

## The DB2 Net Search Extender Version 8.2

Many of today's applications, especially in the e-business and content management areas, call for full-text search capabilities. You'll rarely find a Web site doing serious business unless it offers text search capabilities so customers can quickly locate the products and services in which they are interested. With the DB2 Net Search Extender, the textual documents and textual information for the Web site could be stored in a DB2 relational database for easy searching. Examples of a text document would include a product description on an e-commerce site, or meta data about this text document, such as the author, publication date, and the abstract of the book. An avid reader might want to do a full-text search on thousands of book abstracts on a retail publication site. If the Web site had been implemented with the WebSphere Commerce Suite application, DB2 Net Search Extender could be the power behind the fast, reliable, scalable search for the information. The developer of the site could quickly enable very flexible and powerful search using DB2 Net Search Extender. The site could let the potential buyer combine a full-text search (for a book title) with conditions on other data known to the application such as books that are cheaper than a certain price, books with a similar theme using a thesaurus, or books that are available in a book store that is located near the potential buyer. The end result would be a satisfied customer and a completed sale.

### DB2 Net Search Extender for your e-business

DB2 Net Search Extender is a DB2 database extension which provides a rich set of full-text search functions. DB2 Net Search Extender offers query performance and scalability by integrating caching and optimization technologies. It is most appropriate for high end e-business applications that are database intensive, since it is designed to work closely with DB2 Universal Database applications.

The DB2 Net Search Extender:

- . □ Provides a parallel, scalable, full-text search to handle the heavy text search demands of e-Business Web applications.
- . □ Works seamlessly with text documents contained in DB2 as well as other relational databases, federated databases, and documents stored in file systems.
- . □ Indexes data rapidly, without impacting other applications by locking database tables during indexing and searching
- . □ Delivers excellent performance and scalability with regard to query load.
- . □ Provides three mechanisms for full-text search, tailored to different types of e-business requirements:
  - a. ○ An SQL scalar search function which is the default search option, and is suitable for a majority of full-text search usages;
  - b. ○ A stored procedure search, which exploits the main memory cache

for extremely fast full-text only results;

- c.
  - An SQL table-valued function, which exploits indexes on views and presorted text indexes.
- .
  - Integrates into the DB2 Control Center for seamless and easy to use administration.
  - Extends existing DB2 applications easily by using standard extensions to SQL. The Structured Query Language/Multimedia (SQL/MM) proposed international standard extends SQL with full-text features.

## **DB2 Net Search Extender Key Features**

The DB2 Net Search Extender very fast indexing, dynamic index update, and high search speed. The key features of the DB2 Net Search Extender are:

### **Indexing**

- .
  - Provides fast indexing of large data volumes. Indexing speed approximates 1 -2 gigabytes (GB) per hour on a medium server, depending on the document size and format. Outstanding response time optimization is seen in scenarios such as obtaining the first number of hits from a potentially large list in the shortest time possible.
  - Provides incremental update of indexes which is done asynchronously so as not to impact performance of search applications.
  - Allows the caching of table columns in main memory at indexing time to avoid expensive physical read operations at search time
  - Offers a choice of command line or interface via the DB2 Control Center for index administration
  - Supports different text formats, for example HTML and XML. Support of indexing on large objects (LOB) and user-defined data types is also provided.
  - Supports index and search of documents in other relational databases (enabled with nickname table support).
  - Supports index and search of external files (enabled with DB2 Data Link Manager support).
  - Supports presorted text indexes. This allows the initial update to index the text document in the order specified; search results will be returned in this order.
  - Provides index and search on Double Byte Character Sets or Unicode Databases for flexibility
  - Provides language specific *stopwords* to reduce the index size and improve performance. Stopwords are common language words such as “and”, “or”; not indexing common words results is a smaller and faster index.
  - Monitors the progress of indexing which is important for administrators in managing long running index builds.

### **Search**

SQL scalar search functions, a stored procedure search, and an SQL table-valued function support the following search functions:

- .  Allows Boolean operations for conjunction, disjunction, and exclusion of search terms. Individual search terms may be single words or phrases. Information needs can be specified very accurately.
- .  Provides proximity search for words in the same sentence or paragraph
- .  Supports “fuzzy” searches for words having a similar spelling as the search term. Fuzzy search can be used to find names that have been incorrectly entered into a table or if the correct spelling is not known. For example, a search for "Andrew" can find "Andrews", "Andraw" and "Andru".
- .  Provides wildcard searches, using front, middle, and end masking, for whole words and single characters. Wildcard searches make a search more flexible.
- .  Enables free-text searches, for documents containing specific text. The search argument is expressed in natural language, and is another element of flexibility.
- .  Supports a thesaurus for broader queries by searching not only for a specific search term, but for terms that are related to it. This can be automated with a thesaurus, which is a controlled vocabulary of semantically related terms that cover a specific subject area.
  
- .  Provides search restricted to sections within documents
- .  Search on numeric attributes which could be either in structured documents or within columns

### **Search Results and Responsiveness**

- .  Identifies how the search results are sorted at indexing time (with Stored Procedure Search)
- .  Specifies search result subsets when searching large data volumes where large result lists are expected
- .  Sets a result limit on queries where a high hit count is anticipated.
- .  Builds-in SQL functionality. Scalar search functions are combined with the DB2 optimizer to automatically select the best optimization plan according to the expected search results
- .  Supports 37 languages, including English, German, Spanish, French, Japanese and Chinese.
- .  Provides sub-second search response times. The performance improvement compared to an SQL “like” query can be a factor of 10 up to 100 or even more.

### **An example**

A company is outsourcing an e-mail archive with several million English e-mails. The archive is growing at a fast rate. The requirement is to implement an

application that allows many concurrent users to search on the data with a sub-second response time, and to provide a high indexing speed to allow the user to search for new e-mails as quickly as possible. DB2 Net Search Extender provides the necessary performance and scalability.

### **Text document formats supported**

- .  Text (flat ASCII)
- .  XML Documents
- .  HTML Documents
- .  An interface to plug-in additional filters

### **Platforms supported**

DB2 Net Search Extender V8.2 is available for the following Windows and UNIX operating environments:

Windows XP, Windows NT, Windows 2000, Windows Server 2003, AIX, HP-UX, Linux on Intel, and Solaris.

Support for Linux for S/390 and zSeries will be made available at a later date.