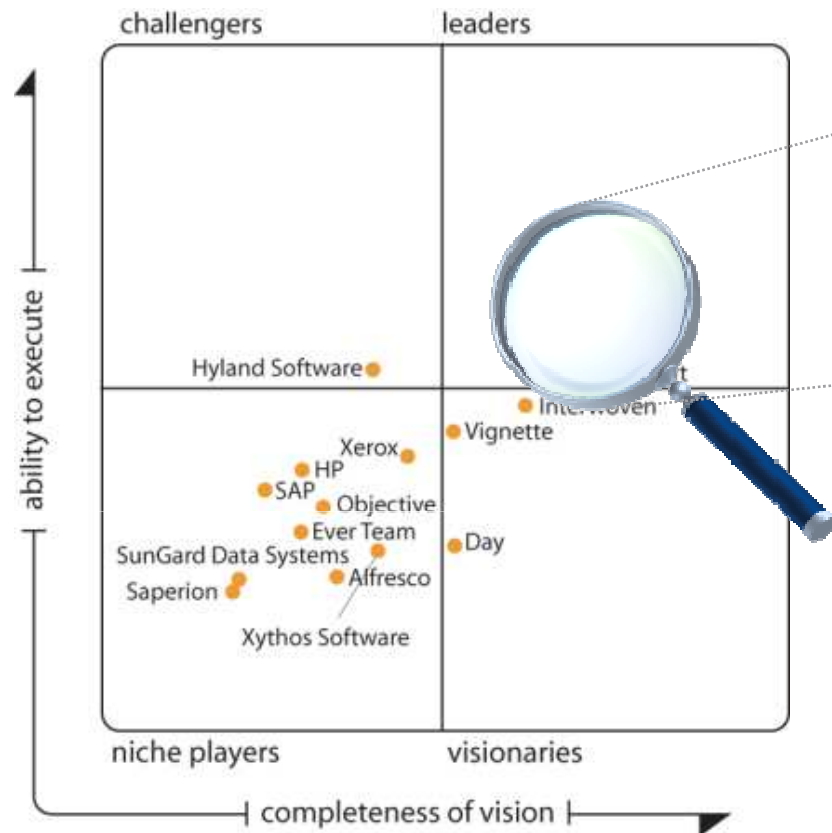
- More than 3000 staff globally
- Publicly Traded: NASDAQ, TSX
- FY2008: Revenue \$725.5 million



Market Share: Enterprise Content Management Software, Worldwide, 2007, Tom Eid and Bianca Granetto 28 May 2008. Gartner Inc.

- 16% Global Market Share
- Largest Independent ECM Vendor
- Recognized as the leading ECM market visionary

ECM 2008 Magic Quadrant Leader



- Open Text has the most “complete vision” in the market second only to Microsoft
- For the third year in a row, Open Text is one of the top three ECM vendors in “ability to execute”
- Open Text is the ONLY independent pure-play ECM vendor in the leaders’ quadrant. ECM is not a sideline ... it’s our only business


Gartner

Magic Quadrant for Enterprise Content Management, 23 September 2008 Karen Shegda, Toby Bell, Kenneth Chin, Mark R. Gilbert and Mick MacComascaigh Gartner, Inc.

About Search



Search is a commodity... and other lies

- Navigation is all you need
 - Keyword search is good enough
 - Enterprise search tools are all alike
 - It is an integrated part of business applications
 - Search is search is search
 - All search engines are the same
-
- 


About Search, Content, and Users

- Many types of content
- Many reasons to search
- Many users with different needs

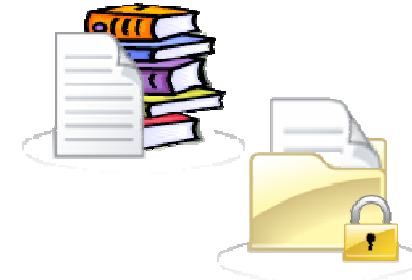
Structured

- Databases
- Forms
- Bibliographic records

Unstructured

- No tags
 - No organization
 - No fields
 - Full text
 - Digital objects
-
- 

What Content?




Collaborative Content <Work Product>	Business Documents	Expert Content <Intellectual Capital>
Created by and consulted when knowledge workers collaborate	High-throughput and highly repetitive business processes	Largely unstructured content; rigorous approval/release requirements
Project plans, Meeting notes	Invoices, Orders, Claim forms	Technical reports, Conference papers
Email, IM, Blog	Payroll, Personnel	White papers, Program summaries
Workflows & Discussions	Policy & procedures	Commercial content: analysts reports, journals, articles
Ephemeral	Archive for compliance	Archive & preserve to sustain research, innovation, culture

Why Search?

- Navigational Search
 - Find resources to do your job
 - Improve productivity
 - Provide access to knowledge

 - Research

 - Business Intelligence
 - Consumer Search
 - E-Discovery
-
- 

Business Content

- Answer a question
 - What is the per diem for food in Washington, DC?
 - What are the terms of this agreement?
 - How do I change the online Help?
 - What is the procedure for computer purchases?

<FIND>

Expert Content

- Gain knowledge
- Develop skills
- Invent products
- Understand complex topics
- Cure disease

<SEARCH>

Keyword Index

Term id	Members	References	Term
1	2	2	ACCOMPLISHED
2	4	4	ACCORDING
3	2	4	ACCOUNT
4	3	3	ACCOUNTING
5	1	1	ACCROSS
6	5	5	ACCURATE
7	1	1	ACCUSTOMED
8	3	3	ACITIVITY
9	1	1	ACITIUVY
10	1	1	ACM
11	2	2	ACQ
12	6	6	ACQUISITION
13	2	2	ACQUISTION
14	4	4	ACRIVITY
15	2	2	ACROBAT
16	2	2	ACROSS

Term id	Members	References	Term
1	1	1	MAINAPPLICATION
2	2	2	MAINFRAME
3	2	2	MAINLY
4	1	1	MAINT
5	4	5	MAINTAIN
6	1	1	MAINTAINABILITY
7	3	3	MAINTAINED
8	3	3	MAINTAINING
9	1	1	MAINTANENCE

Term id	Members	References	Term
1	13	13	M
2	2	2	M204
3	1	1	MA
4	8	8	MAB
5	10	12	MAC
6	1	1	MACC
7	1	1	MACCHIA
8	13	13	MACHINE
9	1	1	MACHINE'
10	4	5	MACINTOSH
11	1	1	MACO
12	1	1	MAD
13	2	3	MAGAZINE
14	1	1	MAGMT
15	1	1	MAGNITUDE
16	4	4	MAI
17	11	11	MAIL
18	1	1	MAILED
19	3	4	MAILING
20	1042	1059	MAIN

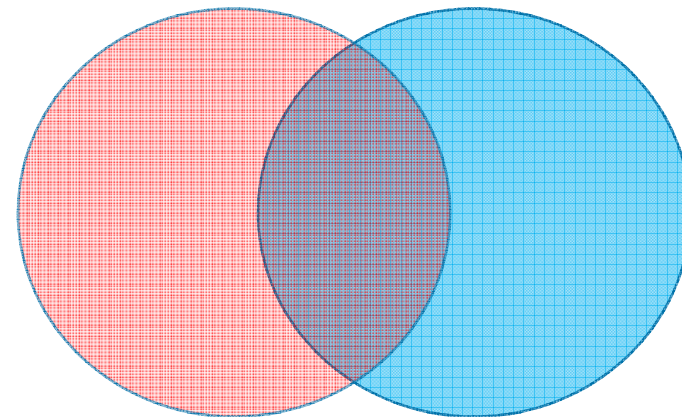
Term id	Members	References	Term
15			MAINTENANCE
1			MAJCOM
1			MAJCON
3			MAJOR
4			MAKING
3			MALAYSIA
2			MALAYSIAN
2			MANAG
13			MANAGE
4			MANAGED
1			MANAGEM

FQM> more

Term id	Members	References	Term
1	1409	1435	ACTIVITY
2	67	82	ACTUAL
3	1	1	ACTUALLY
4	1	1	ACUATIC
5	1	1	AD
6	1	1	ADAPTATIO
7	1	1	ADBSOFTWA
8	3	4	ADD
9	5	6	ADDED
10	3	3	ADDITION
11	4	4	ADDITIONA
12	1	2	ADDRESS
13	1	2	ADEQUATE
14	1	1	ADEQUATELY
15	2	2	ADMIN
16	1	1	ADMINISTRATION
17	1	1	ADMINISTRATION
18	1	1	ADMINISTRATION
19	123	125	ADMINISTRATION
20	7	7	ADMINISTRATIVE

Term id	Members	References	Term
1	1	1	Z39
2	1	1	ZEALAND
3	1	1	ZIPCODE
4	1	1	ZONING
5	1	1	ZSOFT

- Algorithms
 - Boolean logic
 - Bayesian method of probability & statistics
 - Inference
- Methodologies
 - Natural Language
 - Semantic
- Add-Ons
 - Taxonomies
 - Thesaurus
 - Federated search



Natural Language Search

- Attempts to answer a question
- Understands words on page
- Understands language in search
- Interprets word meaning and order in query
- Works best with preselected resources answering anticipated questions.
- Ask Jeeves / ask.com
- Where is Mexico?



- Analysis of words to understand meaning
- Content needs to be structured or include metadata
 - RDF, OWL, XML (technical)
 - Semantic tagging (popular)
- Semantic engine processes metadata to gather knowledge about content to act upon at search
- Semantic Navigation
 - Create connections between content based on associations
 - Map so “this” relates to “that”

Search Add-Ons: Taxonomy

- Impose structure & organization on content
- Can be feature of search engine
- Can be applied as metadata using categories and attributes
 - Virtual folders
 - Multiple access points
- Use to facilitate browsing

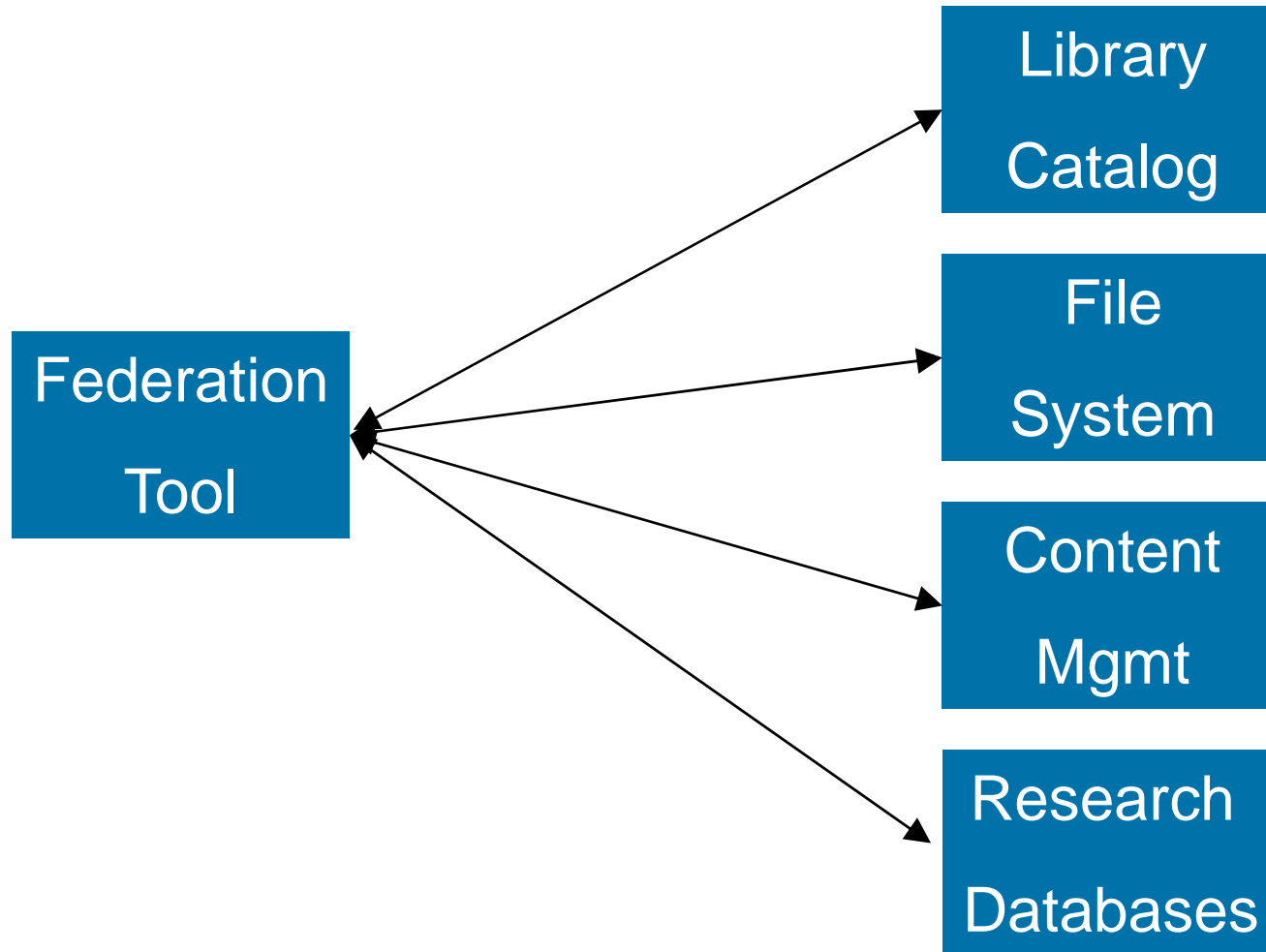
Search Add-Ons: Thesaurus

- Browse hierarchy
- Cross references
- Term switching
- Expand search
- Focus search
- Disambiguate

The screenshot shows a web interface for a thesaurus. At the top, there is a blue tab labeled "Relations". Below the tab is a search input field containing the text "Actions...". Underneath the search field is a list of related terms, each preceded by a small square icon with a minus sign. The terms are:

- LT / Knowledge Management
 - UF / CKM
 - UF / Collaborative knowledge management
 - UF / KM
 - BT / Organizational Culture
 - NT / ELearning
 - RT / Organizational Learning
 - RT / Organizational learning -- Management.
 - SN / A practice of applying formal rules and processes around the co


Search Add-Ons: Federated Search



- The process of analyzing text, extracting relevant information then transforming it into structured information that can be leveraged in different ways.

<Hurwitz Associates>

- In Search

- Analyze content and derive attributes to index as metadata
 - Analyze content on retrieval to rank or cluster results
 - Analyze user search to discern intended meaning
 - Provide summarization to aid user in identifying relevant results
-
- 

Text Analytics || Entity Extraction – Example

Information Professionals work for information organizations, which are defined as those entities that deliver information-based solutions to a given market. Some commonly used names for these organizations include libraries, information centers, competitive intelligence units, intranet departments, knowledge resource centers, content management organizations, and others.


4 concepts extracted		
Concept Nr	Weight	Value
1	100.00	Organization
2	93.49	information
3	87.49	intranet
4	43.24	content

4 concepts extracted		
Concept Nr	Weight	Value
1	100.00	Information Professionals
2	57.15	Organization
3	50.00	Content management
4	36.57	intranet


Function	Definition	Use
Affinity	Identify relationship or coincidence or entities	-Predict buying behavior -Cause and effect
Reputation	Assess attitudes toward person or organization	-Company reliability -Client satisfaction
Sentiment Analysis	Assess individual/group attitude toward product, person, any specific entity	-Evaluate travel destinations -Political predictions
Speech analytics	Process to text but can infer attributes like gender and emotion	-Call center support
Summarization	Abstract to capture essence of the content	-Reduce reading time -Identify key points -Cluster content

Discovery & Navigation



- Sort
 - Relevance Ranking
 - Topical relevance (term occurrence)
 - User relevance (timeliness, authority or novelty)
 - Link ranking
 - User activity ranking
-
- 

Faceted Navigation

**Search Results**POWERED BY
Livelink FCM

Quick Search: Go to...

Search within these results

Subject
Management (88)
Knowledge Management (22)
Industrial Management (10)
Personnel Management (10)
[More...](#)

Material Type
Book (53)
CD/DVD (9)
Video (8)
Document (6)
[More...](#)

Author

Year
1994 (5)
1999 (4)
1998 (4)
2004 (3)
[More...](#)

Search Results 1 - 10 of 88 for: manage* Sort by...

Information ecology : mastering the information and knowledge environment; and then there was more
Author: Davenport, Thomas H.
Published: 1997
Material Type: Book

Corporate memory : strategies for knowledge management
Author: Brooking, Annie
Published: 1999
Material Type: Book

Total time management [sound recording]
Author: Moskowitz, Robert
Published: 1975
Material Type: CD/DVD

Enterprise Content Management: What you need to know
Author: Jenkins, Tom
Published: 2004
Material Type: Book


The secrets of a dynamic subordinate that every manager should know : a success handbook for managers
Author: Crockett, William J.
Material Type: Book

Managing effectively in the world marketplace
Author: Dickson, Douglas N.
Published: 1983
Material Type: Book

Knowledge management and its integrative elements
Author: Liebowitz, Jay
Published: 1997
Material Type: Book

Magic Quadrant for Enterprise Content Management, 2004
Author: Shegda, K.
Published: 2004

Records management in the computer age
Author: Lundgren, Terry D.

**Search Results**POWERED BY
LiveLink FCM

Go to... ▾

Clusters

- Heart Disease** (36)
- Lung Cancer** (75)
- Laryngeal Cancer** (15)
- Cancer** (145)
 - Nicotine
 - Smoking
 - Tobacco
 - Health Care
- Health Care** (23)

Search Results 1 - 10 of 213 for: smoking cancer

- Effect of Diet and Smoking Invervention on the Incidence of Coronary Heart Disease**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- Secondhand Smoke and Lung Cancer**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- Tobacco, Alcohol, Asbestos, and Occupational Risk Factors for Laryngeal Cancer**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- Study Ties Nicotine to Cancer Spread**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- United States Sues Cigarette Companies to Recover Federal Health Care Costs**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- Risks, costs, and lives saved: getting better results from regulation**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- Lung cancer: principles and practice**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- Oxidation: The Cornerstone of Carcinogenesis: Oxidation and Tobacco Smoke**

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor ...
- Cancer: principles & practice of oncology**

Human Behavior: Do you visualize?

Do a shopping list right know !

How does it look like ?

- Tomatoes
- Bread
- Bottled Water
- Socks



Baumgartel, Martin. Advanced Visualizaiton of Search Results: More Risks, or More Chances?. Infonortics Search Engine Meeting, Boston, 17-18 April 2009.

Visualizing Search Results

The image displays three different ways to visualize search results for the query "Pittsburgh events".

Left Panel: Map View
This view shows a circular map with various categories represented by colored circles. A central green box contains the text: "Larger category circles contain more results. Click on any category circle to explore its contents. Click to continue...". Categories include "Information", "Pittsburgh Hotels", "Pittsburgh Steelers", "Pittsburgh Penguins", "Pittsburgh Pennsylvania", "Steel City", "Local", "Pittsburgh Travel", "Pittsburgh", and "Pittsburgh.net".

Right Panel: Detail View
This panel shows a search bar with "Pittsburgh events" and a "Search" button. Below the search bar, there is a list of search results with snippets and source information. The results include:
- **Rue21**: "Add to Working List | Post to del.icio.us | Bookmark | Email" followed by a snippet about offices and headquarters.
- **City of Pittsburgh**: "Add to Working List" followed by a snippet about the city's economy.
- **Andrew Carnegie**: "Add to Working List" followed by a snippet about the city's history.
- **Pittsburgh.net**: "Add to Working List" followed by a snippet about the city's economy.


Bottom Right Panel: Network Visualization
This panel shows a network visualization of search results. The central node is "events", which is connected to various other nodes representing related search results. The nodes include:
- "entertainment news", "sports", "guide", "city", "citysearch.com", "restaurants", "attractions", "information", "travel", "compare", "rates", "hotel", "Hotwire.com", "Pittsburgh.Hyatt.com", "www.visitpittsburgh.com", "en.wikipedia.org", "wikitravel.org", "wikipedia", "action", "pennsylvania", "features", "pittsburgh.net", "pittsburghlive.com", "review", "www.pittsburgh.com", "entertainment news", "sports", "penguins.nhl.com", and "11 700".

Search 2.0 – Social Search



Potential for Web Search

- Increase relevance
- Leverage network of trusted individuals
- Reduce impact of link spam
- Provide human guide through semantic tagging

- Relevance: collective intelligence vs. topical content
 - Social bookmarks, tagging, voting, and reviews
 - Privacy and competitive advantage
 - Ambiguity
 - Control
-
- 

What Library Users Want


- Direct links to online content, text, and media formats
- Evaluative content (i.e. abstracts, TOCs)
- Search results must be relevant
- Relevance must be obvious
- Item availability information
- Simple keyword search with an advanced search option


Online Catalogs: What Users and Librarians Want. An OCLC Report, 2009.

- Undergraduates were more favorable to user-contributed content and ‘quite discerning in their ability to distinguish authoritative from non-authoritative reviews.
- Academic or professional users tended to favor editorial reviews over end-user contributed reviews.
- Scholars were more interested in professional reviews and less interested in other user’s opinions on materials *unless they were experts*.

Online Catalogs: What Users and Librarians Want. An OCLC Report, 2009.

Trends in Search for Libraries

- Next generation user experience
 - New tools to manage digital collections
 - Incorporate Web 2.0 features
 - RSS
 - User tagging
 - Relevance ranking
 - Faceted navigation
 - Evolve to more closely resemble KM/ECM than the traditional ILS (in special libraries)
 - Add-on discovery tools
-
- 

- About Search
 - About Content
 - About User Behavior
 - Search Technology
 - Social Search vs Semantic Search
 - Next Generation Search in Libraries
-
- 



*Carol Knoblauch,
Product Manager,
Libraries & Collections
knoblauch@opentext.com*

www.opentext.com/library