Master Degree Project:

Integrated Bibliometric Analyses by Combining Web Services

Who are we looking for? We are searching for a student with a background in biomedical text mining and/or bioinformatics to help explore the future of systematic literature searching. Programming experience in R, Python and/or Java, and at least a strong motivation to learn about web services, XML and JSON parsing are required. Experience with pathway, network analysis and visualization tools such as Cytoscape will certainly be beneficial, but can also be learned during the project.

What is it all about? Careful study of the published scientific literature is an essential first step when entering a new field of research or when reviewing a particular topic. To make such literature studies systematic, efficient and reproducible, we have recently developed scientific workflows integrating Europe PubMed Central (PMC) (1-2) web services and combined them with other RESTful web services, such as gene symbol lookups, named-entity recognition (NER) and biological pathway analysis. These services can be combined in scientific workflows or scripts. In this project you will investigate how to combine these web services in large-scale bibliometric analyses.

Research questions: How well do the newly developed bibliometric web services interact? What are the advantages and disadvantages of using web services compared to local databases and locally running software? Which services create bottlenecks in bibliometric analyses workflows, and how can these be handled? How do different NERs perform on different corpora (bodies of text, e.g. abstracts and full texts from the scientific literature or patent documents)? These are questions that should be addressed in this project.

With whom do you work? In this project you will work closely with the team of Dr. Magnus Palmblad, Head of Bioinformatics at the LUMC Center for Proteomics and Metabolomics, extending the knowledge of available and developing web services that can be used to support researchers in systematic literature searches. The work will consist of designing bibliometric analyses by connecting existing Europe PMC and other web services, and integrating these with NER and database services for genes, proteins, small molecules and diseases. The work can be performed in Uppsala or Leiden. Supervision can be in Swedish or English.

For more information, contact Dr. Magnus Palmblad by e-mail: n.m.palmblad@lumc.nl.

References

1.   Automating bibliometric analyses using Taverna scientific workflows: A tutorial on integrating Web Services, Güler AT, Waaijer CJF, Mohammed Y and Palmblad M, J. *Informetrics* 10 (3), 830–841 (2016)

2.   Europe PMC: a full-text literature database for the life sciences and platform for innovation, The Europe PMC Consortium, *Nucleic Acids Res*. 43, D1042–D1048 (2015)