

Enterprise Search: What You Must Know about Information Retrieval and the “Google Effect”

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



**ARNOLD
INFORMATION
TECHNOLOGY**

*Postal Box 320
Harrod's Creek
Kentucky 40027*

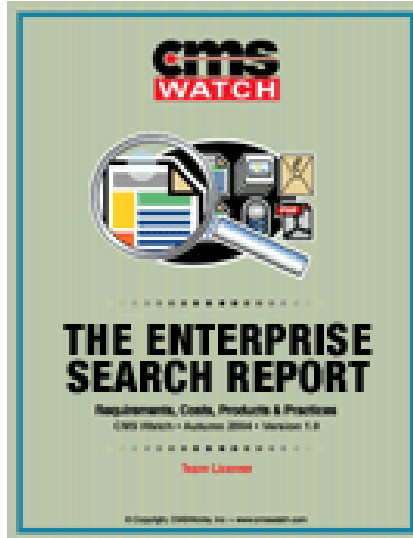
Management consulting and strategy

“The Google Legacy” became Available earlier this month
Order at www.infonortics.com

“The Enterprise Search Report” 2nd edition now out. Order at www.cmswatch.com

Additional information at www.arnoldit.com/sitemap.html
Contact: sa@arnoldit.com

Limited Access Document



6 Costs

Key Points

- Cost—No free lunch
- Effectiveness—Depends on staff, system, and set up
- Control
 - The more data the higher the return
 - Adding subjectivity softens results
- Red dot = hockey-stick cost factor

Costs for Search

- Infrastructure
 - Small: \$5,000 (more of a “test”)
 - Medium: \$100,000 (risky even with a Google Appliance)
 - Large: \$1,000,000 (less risk)
- Licenses / code: requirements drive cost
- Staff (FTE or outsourced)
 - Small: 1.5 persons
 - Medium: 3 – 5 persons
 - Large: > 6 persons

Low Cost Framework (1)

<i>Software</i>	\$900
 <i>People</i>	\$7,200
<i>Hardware</i>	\$340
<i>Network Infrastructure</i>	\$3,000
<i>Services and Consultants</i>	\$480
<hr/> Total	\$11,920

People Example

**Don't know what we don't know ...
until it is too late.**

Low Cost Framework (2)


<i>Software</i>				
<i>Item</i>	<i>Usually Required?</i>	<i>Comment</i>	<i>Est. Cost</i>	<i>Assumptions</i>
License fee	Yes	Standard license fee	\$800	
Maintenance	Maybe	No		
Utilities	Maybe	Nero DVD burning software	\$100	
Database	Maybe	Company has SQL Server and is Microsoft-centric		
Visualization tool	No	No		
Hit boosting tool	No	No		
Upgrades	Yes	No		
Total			\$900	



Database

It is not the software; it is the design, administration, and upkeep that costs money.

Low Cost Framework (3)



<i>People</i>				
<i>Item</i>	<i>Usually Required?</i>	<i>Comment</i>	<i>Est. Cost</i>	<i>Assumptions</i>
Procurement team	No	No		
Search administrator	Yes	One IT professional	\$6,000	\$60,000 per year at 10% of time
Engineer	Maybe	No		
Programmer	Yes	No		
Taxonomy specialist	No	No		
Indexing specialist	No	No		
Business process engineer	No	No		
System architect	No	No		
Security engineer	Maybe	Uses Microsoft server and XP security	\$1,200	\$60,000 per year at 2% of time for one-week security review.
Total			\$7,200	

Indexing

There is a reason why people get degrees in indexing, taxonomy design, and index maintenance...

Bad indexing means bad search results and very unhappy users.

Low Cost Framework (4)




<i>Hardware</i>				
<i>Item</i>	<i>Usually Required?</i>	<i>Comment</i>	<i>Est. Cost</i>	<i>Assumptions</i>
Servers for document processing	Yes	Repurpose one dual processor Xeon with one gigabyte of RAM and two 300 Gb IDE drives	\$240	Set up for one day of IT professional
Servers for query processing	No	No		
Servers for staging	No	No		
Servers for development	No	No		
Storage for document processing	No	No		
Storage for value-added indexing	No	No		
Storage for index or indexes	Yes	No		
Load balancers	No	No		
Other hardware	Maybe	No	\$100	DVD burner
Administrative workstation	No	System administrator can access the Xeon running search from their XP workstation		Maintenance covered in 10% time allocation above
Total			\$340	

Hardware

Search consumes huge amounts of computing resources. Without resources, search systems fall over.

Low Cost Framework (5)

<i>Network Infrastructure</i>				
<i>Item</i>	<i>Usually Required?</i>	<i>Comment</i>	<i>Est. Cost</i>	<i>Assumptions</i>
Intrusion detection	Maybe	No		
Firewall	Maybe	No		
Routers	Maybe	Already in place		
Hubs	Yes	Already in place		
Switches	Maybe	No		
Back up system	Yes	Back up to DVD-R disc on a weekly basis	\$3,000	Time to change media, label, and verify 50 times per year; two hours at \$30 per hour
				
Total			\$3,000	

Back Up Systems

Search systems will fall over... A redundant system is essential.

Low Cost Framework (6)

<i>Services and Consultants</i>				
<i>Item</i>	<i>Usually Required?</i>	<i>Comment</i>	<i>Est. Cost</i>	<i>Assumptions</i>
Customization	No	Yes, system administrator will set fuzziness and other attributes and identify content to be indexed		Set-up covered in initial allocation
Technical staff training	Maybe	No		
User training	No	No		
Documentation	No	Yes, Web page with instructions will be prepared	\$240	Time to write, code and post search documentation
Cost analysis	No	Yes, rough direct cost estimate		CFO and IT manager discussed costs
Performance and optimization	No	No		
Graphic design	Maybe	No		
User interface	Maybe	Yes, search box will be added to company's intranet splash page. However, use default results page.	\$240	Time to add search box to intranet splash page and add corporate logo to results page
Business process analysis	No	No		
Total			\$480	

Extras

Every search system needs customization, faster response time, and integration into work processes. Skip these and no one will use the system.

Low Cost Framework (7)

<i>Miscellaneous</i>				
<i>Item</i>	<i>Usually Required?</i>	<i>Comment</i>	<i>Est. Cost</i>	<i>Assumptions</i>
Allocation for space	Maybe	No, covered in IT overhead		
Travel	No	No		
Metrics	No	Possibly use log files		
User satisfaction	No			
Other over-head	Maybe	No		



Summary of Cost Hot Spots

- Access control lists
- Performance
- “Hand tuning” of
 - Results lists (boost certain content for certain queries)
 - Indexing of content
 - Classification (if offered)
- Interfaces particularly for:
 - Work flow
 - Stored queries
- Cleaning up data to be indexed (e.g., email, contracts, legal information, information related to patents, etc.)



**Costs increase...
Get your budgets nailed down...
Plan for “spikes”**

Watch for Obvious Errors...



Take Aways

- Find an advisor
- What's indexed ... very sensitive
- Technology evolving
- Search is a work in progress.