

*Cost-effective, reliable,
multilingual search that's easy
to deploy*

The same multilingual text processing technology used by industry-leading search engines FAST, Yahoo!, Bing.com, Google, Endeca and others is now available for the open source search solution, Apache Lucene and Solr.

Deploy in days

Out-of-the-box you can connect Basis Technology's Rosette Linguistics Platform (RLP) to Apache Lucene and have robust and accurate multilingual search up and running on your website, intranet or internal network. This combination is an indispensable and low-cost search solution for enterprises of any size.

Dependability you can bank on

Lucene — a high-performance open source search toolkit — is a popular search solution with over 3,000 installations in organizations including IBM, CNET, and Wikipedia. RLP has a ten year track record of providing linguistic intelligence to meet the demanding accuracy and performance required by major search and text mining providers.

Search to the standards of enterprise search vendors

- Language identification and full-text search in 54 languages
- Linguistically improved search in 19 languages including major European languages, Arabic, Japanese, Chinese, and Korean.
- Entity extraction and faceted navigation in 12 languages
- A scalable, high performance architecture

Basis Technology offers a fixed pricing model for RLP -- no matter the number of users or servers.

Users enjoy the same quality of experience with Lucene they have come to expect from their favorite web and enterprise search engines.

All You Need To Do

Download and install the RLP SDK or runtime package. Lucene leverages RLP functions and passes along information such as the location of documents to be indexed. The "RLP for Lucene" module enables RLP to connect to Lucene out of the box. No additional work is needed for Lucene to search text in any language RLP supports. The "RLP for Lucene" module comes with RLP at no additional cost.

RLP linguistic capabilities

- **Language Identification:** Identifies the language a document is written in.
- **Language-specific processing:** The base linguistics function of RLP is the starting point for building a search index and refining queries. Advanced linguistic features improve precision and recall of search results.
- **Segmentation and tokenization:** Separates streams of text into unique word tokens, especially needed for languages -- such as Chinese and Japanese -- written without spaces between words.
- **Lemmatization:** Provides the dictionary form for an inflected word to improve recall.
- **Noun decomposing:** Separates compound words (such as in German and Dutch) into their separate components to improve recall.
- **Part-of-speech tagging:** Improves precision and recall.
- **Entity extraction:** Extracts entities to enable faceted search on key names and entities in search results.

Apache Lucene Performance & Scalability

- Thread safe
- Cross-platform solution
- Support for multiple cores
- Small RAM requirements
- Incremental indexing as speedy as batch indexing
- Index is only 20-30% the size of text indexed
- Powerful search algorithms

For more information

BROWSE: www.basistech.com **WRITE:** info2009@basistech.com **CALL:** (617) 386-2090