

Search for a Search Engine: Selecting and Implementing Corporate Search

Neil Margolis, Ph.D.

Manager, Information Resources

Wyeth Research

April 16, 2002

The Information Retrieval Challenge

Searching is not the same as finding.

Finding Information at Wyeth

- **Wyeth staff spend an average of 4 hrs/week looking for information they need to do their jobs.**

Internal survey

- **Finding and using information is a major part of many people's jobs.**
- **Corporate employees in the pharmaceutical industry spend 4 hours/week on average.**

Outsell, Inc, Super I-AIM report, 2001

The problems with search

- How do I find what I'm looking for?
- Where is the information I need?
- How do I know where to look?

- What's my password?
- How do I use *this* system?
- What terms should I use in my query?

Information is everywhere

**2. Organize
Intelligently**
*contextual
focus,
profiles,
relevance*

Hitlist

[A Prescription for Knowledge Management
What Hoffman-Laroche's Case Can Teach
Others](#), Seemann, Patricia, Ernst and Young

[Creating Fertile Ground for Knowledge at
Monsanto](#), Junnarkar, Bipin, Director of
Knowledge Management, Monsanto

[Working Knowledge: How Organizations
Manage What They Know](#), Davenport, T.H.
and L. Prusak

Davenport, T.H. and L. Prusak. 1998. [IBM](http://www.uk.ibm.com/services/news/km.html)
<http://www.uk.ibm.com/services/news/km.html>

[Knowledge Research Institute, Inc](#)
<http://www.knowledgeresearch.com>

[The Gartner Group](#)
<http://advisor.gartner.com>

[The Delphi Group](#),
<http://www.delphigroup.com/km/index.html>
Anderson, Janice Capel and Moore, Stephen
D. *Mark & Method: New Approaches in
Biblical Studies*, Fortress Press, 1992.

*Apicella, Raymond, *Journeys Into Mark*, St.
Anthony Messenger Press, 1990.

*Apicella, Raymond, *Journeys Into Mark*, St.
Anthony Messenger Press, 1990.

Bausch, William. *Pilgrim Church: A Popular
History of Catholic Christianity*, Fides
Publishers, 1977.

1. Search across repositories

Other
???

Expert
DB

E-mail

Lotus
Notes

OVID

EDMS

Inter-
net

intra-
nets

DBs

Undefined repositories

Minor repositories

Major repositories

Databases

In the beginning...

...there were requirements.

Search requirements

- **Multiple repositories**
 - Intranet, Documentum, Internet, Oracle, Lotus Notes
- **Concept-based**
- **Full-text/ unstructured text**
- **Relevance ranking**
- **Security**
- **Ease of use**

I just bought a new car...



- **Decide what features you want/ need**
- **Look at the options to see which cars fit your needs**
- **Test drive those cars**
- **Pick the one you like best**

Search Engine Selection

It depends on your perspective...



- Requirements prioritized for vendors and evaluators
 - user needs, technical info, financials
- Vendor selection
 - RFP
 - Initial Visit
 - Designed test for evaluation
 - Reference calls and visits
 - Final visit: evaluation test

Pilot Design and Execution

Pilot Objectives

- **Can users search across multiple repositories with a single query?**
- **Does the tool return relevant search results?**
- **Is this tool easy to use?**
- **Does the tool save users time?**
- **Technical needs for production**
- **User needs for production**
- **Purchase Recommendations**

Choosing the Pilot Team

- Range of information needs
- Range of computer/ search experience
- Multisite teams
- Expressed need for multi-repository search
- Indexed resources were chosen with the pilot teams
 - Intranet, selected Internet sites, Documentum, Wyeth's Product Literature, e-journals

Pilot Design

- **4 week pilot**
- **Pilot kickoff meeting**
- **2 office visits**
 - successes, comfort level, skill evaluation
- **Survey for each search performed**
 - satisfaction, time-savings, information found?
- **Post-pilot survey**
 - ease of use, relevance of information, satisfaction, time-savings, recommendations

Pilot Metrics

- **Usability and Usefulness:**

- search logs
- interviews
- post-pilot survey

- **Technical:**

- stability
- speed
- customizability
- administrative requirements

Pilot Results

- **The users told us what we need to bring this system into production.**
- **The IT group told us what we need to support the system.**
- **Our recommendation was based on fulfillment of these needs.**

Pilot Success Factors

- **Pick representative pilot teams**
- **Meet with teams to understand user needs**
- **Communicate your testing needs to pilot teams**
- **Manage user expectations**
- **Heed user feedback**
- **Make purchase recommendations based on pilot results, not vendor selection results**

End User Expectations

“Users are stupid, lazy and evil.”

Anonymous information professional

Familiarity Breeds Contempt

- Everyone has search engine experience on the Internet, so users will be hard to impress. They expect a search engine to find useful information.



Why do I keep getting 500 results?

- It bothers many users to see how many results they get.
- It bothers many users when the best hit they find is #4 on the hit list, rather than #1.

Search of Dreams?

If you query, they (the right answers) will come...



Many users believe their query will identify the most relevant result(s).

If they don't find what they need, they give up.

Search Engine Deployment

**Core Strategy
Team
(People)**

**Deployment
Team
(Technical)**

- **Designed UI's**
- **Developed dictionary**
- **Selected info sources**
- **Tested relevance ranking algorithm (acceptance criteria)**
- **Marketing/ education**
- **Identified metrics**

- **Designed and implemented architecture**
- **Created working UI's with dictionary, information sources, algorithm**
- **Developed maintenance and operational processes**

The Outcome

- **The search engine is now live:**
 - **in the first month, the number of queries has increase compared to the previous search engine.**
 - **search engine speed is impressive to people.**
 - **some have reported that they can now find what they're looking for.**

The Future: Finding, not Searching

- **Taxonomy**
- **Clustering/ Categorization**
- **Did you mean...?**
- **Education**
- **More user control**
 - **term definitions, dates, pattern matching, fielded search, etc.**