
DataLexica

DataLexica is a component which provides two main services:

- recall improvement: by expanding the user's query
- reduction of index sizes: indexing based on linguistic knowledge

Both services can be used, together or separately, to improve the performance of any information retrieval system.

With the technology provided by *DataLexica*, the recall of your queries will increase dramatically; you will be able to retrieve all relevant documents, and there will not be any noise in your results.

Using *DataLexica*, your queries will be expanded in order to cover all the grammatical forms of each word. Therefore, you do not need to worry about word endings (like "starting", "starts", "started"), irregularities (like "may", "might") or special forms (like "good" and "best").

DataLexica consists of two components: a lexical database and an application which extracts data.

DataLexica is fully developed for Spanish, English, French, Basque, Catalan, and Valencian. The Spanish lexical database has over 3,000,000 words classified according to linguistic features. Portuguese, Galician, German, and Italian are under development.

DataLexica can be used with any commercial search engine. *DataLexica* is distributed in DLLs (Dynamic Link Libraries), for Windows, and SO for Linux, which are written in C++.

Bitext
The Bits and Text Company
Edificio Prisma, 1-1 - Cólquide 6
28230 - Las Rozas – Madrid - Spain
<http://www.bitext.com>
info@bitext.com

