

# Achieving “Eureka!”

**CMF2005 - Århus  
November 2005**

**Martin White**

Managing Director, Intranet Focus Ltd. UK

[Martin.white@intranetfocus.com](mailto:Martin.white@intranetfocus.com)

[www.intranetfocus.com](http://www.intranetfocus.com)

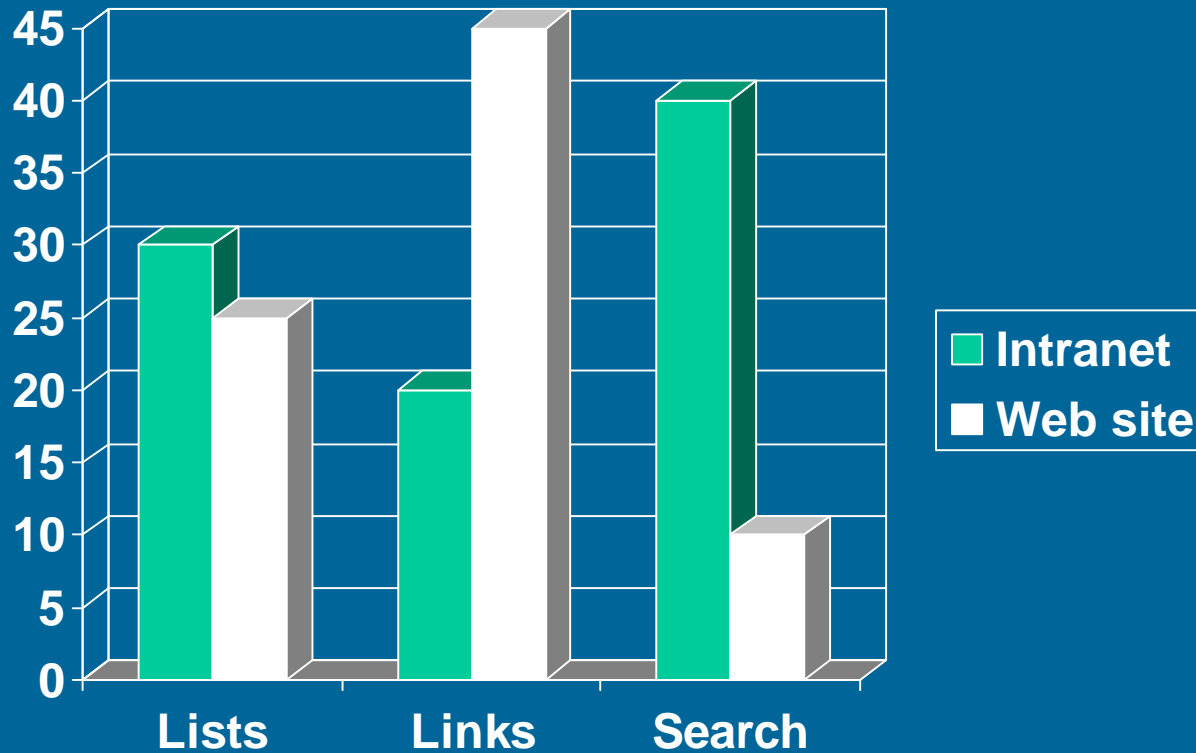
# Agenda

- Looking but not finding
- Identifying user requirements
- The value of personas
- ROI does not work – risk assessment does!
- The technology of search
- The search industry
- Autonomy and Verity
- Implementation never stops!
- Ten critical success factors
- Resources

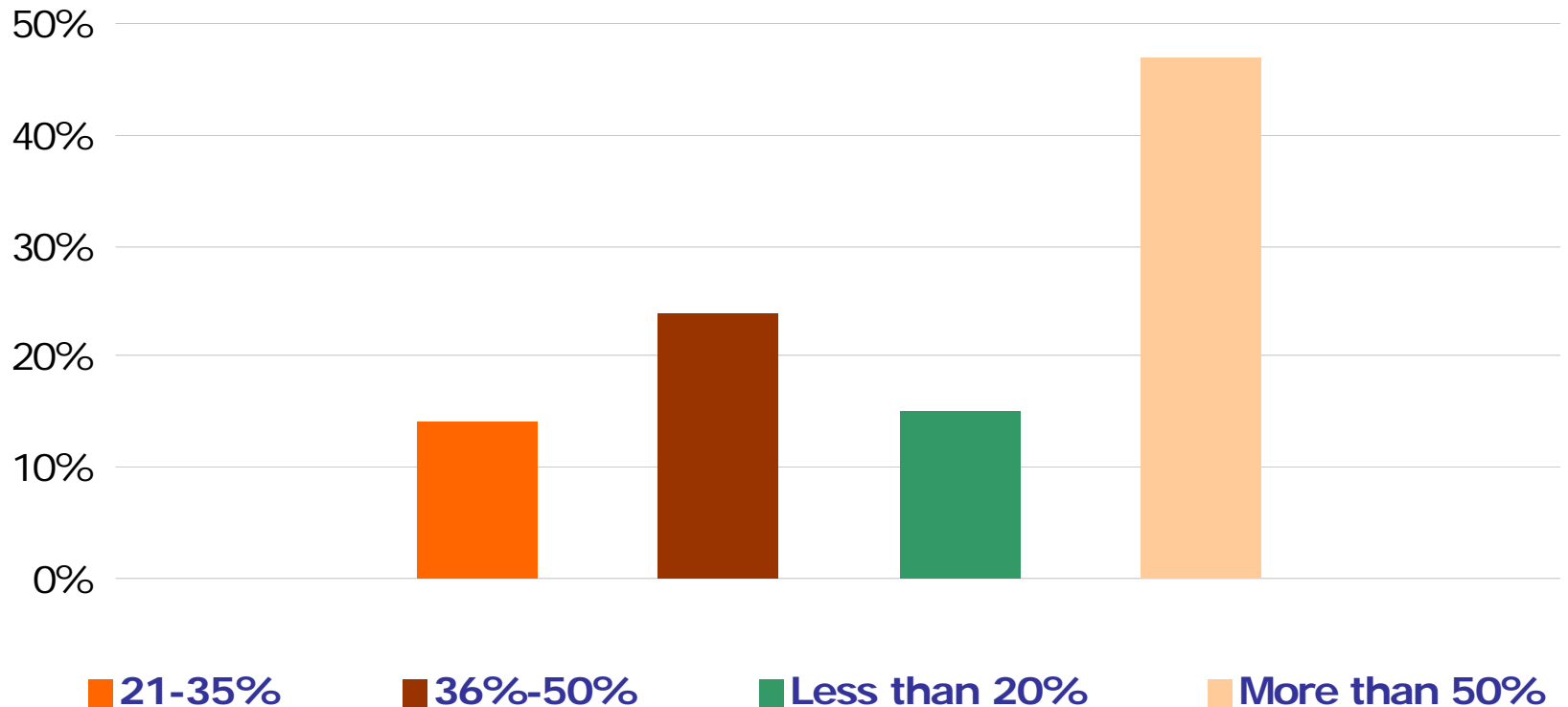
# Looking but not finding

- McKinsey Quarterly executive survey July 2005
  - Survey of 7800 executives in 132 countries
  - 20% of respondents at 'C' level
  - 29% of C-level respondents reported that they found it difficult to find information to make company-wide decisions
  - 40% of senior managers reported the same problem
- IDC research
  - 9.5 hours a week being spent searching for information
  - Of this time 3.5 hours is wasted not being able to find relevant information

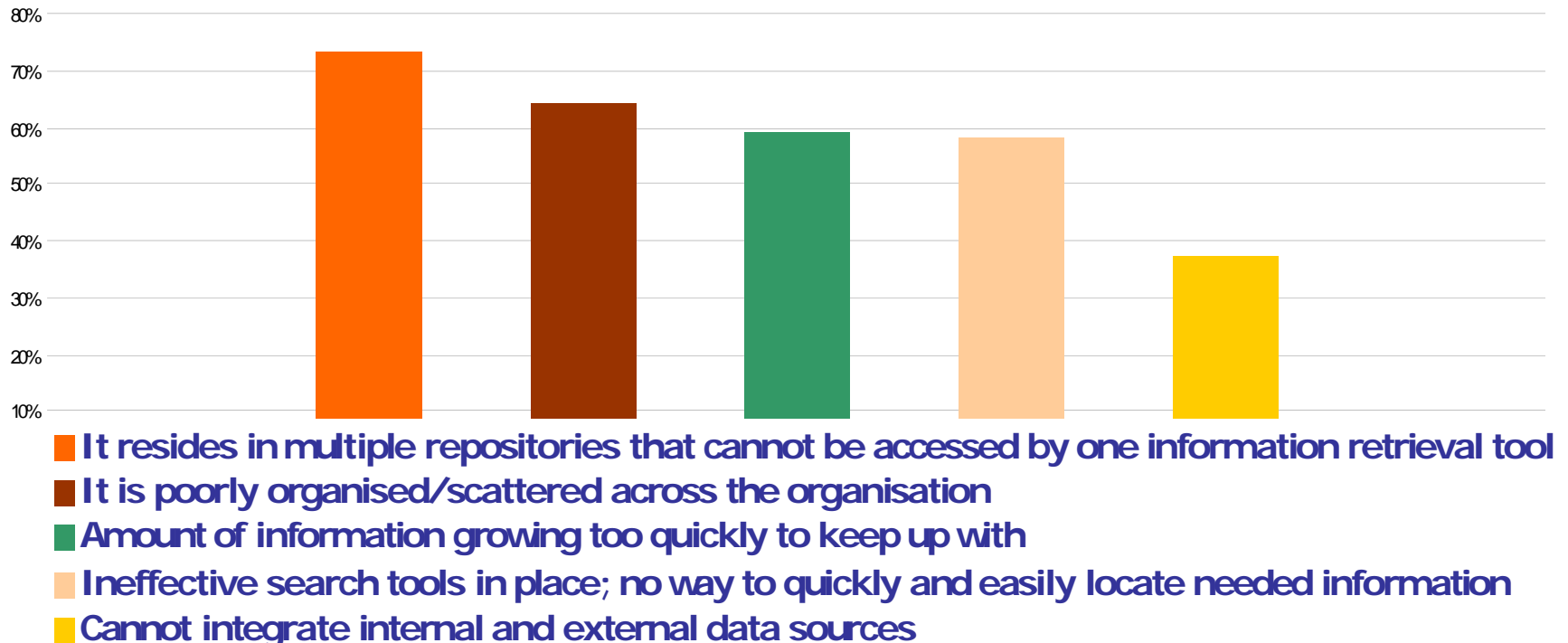
# Access routes – Intranet vs web



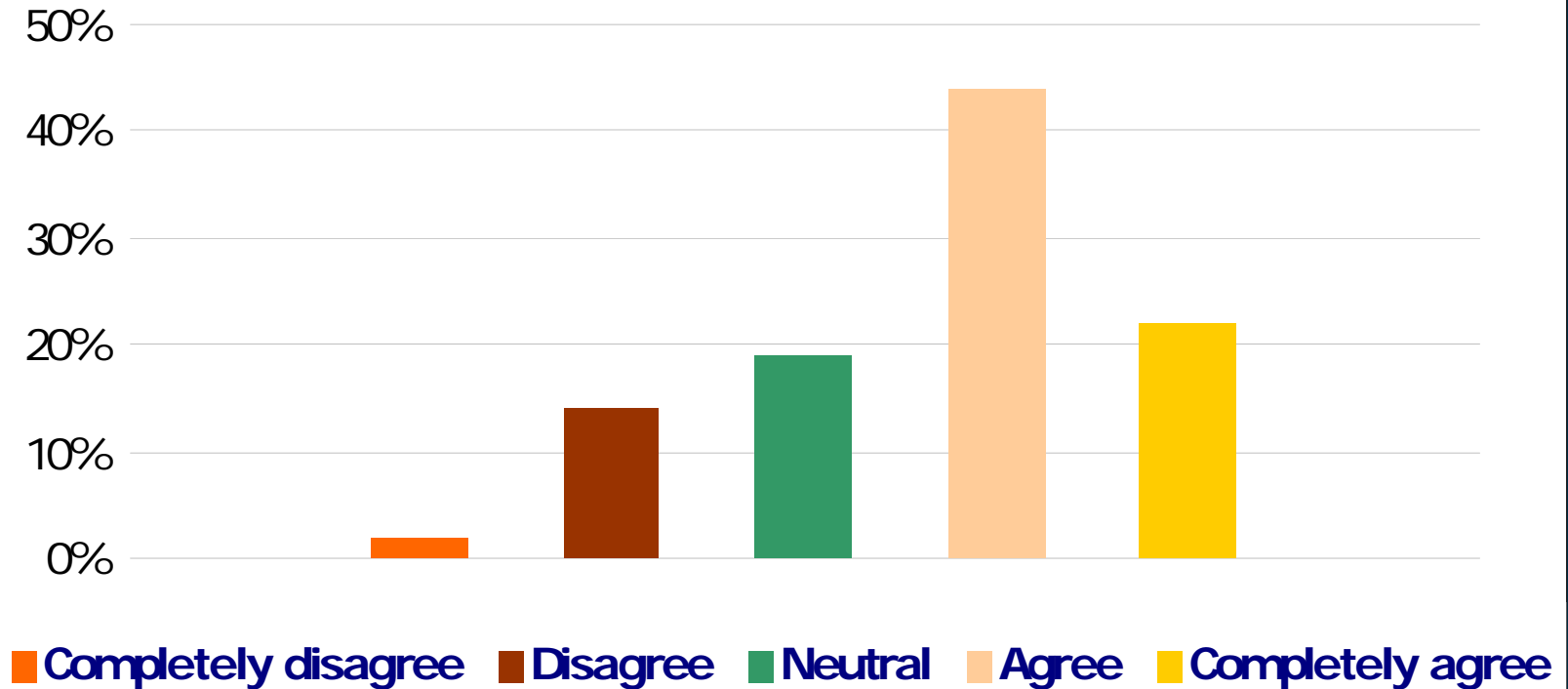
## What percentage of your employees need to search for information as part of their job?



## Which of the following challenges do knowledge workers in your company have when it comes to using the wealth of corporate knowledge and information?



**Workers re-invent or re-generate information because they don't know if that information already exists in the business**



Inight Software/Vascom Bourne - IT  
Managers in the UK Financial Sector 2004

# The magic of search

- Search must be magical
- The search engine has to do all the work
- The usability of the search interface has to be intuitive to users with different search experiences and search requirements
- If people have to be trained to use a search engine then you have lost before you begin
- Relevance is intensely personal
- Getting to “Eureka!” is not easy for the user or for the intranet search team

(From James Robertson, Step Two Designs)

# Trust is the primary metric

- Users want to know that they can trust that
  - That the search engine has indexed all the relevant documents
  - That the desired information can be found if it has been indexed
  - That it will be easier and quicker than any other approach
  - That if the information can't be found it is because there is no relevant document in the system
  - That the search engine is working properly

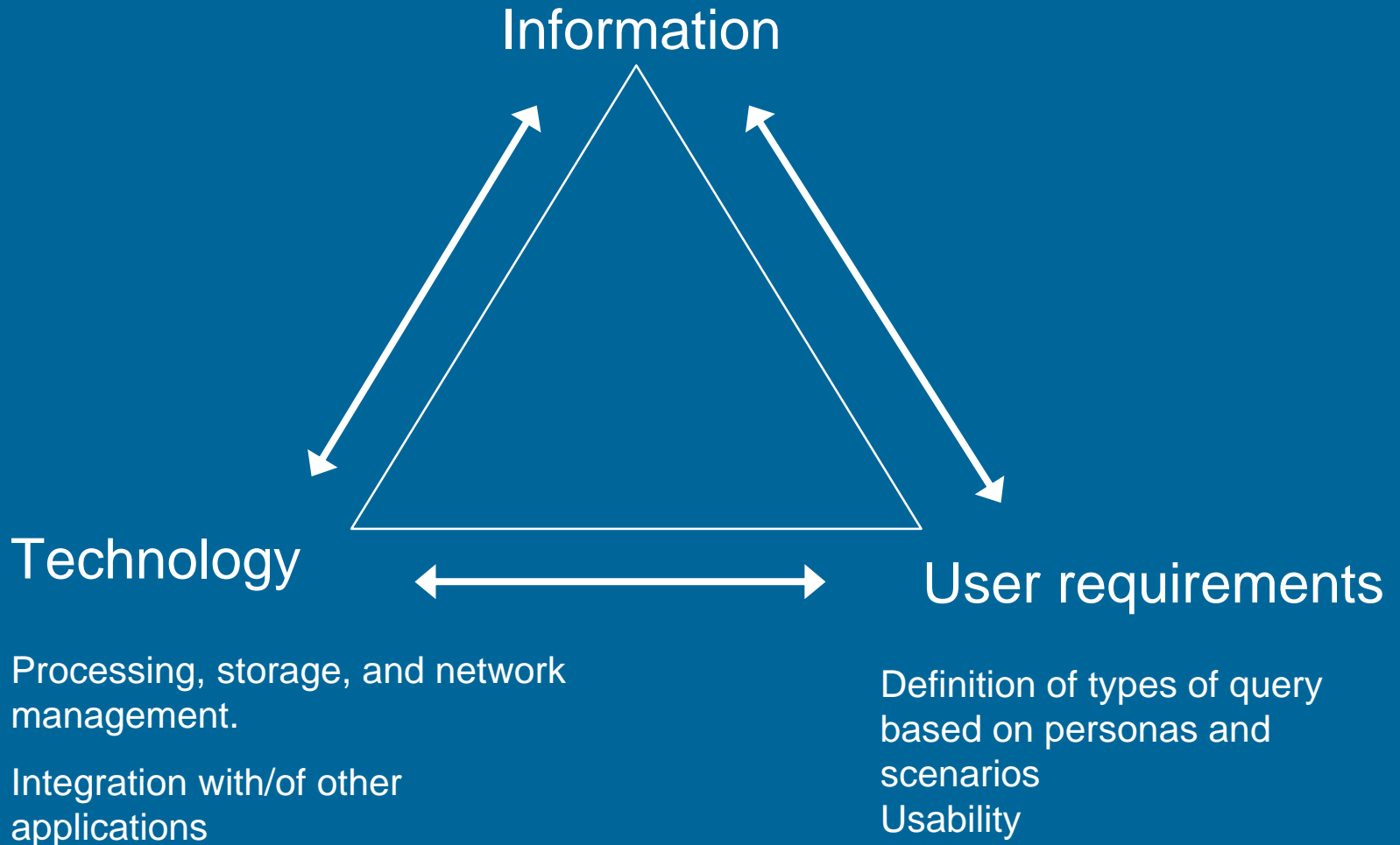
# Intranet or enterprise?

- The quick fix may well be to implement an intranet search engine to cope with html, Office and pdf files
- The better this works the more there will be a demand to extend the intranet search to be an enterprise search, where there will be a much wider range of applications, file types, and governance issues
- At the very least it would be prudent to develop 'what if' scenarios for upgrades in scope and power
- The additional front-end cost may be lower than anticipated

# Enterprise search

- Enterprise search solutions need to cope with
  - Structured (SQL) and unstructured information
  - Information held in specialised applications such as a document management system or a client management system
  - Wide range of document types
  - Wide range of language issues
  - Complex and poorly-defined access privileges
  - IT infrastructure and network issues
  - Disarray in metadata
  - Availability of local support

Size of document, file format, and rate of addition of new documents, language, confidentiality and metadata tagging



# Personas in search development

- Multiple benefits from persona development
  - Use in the RFP to illustrate the way in which searches are carried out
  - Use to help define pilot document set for product evaluation
  - Use to define user interfaces
  - Use to identify application integration requirements
  - Use as the basis for usability testing
  - Use to assess emergence of new search groups and requirements

# There is no such person as a “user”

- Case study - Anglo-German pharmaceutical company
  - Research scientists working on new drug formulations, needing access to highly confidential internal documents and also to external databases
  - Clinical trials teams working through the drug registration process, again involving highly confidential information, much of it in a numeric format
  - Medical information specialists dealing with enquires from doctors and pharmacists about the correct use of the drugs, or initially reporting on adverse reactions. Speed of response is essential to deal with these enquiries

# And yet more!

- Further categories
  - **Product development teams** working on formulations and product packaging, which may need to look at images of packing, and need to be aware of what is happening in the market place
  - **Sales and marketing teams** looking for information on sales channels, sales reports, and the activities of competitors.
  - **Administration and support staff** looking for policy and procedure documents. Unlike the groups above these documents may be in both German and English.

# Making a business case

- It is impossible to make a business case based only return on investment
- It is difficult enough for a CMS, where there are a high volume of standard processes
- This is not the situation with search, and any case based on productivity or efficiency is unlikely to make the cut
- In most companies there is no reliable information on the volume, duration etc of searches

# Risk management approach

- Companies are increasingly risk sensitive, and this can be a very powerful way of making a business case in reverse – “Look what may happen if we don’t make the investment”
- Some typical risks
  - Not being able to respond to a request for a document from a regulatory authority
  - Having to duplicate work because the original documents could not be found
  - Not being able to find documents that contained out-of-date or erroneous information in order to remove them
  - Not being able to respond quickly to requests from customers and clients for information that they would expect the organization to have immediate access to

# Our CMS comes with search!

- All that glistens is not gold
- CMS search comes in three flavours
  - Enables authors to undertake a basic search of a repository for web content for subsequent re-purposing
  - Is an OEM version of an enterprise search product (often Verity K2) that does provide search functionality on the web server
  - Is a proprietary fully-featured search product developed by the CMS vendor
- Perhaps a short term solution but what is the upgrade path?
- What happens if you add/delete a CMS product?

# The main elements

- Content acquisition
  - Spidering and scripts
- Indexing
  - Categorisation and taxonomies, synonyms, numeric ranges
- Query management
  - Boolean, natural language, 'same as'
- Result formatting
  - Wide range of options requiring careful usability testing
- Document access
  - IT capacity management issues

# CMSWatch list A-L

- Arikus: Aire
- Autonomy: IDOL Server
- Blossom: Blossom Enterprise Search
- Coveo: Coveo Enterprise Search
- dtSearch: dtSearch
- Endeca: Profind
- FAST Search & Transfer: Enterprise Search Platform (ESP)
- Google: Google Search Appliance
- Hummingbird: Hummingbird Search Server
- Innerprise: ES.NET 2004
- InQuira: InQuira
- iPhrase: OneStep
- ISYS: ISYS Search Suite
- Lextek International: Onix

# CMSWatch list L-Y

- Microsoft: SharePoint Search Services
- Mondosoft: MondoSearch
- OpenText: Livelink
- Oracle: Oracle Text
- SAP: TREX
- Speed of Mind: Speed of Mind Index Server
- Stratify: Discovery System
- TeraText: TeraText Suite
- Thunderstone: Thunderstone Search Appliance
- Triplehop: MatchPoint
- Verity: K2 Enterprise
- Verity: Ultraseek
- Vivisimo: Vivisimo Clustering Engine
- WebSideStory: Search (formerly Atomz)
- YourAmigo: YourAmigo Enterprise Search

# Desk top search

- Rapid development over the last 18 months
- Early entrants were from the web search sector
  - Google
  - Yahoo
  - MSN
- Now the search engine vendors are moving in
  - Isys
  - Verity
  - Fast Search
- Good for demonstrating the principles of search

# Search engine appliances

- 'Black box' for insertion directly into a server rack
  - Google
  - Thunderstone
  - BA-Insight
- Typical set up time can be as low as 30 minutes
- Pricing based on number of documents, and increases in steps
- Search software is totally proprietary
- Google does not use the same algorithms as the public search engine



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

Search

[Advanced Groups Search](#)

[Preferences](#)

## Google Search Appliance

[Start a new topic](#) - [Unsubscribe or change membership](#) - [About group](#)

[View titles only](#) [View](#)

### ☆ Can anybody tell me how to access the GSA admin panel? What is the URL to login?

**infosistemas** - Nov 8, 8:06 pm

I have bought a used GSA at eBay, I have its IP address and login name, but the old owner does not remember what is the URL to access the admin control panel... he told me to use [\[link\]](#) but this is not correct, it does not work. Please let me know the URL to access... http or https? Thank you, you can reply directly to my email [jsanc...@infosistemas.net](mailto:jsanc...@infosistemas.net)

[read more »](#) 2 new of 2 messages - 2 authors

### ☆ Questions: GAS and XML

**Derrick** - Nov 3, 3:21 pm

I know from attending the GAS demo the other day that the Appliance definitely does support returning search results as XML. My questions are as follows: 1. What does this XML look like? Does it follow a DTD? 2. If a Google search returns paged results, can one set this return XML to return all the results in one result set or does the Google XML

[read more »](#) 7 new of 7 messages - 3 authors

### ☆ ACLs and authorization

**Sam Yates** - Nov 1, 9:04 pm

I have tried to search the various threads that have anything to say about this, but it's been difficult to find a straight answer. Say you have a basic requirement that says the user executing the search must not see any search results that they would not be able to see by accessing the "source" system directly. This would hold true

[groups](#)  
[starred](#) [topics](#)

[Recent groups](#)

[Google Search Appliance](#)

[Groups Alerts](#)

[Create a new group](#)

[Join Google Groups](#)

[Active older topics](#)

☆ [Searching for websites](#)  
1 new of 5

[Related groups](#)

[Google Mini](#)  
This is a discussion where you can share information with Mini customers and u...  
268 messages

[Google Desktop Enterprise](#)  
The Google Desktop Enterprise group is an unmoderated forum for discussion about Desktop Search...  
278 messages

# SharePoint

- SharePoint comes in two versions
  - Web Server
    - Uses WDMSE and has no search
  - Portal Server
    - Uses SQL2000 and has reasonable search
- Products available that integrate with SharePoint include
  - BA-Insight Longitude appliance
  - Coveo search software
  - ISYS search software
- More on the way

# Open source search

- There are a number of open-source search products, of which probably the most widely used and robust is Lucene
- This is supported by the Apache Consortium and there is an excellent global development community
- Much more limited array of open-source enterprise search products than is the case with CMS

# Enterprise search

- Enterprise search solutions may need to cope with
  - Structured (SQL) and unstructured information
  - Information held in specialised applications such as a document management system or a client management system
  - Wider range of document types
  - Wider range of language issues
  - IT infrastructure and network issues
  - Even more disarray in metadata
  - Availability of local support

# Autonomy and Verity

- The winners
  - Fast Search
  - Google
  - Any venture-capital funded search engine vendor
- The losers
  - Verity customers
  - OEM vendors
  - Convera
  - Autonomy shareholders

# Implementation

- Implementation never stops!
- There will always need to be
  - Fine tuning of the relevance algorithms
  - Dealing with feedback about broken searches
  - Undertaking usability tests
  - Modifying the user interfaces
  - Developing and managing application integration
  - Managing metadata and taxonomies
  - Deciding on whether to upgrade
  - Enhancing system performance

# Ten critical success factors

- Excellent project management
- Knowing what there is to be searched
- Rigorous testing as part of the product evaluation
- Understanding user requirements
- Understanding the technology of search
- Anticipating scalability and extensibility issues
- Assessing integration issues
- Testing for usability
- Meeting changing user requirements
- Appreciating that 'trust' is the best metric of search performance

# Resources

- CMSWatch report on Enterprise Search, written by Steve Arnold
  - [www.cmswatch.com](http://www.cmswatch.com)
- Enterprise Search Guidebook, written by Martin White
  - <http://www.freepint.com/shop/report/>

The fire of progress is lit by inspiration  
Fuelled by information  
And sustained by hope and hard work