Enterprise search trends and challenges

Gregory Grefenstette
Exalead

Final CHORUS Conference
Brussels, May 26-27, 2009
Enterprise search

From Wikipedia, the free encyclopedia

Enterprise search is the practice of identifying and enabling specific content across the enterprise to be indexed, searched, and displayed to authorized users.

Contents

1 Enterprise search summary
2 Differences from web search
3 Relevance factors for enterprise search
4 See also

Enterprise search summary

The term Enterprise search is used to describe the application of search technology to information within an organization. This is in contrast to the other two main type of horizontal search environment: web search and desktop search.
Enterprise Search represents a huge market opportunity for Search Technology

- $1.75B in 2007 → $2.55B in 2010 (Forrester 2008)
- 23%-25% annual growth

“Biggest growth will be from new breed of search vendors who develop search platforms that can scale for new paradigms of information access.”

These vendors will bring together vast amounts of structured and unstructured data in ways that allow users to make faster, better decisions.”
Enterprise Search Motivators

- Information Workers
- Compliance, eDiscovery
- Database Access Democratization
Information everywhere...

210 Billion
Emails per day

4.5 trillion
pages copied/printed per year

300TB
average size of a
data warehouse

80% of enterprise information unstructured

988 Exabytes
worldwide data volume in 2010

Source: IDC
Enterprise Search Motivators

- Compliance, eDiscovery
  - Sarbanes-Oxley, 2002 post-Enron
  - HIPAA, Health Insurance Portability and Accountability Act, 1996
  - FRCP, Federal Rules of Civil Procedure
- email equal to paper-based documents as valid and admissible in a court of law
- Obligation to store all business communication
- 40,000 email per year per employee
Enterprise Search Motivators

• Company Information in Databases
  – Vital company data protected in transactional systems
  – Optimized for protection not search
  – Problem: skilled users formulate DB queries
  – Solution 1: Data Cubes -- canned queries
  – Solution 2: Database offloading
Enterprise Search Motivators

- Compliance, eDiscovery: legal requirements
- Information Workers: masses of local information
- Database Access Democratization: access, silo smashing
Database Offloading Example, 1

- Only one search request to define

ODBC Connector (SQL request)

EXALEAD INDEX
Database Offloading Example, 2

Administrative interface, suite
outline

- What is Enterprise Search
- Differences Enterprise/Web Search
- Enterprise Search Needs
Difference 1: The Enterprise is not just "a small Internet"
Not just “web pages in – web pages out”
Difference 2: Not “web pages in – Web pages out”
Difference 4: Page Rank deflates behind firewall
Redzai Suroi was 59. Suroi died Thursday near Guadalajara as he returned alone to Madrid from a private trip, said embassy counselor Zoran Raicevic. Suroi, a native of Erizren in the autonomous province of Kosovo, had held the post of adjutant to the Yugoslav foreign minister before coming to Spain in October 1985. He also served as ambassador to Mexico from 1978 to 1982 and to Bolivia from 1970 to 1974, Raicevic said. Suroi, a law graduate of Belgrade University, worked as a journalist for 15 years, and became director of Radio Pristina in Kosovo before beginning his diplomatic career in 1970, Raicevic said. Survivors include his wife, a son and a daughter, the counselor said.
Blog buzz vs Bug reports
Difference 4: Single Shot Relevancy vs Faceted Navigation
# Difference 4: Single Shot Relevancy vs Faceted Navigation

**PREZZO**
- 700 €
- 15 000 €
- 30 000 €
- 50 000 €
- 70 000 €
- 100 000 €
- Da concordare

**CHIOMETRI**
- 10 000 km
- 20 000 km
- 40 000 km
- 60 000 km
- 80 000 km
- > 100 000 km
- Non specificato

**ANNO**
- Ante 1990
- 1991-1995
- 1996-2000
- 2001-2002
- 2003-2004
- 2005
- 2006
- 2007
- Dopo 2007
- Non specificato

**MARCA**
- Ferrari
- Maserati
- Lamborghini
- Mercedes
- Porsche

**COLORE**
- Rosso
- Grigio
- Blu
- Argento
- Nero
- Giallo
- Bianco

**ALIMENTAZIONE**
- Benzina
- Diesel

**FOTO**
- Solo con foto

**DISTANZA**
- 5 km
- 10 km
- 20 km
- 50 km
- Provincia
- Regione
- Tutta Italia
Difference 5: User Experience: High Level
Difference 6: identified users
Difference 7: sliders, bar graphs, pie charts
What is Enterprise Search
Differences Enterprise/Web Search
Enterprise Search Needs
Enterprise Search Needs

• Depth and breadth ➔ connectors
• More varied semantic typing / text layers
• Database offloading
• Multimedia processing
• Proactive search
Connectors
# The Data

## Open Architecture of Connectors

### INTERNAL SOURCES
- HTTP, File system, Databases (ODBC)
- Lotus Notes, Exchange, LDAP
- Mail server, Newsgroup
- Documentum and E-room
- Livelink (OpenText), Webdav
- Microsoft Sharepoint
- **Universal APIs**

### EXTERNAL SOURCES
- www.exalead.com
- External databases
- A universe of pre-defined clients

## A Universal Model, Data

- More than 400 Formats
  - File types
    - Word, Excel, PowerPoint, Access, WordPerfect, Adobe PDF, RTF, JPEG, PNG, MP3, ZIP, RAR, etc.
  - Types of data:
    - Unicode (UTF), specific coding (Arabe, Chinese, Korean, Japanese and Russian)

---

**Exalead Desktop**: Microsoft, Lotus Notes, etc.
Enterprise Search Needs

- Depth and breadth ➔ connectors
- More varied semantic typing / text layers
- Database offloading
- Multimedia processing
- Proactive search
Enterprise Search Needs

• Depth and breadth ➔ connectors
• More varied semantic typing
  – each type is a facet dimension
  – merging policy for types
    • addition, counts, list, sorted list
  – Mashups
• Database offloading
• Multimedia processing
• Proactive search
Nicolas Sarkozy: Country Leader, Lawyer, and Nobleman

Nicolas Sarkozy is the 23rd President of the French Republic and ex-officio Co-Prince of Andorra. He assumed the office on 16 May 2007 after defeating Socialist Party candidate Ségolène Royal ten days earlier.

Description from Wikipedia

Top connections

NATO
France
Felipe Calderon
Carla Bruni
Angela Merkel
Mexico
Silvio Berlusconi

Dig Deeper

Filter by top connection
France
NATO
Felipe Calderon
Carla Bruni
Angela Merkel
Mexico
Silvio Berlusconi

Filter by category
Country Leader
International organization
European Union
NATO
European Parliament
Mediterranean Union
United Nations

Country
Model
Politician
Defense Company
Newspaper

Filter by activity
returning
meeting with
supports
holding
leading
finishing
announce

French govt faces no confidence vote over NATO

...tar the debate, the parliament will vote whether it has confidence in the government. Sarkozy's plan to rejoin the NATO command will not be put to a vote, however. If it had, the vote might have been close, because a few dozen UMP members are opposed...

1 hour ago | AP Online | See the Evri Widget in Action

Sarkozy to endorse French return to NATO

Sarkozy to endorse French return to NATO Sarkozy: France will return to NATO...

6 days ago | CNN

Sarkozy to end France's 40 year NATO feud

Sarkozy to end France's 40 year NATO feud Sarkozy to end France's 40 year NATO feud...

5 days ago | The Hindu - News

Show more

Pictures

Videos
Blue Smoke
110 E 27th St, New York, NY 10016, USA
Phone: (212) 447-7733
Subway: 6 at 23rd St.
Website: http://www.bluesmoke.com - Book Online!
Kind of food: Barbecue, American, American Traditional, Southern & Soul, American (Traditional)

Overview

Review
Forget what the clue quibblers say? Blue Smoke is an ingenious Manhattan barbecue joint with a woody barroom full of red-vinyl booths, a skylit dining room, and a hot jazz club downstairs. True, some of the barbecue isn’t there yet, but each visit shows improvement, as pit master Kenny Callaghan breaks in those Missouri-made smokers. Salt-and-pepper beef ribs are terrific; so are sides and desserts, and as far as pit-smoked fare goes, this ain’t least Freestump, Texas, after all.

Most viewed comment
I went to Blue Smoke tonight. The waitstaff was friendly, even if the last sip of my beer was whisked away without so much as an “Are you finished with that?” Had the baby back ribs, with a side of fries, mac & cheese and hash-puppies. The remodel that comes with the hash-puppies makes this dish a real specialty. The mac & cheese was creamy, to put it mildly; very tasty, but not for the hint of chemically Velvita-esque after-taste. The fries - excellent; they arrived fresh, hot, with a crispy exterior (no batter on these little critters) and with the genuine taste of potato without . . . More >>

From New York Magazine

Bloggers entries about Blue Smoke
• Restaurant Week '06: Chase away the blues at Blue Smoke (New York City's Hottest Restaurant Deals)
• Summer Restaurant Week 2007 Pick: Blue Smoke (New York City's Hottest Restaurant Deals)
• Hill Country vs. Blue Smoke, Run On Their Own Turf (Grub Street)
• Restaurant Week Picks: Winter 2003 (Part 2) (New York City’s Hottest Restaurant Deals)
• The Big Apple BBQ Block Party Demands a Lot From its Vendors (Grub Street)
• Meyer Makes a Move on Off Field (Grub Street)
• Robbie Richter and Hill Country Part Ways (Grub Street)
• B.K. Guest’s New BBQ: Will Wildwood Succeed? (Grub Street)
• Heroic Blogger Saves Burger Lovers From Winter of Discontent (Grub Street)
• Profile: Midtown Luncher “vanessa” & Big Apple BBQ Fast Pass Giveaway! (Midtown Lunch)
• PROFILE: Midtown Luncher “Chris” (Midtown Lunch)
• Daisy May’s BBQ Cart... the best-scratch-think-ONLY BBQ in Midtown (Midtown Lunch)

Details
Payment options: AMEX, Diners Club, Discover, MasterCard, Visa

Opening Hours

<table>
<thead>
<tr>
<th>Day</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
<th>Sunday</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11:30am - 11pm</td>
<td>11:30am - 11pm</td>
<td>11:30am - 11pm</td>
<td>11:30am - 11pm</td>
<td>11am - 1pm</td>
<td>12pm - 1pm</td>
<td>12pm - 11pm</td>
</tr>
</tbody>
</table>

People sentiments

reasonable okay spicy inspired hot baked cheap habilat familiar good famous improved boiled impressed key amazing consistent negative expensive superb efficient endless tiny meaty good nasty available genuine not typical picky yummy good popular outer delicious tasty
Enterprise Search Needs

- Depth and breadth ➔ connectors
- More varied semantic typing / text layers
- Database offloading
- Multimedia processing
- Proactive search
Enterprise Search Needs

• Depth and breadth ➔ connectors
• More varied semantic typing / text layers
• Database offloading
• Multimedia processing
• Proactive search
 BBC News US (29/7/2007)  
0m 1s ... every passing day and suicide attack the pressure on George Bush to get council Iraq proves that America's middle eastern allies here literally would lead ...

0m 6s ... stuck in this part of this the push by George Bush to rescue his Iraq strategy by putting more American boots on Iraqi Graham ...

 CNN (4/9/2007)  
0m 16s ... less than two weeks before he delivers a crucial Iraq report to Congress President Bush hinted at a possible future troop drawdown general conterras in ...

 CNN (16/8/2007)  
0m 23s ... as you may have heard the Bush administration is getting ready to add Iran's revolutionary guards a ...
0m 47s ... terrorists it's accurate they're after all killing our troops in Iraq but I'd also send a message that this administration has and forget about the policy that brought us to Iraq in the first place that policy is the Bush doctrine ...

 CNN (13/8/2007)  
1m 11s ... the final in a year and a half of the Bush white house ...
1m 23s ... a time of the war on terror the war in Iraq but I don't think any worse than before but he's ...
Enterprise Search Needs

- Depth and breadth ➔ connectors
- More varied semantic typing / text layers
- Database offloading
- Multimedia processing
- Proactive search / hidden search
Each click generates a search
- What is Enterprise Search
- Differences Enterprise/Web Search
- Enterprise Search Needs
Conclusions
More semantics