Government Printing Office

Future Digital System

Search Product Capability Questions

August 31, 2006

This document contains the initial set of questions in support of the Future Digital System search product evaluation process. Additional information and clarification questions may be provided to vendors in the future if needed.

1.0 Requirements Compliance

The accompanying spreadsheet contains selected requirements from the FDsys Requirements Document Version 2.1 that pertain to search. Please evaluate each requirement in the spreadsheet against the capabilities of your product. Fill in column C with either a 1 (product meets the requirement) or 0 (product does not meet the requirement). If you have multiple products that are separately priced, then fill in the name of the product that meets the requirement in column D. If you have any questions or comments about a requirement then record them in column E. Please limit questions and comments to technical information and refrain from including marketing statements.

2.0 Technical Product Questions

Please answer the following questions as briefly and concisely as possible. Please limit your response to each question to no more than 500 words and an optional diagram or diagrams.

2.1 Architecture and Performance

1. Please describe your core technology architecture.

2. How many total documents can be indexed in the system (index capacity)?

3. How rapidly can documents be added or re-processed into the index (index rate)?

4. How many queries per second (QPS) can the product process (query processing)?

5. What constraints does your product place on choices for operating system, server type, networking, and storage architecture?

6. What is the recommended operating system, server type (PC, blade, SMP, etc.), networking, and storage architecture for hosting your product?

7. What is the recommended number and type of servers and storage capacity for supporting your product assuming 20,000 users logged on to the system, with 5% concurrent requests, less than 2 second search response time is required, and the system contains 500,000 source documents of 1MB in size that are full text indexed as well as 20 text fields each of no more than 512 characters that are also indexed? How does the server and storage sizing change if we double the size of the indexed documents? How about quadrupling? How would this system architecture be affected if the corpus increases by 5,000 documents a month over the next 5 years?

8. How large is your index size compared to the indexed document size with every option that is required for precision searching rather than relevant searching (i.e. every document that is essentially about "the GPO fiscal requirement" rather than simply mentioning "the GPO fiscal requirement"?

9. How does your product support the availability requirements for FDsys (see spreadsheet)?

10. Does your product support replication of its indexes to standby systems?

2.2 Search Processes

1. Please describe your content acquisition capabilities. (e.g., spider, FTP, conversion scripts, import filters, APIs, file support, ECM support).

2. Please describe your indexing capabilities (e.g., keyword, concept extraction, categorization, hyperlinking, advanced linguistics, real-time updating).

3. Please describe query processing capabilities including parsing, matching, and post-processing (e.g., proprietary parsers, matching techniques, relevancy, clustering, "see also" references, saved searches).

4. Does your product support Xpath and Xquery searching for XML documents? If so, please describe?

5. Does your product support the Z39.50 search interface? Does your product support federated searching of Z39.50 systems?

6. Does your product support a web services interface? Java interface (please describe the type of interface)? Other language interfaces?

2.3 Results and Delivery to End Users

1. Please describe your capabilities for displaying and manipulating results and delivering search to End Users (e.g., WYSIWYG controls, templates, user interfaces, formatting results).

2.4 Administration

1. Describe your administrative features (e.g., metrics, performance controls, usage tracking, system dashboards, rule creation tools).

2. Please describe capabilities that allow administrators to control search processes (e.g. search algorithm, business rules, relevancy weighting profiles for specific collections, indexing process, how repositories are indexed).

2.5 Security

1. Please describe your security model (e.g., the engine stores access control lists in the index itself, the engine checks with the host repository before results are presented, front-end application parameters are added to the user's search query at execution time to ensure appropriate access rights).

3.0 Cost and Support Questions

1. What are your licensing models, including support?

2. Which licensing model do you believe is the most cost effective for FDsys?

3. What is your estimated license cost, including support per year, (rough order of magnitude) for your product for FDsys?

4. Do you provide integration and administration training classes? What is the cost and frequency?