Chapter XIX
Seekbio:
Retrieval of Spatial Relations for System Biology

Christophe Jouis
Université Paris III – Sorbonne, France, & LIP6 (Laboratoire d'Informatique de Paris VI – Université Pierre et Marie Curie), ACASA team, France

Magali Roux-Rouquié
LIP6 (Laboratoire d'Informatique de Paris VI – Université Pierre et Marie Curie), ACASA team, France

Jean-Gabriel Ganascia
LIP6 (Laboratoire d'Informatique de Paris VI – Université Pierre et Marie Curie), ACASA team, France

ABSTRACT

Identical molecules could play different roles depending of the relations they may have with different partners embedded in different processes, at different time and/or localization. To address such intricate networks that account for the complexity of living systems, systems biology is an emerging field that aims at understanding such dynamic interactions from the knowledge of their components and the relations between these components. Among main issues in system biology, knowledge on entities spatial relations is of importance to assess the topology of biological networks. In this perspective, mining data and texts could afford specific clues. To address this issue we examine the use of contextual exploration method to develop extraction rules that can retrieve information on relations between biological entities in scientific literature. We propose the system Seekbio that could be plugged at Pubmed output as an interface between results of PubMed query and articles selection following spatial relationships requests.

INTRODUCTION

For decades, it was thought that more an organism was complex, more the number of genes they contain had to be high. At the completion of the Human Genome Project, it was found that living systems were having approximately the same number of genes coding for proteins (about 20,000) and these assessments resulted in biology's big bang undergoing a paradigmatic change to address biological complexity. Accordingly, the universe of biologists was changing and the challenge of unravelling complexity of living organisms was stipulated on the spatio-temporal variety of the biological components and their relations. Countless examples show that identical molecules could play different roles depending of the relations they may have with different partners embedded in different processes, at different time and/or localization; let mention, muscle differentiation that is orchestrated by the differential localization of a molecular species (Misca et al., 2001) or molecular gradients that regulate the developing eye polarity (Strutt, 1999). To address such intricate networks that account for the complexity of living systems, systems biology is an emerging field that aims at understanding such dynamic interactions from the knowledge of their components and the relations between these components. Among main issues in systems biology, knowledge on entities spatial relations is of importance to assess the topology of biological networks; in this perspective, mining data and texts could afford specific clues. To address this issue, we examine the use of contextual exploration method (Jouis, 1993; Jouis, 2007) to develop extraction rules that can retrieve information on relations between biological entities in scientific literature. To achieve this, we propose SEEK*bio* tool. Among further uses, SEEK*bio* could be plugged at Pubmed (Srinivasan, P., 2001) output as an interface between results of PubMed query and articles selection following spatial relationships requests. These spatial relationships requests would consist of terms graph of the original query and location relations put by the user in the SEEK*bio* graphs editor. The built graph would act as a filter to select relevant articles from a spatial point of view (see figure 1). Enabling this strategy should return selected documents in which relations arguments could be added automatically.

RELATED WORKS

Two different approaches to extracting relationships from biological texts are used (Ananiadou, S., Kell, D. B. & J. Tsujii., 2006).

Co-occurrence

The simplest approach is to identify entities that co-occur within abstracts or sentences in texts. Most of techniques make use of a score based on the apparition frequency of terms to establish and classify terms relationships (Donaldson, I et al., 2003), (Cooper, J.W. & Kershenbaum, A., 2005), (Stapley, B. J. & Benoit, G., 2000). If a couple of terms is regularly repeated, the terms are supposed to have relations although the type of relation is not precised (Wren, J. D. & Garner, H. R., 2004). Learning techniques such as artificial neural networks (ANNs), support vector machines (SVMs), hidden Markov models (HMMs), and naïve Bayes classifiers (NB) are used to detect such relations. They have been adapted for spatial information retrieval in biology. For example, the sub-cellular location of proteins was predicted from text using support vectors machines (Stapley, B. J. et al. 2002).

Otherwise, Protein location was predicted using sequence homology and retrieved information on known or predicted functions, interacting partners, additional sequence motifs, source organism, etc. This approach used a Naïve Bayes classifier to make a prediction about the place where a protein should be located, using only the protein sequence as its input and search against

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/seekbio-retrieval-spatial-relations-system/23070

Related Content

An Overview of Graph Indexing and Querying Techniques
Sherif Sakrand Ghazi Al-Naymat (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications (pp. 222-239).*
www.irma-international.org/chapter/overview-graph-indexing-querying-techniques/76064

A Novel Radial Basis Function Networks Locally Tuned with Differential Evolution for Classification: An Application in Medical Science
Ch. Sanjeev Kumar Dash, Ajit Kumar Behera, Satchidananda Dehuriand Sung-Bae Cho (2013). *International Journal of Systems Biology and Biomedical Technologies (pp. 33-57).*
www.irma-international.org/article/a-novel-radial-basis-function-networks-locally-tuned-with-differential-evolution-for-classification/89399

Neuroscience and Symptoms Related to the CADASIL Disease
Katerina Ioannidou, Dimitrios Vlachakis, George Matsopoulosand Sophia Kossida (2013). *International Journal of Systems Biology and Biomedical Technologies (pp. 17-23).*
www.irma-international.org/article/neuroscience-and-symptoms-related-to-the-cadasil-disease/105595

Normalizing Cross-Border Healthcare in Europe via New E-Prescription Paradigms
Alexander Berlerand Ioannis Apostolakis (2014). *Research Perspectives on the Role of Informatics in Health Policy and Management (pp. 168-208).*
www.irma-international.org/chapter/normalizing-cross-border-healthcare-in-europe-via-new-e-prescription-paradigms/78697

Introduction to Classification Error Estimation
(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations (pp. 334-340).*
www.irma-international.org/chapter/introduction-classification-error-estimation/53912