GeneFetch : enriching gene information with text mining

Sofie Van Landeghem[1,2], Yvan Saeys[1,2] , Chris Cornelis[3], Mohammed Adberrafi Benotmane[4] and Yves Van de Peer[1,2]

[1] Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Gent, Belgium

[2] Department of Molecular Genetics, Ghent University, Gent, Belgium

[3] Computational Web Intelligence, Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium

[4] Laboratory of Radiobiology, Studiecentrum voor Kernenergie/Centre d'etude de l'Energie Nucleaire (SCK-CEN) Boeretang 200, Mol B-2400, Belgium

E-mail : sofie.vanlandeghem@ugent.be
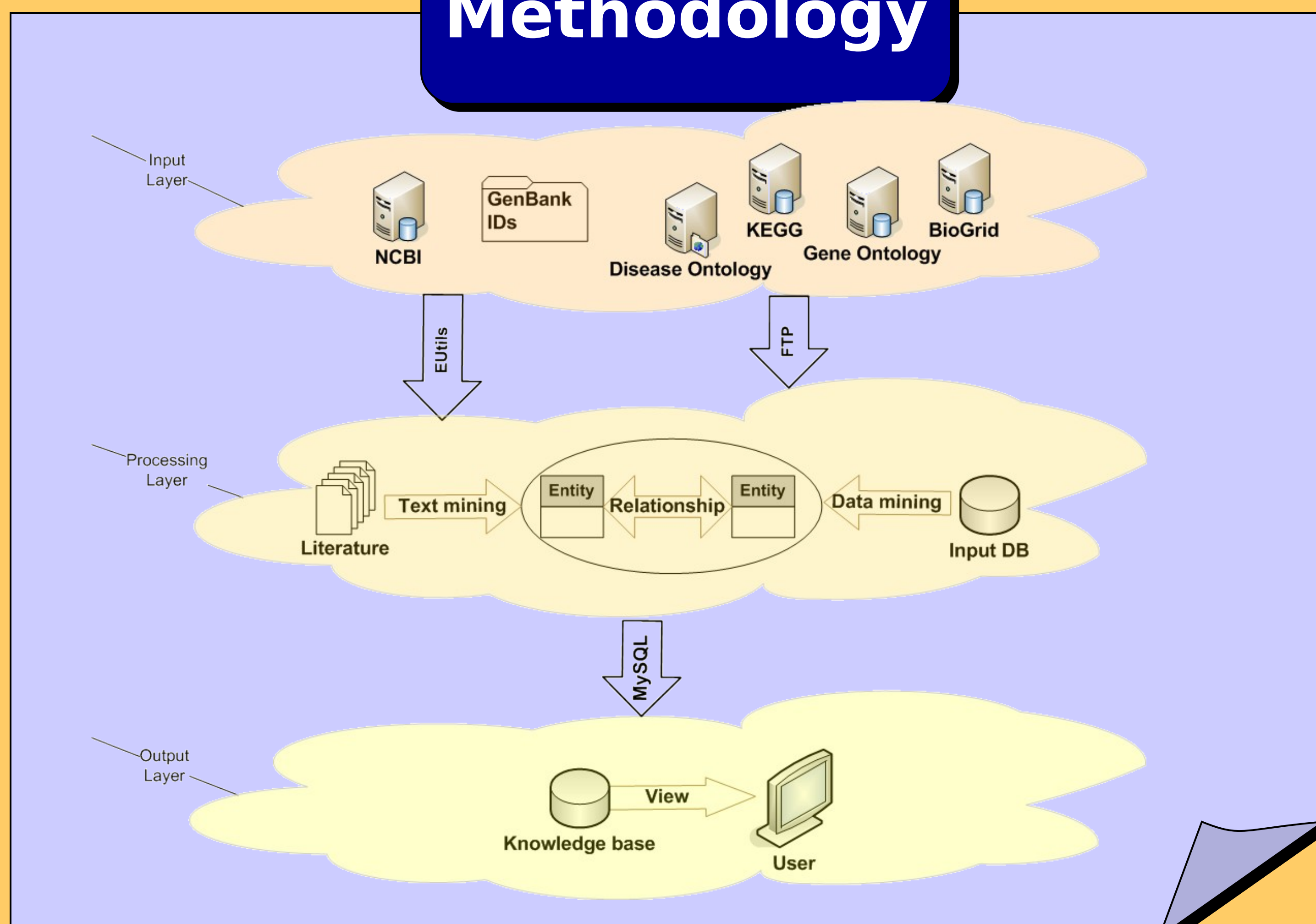
Introduction

As publicly available data in the field of life sciences is increasing continuously, **data mining** techniques that automatically **extract knowledge** from online databases and scientific literature have become a time saving necessity. We present **GeneFetch**, a system that enables biologists to easily **find and browse vital information on any gene or group of genes**. A **knowledge base** is constructed automatically by retrieving facts from established sources like the National Center for Biotechnology Information and Gene Ontology, as well as by performing **text mining** on PubMed articles to find links between genes, diseases and GO terms. We offer up-to-date records by incrementally expanding the database as new data become available. GeneFetch is the first application that **fully integrates facts with text mining results**, offering a full report per gene. This summary can be used by a whole range of applications, including clustering and customized visualization. Our example web application is available at **http://bioinformatics.psb.ugent.be/webtools/genefetch**.

GeneFetch

Find common properties of a group of genes

Look at the genes involved in a certain process

Show origin of extracted relationships

Methodology

Features

- Relationships between genes, diseases, pathways, molecular functions, biological processes, cellular components and articles
- Data mining of NCBI and GeneOntology databases
- Text mining on PubMed articles
- Results combined and stored in local database
- Fast access to data
- Database updated and incrementally expanded
- Profile for each gene
- Summary for each gene listing related biological entities
- Evidence stored to show origin of extracted relationships
- General properties of a group of genes
- Organism independant
- Filter functionality

Applications in systems biology

- Constructing protein-protein **interaction networks**
- **Clustering** of genes using annotations extracted from literature and through data mining
- **Prediction** of protein function and other properties through the **analysis of homologues**