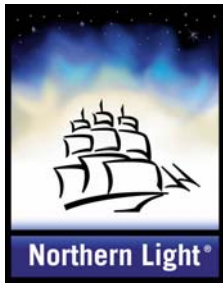


Northern Light® Enterprise Search Engine



No compromises: customization +
security + performance

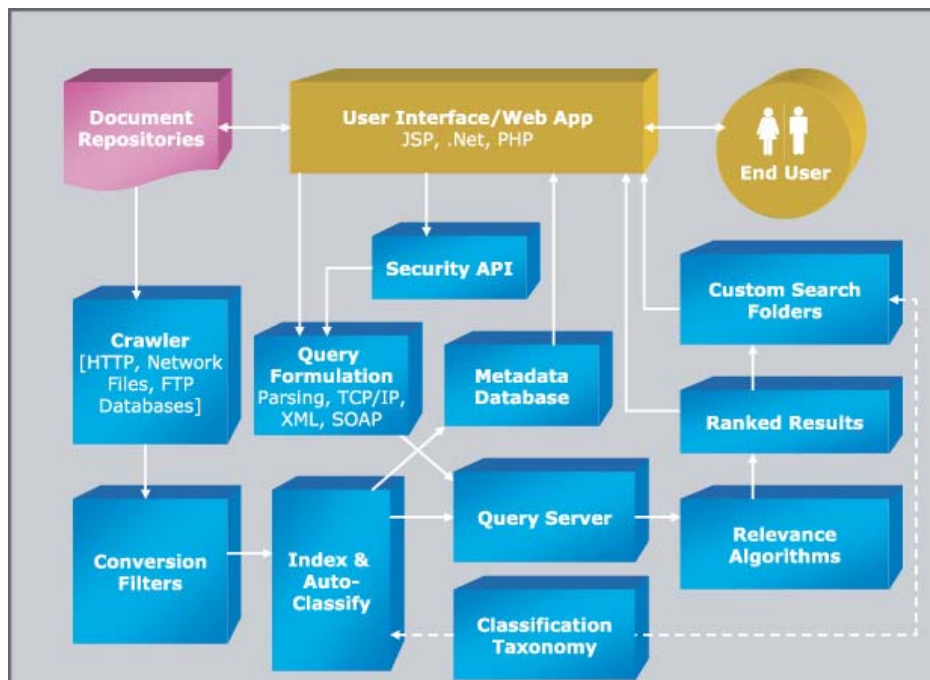
Northern Light Enterprise Search Engine is a customizable, secure, fast, and scalable Linux-based enterprise search engine. Northern Light features open APIs optimized for customization to the business requirements of enterprise applications. APIs and support for industry standard protocols ensure flexibility and ease of integration into portal environments, websites, enterprise applications, and organizational workflows. Northern Light offers the Enterprise Search Engine for an annual document count-based license that includes support, maintenance, and upgrades. There are no additional licenses required for servers used for redundancy or development.

NORTHERN LIGHT ENTERPRISE SEARCH ENGINE

At a glance:

- Open application, API's, command line options, programmable components
- Security API for insuring users only see documents they are authorized for
- Scales to over a hundred million documents in one search index
- Superior query performance, less than 100 milliseconds
- Runs on LINUX
- Award winning, tunable relevancy ranking
- Content acquisition, filtering and integration subsystem
- Pre-built and user definable taxonomies, and automatic classification engine

See the Northern Light Enterprise Search Engine in action at
www.epa.gov
www.marketingpower.com
www.nlresearch.com



Open API: Well-documented API's, XML search results, J2EE, SOAP for Web Services integration, TCP/IP transactions, and sample user interface code supports the integration of Northern Light into corporate and Web applications. Supports direct manipulation of all search functions from scripts and other applications.

Programmable content filters: Published load format specification allows the full-text of any file type, from any source, located anywhere to be filtered, indexed, and searched. The data conversion system includes filters for Microsoft Office, PDF, HTML, XML, RTF, and text formats. Our filter framework is completely object-oriented. We provide all the source code to facilitate customization and capture, application, or creation of metadata.

Programmable, configurable, crawler and content acquisition subsystem: Connects via HTTP, HTTPS, FTP, SFTP, NFS and SMB (Windows) protocols and navigates websites and networks to acquire content, including websites requiring a login and network file servers. The crawler can acquire content in multiple formats: HTML, MS Office, PDF, XML, and content residing in content management systems and relational databases. It can be programmed to perform tests for inclusion or rejection of content and to assign metadata to documents.

Capture and search on any metadata: Programmable crawlers and content filters can capture any metadata available in the content source, look up external metadata, use extracted text, or perform logical operations to apply missing metadata to documents before they are indexed. All metadata is represented in the index, which means you can use search forms or syntax to qualify the results. Any metadata can be returned in the search results or used in results list sorting.

Document text snippets in results list items: If high quality summaries are not available, search results can show selections of text from the documents with query key words highlighted.

Indexing: Indexes every word of the full-text of every document and every two-word pair (to support phrase searching), as well as all the metadata associated with a document. You can create new index fields to facilitate custom search options reflecting the best way your content can be organized.

Search Alerts: Allow users to save searches and trigger alerts when new content is indexed. Alerts can be delivered as emails and as RSS feeds.

Performance: 216 queries per second with a query response time of 80 milliseconds on an index of 5 million documents (original webpage content file of 25 gigabytes) running on a LINUX PC-scale server.

Scalability: 64-bit file access functions so you can create databases of over 100 million documents with a single software installation on a single server.

Security API: Filters results so users can only see information for which they are authorized. Designed to avoid any performance penalty at run time.

Query Syntax and Languages: Keywords, phrases, natural language, simple, compound and nested Boolean expressions, wild cards. Queries search all document full-text and/or metadata. Optional query optimization for part of speech stemming. Index documents and search in any of 13 languages.

Relevance ranking effectiveness and tuning: Unique multi-factored approach to relevance ranking that produces the best results possible for enterprise and published content. Sub-collections of documents that have been determined to be of high quality can be boosted to the top of search results.

Automatic classification and application-owner defined taxonomies: Classifies every document in the database to our subject taxonomy automatically (by examining the document text) or programmatically (by using the metadata applied by the crawler or filters). We provide a complete 17,000-node subject taxonomy developed by our expert librarians. Classification powers advanced search forms, vertical search applications, and our patented Custom Search Folders™ for results navigation. Definable taxonomies can be used to create custom classification solutions.

Custom Search Folders™: Patented functionality clusters the results into relevant categories for presentation so that users can efficiently drill down into the most relevant part of the search results.

Administration tools: Operations can be controlled through the browser-based Control Center, command line, or scripts. The Control Center includes a test harness search UI, a scheduler to manage crawling, data conversion, database loading, and a system configuration manager. All events are logged for maintenance, tuning and reporting.