

NXT[®] 3 SEARCH ENGINE[®]

TECHNICAL OVERVIEW

INTRODUCTION

Quickly and easily locating the right information is critically key to any business organization. Business managers, knowledge workers, customers and suppliers should all have simple and timely access to relevant information. Whether information is organized, proximally located or in a particular format should not complicate or lengthen the search or be a barrier to access.

NXT 3 Search Engine[™] from NextPage enables users to quickly and easily search millions of digital assets and find (with precision) information that is relevant to a specific request. Based on powerful and proven search technology, NXT 3 Search Engine provides a platform that searches all content available in a network, ranks the relevance, filters the search throughout the entire document and subdocument collection and truly delivers what the user requests.

NXT 3 Search Engine includes a comprehensive set of search features that enable any search from a simple key word to a complex customized query to take place. A short list of NXT 3 search functions includes: keyword, Boolean, wildcard, phrase, stem, thesaurus, proximity, nested phrase and many more. The types of documents that can be simultaneously aggregated in a search include general business documents such as word processing, spreadsheet, and presentation files as well as PDF, HTML, XML and database files.

Searched information is made more relevant for users through NextPage retrieval and presentation technology. Retrieved information can be presented in results tables ordered by specified fields, relevance or metadata elements. NXT 3 conceptual analysis extrapolates 'key concepts' from documents and can dynamically generate document summaries for simplified analysis. Other search result features include relevance feedback ('more like this'), table of content generation, topic maps and categorical and hierarchical ordering.

The benefits of the NXT 3 Search Engine to users are significant. E-businesses struggling to strike a balance between escalating volumes of information and partner demands for precision access to relevant information now have the means to deliver a solution. Self-service information simplifies queries, lowers costs and increases customer and partner loyalty.

NXT 3 SEARCH ENGINE FEATURES

Search engines have become widely used and understood with the evolution of the Internet. The NXT 3 Search Engine takes the basic functionality of search and

significantly augments it with a complete set of features that result in the most powerful, flexible and comprehensive search solution available. Search functions can be used individually or combined to create customized and unique queries that can filter down to the most difficult to find information. NXT 3 search functions include the following.

Keyword Search - This basic function locates and identifies every document that includes one or more instances of a specified key word. Documents searched can be from multiple sources, locations or formats. Keyword search incorporates the industry's best technology for indexing, parsing and identifying key words in the optimal amount of time. Examples of simple keyword search include searching an entire Web site for any documents containing the word 'performance'.

Boolean Operator Search - Boolean operators (AND, OR, NOT, XOR) enable users to refine queries to more specific or expand queries to filter more general information. Boolean operators may be combined together creating powerful filters that can locate specific information with precision. A Boolean search expression including a date AND pricing AND a specific vendor would return a current price list for the specified vendor.

Phrase Search - A phrase search finds all instances of a specified phrase. The search exactly matches a string of words in a particular order. An example would be to search using the complete title of an article or a sentence from the abstract of a research report. Phrase search provides the ability to filter out common information that may be related but not relevant.

Ranked Search - A ranked search displays the information or documents located in order of relevancy. Documents that contain keywords in the same order or more instances of the keywords are displayed at top of the results list. Ranked search results are especially valuable when using complex searches across large volumes of similar documents.

Wildcard Search - In many cases, users may not know a complete name or the actual spelling of a word. Wildcard searches allow them to substitute or replace unknown letters with generic placeholders that return all search hits with the known word pattern. Wildcards can include single letters, multiple letters or beginning and ending segments of words.

Stem Search - Stem search allows users to find words that match truncated 'stems' of words. For example, 'congres' would return all documents containing either 'congress' or 'congressional'.

Thesaurus Search - Searching through large volumes of data for peripherally related information can be assisted using Thesaurus Search. NXT 3 expands the search parameters to include keywords that are synonyms to the specific word entered. For example, a search on 'retrograde' would return documents including 'retrogressive', 'recessive', 'backward', 'degenerate', 'decrepit', 'rotting', 'decaying', etc.,

Field Search - Field search allows you to search using specified parameters on formatted, tagged or relational data. An example would be to search on date-stamped data that is between two specific dates (i.e. return all documents dated between 9/9/00 and 9/15/00).

Proximity Search - Proximity search allows you to specify how close words must be to each other in order to register a result. The proximity range can be specified as number of words, sentences, paragraphs, fields or partitions. For example, return documents where the word 'overstock' and the word 'inventory' appear in the same paragraph.

Nested Phrase - A nested phrase searches for two phrases within a designated number of words from each other. For example, searching for 'United States' within fifty words of 'Congressional Act' would return all instances of these two phrases within the specified proximity.

Search Within Results - Narrowing search results is often necessary to limit selection and pinpoint relevant information. Users can further 'refine' a search result set by including additional keywords that eliminate irrelevant material and filter down to the desired information.

More Like This/These - This option enables the use of other categorized or classified parameters to return 'similar' documents that may not contain the specified keywords. For example, a keyword search may produce a 'technical support' document; 'more like this' would produce additional support documents around the same subject.

Not Like This/These - This option intelligently narrows a search to exclude documents similar to those selected and provides an exclusion filter.

Word Wheel - The word wheel component displays a set of valid fields and values that can be selected in a guided search. A word wheel allows users to select values and relational type fields from a dynamic, interactive list, providing meaningful feedback with each keystroke.

Concept Based Search - Using intelligent conceptual analysis, NXT 3 search examines content of a document to determine 'key concepts'. These concepts are identified and can be used to identify other related documents or similar topics.

Natural Language Search - Allows users to enter queries in a conversational or natural language format (i.e. 'find documents with information about the country of Sudan').

NXT 3 provides these search functions across a wide spectrum of data types. Data repositories with mixed formats and differing document types can be included with any search. Document types which are supported using NXT 3 search include:

- Word processing - all standard formats including Microsoft Word, ASCII text, and other common formats
- Spreadsheet - all standard formats including MS Excel
- Presentation - all standard formats including MS PowerPoint
- Standards based document formats including Adobe PDF
- Internet documents including HTML and XML
- Database records - all fields within a database (ODBC, LDAP databases)
- Commercial applications - additional services such as MS Exchange and Lotus Notes

In addition, document metatags or metadata may be searched with hits included in the search results. Metadata includes information such as document author, creation date, document summary, etc. This information combined with the 'field search' described above can be very valuable for automated and programmatic queries.

INFORMATION RETRIEVAL AND PRESENTATION

Increased connectivity has radically expanded the volume of available information without making the information easier to navigate, search, or use. The real value of relevant information is only realized if that information can be easily retrieved and presented in a format that is simple, accessible and presented in a manner that efficiently displays search results. NXT 3 provides several methods for information retrieval and presentation that set the standard for information 'quality of access'. Several of these presentation features include the following.

Native Document Viewing - Documents may be viewed directly in the browser (using standard plug-ins) or in the native application.

Browser Viewing - All Internet presentation standards are supported including HTML, XML, extensible style sheets for XML (XSL) and cascading style sheets (CSS) for XML and HTML.

Sorted Results - NXT 3 displays search results in a table that includes a link and a set of properties or metadata for each matched document. The table may be sorted by relevance, by any field or metadata element, or in hierarchical order.

Hit Highlighting - Search hits are identified by highlighted color to assist in keyword location in HTML, XML and PDF documents.

XML to HTML Conversion - NextPage provides a conversion filter to translate XML to HTML on the *server* in real time.

NXT 3 also provides a significant degree of flexibility for information presentation. At the simple end of the scale, NXT 3 provides several predefined program templates that can be modified to meet a specific organizations needs. These templates simplify the selection of layout, search functions, result display and color schemes. Templates enable searching by data field, category, keyword or by a dynamically generated topic lists. Individual users can also select layout formats and whether they wish to use Java components, frames or access one of the standard template forms.

At a more technical level, NXT 3 searches can programmatically initiated and results can be captured for presentation through any API or interface including COM interfaces accessible from C++. Content services are easily accessible to Active Server Pages (ASP) and other scripting engines. Support for XML enables the use of elements, parameterized fields and index sheets to flexibly present information in any configuration. NXT 3 presentation capabilities make it easy to incorporate search in any portal, subportal, or Web site.

Access to search services can be controlled through role-based content views. In order to secure classified information and prevent information overload, NXT 3 carefully delivers appropriate information to the appropriate user. Administrators associate user identities or roles with specific information sets. Individual users only access those views and information sets for which they have authorized access.

CONCLUSION

The benefits of powerful search capabilities for information intensive organizations are significant. Any company that has distributed, dynamic and abundant information can profit from the ability to search, sort, categorize and present that information to any interested stakeholder including employees, managers, customers, partners, suppliers and contractors.

Growing E-businesses must balance the increase in information available with customer demands to access the 'right' information. The information search and presentation capabilities available with NXT 3 provide a comprehensive and powerful solution for companies wishing to strike this balance. The NXT 3 Search Engine provides a range from simple to complex search functions that enable the precise and quick location distributed information in any format. NextPage's information retrieval and presentation features allow this information to be easily accessed by any type of user.

The overall result of faster information access is a boon to both companies and their customers and partners. Companies benefit through more efficient information flows, reduced information management resources and lower information maintenance and administration costs. Customers and partners benefit through direct and timely access to relevant information which not only reduces their costs but serves to create greater customer satisfaction and loyalty.