

Preface

With the emergence of the World Wide Web and recent growth in Digital Libraries, the amount of on-line text that is now available is staggering. As a result, the need for computational tools to help users deal with the vast amounts of available textual information is also growing quickly. Specifically, automated learning methods have the potential to help users filter relevant from non-relevant information, adapt to users' interests over time, and automatically find new and interesting pieces of information. Thus, the success of such learning methods has the potential to play a significant role in how end users interact with information in the future.

The goal of this workshop is to gather together researchers interested in the theory and application of learning methods aimed at addressing the problems of text categorization. In doing so, we hope to foster further understanding and interaction between the various research communities examining these problems. More specifically, by seeking possible synergies between fields such as Machine Learning, Bayesian Networks, Information Retrieval, Natural Language Processing, Case-Based Reasoning, Language Modeling and Speech Recognition, it may be possible to gain better insights into building more powerful tools for text categorization.

Some of the particular issues we hope to address in this workshop include:

- The use of natural language processing and knowledge-based systems in text categorization.
- A better understanding of machine learning models used with text and their possible extensions.
- Issues related to categorizing documents over time.
- Considering domain characteristics of text collections.
- Agents that act and learning in text domains.
- Innovative applications of text categorization.

This is by no means an exhaustive list of topics. Rather, we hope that the papers presented at this workshop serve as a starting point for new research directions and collaborations between researchers interested in these and related problems.

The immense research interest in this area led to a very large number of full papers being submitted to this workshop. Consequently, we were able to accept less than 40% of these papers for full presentation, given the time constraints and the desire to allow ample time for discussion and interaction. As a result, however, we believe that the workshop has a very exciting technical program, outstanding invited speakers and will generate a great deal of lively discussion.

We would like to thank all the contributors of papers to the workshop for their work and all the workshop participants for their interest. We are especially indebted to our invited speakers Jaime Carbonell and Fernando Pereira. Finally, we would like to thank the AAAI-98 workshop chairs David Leake and Raymond Mooney for their help and the staff at AAAI and ICML for organizational support.

Mehran Sahami, Chair
Mark Craven
Thorsten Joachims
Andrew McCallum