

Chapter XV

Mining BioLiterature: Towards Anotations of Genes and Proteins

Francisco M. Couto, Universidade de Lisboa, Portugal
Mário J. Silva, Universidade de Lisboa, Portugal

Abstract

This chapter introduces the use of Text Mining in scientific literature for biological research, with a special focus on automatic gene and protein annotation. This field became recently a major topic in Bioinformatics, motivated by the opportunity brought by tapping the BioLiterature with automatic text processing software. The chapter describes the main approaches adopted and analyzes systems that have been developed for automatically annotating genes or proteins. To illustrate how text-mining tools fit in biological databases curation processes, the chapter presents a tool that assists protein annotation. Besides the promising advances of Text Mining of BioLiterature, many problems need to be addressed. This chapter presents the main open problems in using text-mining tools for automatic annotation of genes and proteins, and discusses how a more efficient integration of existing domain knowledge can improve the performance of these tools.