

Tilburg University

Text mining for generating biological pathways (Abstract)

Waagmeester, A.S.; Donkers, H.H.L.M.; Evelo, C.T.A.; van den Herik, H.J.

Published in:

Third international symposium on Networks in Bioinformatics, Amsterdam, May 29-31, 2006

Publication date:

2006

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Waagmeester, A. S., Donkers, H. H. L. M., Evelo, C. T. A., & van den Herik, H. J. (2006). Text mining for generating biological pathways (Abstract). In *Third international symposium on Networks in Bioinformatics, Amsterdam, May 29-31, 2006* (Vol. 3, pp. 1). [s.n.].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Text mining for generating biological pathways

A.S. Waagmeester*[†] H.H.L.M. Donkers* C.T.A. Evelo[†] H.J. van den Herik*

Pathways are graphical representations of interactions between biochemical components in a cell. They are currently used (1) to facilitate the exchange of knowledge among colleagues and (2) to achieve a better understanding of biological systems. Most biologists will be able to reproduce the well-known pathways, such as the one for glycolysis or the citric acid cycle. The third reason to apply pathways is rather new, that is in the analysis of (gen)omics data. Methods such as Micro-array analysis produce huge amounts of data that are hard to analyze. Pathways can provide a way to structure and relate these data. Since increasingly more micro-array data comes available that asks to be analyzed, there is a need to generate appropriate pathways. The atomic parts of pathways are the *interactions* between biochemical components. In order to generate a pathway it is necessary to know these interactions, which usually come from three sources: (a) the intrinsic knowledge of a scientist, (b) the scientific literature, and (c) databases, such as REACTOME and GENMAPP [2, 4].

A pathway is usually generated by merging knowledge from these three sources. So far, it is a manual process in which a scientist translates the knowledge captured in the literature, into his own knowledge framework. So, the knowledge becomes intrinsic and is used to complete the generation of pathways, meanwhile partly supported by software tools which visualize database knowledge.

An intriguing question is: can we automatically make pathways from knowledge published in the literature, i.e., with the help of tools that currently are hand-driven and soon will be replaced by intelligent support tools?

This question is also vital for another reason: the amount of knowledge in scientific literature is growing almost exponentially [3] and that the knowledge is growing in every sub-domain of science. This is why we previously raised the question whether all known biochemical interactions can be extracted from literature by translating it first into the knowledge framework of the scientist. Here the danger was: are not we overlooking any knowledge? Therefore our research should focus on complete knowledge extraction and automatic translation.

We believe that text-mining methods may be an aid in the discovery and visualization of interactions between biochemical components captured in the literature. Text mining is the collective term for information-retrieval methods in which the source of information is unstructured text such as scientific literature.

In the poster we first present the overall process of visualizing knowledge needed for constructing a pathway. Subsequently we focus on the value of text mining methods in generating pathways from literature. Since text mining is a collective term, we have created a taxonomy to classify the specific text-mining methods needed for visualizing pathway knowledge in literature. This taxonomy is based on earlier published taxonomies [1, 3]. It provides an overview of existing methods and those that need yet to be developed in relation with our task. We end our poster with an overview of this taxonomy.

References

- [1] A. M. Cohen and W. R. Hersh. A survey of current work in biomedical text mining. *Briefings In Bioinformatics*, 6(1):57–71, 2005.
- [2] K. D. Dahlquist, N. Salomonis, K. Vranizan, S. C. Lawlor, and B. R. Conklin. Genmapp, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genetics*, 31(1):19–20, 2002.
- [3] L. J. Jensen, J. Saric, and P. Bork. Literature mining for the biologist: from information retrieval to biological discovery. *Nature Reviews Genetics*, 7(2):119–129, 2006.
- [4] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D’Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research*, 33:D428–D432, 2005.

*BioMICC, MICC - IKAT, Universiteit Maastricht, 6200 MD Maastricht, the Netherlands

[†]Bigcat bioinformatics, Universiteit Maastricht, 6200 MD Maastricht, the Netherlands