



Open Source Search Engine

OpenSearchServer is a powerful, enterprise-class, search engine program. Using its web user interface, the crawlers (web, file, database, ...) and its REST/RESTful API you will be able to integrate quickly and easily advanced full-text search capabilities in your application. OpenSearchServer runs on Windows and Linux/Unix/BSD.

Contents

- [1 Search functions](#)
- [2 Indexation](#)
- [3 Crawlers](#)
- [4 Parsers](#)
- [5 General](#)

Search functions

- Advanced full-text search features
- Phonetic search
- Advanced boolean search with query language
- Clustered results with faceting and collapsing
- Filter search using sub-requests (including negative filters)
- Geolocation
- Spell-checking
- Relevance customization using algebraic functions
- Search suggestion facility (auto-completion)

Indexation

- Supports 17 languages
- Fields schema with analyzers in each language
- Several filters: n-gram, lemmatization, shingle, stripping diacritic from words,...

- Automatic language recognition
- Named entity recognition
- Word synonyms and expression synonyms
- Export indexed terms with frequencies

Crawlers

This is the module which creates the index that will process the queries and return answers.

OpenSearchServer is equipped with several crawlers that each allow you to browse and index different categories of content:

- The web crawler for internet, extranet and intranet
- The file systems crawler for local and remote files (NFS, SMB/CIFS, FTP, FTPS)
- The database crawler for all JDBC databases (MySQL, PostgreSQL, Oracle, SQL Server, ...)

Each crawler offers a list of parameters allowing developers to customize its behavior.

- Filter inclusion or exclusion with wildcards
- Session parameters removal
- SQL join and linked files support
- Screenshot capture
- Sitemap import

Parsers

Parsers recognize and analyze the MIME type of crawled documents or file extensions and then automatically extract information necessary for indexing (Title, Text, Author, hypertext links, etc.).

Supported formats are:

- HTML / XHTML
- MS Office documents (Word, Excel, Powerpoint)
- OpenOffice documents
- Adobe PDF (with OCR)
- RTF, Plaintext
- Audio files metadata (wav, mp3, AIFF, Ogg)
- Torrent files
- OCR over images

General

- [REST API](#) (XML and JSON)
- SOAP Web Service
- Monitoring module
- Index replication
- Scheduler for management of periodic tasks
- [WordPress plugin](#) and Drupal module

[About us](#) | [Contact](#) | [Newsletter](#) | [References](#)

