

---

## DataLexica

---

*DataLexica* es un componente que proporciona dos servicios principales:

- Mejora de la cobertura: expansión de la búsqueda del usuario
- Reducción del tamaño de los índices: indexación basada en conocimiento lingüístico

Ambos servicios pueden utilizarse, conjuntamente o por separado, para mejorar el rendimiento de cualquier sistema de recuperación de información.

Con la tecnología *DataLexica*, la cobertura de sus búsquedas aumentará drásticamente; podrá recuperar todos los documentos relevantes, y no se añadirá ruido a sus resultados.

Por medio de las búsquedas con *DataLexica*, sus consultas serán expandidas para cubrir todas las formas de cada palabra. Así, no necesitará preocuparse sobre finales de palabras (como "comenzó", "comenzaba", "comenzándolo"), irregularidades (como "ir", "vamos", "fueron") o formas especiales (como "bueno" y "mejor").

*DataLexica* está integrado por dos componentes: una base de datos léxica y una aplicación que la explota.

*DataLexica* está completamente desarrollado para el español, inglés, francés, euskera, catalán y valenciano. La base de datos del español contiene más de 3.000.000 de palabras clasificadas según sus propiedades lingüísticas. Estamos trabajando en versiones para portugués, gallego, alemán e italiano.

*DataLexica* puede utilizarse con cualquier buscador del mercado. Se distribuye como librerías: DLLs -Dynamic Link Library, en su versión Windows, y SO en su versión Linux), escritas en C++.

**Bitext**  
**The Bits and Text Company**  
Edificio Prisma, 1-1 - Cólquide 6  
28230 - Las Rozas - Madrid - España  
<http://www.bitext.com>  
[info@bitext.com](mailto:info@bitext.com)

