

Using Information Extraction to Build a Directory of Conference Announcements

Karl-Michael Schneider

University of Passau, Department of General Linguistics
Innstr. 40, 94032 Passau, Germany
schneide@phil.uni-passau.de

WWW home page: <http://www.phil.uni-passau.de/linguistik/schneider/>

Abstract. We describe an application of information extraction for building a directory of announcements of scientific conferences. We employ a cascaded finite-state transducer to identify possible conference names, titles, dates, locations and URLs in a conference announcement. In order to cope with agrammatical text that is typical for conference announcements, our system uses orthographic features of the text and a domain-specific tag set, rather than general purpose part-of-speech tags. Extraction accuracy is improved by recognizing other entities in the text that are not extracted but could be confused with slot values. A scoring scheme based on some simple heuristics is used to select among multiple extraction candidates. We also present an evaluation of our system.

1 Introduction

A conference directory lists conferences, workshops and other meetings, often in one scientific field, such as Computer Science, Computational Linguistics or Artificial Intelligence.¹ A conference directory is a valuable resource for researchers who want to present their research to an interested audience, or are themselves interested in the research of others in their field, and for this purpose are looking for an appropriate conference.

An entry in a conference directory consists of at least the conference name (e.g. *CICLing 2004*), title (*International Conference on Intelligent Text Processing and Computational Linguistics*), date, location and a link to the conference web page that contains the full announcement and call for papers. New conference announcements are posted in newsgroups and distributed by electronic mail through mailing lists. Upcoming conferences should be included in the directory with little or no delay in order to keep the directory up-to-date.

Maintaining a conference directory requires to extract the relevant information from each incoming conference announcement. Doing this automatically

¹ See for example:

<http://www.aaai.org/Magazine/Departments/calendar.html>

<http://www.conferencealerts.com/>

<http://www.cs.jhu.edu/~yarowsky/sigdat.html>