

Selecting a search engine for enterprise applications

Martin White

Managing Director, Intranet Focus Ltd.

Martin.White@intranetfocus.com

Search Engine Meeting, Boston 2003

www.intranetfocus.com

Agenda

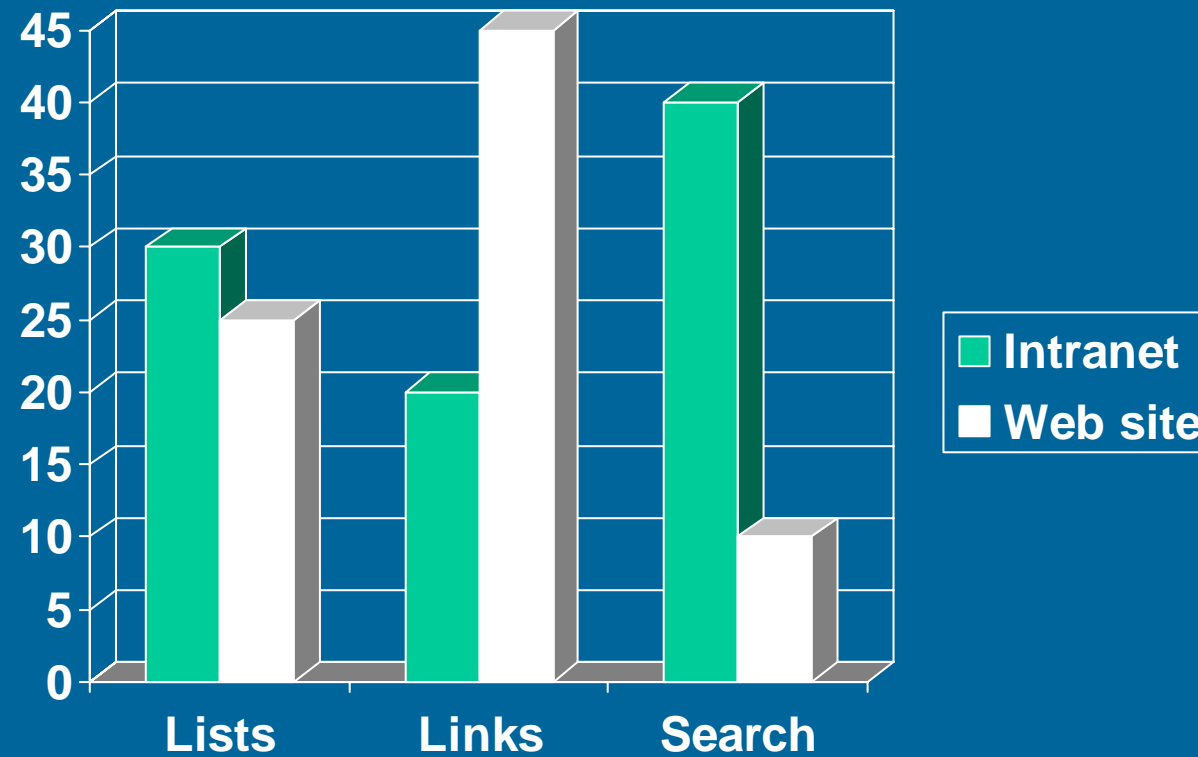
- Some reflections on
 - The content management paradox
 - Search engine vendors
 - The elements of a search solution
 - Why intranets are different
 - Taxonomy
 - Selecting a vendor
 - Implementation issues

To begin with a quote

“The critical component in quick access to intranet content is the search engine. People use a search engine to cut through navigational red tape, and to locate information when lost in a site”

Dennis Deacon, Bank One intranet manager, at an IQPC conference, Chicago, April 2002

Access routes – Intranet vs web



UK government showing the way?

A search on 'leaflet AND policy' gave three identical hits (reading as below) as the three most relevant items

Iia5.2: IIA5.2 – Cabinet Office Regulatory Impact Unit
(No version date)

Policy/Legislation and devolution Bottom of Page Main Contents Previous Page Next Page If you are involved in developing new policy or amending current legislation you need to think carefully about whether it has any particular impacts on Scotland, Northern Ireland, Wales or London.

The content management paradox

- Content management systems do some things well (hopefully!)
 - They enable content to be added to a repository
 - They enable content to be published in a range of formats
- However they usually fail to realise that contributors contribute content but users are looking for information. And there are more users than contributors!
- And if there is not 100% trust in an intranet no clever usability tricks will fix the problem

Understanding search

- The level of understanding of search in most organisations is very limited
- Ask most CIOs about the difference between recall, precision and relevance!
- CIOs in general are dealing with highly-structured data with very standard repeat calls on this data developed through business process analysis
- The situation is not helped by a lack of good marketing communications from vendors, and “The Google Factor”

Mission-Critical Enterprise Search & Categorization Solutions

across more than 200 forms of text, image and video ...

Hot Topics

Convera and IXOS Partner to Deliver Highly Scalable Search and Archiving Solution

Convera Launches Taxonomy Developer Certification Program with Ten Partners

Convera and T-Systems Launch First Online Film Distribution Platform with Worldwide Access

EDS Selects Convera to Provide Advanced Search and Information Discovery to the U.S. Navy and Marine Corps



More News →

Events

Gartner Symposium ITxpo
03.23-27 San Diego, CA

Pharma-IT Summit 2003
03.31 New York, NY

SecurE-Biz CxO Summit
04.01-02 Arlington, VA

Convera Partner Summit
04.02 Tysons Corner, VA

More Events →

Solution Centers

Convera's dedicated Solution Centers are built to address your industry's specific needs:

Financial

Government

Life Sciences

Media & Entertainment

Recent Awards:

- ♦ The EContent 100
- ♦ Frost & Sullivan 2002 Market Engineering Leadership
- ♦ KM World 100 Companies that Matter



Home

Company

Customers

Products

Partners



Search

Federator^{Verity}

Learn more about Verity Federator:

[Download](#) our free White Paper

or

[Browse](#) the Complete Product Information

Most Visited Links:

Choose...

What Kind of Solution Do You Need?

- E-commerce
- Enterprise
- Departmental Search
- Federated Search
- OEM Technology
- Offline Publishing

Catalog^{Verity.K2}

Turn browsers into buyers, and keep them coming back for more. Add intelligent merchandising, profitable site stickiness, adaptive personalization to your e-commerce portal.

Recent News

- Verity Introduces Verity® Federator to Integrate Multiple Enterprise Search Systems and Sources
- Verity Provides Certified Content Categorization for Stellent Content Management System
- Verizon Information Services Turns to Verity to Boost Performance of SuperPages.com
- Verity Reports Third Quarter Results, Sequential Revenue and Profitable Earnings Growth

Download Center

Verity Connector

Subscribe to Verity's Monthly Newsletter

Upcoming Events

- Gartner ITXPO 2003**
23-27 March 2003
- Gartner Retailvision**
25-28 March 2003
- [Other Events](#)

Partner Webinars

- Oracle Joint Webinar
- Stellent Joint Webinar

Verity In Action

- [WACKER](#)
- [TelMex](#)
- [American Hospital Association](#)
- [Other Case Studies](#)

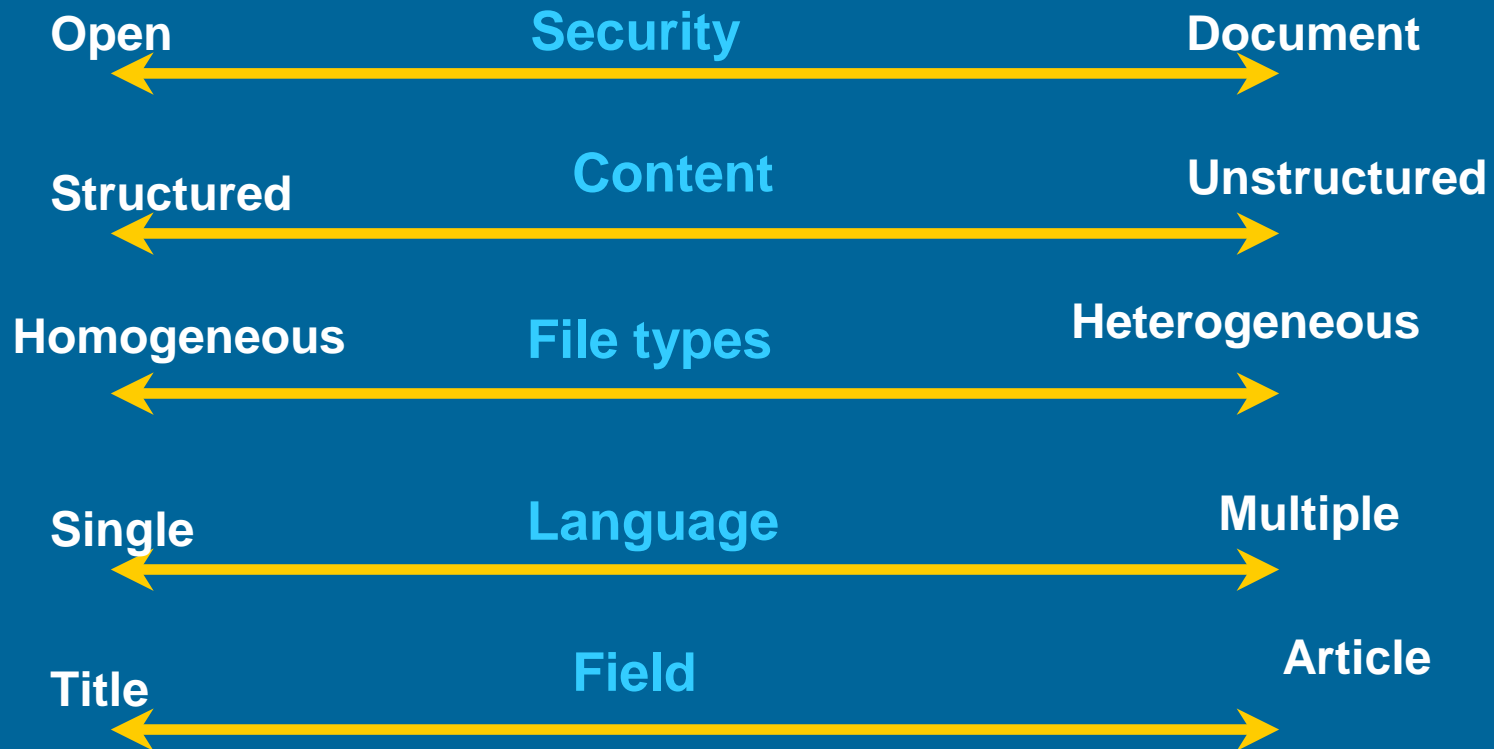
Services

- Customer Support
- Education

The elements of a search solution

- Who is going to use 'search', and for what purpose
- How will this affect
 - What content elements will need to be indexed?
 - How often the indexes are updated?
 - How the results will be ranked?
 - How will the results be sorted?
 - How will each "hit" be presented?
 - What will the user do next?
 - What happens if they can't find what they are looking for?

Information access parameters



Intranet-specific issues

- Document level security
- Very high level of trust required in the results
- Wide range of user skills and expectations
- Network performance issues, especially with remote access and scanned images
- Rich language set, even without multilingual requirements
- Distributed content contribution leading to variations in metadata quality
- Intranet vs. web site vs. extranets vs. everything else

Taxonomy

- Taxonomy requirements is usually underestimated
- Although there are some excellent software tools that can assist in developing a bottom-up taxonomy they all need to be managed by staff who understand the edges of the taxonomy envelope
- It is important that the selection of a search engine and the requirements for taxonomy development are considered separately, and then integrated into one vendor only when this represents the best, rather than the most convenient, solution.

Selecting a vendor – initial steps

- Clarify the requirements for search vs taxonomy creation
- Talk to the vendors – there are not that many of them
- Do an information audit focused on the use that will be made of the information, not where it will be stored
- Segment and profile user requirements
- Understand the work involved in enhancing metadata and security in legacy documents
- Go and look intelligently at some search solutions in action

Develop a 'Statement of Requirements'

- Do not write the SoR in terms of functionality
- It should be written in terms of organisational needs, for example
 - Document types
 - Metadata schemes
 - Taxonomy management
 - End-user profiles
 - Search scenarios
- It is also important to set out development scenarios for future search requirements

Search scenario

“Policy development staff will want to search for reports and internal briefings in order to identify research resources. The reports will be in Word and pdf, and there will be PowerPoint presentations linked to the briefings

They will want to have the hits listed by file type, and by date of origination. For larger reports they will want to know the occurrence of the term in the document, and be able to look at each occurrence in the context of 50 words each side of the term.

They will want to store the search profile to be able to run it as required during the course of the project, with new items added to the collection being flagged”

Implementation issues

- Implementing a search solution is even more complex than a CMS
 - What would be a realistic pilot phase
 - How will the search solution be evaluated at the pilot phase, and is this scalable?
 - What will be the role of the vendor in the implementation phase?
 - How will the APIs be written and documented?
 - What is a realistic implementation schedule?

Summary

- Search purchase must be based on a Statement of Requirements that reflects user needs
- Do not confuse taxonomy development and search
- A search engine is not the answer to poor information architecture
- Ensure that the implementation schedule is very clearly set out
- At all times manage the expectations of everyone concerned

Questions?

www.intranetfocus.com