

# Identifying Effectual Indexing Terms for the Automatic Annotation of Biomedical Illustrations

*M. Simpson, D. Demner-Fushman, C. Sneiderman, S. Antani, G. Thoma*

## **Abstract:**

The automatic annotation of biomedical images is an attractive approach for enabling convenient access to a plethora of images found in biomedical articles. Such images have relevance in support of clinical decisions and for instructional purposes. However, extracting and selecting appropriate indexing terms from the publication is difficult given the domain-specific nature typical of biomedical images and the large number of vocabularies in the biomedical sciences. The goal of this work is to improve image annotation quality by selecting for indexing only the extracted terms which are most likely to be useful. We do this by training a binary Naive-Bayes classifier with non-lexical features generated from the extracted terms. Experimental results indicate we are able to significantly reduce the number of ineffectual indexing terms.