

EMBL

# Cross-disciplinary knowledge integration

Evangelos Pafilis, Georgios Pavlopoulos, Sean O'Donoghue, Reinhard Schneider

European Molecular Biology Laboratory, EMBL, Heidelberg, Germany

{pafilis,pavlopou,odonoghu,rschneid}@embl.de

Data Integration and Knowledge Management



Evangelos Pafilis  
Services developer



Georgios Pavlopoulos  
3D visualisation



Dr Sean O'Donoghue  
Project co-supervisor

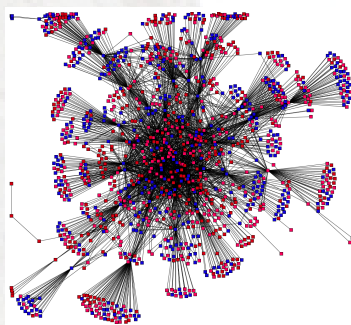


Dr Reinhard Schneider  
Project supervisor

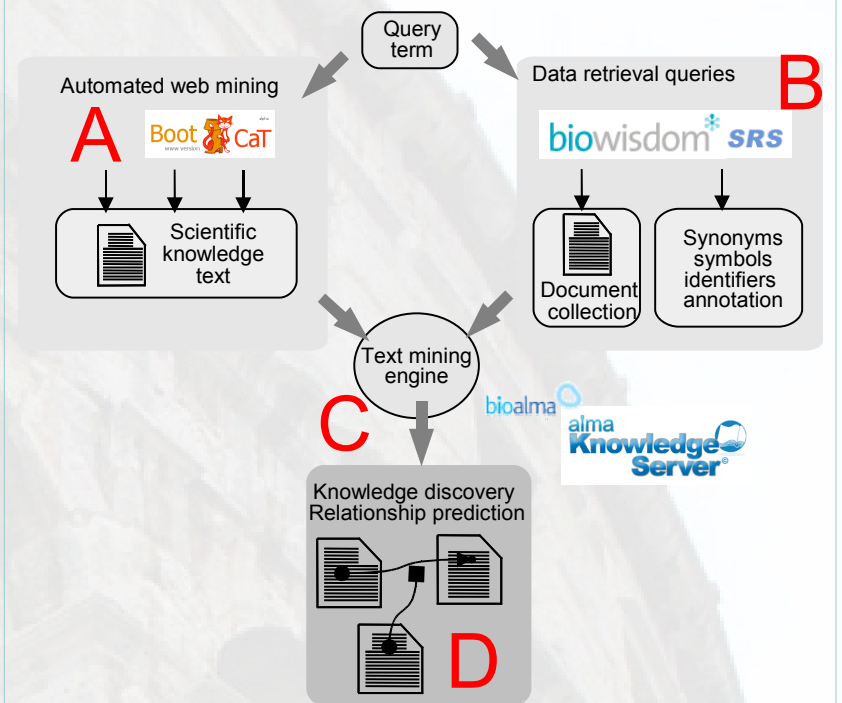
## Introduction

- Increasingly many biological-related databases and services are available via web services.
- This enables the creation of new services by combining existing services.
- Our goal is to combine different services into a single view for each entity, e.g. proteins, chemicals, diseases, etc.
- Current methods for visualizing cross-disciplinary data are mostly 2D, such as Medusa shown here.

- Our goal is also to visualize interconnectivity between entities, and allow intuitive navigation to related entities.
- The result will be a novel and useful web-based service for exploring biological data.

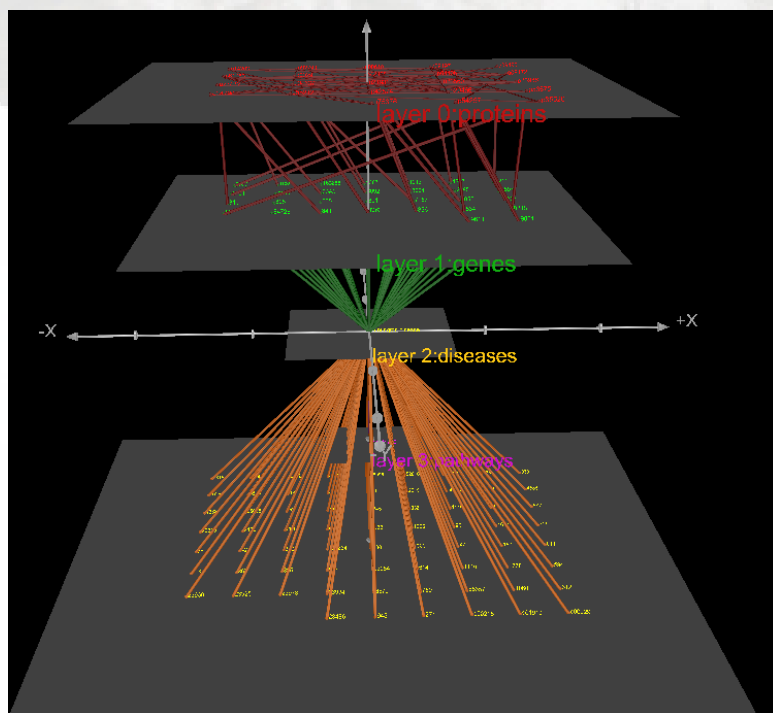


## Methods



**Figure 1:** System architecture: **a.** web-based document collector based on BootCaT, **b.** biological database integration system interface based on SRS, **c.** text mining interface based on Bioalma **d.** 3D visualisation (Fig. 2).

## Results



**Figure 2:** The figure (left) shows a multi-layer 3D representation of proteins, genes, diseases, and pathways related to "huntington disease" displayed on our Arena3D tool.

- The figure was automatically generated from the output of the text mining method shown in Figure 1.
- Our system also has Taverna-compliant interfaces to the SRS data integration platform, based on SRS web service objects.
- We are using our system in collaboration with systems biology groups to visualize and investigate new experimental data.

## Future work

- We are continuing to develop and integrate our text-mining, database-access, and 3D visualization tools.
- We plan to enable 3D visualization of queries, as well as comparative analysis of whole genomes and proteomes.
- We also plan to automatically add additional text documents from the web, e.g. from the FDA site.
- Our system will be available in the near future, as a web server, as well as client-based tools.

## References

- Web services: <http://en.wikipedia.org/wiki/Webservice>
- Workflow: <http://en.wikipedia.org/wiki/Workflow>
- Taverna: <http://taverna.sourceforge.net/>
- SRS: <http://www.biowisdom.com/solutions/srs/>
- Bioalma, almaKnowledgeServer: <http://bioalma.com/>
- BootCat: <http://sslmit.unibo.it/~baroni/bootcat.html>
- Medusa: <http://coot.embl.de/medusa/>