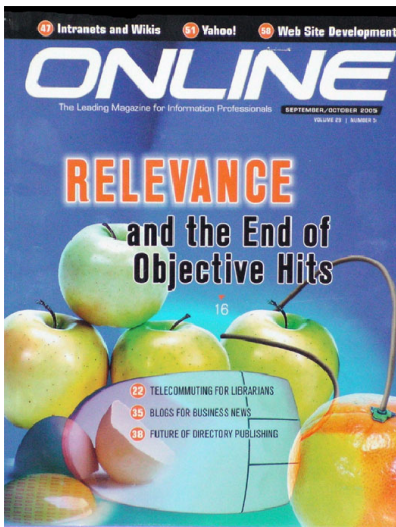


The End of Relevance and Objective Search Results

© Stephen E. Arnold, Postal Box 320, Harrod's Creek, KY 40027
Email: sa@arnoldit.com – Voice: 502 228 1966

Cover Story: Relevance Gone?



Google Legacy News Search



[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#) [Desktop](#) [more »](#)

[Advanced News Search](#)
[Preferences](#)

News

- Top Stories
- World
- U.S.
- Business
- Sci/Tech
- Sports
- Entertainment
- Health

News Alerts

[RSS](#) | [Atom](#)
[About Feeds](#)

[About](#)
[Google News](#)

Results 1 - 8 of 8 for **google.legacy** (0.12 seconds)

Sorted by relevance [Sort by date](#)


[Google's plan for a computing platform to dethrone Microsoft](#)
InfoWorld - Sep 21, 2005
 ... In the ebook, "The **Google Legacy**, How Google's Internet Search is Transforming Application Software," Arnold says Google is about much more than search. ...

[GoogleNet and the Internet Age](#)
ZDNet - Sep 21, 2005
 GoogleNet and the Internet Age by ZDNet's Dan Farber -- Elinor Mills of news.com writes about Stephen Arnold's "The **Google Legacy**: How Google's Internet Search ...

[Google's grand ambition](#)
ZDNet - Sep 21, 2005
 ... computing platform and could be about to unseat Microsoft from its throne," Arnold writes in a summary of his book, "The **Google Legacy**: How Google's Internet ...

[Google: Time for world domination?](#)
ZDNet UK, UK - Sep 21, 2005
 ... computing platform and could be about to unseat Microsoft from its throne," Arnold writes in a summary of his ebook, The **Google Legacy**: How Google's Internet ...

Google Legacy Web Search



[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#) [Desktop](#) [more »](#)

[Advanced Search](#)
[Preferences](#)

Web

Results 1 - 10 of about 562 for "**google legacy**". (0.23 seconds)

[Arnold: The **Google Legacy** \(ebook\)](#)
 Google, search, internet future, ebook, e-book, search engines, email, office software, mapping software.
www.infonortics.com/publications/google/google-legacy.html - 19k - [Cached](#) - [Similar pages](#) - [Remove result](#)

[Google Effect: Search Engine Meeting 2005 Tutorial](#)
 Tutorial: Sunday April 10, 2005 (afternoon). The Google Effect: Redefining the Landscape of Search. Stephen E. Arnold President ...
www.infonortics.com/searchengines/sh05/slides/05tutorial-arnold.html - 3k - [Cached](#) - [Similar pages](#) - [Remove result](#)
[\[More results from www.infonortics.com \]](#)

[The **Google Legacy** - eBook Released | Threadwatch.org](#)
 The **Google Legacy** - eBook Released. By: Nick W [privmsg - website] On 31st Aug 2005 In Publishing ... Source: Ifnonortics release e-book: The **Google Legacy** ...
www.threadwatch.org/node/3662 - 13k - [Cached](#) - [Similar pages](#) - [Remove result](#)

[Google Legacy order form](#)
 The **Google Legacy**: How Google's Internet Search is Transforming Application Software. Stephen E. Arnold ... I wish to order access to The **Google Legacy** ...
<https://site2611.mysite4now.com/samold2/https/google-order.html> - 6k - [Cached](#) - [Similar pages](#) - [Remove result](#)

What We'll Cover

- Situational relevance
- Interfaces... growing importance
- Internet vs Intranet relevance
- You can cheat

Forces for Relevance

- Concerns about time wasted looking for information
- Under pressure, users accept “good enough”
- Web search
 - Voting
 - Popularity
- Intranet search
 - Algorithms that “read”
 - Human intervention

Relevance... Underseige

- *Dot Net* reported in August 2005:
 - Overlap among Ask Jeeves, Google, and Yahoo on 10,000 queries was as low as three percent
 - Coverage of search engines and their query matching algorithms are diverging
- 90 percent of search users
 - Cannot differentiate between “organic” search results and ads (“inorganic search results)
 - Click on “what seems to be the best answer”
 - Do not recognize when searches run automatically and display results in a work-related system
- What is “relevance”? – What is beauty?

Relevance... A Fuzzy Black Box



Precision and recall give way to “situational relevance”

The Black Cube's Needs

Need for users

Need for
traffic (clicks)

Need for
partners

Need for
revenue



Need to respond to
competition

Need for
buzz

Intranet Search

Content must be in the system

+

Suggestions in the interface

+

Yahoo style plus work flow
hooks

Internet Search
Web sites must be in Google
+
Appear in first page of results
and
In ads on the first page of results
but
Tricky and resource intensive

Web Search

- **Best results:**
 - Click a hard wired “topic” like News, Weather, etc.
 - Use Boolean syntax
- Google / MSN / Yahoo improving deduping
- Becoming more alike even though overlap among results lists range from five percent to 50 percent in our tests
- Conclusion...

The Obvious

What we've known all along...

If you know what you are looking for, a trained searcher can find relevant information

The Less Obvious

What we've known all along...

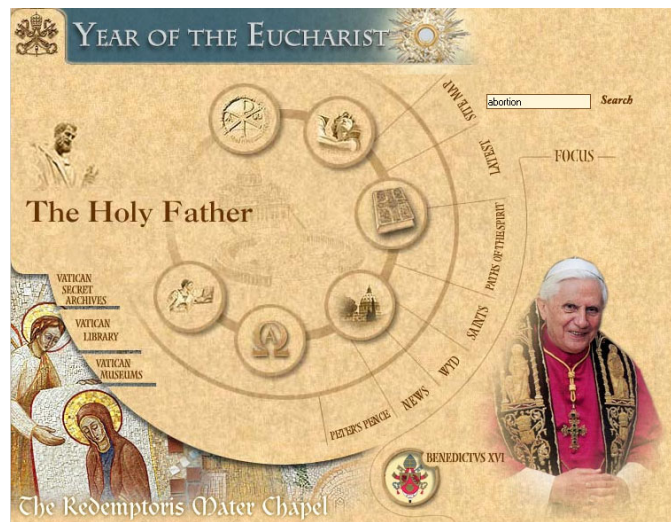
Users who take what is provided accept "good enough" information

Good enough is now dominant in search

The User Interface Gets Training Wheels

- Search takes time
- People have a hard time coming up with a query
- Long results lists are ignored
 - Clustering: FirstGov.gov will use MSN spider and Vivisimo federation engine with clustering
 - Canned queries: Push results so no search is needed
 - Appliance solution: Google Appliance, Index Engines, or EPI Thunderstone appliance

Mondosoft “Hybrid” Interface



Hybrid Search: Facets, Hard Coding, Synonym Expansion

The screenshot displays a search interface with several facets and filters. On the left, there are two main facet categories: 'All categories' with 298 items and 'Vatican Offices' with 119 items. Under 'All categories', there are sub-facets for 'The Holy Father' (144), 'Reference' (13), 'Calendar' (8), and 'News' (14). Below these, there are search results for 'partial-birth abortion' and 'evangelium vitae' with their respective dates and sizes. On the right side, there is a search bar with the word 'abortion' entered and a 'Search' button. Below the search bar, there are filters for 'Return results from' (last week, last month, All), 'Select category' (All Categories), 'Return documents by author' (Any author), and 'Return documents by title'. At the bottom right, there is a 'Language' filter showing counts for German (2), Spanish (4), French (3), Italian (4), Portuguese (2), and English (90%).

What's Needed for "Faceting"

- Human involvement
 - Stop word list essential
 - Synonym list
- Software tools
 - Data Harmony's controlled vocabulary editor
 - Tools included in ISYS 7, Stratify, Endeca, Mondosoft and others for interacting with the search system
- Metrics and usage analysis
- Time
- Resources

“Relevance” via Cheating

- Linkfarms
- Fool the Googlebot with a doorway page
- Blog spam
- Duplicate content
- Etc.

Summary

- Many options in search... more coming
- Focus on basics
- Remember the principles of good librarianship:
 - Collections
 - Accuracy
 - Currency
 - Selective depth
- Read authoritative material; short cuts will:
 - Get you in the sandbox or blacklisted
 - Lower your ranking
 - Waste your money unless you are informed



**ARNOLD
INFORMATION
TECHNOLOGY**

*Postal Box 320
Harrod's Creek
Kentucky 40027*

**Management consulting and
strategy**

**"The Enterprise Search Report"
2nd edition now out. Order at
www.cmswatch.com**

**"The Google Legacy" became
Available earlier this month
Order at www.infonortics.com**

**Additional information at
www.arnoldit.com/sitemap.html**

Contact: sa@arnoldit.com