PROGRAM COMMITTEE:

Stephen Johnson, Chair (Columbia University)
Judith Klavans, Co-chair (Columbia University)
Udo Hahn, Co-chair (Freiburg University)

Robert Baud (University Hospital of Geneva)
Carol Friedman (Columbia University)
Robert Futrelle (Northeastern University)
Lynette Hirschman (MITRE Corporation)
Jun'ichi Tsujii (University of Tokyo)
Alexa McCray (National Library of Medicine)
Tom Rindflesch (National Library of Medicine)
Donia Scott (University of Brighton)
Nina Wacholder (Rutgers University)
Bonnie Webber (University of Edinburgh)
W. John Wilbur (National Center for Biotechnology Information)
Pierre Zweigenbaum (Assistance Publique - Hôpitaux de Paris)

CONFERENCE WEBSITE:

http://www.dmi.columbia.edu/nlpwg/ACL02.html

INTRODUCTION

This volume contains the papers accepted for presentation at the Workshop on Natural Language Processing in the Biomedical Domain, held at the University of Pennsylvania on July 11, 2002, just following the 40th Meeting of the ACL.

Biomedicine is a large domain, comprising biological sciences, clinical medicine, and public health. While this subject area is vast, spanning from the molecular level to whole populations, there is a unifying focus on health and disease. These characteristics imbue biomedical language with unique properties: an enormous number of lexical items but relatively small number of semantic patterns.

Biomedicine presents many opportunities for application of NLP technologies such as information extraction from biomedical texts, document and answer retrieval from large, unstructured text collections (such as the biomedical literature and the World Wide Web), and interaction with users through natural language.

The principal purpose of the workshop was to explore challenges in processing biomedical language and to present results in developing techniques for this domain. Another important motivation was to bring researchers together from computational linguistics, bioinformatics and medical informatics. Until recently, the level of collaboration among these disciplines has been limited. Indeed, this was the first workshop under the auspices of the ACL entirely devoted to biomedical language processing.

We received a total of 26 submissions, from which 12 were selected for presentation, using a double-blind refereeing process. The papers in this volume represent work from 5 countries in Asia, Europe, and North America, which provides some evidence of the growing interest in this domain. The submissions also demonstrate the considerable breadth of research in this area.

The papers are grouped into four themes. These illustrate some of the key issues currently being explored in biomedical language research. The first group, "Biomedical Name Recognition" deals with the challenge that biomedical names (eg genes and proteins) are not only myriad, but also constantly growing. The second group "Machine Learning of Biomedical Language" investigates the use of machine learning techniques by exploiting various aspects of the restricted nature of biomedical language. "Biomedical Indexing" returns to the large vocabulary problem in the context of information retrieval, a vital application in medicine and biology. Finally, "Biomedical Information Resources" explores various ways in which NLP can help biomedical researchers and clinicians access the knowledge emerging in the scientific literature.

The idea for this workshop originated at the annual meeting of the American Medical Informatics Association in November, 2001. I am indebted to my co-chair Judith Klavans for the suggestion, and also for her encouragement and advice. I would also like to thank the program committee for their assistance in refereeing the papers, and also for many helpful suggestions.

Stephen Johnson
May 2002