CfP Scanner:
Extracting Conference Attributes from Calls for Papers

A Proposed Project

Jochen L Leidner / Mark Stevenson

Motivation

- Researchers publish their research at conferences
- To do so, they subscribe mailing lists where conference organizers post **Calls for Papers (CfPs)**

Example CfP

Call for Papers
==========

-- Apologies for cross-posting --

78th ACL Conference, Honululu, Hawai, USA, 3-7 July 2052.

We invite novel papers for publication in the areas of Natural Language Processing, Dialog, Summarization, Syntax & Semantics.

Important dates:
- 28 Jan 2052 Deadline for submitting main papers
- 1 Mar 2052 Notification of acceptance
- 3-7 July 2052 Main Conference

Jack Chomsky, Jr., General Chair

Program Committee
Bob Smith
Nina Miller
John McCann
Fred E. Llipsis

Idea: CfP Scanner

- Let’s automate it!
- Avoid:
  - manual reading of mailing lists
  - copy & paste (of conference names, deadlines) from emails to calendars or databases

Input, Extraction Processing and Output

- **Input**: An e-mail account comprising a set of incoming emails
- **Processing**: Scan incoming emails for calls for papers for conferences
  - Extract key relevant attributes from incoming e-mail mailbox
    - name of the conference
    - date and location of the conference
    - submission deadline
    - date of notification whether a contribution was accepted or not
    - conference date
- **Output:** Extracted conference attributes, one conference or workshop per record
  - comma-separated vertical (CSV) file format,
  - where each field of the record corresponds to an attribute

Collecting Data

- The Internet is full of mailing list archives containing CfPs
- Find a few of these and put together a data-set for conferences like e.g. ACL, EACL, COLING, CoNLL, EMNLP, *SEM, ECML/PKDD, ICML, NIPS, RANLP, CICling, LREC, SIGIR, ECIR.
- Explore WikiCFP as a source

Annotation & Evaluation

Evaluate your system:
- construct a data-set (at least a few dozen CfP e-mails), where you will
- manually label the relevant attributes (text span + class)

so you can use them for **training** and **testing** your CfP information extraction.

Student Prerequisites

You should have:

- Solid Python or Java programming skills
- Familiarity with regular expressions and the concepts of machine learning
- an interest in information extraction (IE)

References

[1] Liu, Bing (2011) *Web Mining. 2nd ed.* New York: Springer.