

# Terrier



## Terrier – Terabyte Retriever

The Terrier Information Retrieval (IR) platform is a modular framework for large-scale IR applications. It is based on **Divergence From Randomness** **parameter-free** probabilistic models for automatic document ranking and query expansion.

Terrier supports the indexing of various types of collections, including TREC and Web collections. It can also be used to develop Web, intranet and desktop search engines for different types of documents.

Terrier offers a modular API for querying, allowing the development of applications as diverse as experimenting with standard test collections, or building Web search engines and XML applications.

## Features

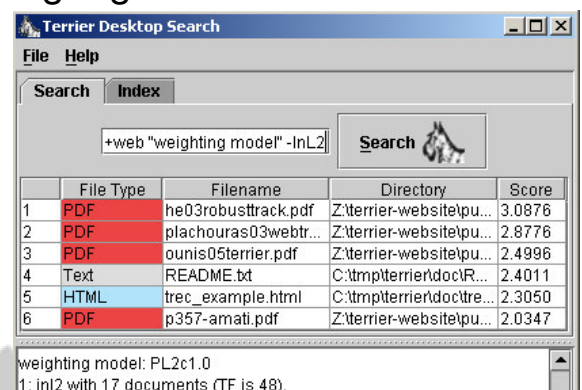
- Cross-platform, written in **Java**
- Highly **compressed** indices
- **Parameter-free** query expansion with **highly compressed** direct file
- Handling **large-scale document collections**
- **Test-bed** for experiments with standard TREC collections
- Powerful **desktop search** application for various common types of documents
- **Modular and extensible APIs** for querying and indexing
- Provides DFR & various other well-known IR models for easy cross-comparison



[terrier@dcs.gla.ac.uk](mailto:terrier@dcs.gla.ac.uk)

## Terrier Desktop Search

- Stand-alone Java application
- Easy to select the directories/files to index
- Supports HTML files, PDF files, Microsoft Office files & source code files, easily extensible for new document types
- Integrated with Terrier's query language



- Easy addition of new IR models
- **Advanced query language** for phrases/proximity, fields, Boolean and +/- operators
- Standard evaluation of results
- Ideal for Internet search
- **Open source** software

## Getting Terrier

- Terrier is distributed as open source software, under the Mozilla Public License (MPL).
- Terrier can be downloaded from <http://ir.dcs.gla.ac.uk/terrier/>



<http://ir.dcs.gla.ac.uk/terrier/>