A Hybrid Approach incorporating XML and NLP Techniques for Focused Information Retrieval in the Biomedical Domain

Thilky Perera
Faculty of IT, Queensland University of Technology, Brisbane, Australia
mt.perera@student.qut.edu.au

Abstract: Research in Biomedical IR has gained momentum in recent years with the advent of gene identification techniques and the subsequent promise of personalized medication. This in turn has triggered a need for domain-specific IR techniques to be adopted in order to satiate the various information needs of the biomedical research fraternity. We propose here, such a domain-specific approach, which address, issues pertinent to the biomedical domain; mainly those specific to the end-users of the domain and the issues due to the inherent nature of the biomedical literature itself. To this end, we have analyzed the various approaches pursued in this direction with due attention to their claimed merits as well as the explicit/implicit shortcomings. This analysis is then piped in as the ‘requirement analysis’ for our next step towards a plausible solution pathway; a hybrid model encompassing XML and NLP-related techniques for Biomedical IR. We perceive that, if brought together in the right order and depth, XML and NLP could complement each other by yielding a most natural and intuitive pathway in resolving the issues identified so far in our literature review. Through this hybrid approach, we intend to contribute to a focused information retrieval strategy for the biomedical domain.