# The LAILAPS Search Engine: New Features

Jinbo Chen[1], Matthias Lange[1], Uwe Scholz[1]

[1] Research Group Bioinformatics and Information Technology – BIT
Leibniz Institute of Plant Genetics and Crop Plant Research – IPK
Corrensstr. 3, 06466 Stadt Seeland, OT Gatersleben, Germany

Website:   http://lailaps.ipk-gatersleben.de
Contact:   chenj@ipk-gatersleben.de

IB 2012

## Motivation

LAILAPS is a combined relevance ranking system and integrated search engine for life science databases. The concept is to combine a feature model for semantic relevance ranking, a machine learning approach to interacts with user relevance profiles and user feedback tracking for a self-trained ranking improvement. The ultra fast query response, features like recommendation of related data records, query suggestion, interactive ranking optimizer, and an wizard style installation software provide a full featured search appliance for an user customized set-up of individual search portals.

## LAILAPS Overview
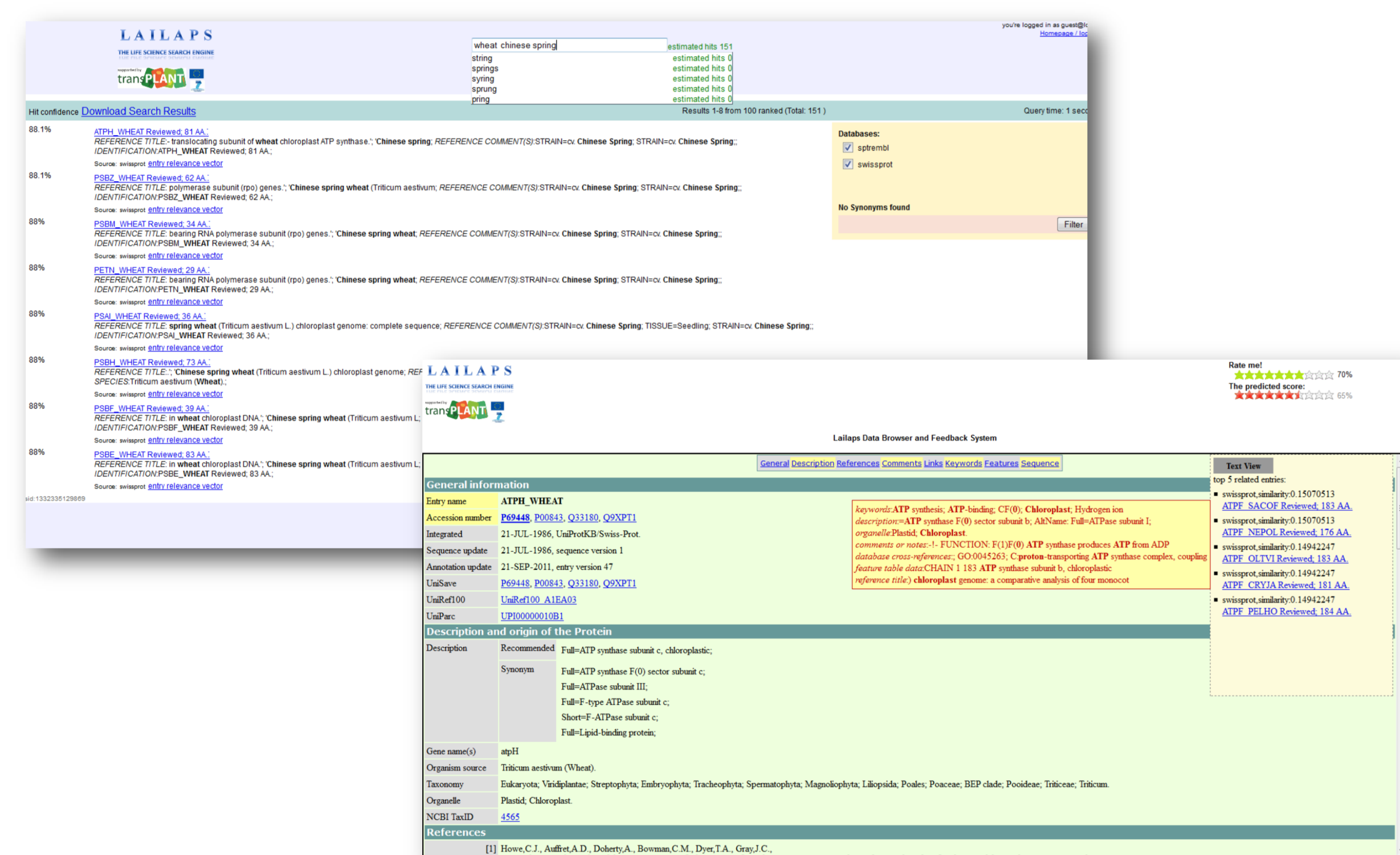


- fast keyword based search
- non-static relevance ranking
- self learning by user tracking
- installer for in-house deployment
- user specific relevance profiles
- suggestion of related entries
- deployable at standard desktop PC
- 100% JAVA

## LAILAPS System Architecture
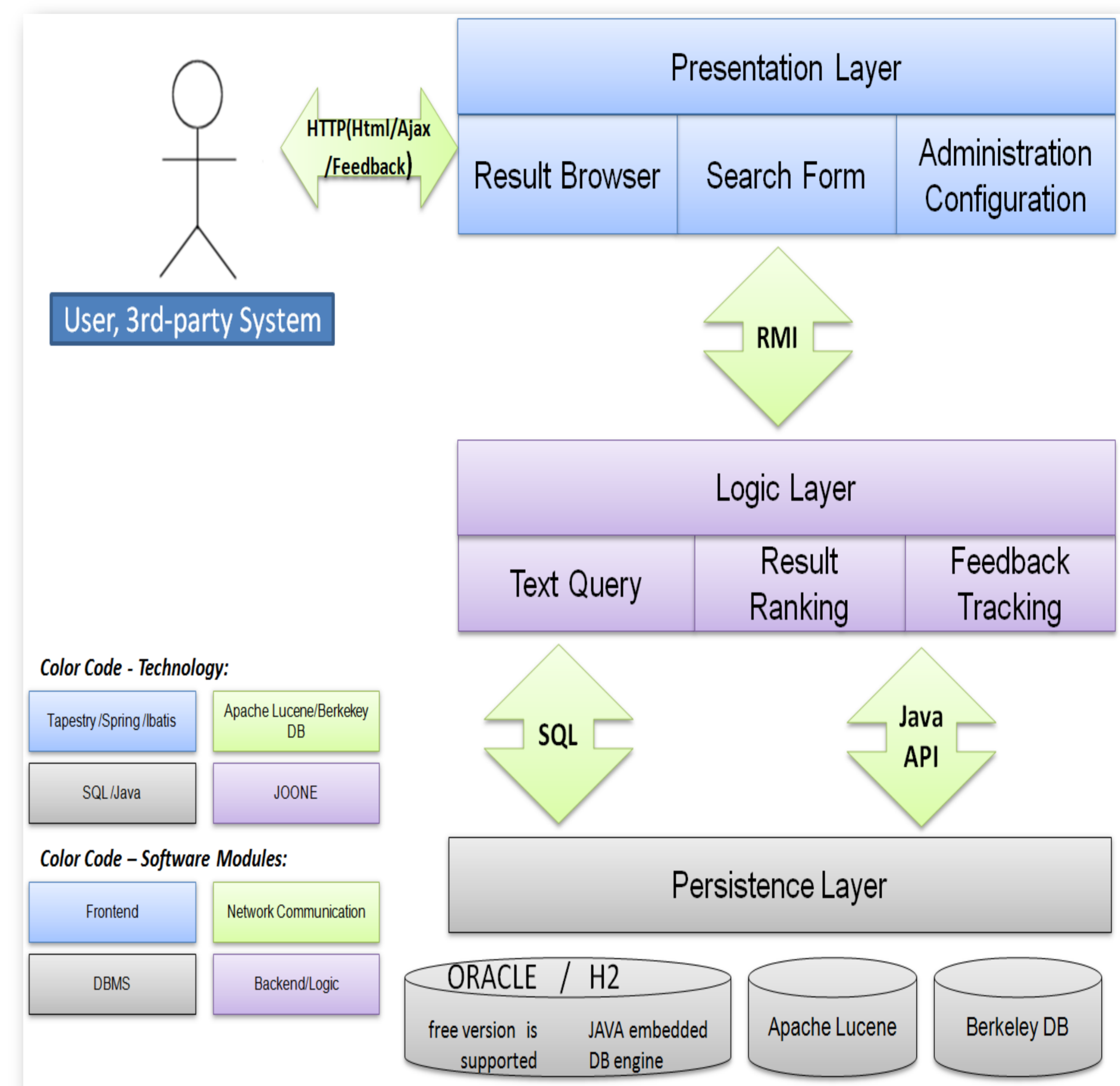


LAILAPS is implemented as 3-tier system, consist of frontend, business logic and database backend. The frontend is a J2EE web application, The business server is implemented as open API using JAVA RMI. This enables software developer to program customized, distributed JAVA based search applications. Using optimized storage backend, we guarantee a very well scaling management of very big text indexes over dozens of integrated life science databases and as well as ultra fast query suggestion by bloom filter technology.
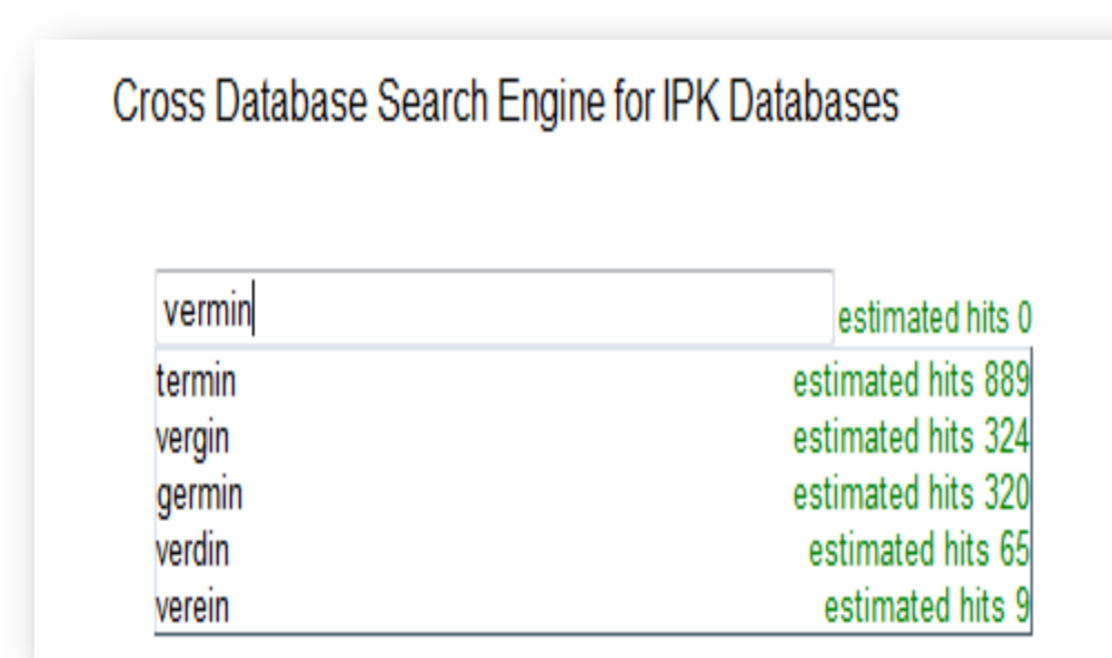
## LAILAPS New Features

### Installation program
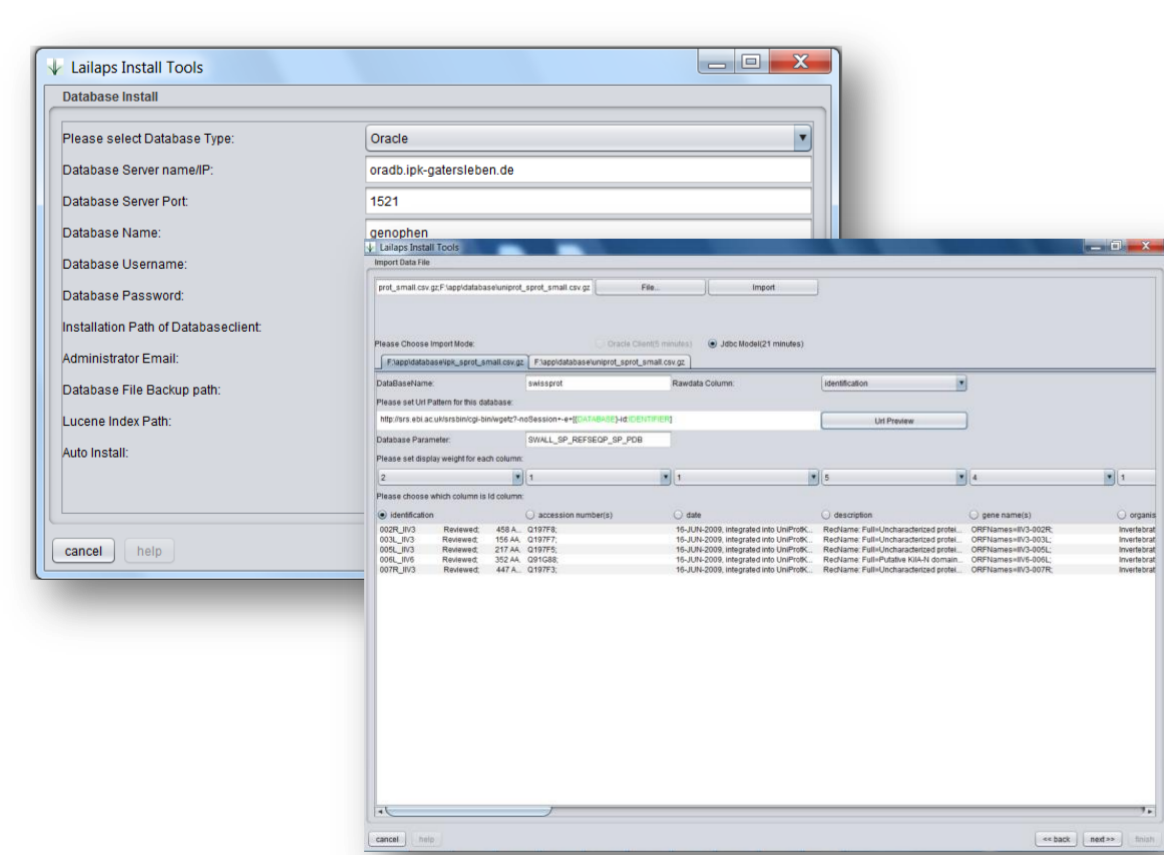


wizard style installation program

- import user-defined CSV format database, URL, synonym lists, keyword lists
- train neural network
- define weights for different databases and database fields
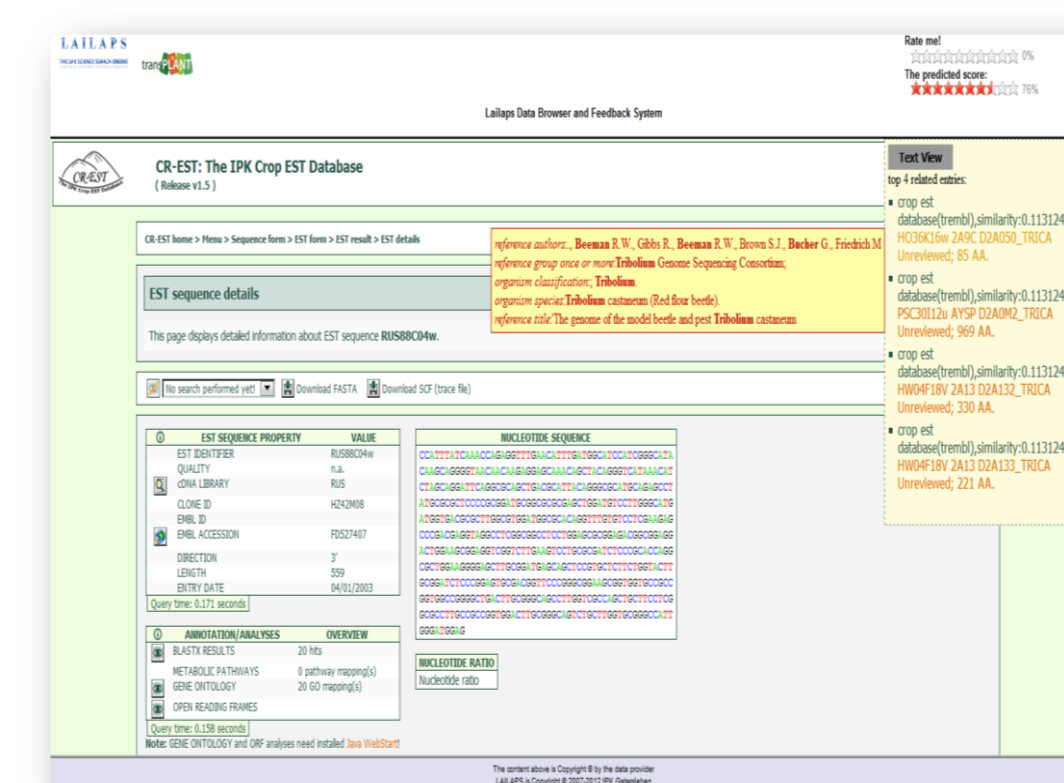- support any JAVA platform

### Auto completion



query suggestion

- ultra fast
- based on Bayes' theorem and edit distance
- support mass corpus with limited memory
- efficient data structure based on Bloom Filter
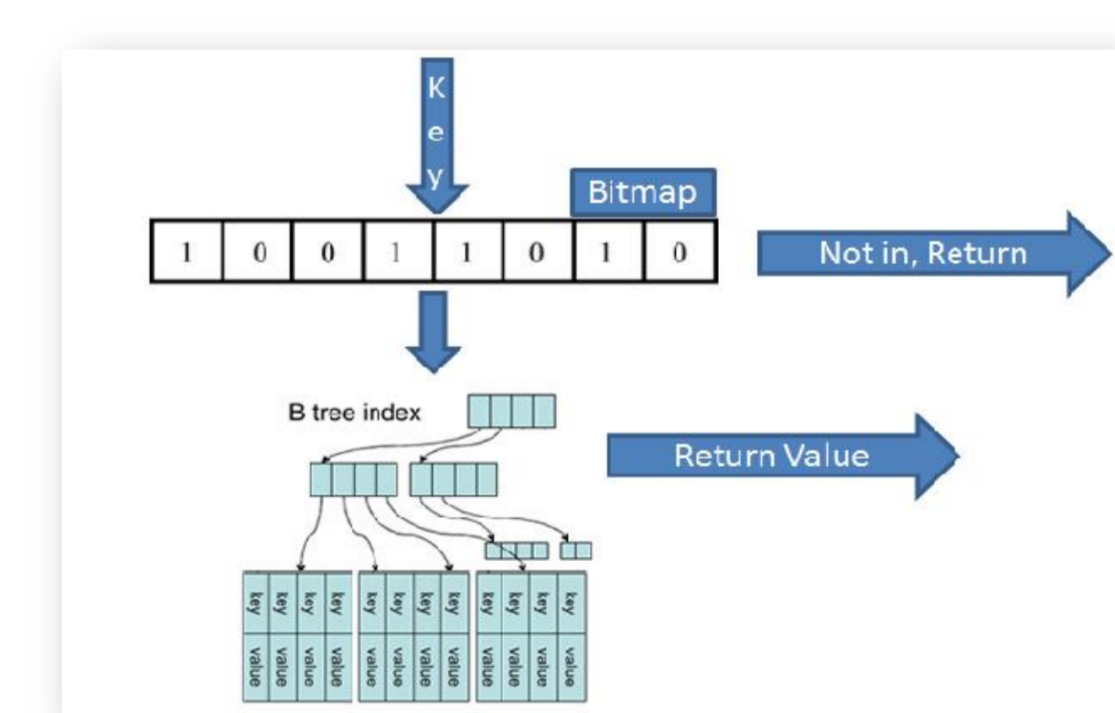- automated, non-static corpus extraction from the data that is indexed by LAILAPS

### Page like this



recommendation of similar data records

- based on TF-IDF (Term Frequency - Inverse Document Frequency)
- enhanced implementation in progress (e.g. improved Simhash)

### Key-value database



two level key-value storage

- indexed data content is stored in fast key-value database
- two level key value query:
  1. in-memory bitmap cache for ultra fast key lookup
  2. key-value database for querying data
- reduce IO operation using compression based data serialization and deserialization

## References

M. LANGE et al. (2010) The LAILAPS Search Engine: Relevance Ranking in Life Science Databases. Journal of Integrative Bioinformatics, 7(2):e110, 2010.

Leibniz Gemeinschaft     IPK GATERSLEBEN