

## Table of Contents

<b>Background, License, &amp; Acknowledgements</b> .....	<b>i</b>
<b>Table of Contents</b> .....	<b>iii</b>
<b>Frequently Asked Questions About Enterprise Search</b> .....	<b>vii</b>
<b>I. Introduction: What is Search and How Does it Work</b> .....	<b>1</b>
The Knowledge Management Machine? .....	1
The Google Effect .....	2
Web Search and Enterprise Search .....	3
What is Enterprise Search .....	4
Core Features of Enterprise Search .....	5
Advanced Features .....	5
One Search Engine to Rule Them All? .....	7
Diverse Content Types .....	8
How Enterprise Search Works .....	9
Components of a Search System .....	10
Content Acquisition .....	11
Indexing .....	11
Query Processing .....	13
Formatting Results .....	14
Retrieving the Document .....	15
Roles in Enterprise Search .....	15
The Project Manager .....	16
The Information Professional .....	16
The Information Technology Professional .....	16
The Financial Professional .....	16
The Consultant .....	17
The Searcher .....	17
Managing Expectations .....	18
Does IT Matter in Enterprise Search? .....	18
Goals of This Report .....	19
<b>II. The Business Case for Enterprise Search</b> .....	<b>21</b>
The Omniscient Knowledge Worker .....	21
The Informed Transaction Processor .....	22
Your Business Case .....	23
Which Rationale? .....	23
An ROI-based approach .....	24
The Cost of Enterprise Search .....	26
Variables Affecting Cost Estimates for Enterprise Search .....	26

To purchase this report, go to <http://store.yahoo.com/cmsworks/>

How to Estimate Search Costs .....	27
Costs for a Small Enterprise Search Installation .....	32
Costs for a Large Search Installation .....	36
Cost and Value: A Never Ending Challenge .....	41
Frequently-Asked Questions About Enterprise Search Costs .....	42
<b>III. Enterprise Search Requirements .....</b>	<b>46</b>
Project Phases .....	46
Accelerated Search Deployment .....	48
Budget: Costs and Financial Impact .....	49
Documents to Be Indexed .....	49
Staff Member to “Own” the Project .....	49
Minimum Security Requirements .....	50
Search System inside the Organization or a Hosted Solution .....	50
Update Cycle .....	50
Problem Content .....	51
Time Available for Implementation .....	51
An Available Search System .....	52
Some Caveats about Risks of Accelerated Search Deployment .....	52
Fast-Cycle Summary .....	53
A More Expansive Search Requirement Checklist .....	54
Architecture and Technical Infrastructure .....	54
Content: Obtaining and Indexing .....	56
Security .....	58
Content Security .....	59
Controlled Access .....	60
Access from Off Premises Locations .....	61
Search Features .....	61
Prioritizing Features .....	62
Asking the Right Questions .....	63
Search Management .....	63
System and Architecture .....	64
Content .....	66
Indexing .....	68
Search .....	69
Other Features .....	71
Frequently-Asked Questions About Metrics, Performance, and Planning .....	72
<b>IV. Enterprise Search Vendors .....</b>	<b>77</b>
The Search Vendor Landscape .....	77
About our Categorization Scheme .....	77
Other Approaches to Categorizing Vendors .....	78
Brief Synopses .....	79

To purchase this report, go to <http://store.yahoo.com/cmsworks/>

About the Product Reviews .....	80
Caveats .....	81
Arikus: Aire .....	82
Atomz: Atomz Search .....	91
Autonomy: IDOL Server .....	101
Blossom Software: Enterprise Search .....	112
Convera: RetrievalWare 8.0 .....	122
Copernic: Enterprise Search 2.0 .....	137
dtSearch: dtSearch .....	146
Endeca Profind .....	156
Fast Search & Transfer (FAST): Enterprise Search Platform .....	165
Google: Google Appliance .....	177
Hummingbird: Search Server .....	194
Innerprise Corp: Enterprise Search .....	204
InQuira, Inc.: InQuira 6.5 .....	212
iPhrase: One Step .....	222
Lextek International: Onix .....	232
Microsoft: SharePoint Search .....	242
Mondosoft A/S: MondoSearch .....	263
Odyssey Development, Inc.: ISYS 6 .....	273
Open Text: Livelink .....	281
Oracle: Oracle Text 10.x .....	294
Speed of Mind: Index Server .....	309
Stratify: Discovery System 3.x .....	318
TeraText: TeraText Suite .....	333
Thunderstone: Taxis and Webinator .....	340
TripleHop Technologies: MatchPoint 3.0 .....	352
Verity: K2 and Ultraseek .....	365
Vivisimo Corp.: Vivisimo Clustering Engine .....	382
YourAmigo: Enterprise Search .....	390
Open-Source Alternatives .....	408
Frequently Asked Questions About Vendors and Tool Selection .....	409
<b>V. Advice, Pitfalls, and Best Practices .....</b>	<b>421</b>
How to Select a Search System: A 12-Step Process .....	421
Vendor Pricing .....	430
Rules of Thumb for Minimum Costs Associated with Search .....	430

To purchase this report, go to <http://store.yahoo.com/cmsworks/>

Cost Parameters for Typical Enhancements .....	431
Vendor's Pricing Approaches .....	432
General Pricing Lessons Learned .....	436
Pitfalls to Avoid and Best Practices to Follow .....	438
Trends in Search Systems .....	444
A Business Case Is a "Must" .....	445
Do It Right, Instead of "Just Do It" .....	445
XML Will Continue to Gain Importance .....	446
Search Will Require Access to Internal and External Content .....	446
Buyers Will See Intensified Competition in the Market .....	447
Workflow Integration to Become More Important .....	448
Conclusion .....	448
Frequently Asked Questions About Security and Enterprise Deployments .....	449
<b>VI Glossary .....</b>	<b>459</b>

## Frequently Asked Questions About Enterprise Search

### II. The Business Case for Enterprise Search

- Q1 What search-related costs are the most difficult to control and why? What types of expenses are we looking at for hardware, programming, network services, third-party add-ons, etc.? (Page 42)
- Q2 What is the ratio of license fee to professional services' charges for an enterprise search system that serves 1,000 users, runs on hardware that has eight processors, and indexes two million documents? (Page 43)
- Q3 How much does it cost to manually index 100,000 documents? How much if you assume commercial automation tools in the mix? (Page 43)
- Q4 What is computational cost? (Page 44)

### III. Enterprise Search Requirements

- Q5 What does full-text search really mean? (Page 72)
- Q6 What is the best way to calculate the amount of storage we need for our enterprise search and retrieval system? We have about two million documents to index, and the number of documents is growing at about 50 percent a year. (Page 72)
- Q7 How much time does it take to perform these basic search set up functions: [a] install the software, [b] index two million documents, [c] tune the spider to minimize its impact on the shared network, [d] test and tune the index, [e] customize the interface? (Page 73)
- Q8 Our system performed well when we first installed it. Over the past three months, however, response time has been slowing. Some queries now take more than a minute to execute and display results. We have monitored the system and discovered that we have no more than 20 simultaneous users. The number of documents we have indexed is fewer than 500,000? What is the problem? (Page 74)

### IV. Enterprise Search Vendors

- Q9 What search system should we buy? (Page 409)
- Q10 We hear about statistical search, natural language search, and other types of search. What are these? (Page 409)
- Q11 How do we search our Web content without going to the time and expense of licensing search software and running it inside our own company? (Page 411)
- Q12 We have heard many seemingly similar descriptions of search software from vendors of search, customer relationship management systems, application server software, and content

To purchase this report, go to <http://store.yahoo.com/cmsworks/>

- management system software. Are these companies selling the same basic capability? Are some of these companies offering something unique? (Page 412)
- Q13 What are the essential differences among Autonomy, FAST, Convera, Verity, Hummingbird (formerly Fulcrum) and Open Text search-and-retrieval products? (Page 413)
- Q14 How many search-and-retrieval solutions are commercially available at this time? (Page 413)
- Q15 Our consultants recommend a “pilot” test of any enterprise search software prior to making a final licensing decision. What does a pilot mean? What are the requirements for a meaningful test? (Page 413)
- Q16 We have an enterprise search system from one of the leading vendors. We want to switch to another product. What are the steps? Where are the pitfalls? What are the risks? (Page 414)
- Q17 The local university has a very fine computer science department. One of the professors has developed an advanced search engine with clustering and automatic indexing. Should our firm consider using this type of software? It appears to be powerful, and it is available at a very attractive licensing fee. (Page 415)
- Q18 We are an IBM-only company. We have heard about Web Fountain from our sales representative. Is this a viable technology for us to use for enterprise search? What is the relationship between Factiva’s search service and Web Fountain? Our sales representative left us with a press release announcing a linkage between these two companies. (Page 415)
- Q19 Will it be possible to use a Google-type of system to get access to all of the digital content our firm creates, licenses, and archives? (Page 416)
- Q20 Do enterprise search systems support queries for mathematical equations, chemical structures, and engineering drawings? If the answer is “no,” what is needed to search these types of content? (Page 416)
- Q21 We have heard that enterprise search systems “own” a customer. What does this mean? How can we retain our independence and avoid vendor lock-in? (Page 416)
- Q22 We have multiple locations and work in English, French, German, and Chinese. Is there an enterprise search system that allows an English speaker to retrieve material in other languages? (Page 416)
- Q23 We have seen some very interesting visualization and graphic search demonstrations. One was from a company called Plumb Design, which offered a visual approach to our thesaurus. Kartoo demonstrated a three-dimensional display of search results. Should we consider adding a visual display of search results to our enterprise search to improve its usability and appeal to our users? (Page 417)
- Q24 Why does Yahoo offer a “search box” and a point-and-click list of headings? Do users prefer one way to search for information over another? We do not have a classification system for our documents. Is this direction we should consider moving with our search system? (Page 417)

To purchase this report, go to <http://store.yahoo.com/cmsworks/>

- Q25 We have heard a great deal about “guided navigation.” Is guided navigation search and retrieval, or is it a separate type of search functionality that can be added to our existing system? (Page 418)
- Q26 Our customer relationship management system is based on Siebel Systems technology. The vice president responsible for customer support has licensed iPhrase after looking at InQuira and several other search engines oriented toward customer support. Can iPhrase use the indexes from our Verity system? (Page 418)
- Q27 What search engine is included in Adobe’s software? (Page 419)
- Q28 What search engine is included with BEA Systems application server? (Page 419)
- Q29 We have a search system from one of the Big Four [Autonomy, Convera, FAST Search & Transfer, or Verity]. How do we determine if we should upgrade or buy a different system? (Page 419)
- Q30 How easy is it to move from the search functionality in Adobe, BEA, Vignette or some other software product that includes search technology to an enterprise solution from a different vendor? (Page 420)
- Q31 We need help acquiring an enterprise search system. Where can we turn for impartial help? (Page 420)

## **V. Advice, Pitfalls, and Best Practices**

- Q32 In what parts of the overall search system is human expertise typically most essential? (Page 449)
- Q33 How do we search our Web site and Lotus Notes’ databases? (Page 449)
- Q34 What is the most cost effective, easy-to-operate way to integrate content produced by our employees with content from an outside Web source? Outside commercial sources such as the Associated Press? A mixture of Web and branded third-party data in text and row-and-column format? (Page 449)
- Q35 Our organization has never had an enterprise search capability. Do most organizations index “from this day forward,” or do organizations perform retrospective indexing of older material? (Page 450)
- Q36 We have multiple search engines at our company. We have the search tools that come with Oracle. We have Microsoft’s built-in search service. We have licensed a commercial engine. Do we need multiple search engines? (Page 451)
- Q37 How can we search our database content, the size of whose records are small in comparison with our Word, PowerPoint, and Web documents? The database information is very important. We want to see both the structured database information and unstructured text in one integrated display. Is this possible? (Page 452)
- Q38 Our internal legal counsel says that we must comply with Sarbanes Oxley. We have to make archives of electronic mail and other documents that we have not previously stored or index

To purchase this report, go to <http://store.yahoo.com/cmsworks/>

- for later retrieval. What enterprise search system addresses this need? Can the solution for Sarbanes Oxley be used to provide enterprise wide search? (Page 453)
- Q39 Our internal systems professionals – who are focused on technology – have installed a search system that does not meet our needs in marketing and other departments. How can we show the IT professionals that “pure technology” is not addressing the nuances in our content? (Page 453)
- Q40 Our information center has licensed external content from The Thomson Corporation and United Press International. Is it really true that a single enterprise search system can address both internal- and externally-provided content? If we do integrate some of this external data, what happens if we break our agreement with these third party providers? (Page 453)
- Q41 We have a top notch corporate information center (library). How can we best utilize the expertise of the professionals in this unit for enterprise search? (Page 454)
- Q42 What type of training is needed to optimize the effectiveness of our enterprise search system? (Page 454)
- Q43 We have an AS/400-based (mainframe) accounting system. The basic record contains numbers corresponding to our products and their prices. We also make use of a “Notes” field that has text information about a customer. Can we search these “Notes” fields from our portal? (Page 455)
- Q44 Our company has a lawyer responsible for our iManage-based litigation support system from Interwoven. We have a certified records manager responsible for the system that maintains our engineering documentation associated with our nuclear power plant work. We have an information center with satellite offices in three other locations around the world with its online access catalog system. Our IT department has installed separate search systems for different units of our company, including one Japanese language system in Tokyo. How can one enterprise search system address these different, intensely political power centers? (Page 456)
- Q45 We have a Google appliance. We have indexed the content on 25 servers. Some of the content must not be available to anyone other than our president. How do we [a] remove the protected content from the index and [b] how do we prevent the appliance from indexing material we don't want available via our enterprise search? (Page 456)
- Q46 What is the minimal security procedure necessary to allow certain employees or authorized users to see only content appropriate for them? (Page 457)
- Q47 How secure are enterprise search systems? (Page 457)
- Q48 What are issues associated with having our enterprise search operated by a third party under a managed service contract (outsourcing)? How do we protect our proprietary data? What are the costs for this type of enterprise search service? (Page 458)
- Q49 We have an enterprise search system. Some departments are “hiding” documents so they cannot be indexed. How do we index these documents? (Page 458)
- Q50 We have an enterprise search system. Some departments are “exposing” all of the data on their departmental computers. How can we index only appropriate documents? (Page 458)