Extracting Information from Conference Announcements: High Recall, High Precision

Kevin Cheong
Language Technology Group, Microsoft Research Institute
School of MPCE, Macquarie University
Sydney NSW 2109, Australia
kevinc@mpce.mq.edu.au

Abstract

Conference announcements are distributed widely each day via electronic mail to the research and industrial community. These conferences inform researchers, academics and the industry about the research and development (R & D) work performed in a particular field of interest. There is a wealth of information contained in this multitude of conference announcements. The aim of this research is to extract essential and relevant information from conference announcements and to explore the technologies involved. In this paper we describe an architecture for a system we have developed that extracts relevant and useful information from conference announcement electronic mail messages, with a focus on achieving a high recall and precision rate. We also discuss the extent to which the success of this information extraction task depends on domain and world knowledge.

1 Introduction

Conference announcements encourage the community to participate in an event and share in an ongoing research field. Conference organisers distribute the announcements widely to ensure the community is informed and that interested parties will participate. Conference announcements are a genre of text that present factual information about events — the main event is the advertised conference, accompanied by two sub-events each corresponding to the conference submission and conference registration information. Conference information is structured and presented in an announcement to elicit a response from the reader by way of participation, decision-making or comment. To the human reader equipped with "world" knowledge, extracting information from conference announcements can be a simple task albeit a tedious one.

With the Internet, conference announcements are assured of wide distribution. The community encounters the multitude of announcements arriving in their electronic mailboxes, each possibly presented in a different and creative way. Like any on-line information, readers are faced with the inevitably daunting task of reading the conference announcements to manually extract conference information. It would indeed be useful if relevant information could be automatically extracted from these announcements. Consider a day in the life of a human reader receiving conference announcements[1] (Figure 1).

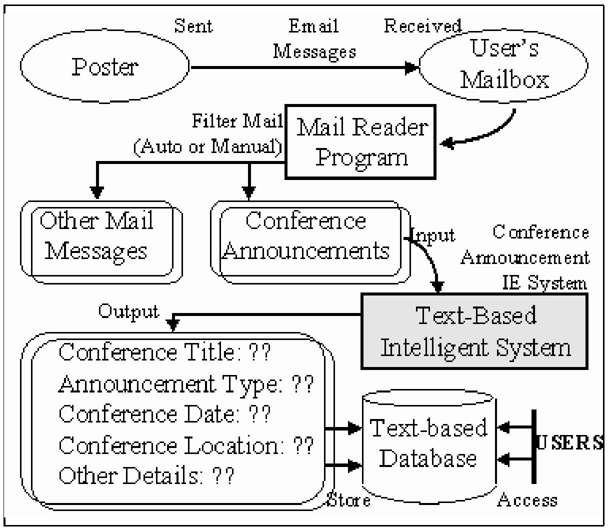

Figure 1: Life with Conference Announcements

Our society is inundated with information, ever increasingly in electronic form, as people continually produce and collect it. Our capacity to retrieve, interpret and use information wisely is challenged by the amount and variety of information available. Information can be packaged succinctly and to the point without any loss of vital information

[1] Appendix A shows a sample marked-up conference announcement.

Kevin Cheong (1998) **Extracting Information from Conference Announcements: High Recall, High Precision.** In D. Estival (ed.) *Abstracts for the ANLP Post Graduate Workshop,* pp 11-20.

and vagueness in conveying the intended message. This is evident in the multitude of announcements, advertisements, product descriptions, product comparisons and the likes, all short, succinct, precise, information-rich and readily available online to the "networked" public for perusal and assimilation.

There is a great need for automatic methods to deal with the large amounts of readily available online text. The common thrust of tasks dwells on using computers to sift through volumes of information, providing access to and synthesis from that information (Jacobs and Rau, 1993). The research of retrieving and extracting relevant and useful information from online (electronic) text attempts to deal with the large amounts of readily available online text by automating the process of information retrieval, assimilation and presentation.

This work is part of an ongoing research in extracting information automatically from conference announcements contained in electronic mail messages. The goal is to extract specific, relevant and useful information from conference announcement email messages, with a high recall and precision rate[2] ($\geq$ 90%), based upon a pre-specified set of IE task definitions. To express the practical aspect of this research work, a Conference Announcement Information Extraction (CAIE) system is built to assist users in extracting information of interest from email messages containing conference announcements.

In this paper we will discus three interesting issues that arise when automating the process of extracting relevant and useful information from conference announcements, which humans do well but find tedious and time-consuming.

1. What information is deemed relevant and useful in a conference announcement?
2. Do we need "world" knowledge to extract useful information from conference announcements? Is "domain" knowledge sufficient?
3. Can a high recall and precision rate be achieved for the task of extracting relevant and useful conference information from conference announcement email messages? More generally, is it feasible to build a high-recall, high-precision IE system for a specific task and domain?

The following section highlights the IE tasks; Section 3 details the genre and information contained in a typical conference announcement and Section 4 presents a simple architecture for building a high-recall, high- precision conference announcement IE system. Section 5 discusses future directions for this research work, while Appendix A shows a marked-up sample electronic mail message containing a conference announcement.

[2] Recall is the percentage of correct answers produced by a system given all possible correct answers. Precision is the percentage of correct answers given all the answers produced by the system.

2 Information Extraction Tasks

Information Extraction (IE) systems are a genre of Text- Based Intelligent Systems (Jacobs, 1992) that extract factual information of interest from a large corpus of text. Not all kinds of text are suited for IE. The primary purpose of these texts is to communicate factual information on events and participant entities. For example, the Message Understanding Conference (MUC)[3] participants have analysed texts from Wall Street Journals, microelectronic fabrication process documents, news reports on terrorist events and telegraphic messages on naval sightings. These texts contain information about events, participants, objects and their associated attributes, and relationships among these entities and events — namely who did what to whom, when and where. Although factual information spans the whole text, not every sentence in the text will be of interest. Hence, the task of extracting information is not well suited to a large corpus of text in which almost every sentence contains information for extraction. The goal of IE is not to understand the text as a whole but to focus on specific factual information of interest as defined by the IE tasks.

The IE tasks are,

1. to analyse the corpus of text and clearly specify a set of IE task definitions,
2. to create a set of information templates (abstract data structures — each template pertaining to a particular type of information for a specific domain) from the specified task definitions,
3. and to fill the templates with the extracted textual information (raw or normalised text form)

[3] The research of IE was the foundation of the Message Understanding Conferences 1–6; see (MUC-4, 1992), (MUC-5, 1993), (MUC-6, 1995) and (Hirschman and Vilain, 1995). These conferences promoted research in analysing volumes of free and unstructured text, extracting a pre-determined set of events and features from the text and storing the information in a structured and consistent manner, like a database. Message understanding is constrained natural language understanding. The task of understanding a message is not merely driven by the need of identifying events, entities and features in the text and extracting them, but also on the need for evaluating the results of an extraction system.

An IE system takes a text input, extracts information from this text with respect to the prespecified task definitions and fills the corresponding templates, created for this domain, with the extracted information. The filled template may be processed further or used to build/populate a new/existing database. It is important to determine and clearly specify a feasible IE task, with respect to the events and entities contained in the corpus of text, to achieve a reasonably good recall and precision rate.

3 Conference Information

Conference announcements are distributed widely each day via electronic mail (email). The initial corpus for analysis consists of 150 email messages containing conference announcements. A typical conference announcement email message is composed of three distinct parts (Figure 2; see Appendix A for a sample conference announcement email message).

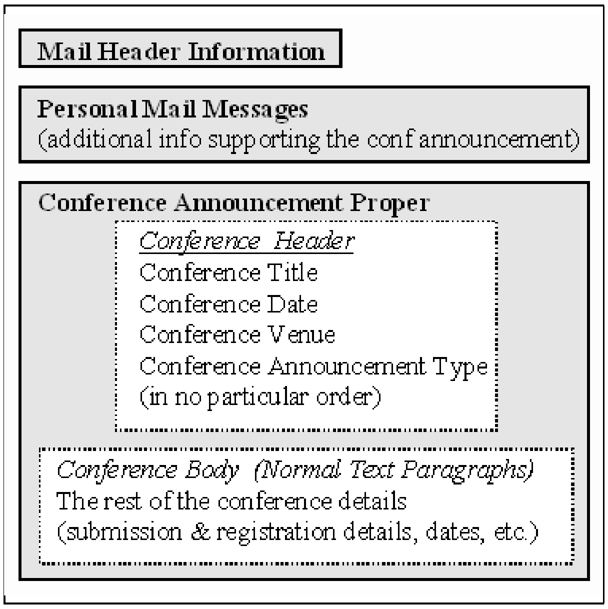

Figure 2: A typical Conference Announcement.

The *mail header information* consists of Internet mail routing information. This occurs at the very top of the conference announcement. Some conference announcements are distributed in digest form automatically by a list server[4]. List server information and commands are also embedded within this portion of the text.

[4]A list server is a small archive system that collects messages from a list of subscribers and routes these messages, single or in digest form, to other subscribers on the same list.

Personal mail messages may follow the mail header information and after the conference announcement proper (sender's signature).

This personal text is additional information accompanying and/or complementing the conference announcement, extra messages as intended by the sender. The information conveyed here is separate from the conference announcement proper, not originally intended by the creator and original poster of the conference announcement. Its purpose is to highlight portions of the announcement, to bring the reader's attention to a particular conference or to inform readers of certain changes.

The focus of extracting information from conference announcement email messages resides in this last part — *conference announcement proper.* Human readers may easily distinguish the *conference header* from the *conference body.* The conference header contains mainly the title of the conference, conference dates, conference location and/or the type of announcement. The subsequent conference body contains information pertaining to either submission details or conference registration. For some cases, the conference location and/or dates may be found in the conference body.

3.1 Conference Announcement Proper

A conference announcement informs the reader of a conference event and the corresponding process one needs to follow for submission and/or registration. Three main categories of relevant and useful information are abstracted from the conference announcement proper. Structure E is abstracted from the conference header; Structure S and R are abstracted from the conference body (see Figure 3).

This information was derived after analysing a corpus of 150 email messages containing conference announcements. 100 email messages sieved from the 150 emails formed the training corpus[5]. Another 150 email messages containing conference announcements were analysed (later) and proved to contain the same types of information, of which another 100 were filtered to form the test corpus (similar to the training corpus). These structures form the basis of the IE task and the information extracted from a conference announcement.

The relevancy and usefulness of information extracted from conference announcements is eventually subjective to the reader given a particular context and information requirement.

[5]50 email messages were eliminated because there were duplicates, conference announcements contained in a digest, or email messages that contained only hyperlinks to conference announcements.

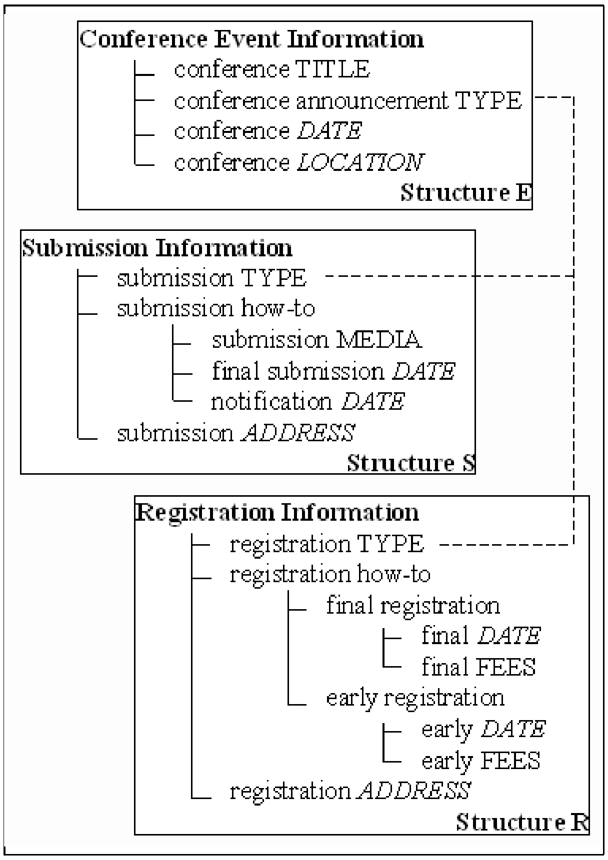

Figure 3: A Simple Structure for conference announcement information.

The information given in Figure 3 is posited to be the bare minimum for a reader to decide if one should participate in the advertised conference. It contains sufficient information for the reader to act upon with the accompanying information on how to act. Figure 4 shows a sample conference announcement header.

3.2 Where's The Information?

An interesting point to note is that the information contained in announcements does not follow a consistent and prescribed presentation style (format) unlike the corpora used by the MUC community. In MUC, event and object information is mainly found at the paragraph levels of the text, independent of how the information is formatted and presented in the text. In fact the presentation style is the same as if not similar to texts of a particular corpus and domain used in MUC. For example, the corpus of terrorist events has a header field containing a subject title, file index and date presented in a fixed location and format. All other information pertaining to reporting a terrorist event is embedded in the text paragraphs that follow, hence current MUC-style IE techniques focus on extracting information from these free-text paragraphs.

```
********************************************************
*                ACL-97     EACL-97                    *
*   C A L L   F O R   S T U D E N T   P A P E R S      *
*                                                      *
*                  Student Sessions                    *
*                       at the                         *
*            35th Annual Meeting of the                *
*      Association for Computational Linguistics       *
*                        and                           *
*   8th Conference of the European Chapter of the      *
*      Association for Computational Linguistics       *
*                                                      *
* Universidad Nacional de Educacion a Distancia (UNED) *
*                   Madrid, Spain                      *
*                  July 7-10, 1997                     *
********************************************************
```

Figure 4: A sample Conference Announcement Header.

Different conference hosts have different ways in highlighting/presenting information to attract participants. For example, indentation, centre justification, the use of capitalised words and character-based boxes surrounding text (see Figure 4) indicate conference header information. Bullets indicate a list of items like related conference topics. Text presented in tables detail the conference program or schedule. Text paragraphs and complete sentences form the rest of the conference body, namely conference submission and/or registration information.

For conference announcements, text layout information (of formatting) is invaluable as it provides cues to where relevant information can be found, thereby reducing the amount of text considered for processing and extraction. Information is extracted both from the head of the conference and the body.

4 A Simple IE Architecture

The focus of the IE task is the conference event and the associated submission or registration information. The nature of the conference announcement text is to inform the reader of an event and to provide details for the interested reader to act upon. A conference is an event with a title, location and date attributes; essentially an event name, where the event is held and when will it be. The information on how to submit papers to the conference or to register with a conference are sub-event processes interested participants have to note and follow. Here is a partial output of the prototype Conference Announcement IE (CAIE) system for the conference announcement header presented in Figure 4.

You've received a Conference Announcement
<u>Titled</u>:

ACL-97 EACL-97 STUDENT SESSIONS AT THE 35TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS AND 8TH CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS
It Is A:
CALL FOR PAPERS
The Conference Is On:
07th of July, 97 - 10th of July, 97 (970702-970710)
And Will Be Held In:
MADRID, SPAIN

4.1 Domain and World Knowledge

There are four main types of IE tasks that have evolved out of the MUCs. These tasks form the Component Technologies (see (Grishman, 1995) and (MUC-6:Appendix C and E, 1995)) as introduced in MUC-6. It stemmed from the need of building domain- independent component technologies, making IE systems more portable and encouraging work on "deeper and structured understanding", once text has been suitably processed.

1. Named Entity (NE) Recognition — Identify entity (person or organisation) and place names, temporal expressions, and some numerical expressions.
2. Coreference Resolution (CO) — Identify identity relations between entities.
3. Template Element (TE) Construction — Add descriptive information related to entities like organisation and person.
4. Scenario Template (ST) Production — Fits TE results into pre-specified event scenarios.

These tasks have been adapted for extracting information from conference announcements. The structures presented in Figure 3 highlight some domain-independent information such as numbers (ordinal and cardinal) and locations. These simple informational elements are constituents of larger informational elements such as dates and money, and addresses respectively. This forms the NE task for conference announcements — identifying domain-independent information and building up simple informational elements into larger and more complex informational elements.

Once informational elements like dates, money and locations have been identified; the next task is to construct TEs. For example, a conference date is distinguished from a conference paper submission date; this date from a conference paper acceptance notification date. It determines what the identified numbers (cardinal and ordinal) refer to — for example, "4 hardcopies, not more than 30 pages and 8000 words, double spaced and 11 point font". Part of this task includes the CO task — for example, identifying location name relations (the Language Technology Group is part of the Microsoft Research Institute, which is part of Macquarie University, Sydney, New South Wales, Australia).

The TE task is analogous to identifying small islands of information — domain independent and dependent — in the text. The idea is to identify islands of text that represent small informational elements, domain dependent or independent. The approach is to "look" at the left and right context of a NE, limited by a local text window, and to identify a larger sequence of text (containing the NE), the island, representing a larger informational element. For example, the island "5-10 pages" consist of two numbers "5" and "10" and it refers to the number of pages. The island itself can grow and be merged with other islands to form event scenarios, the ST task.

Finally, all information that has been identified will be used to fit into pre-specified event scenario templates — the conference event and accompanying information for registration and/or submission. This is the ST task. Note that the conference title and announcement type does not necessarily contain NEs or TEs. These information is domain dependent, hence employ another approach using domain keywords and clauses that signify the information. Domain-relevant patterns are constructed to extract such information. This is also part of the ST task.

As shown, the IE tasks rely heavily on domain knowledge and very little world knowledge. All the information contained in the corpus of conference announcements is sufficient for the specified IE tasks for extracting information from conference announcement — presented in Structures E, S and R.

4.2 A Knowledge Engineering Approach

The CAIE architecture adopts a *Knowledge Engineering* (KE) (Appelt and Israel, 1997) approach with the goal of achieving a high-recall, high-precision rate. The KE approach focuses on the information of interest and attempts to construct grammars that recognise it. It requires intensive corpus analysis to identify domain keywords and patterns from the corpus of text. Robust parsing attempts to parse small (localised) fragments of text correctly, mainly using simple grammars to recognise smaller portions of text. The parser is a finite-state transducer, which accepts a sentence as input and maps (or groups) it into corresponding grammatical constituents. The IE system is a cascade of finite-state transducers, each one using the output of the previous stage (see (Hobbs et al., 1993), (Appelt et al., 1993) and (Appelt et al., 1995)). The grammars (rules) for the knowledge-based IE system need only parse and extract relevant information.

A standard KE approach is to identify clausal patterns in which the relevant information is expressed. Domain-relevant patterns (grammars or sets of rules) are built to cover these instances of information. This approach starts with a high precision and low recall, identifying the most reliable and important information found in the text. The recall is improved when more rules are added to encompass other relevant but less common instances of information, albeit with some precision trade-off. This approach is suitable for conference title and announcement type information where the information depends a lot on domain knowledge. Domain-relevant keywords, clauses and patterns are constructed to model these types of information.

Another KE approach (the "Atomic" approach — a term referred to by Appelt & Israel in the recent ANLP-97 IE Tutorial, (Appelt and Israel, 1997)) focuses on the grammatical constituents of the text and posits a relationship among object constituents and its associated attributes and among local noun groups and verb groups, all initially independent of the IE task definitions. The general idea is to tag all objects or participants (entities) in the text and then identify local (within a small window of text) partial descriptions of possible events and relationships involving these entities. This local information is then merged to form larger and complete events and scenarios pertaining to the domain of the text. This new approach initially aims for a high recall of information but with low precision. The partial descriptions extracted are filtered using domain knowledge thereby increasing the precision of the IE system, possibly with some recall trade-off. This approach is suitable for building TEs from NEs (see Component Technologies in Section 4.1).

The "Atomic" approach was attempted as an experiment by SRI (FASTUS — an IE system built for MUC, (Appelt et al., 1995)) in MUC-6 but did not achieve the desired results as the original FASTUS approach (the standard KE approach). SRI determined that this approach is suitable for IE tasks whose,

1. Entities in the domain have easily determined types
2. The templates are structured so that there is only one of a very small number of possible slots that an entity of a given type can fill, and only entities of a given type can fill those slots

This is true of the information that is abstracted from conference announcements (Structures E, R and S) and the templates that hold the extracted results. SRI also considered a combination of both approaches, combining high-precision and high-recall system strategies, to improve on their MUC-6 results but did not complete their explorations. However, given the pre-specified information abstracted from conference announcements and the intended application of the IE system (to inform the user of a conference and provide supporting submission and/or registration information to elicit an action), we will explore and show that a combination of both KE approaches to build an intelligent conference announcement IE system can achieve both high recall and precision rate.

The goal is to build a high-recall, high-precision conference announcement IE system given a pre-specified conference information for extraction. To achieve this, a combination of the two KE approaches is employed. The former approach extracts information that rely heavily on domain knowledge (modelled by domain-relevant patterns) and the latter extracts information that is less domain dependent (more independent of the domain), as defined by the adapted Component Technologies IE tasks. The hybrid IE system starts with a high recall and low precision, identifying every entity of the right type as a candidate for participation in one of the relationships or events of interest. Using domain knowledge and application- specific criteria, irrelevant information is filtered or eliminated until the desired scenarios and events have been extracted.

4.3 An Intelligent IE System

The conference announcement IE system comprises of three simple stages. These stages were adapted and condensed (for the purpose of extracting conference announcement information) from the Generic IE system introduced by Hobbs (Hobbs, 1993) and the Vanilla Information Extraction (VIE) system, built using the GATE/CREOLE development tool developed by the University of Sheffield (Gaizauskas et al., 1997).

1. Text Pre-processing stage — NE and CO tasks
2. Event Construction stage — TE and ST task
3. Template Filling stage

The input to the conference announcement IE system is an email message containing a conference announcement. The output of one process stage forms the input of the next process stage. The function of each process stage is modular, cohesive and tightly coupled, and must be applied one after another. Each processing stage is further made up of a subset of tightly coupled modules. The final output are templates storing relevant conference information — the conference event (template based on Structure E), conference submission details (template based on Structure S) and/or conference registration procedures (template based on Structure R). Pattern-matching and simple NLP techniques have been used "intelligently" to extract relevant information from conference announcements. There is no "deep" analysis of the text as it is currently deemed unnecessary.

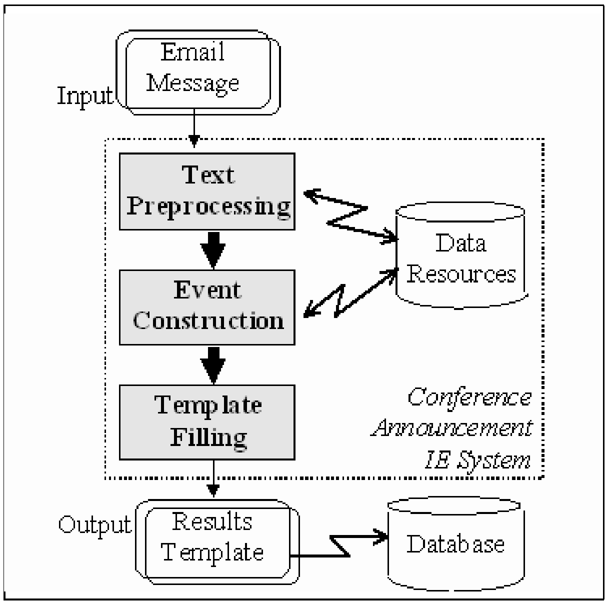

Figure 5: The three distinct Processing Stages.

The *Text Pre-processing* stage segments the text into smaller portions of text. Relevant portions of the text are identified and filtered for processing. Both the Named Entity (NE) and Coreference Resolution (CO) tasks are performed in this stage. Note that these tasks have been adapted for extracting conference announcement information. This stage uses the following data resources — lexicons, gazetteers and domain-independent simple pattern grammars. The information identified in this stage is marked up and passed on to the following stage.

The *Event Construction* stage takes the annotated conference announcement email message as input and applies domain knowledge to filter out candidates of information (identified in the previous stage) that are irrelevant and not important. The domain knowledge are encoded as domain-relevant patterns (rules or grammars) and are the product of intense conference announcement corpus analysis. The goal is to identify events (Template Entities, TE) and scenarios (Scenario Templates, ST) to facilitate filling in the output templates in the last stage. Hence this process stage identifies and marks up the conference event information, conference submission and/or registration information given the annotated input of the previous stage.

The final *Template Filling* stage receives a richly annotated conference announcement email message as input and attempts to fill the required event and scenario templates with the extracted information. There is one template for each Structure E, R and S. Each template is a list of slots, each slot refers to a particular information that has to be extracted from a conference announcement (based on the Structures E, R and S). The IE system is evaluated based on the success of filling the slots in each template.

- Equality — the extracted output is the same as the target output
- Subset — the extracted output is less than the target output (under-extraction)
- Superset — the extracted output is more than the target output (over-extraction)
- False Positive — there is extracted output when there should not be one (target output)
- Empty — no extracted even if there should be one (target output)
- Inequality — the extracted output is not the same as the target output (not any of the above)

5 Future Directions

A prototype Conference Announcement IE system is being developed:

1. To extract relevant and useful information from conference announcement email messages
2. Relying on very little "world" knowledge but dependent on domain knowledge
3. To achieve a high-recall, high precision rate successively given pre-specified IE tasks

A simple architecture combining the two Knowledge Engineering approaches is employed to build the intelligent IE system. A web-based interface to a database of conference events and associated information (extracted information stored in a text-based repository) will help realise the usefulness and user-friendliness of such an application, by providing appropriate hypertext links and indexes to the conference announcements and extracted information.

There's more than meets the eye when extracting conference titles from conference announcement email messages. Unlike other types of information (given by Structures E, R and S), conference titles rely heavily on domain knowledge but do not have well-defined domain-relevant patterns (grammars). Certain domain keywords and phrases do occur in conference titles consistently, allowing the construction of a few domain-relevant patterns for conference titles. However, this is not sufficient for the IE system to achieve both a high recall and precision rate for conference titles. More text knowledge (knowledge derived from the raw email texts, instead of only the conference announcement domain) is required to help achieve the desired goal. For example, the subject line, found in the mail header information, can contain clues about the title of the conference. Abbreviated forms and acronyms can be "expanded" to match possible strings of text (matching only the first capitalised letter of a word in a string of text) to assist in identifying conference titles. These forms and acronyms are likely to be shortened forms of the conference title — e.g. ANLPF.

Another research direction is to perform a "deep" analysis of the content of information contained in a conference announcement and show that the proposed simple architecture can still be used to extract this information — for example, related conference topics for submission of paper.

References

D. Appelt and D. Israel. 1997. Building Information Extraction Systems. *Applied Natural Language Processing Tutorial.*

D. Appelt, J. Hobbs, J. Bear, D. Israel, M. Kameyama and M. Tyson. newblock 1993. SRI: Description of the JV- FASTUS System Used for MUC-5. In *Proceedings of the Fifth Message Understanding Conference (MUC-5)*, pages 221–235.

D. Appelt, J. Hobbs, J. Bear and D. Israel. 1995. SRI International FASTUS System. MUC-6 Test Results And Analysis. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 237–248.

R. Gaizauskas, H. Cunningham, Y. Wilks, P. Rodgers and K. Humphreys. 1996. GATE: An Environment to Support Research and Development in Natural Language Engineering. In *Proceedings of the 8th IEEE International Conference on Tools with Artificial Intelligence*, Toulouse, France.

R. Grishman. 1995. Design of the MUC-6 Evaluation. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 1–11.

L. Hirschman and M Vilain. 1995. *ACL 95 Tutorial on Extracting Information from the MUC.* The Mitre Corporation, MIT Cambridge MA.

J. Hobbs. 1993. The Generic Information Extraction System. In *Proceedings of the Fifth Message Understanding Conference (MUC-5)*, pages 87–91.

J. Hobbs, D. Appelt, J. Bear, D. Israel, M. Kameyama, M. Stickel, and M. Tyson. 1993. FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural Language Text. AI Center, SRI, California, http://www.ai.sri.com/ appelt/fastus-schabes.html

P. Jacobs. 1992. Introduction: Text Power and Intelligent Systems. *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval.* LEA Publishers, New Jersey.

P. Jacobs and L. Rau. 1993. Innovations in text interpretation. *Artificial Intelligence 63*, pages 143–191.

MUC-4. 1992. Proceedings of the Fourth Message Understanding Conference (MUC-4). Sponsored by NRaD and ARPA, Morgan Kaufmann Publishers.

MUC-5. 1993. Proceedings of the Fifth Message Understanding Conference (MUC-5). Sponsored by NRaD and ARPA, Morgan Kaufmann Publishers.

MUC-6. 1995. Proceedings of the Sixth Message Understanding Conference (MUC-6). Sponsored by NRaD, DARPA and the Tipster Text Program, Morgan Kaufmann Publishers.

MUC-6. 1995. Appendix C: Named Entity Task Definition (v2.1) and Appendix E: Information Extraction Task Definition. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 317–332 and 345–359. Sponsored by NRaD, DARPA and the Tipster Text Program, Morgan Kaufmann Publishers.

A A Sample Conference Announcement

Here's a sample conference announcement — The 12th Pacific Asia Conference On Language, Information and Computation — contained in an electronic mail message. This sample conference announcement has been marked up with SGML-like[6] tags, where < **TAG** > denotes the start of a portion or text and < **/TAG** > denotes the end of the same portion of text. The portion of text delimited by **TAG** contains information pertaining to the **TAG** label.

< **EMAIL** >
< **MAIL HEADER** >
Subject: PACLIC 12 – First Call For Paper
From: Guo Jin ¡guojin@iss.nus.sg¿
To: elsnet-list@let.ruu.nl
Cc: paclic12@iss.nus.sg, luakt@iscs.nus.sg, guojin@iss.nus.sg
Date: Fri, 30 May 1997 20:39:01 +0800
< **/MAIL HEADER** >

< **CONFERENCE PROPER** >
< **CONFERENCE HEADER** >
PACLIC 12 — FIRST CALL FOR PAPER
THE 12TH PACIFIC ASIA CONFERENCE ON
LANGUAGE, INFORMATION AND COMPUTATION
< **/CONFERENCE HEADER** >

< **CONFERENCE BODY** >
The Chinese and Oriental Language Information Processing Society (COLIPS – http://www.iscs.nus.sg/ colips/) is pleased to announce that the 12th Pacific Asia Conference on Language, Information and Computation (PACLIC 12) will be held in Singapore on February 18-20, 1998.

The Conference is an annual meeting of scholars with a wide range of interest in theoretical and computational linguistics. The Conference solicits papers treating any field in theoretical and computational linguistics, including syntax, morphology, phonology, semantics, pragmatics, discourse analysis, corpus linguistics, formal grammar theory, natural language processing, and computer applications. We plan to give each paper 30 minutes for presentation and discussion.

< **CONFERENCE SUBMISSION INFO** >
Full paper submission is required. Please mail 4 hard copies of the paper with the title, the author's name, affiliation, mailing address, FAX number (if any) and e-mail address on a separate page to the address shown below. Submission by e-mail is encouraged (but no FAX submissions). Accepted papers will be published in the conference proceedings.

IMPORTANT DATES

Deadline for paper submission: October 1, 1997
Notification of acceptance: November, 15, 1997
Submission of camera-ready due: January 1, 1998
Conference: February 18-20, 1998

CONTACT

Please email your paper submission to:
paclic12@iss.nus.sg

We only accept (1) Microsoft Word Rich Text Format (.rtf); (2) Plain text (.txt); and (3) Postscript (.ps)). Otherwise, please send three hard copies to:

PACLIC 12
Institute of Systems Science
National University of Singapore
Heng Mui Keng Terrace, Kent Ridge
Singapore 119597

For up-to-date information on the conference, please check our web page http://www.iscs.nus.sg/ colips/paclic92.html or write to the addresses above.
< **/CONFERENCE SUBMISSION INFO** >

CONFERENCE CHAIRS
Kim Teng Lua, Chairman
National University of Singapore, and
President of the Chinese and Oriental Language Information
Processing Society (COLIPS)

Benjamin K. T'sou, Co-chairman
City University of Hong Kong
Chu-Ren Huang, Co-chairman
Taiwan Academia Sinica

[6]Standard Generalised Markup Language (SGML) and Hypertext Markup Language (HTML) are two common forms of markup used today; the former is used to store large amounts of documents in an efficient manner and the latter used for building web pages for the World Wide Web.

Young-Hern Lee Co-chairman
Korea Chosun University, and
President of the Korean Society for Language and Information (KSLI)

Akira Ikeya, Co-Chairman
Tokyo Gakuen University

PROGRAM COMMITTEE

Jin Guo, National University of Singapore (Chairman)
Akira Ishikawa, Sophia University
Baosheng Yuan, National University of Singapore
Byung-Soo, Park Kyung Hee University Korea
Changning Huang, Tsinghua University
Chris Tancredi, The Yokohama National University
Chungmin Lee, Seoul National University
Haizhou Li, National University of Singapore
Hongyin Tao, National University of Singapore
Ik-Hwan Lee, Yonsei University
Jerry Seligman, National Chung Cheng University
Jie Xu, National University of Singapore
Kathleen Ahrens National Taiwan University
Jian Su National University of Singapore
Keh-Jiann Chen, Academia Sinica
Kevin Knight, University of Southern California
Kiyoshi Ishikawa, Hosei University
Ling Cao, National University of Singapore
Maosong Sun, Tsinghua University
Masato Ishizaki, ATR
Martha Palmer, University of Pennsylvania
Paul Horng Jyh Wu, National University of Singapore
Richard Sproat, AT&T Bell Labs
Shiwen Yu, Beijing University
Shuichi Yatabe, University of Tokyo
Taiyi Huang, Institute of Automation, Academia Sinica
Von-Wun Soo, National Tsing Hua University

RELATED LINKS
PACLIC11 – know more about the last conference.
COLIPS – the organizer of PACLIC12.
< /CONFERENCE BODY >
< /CONFERENCE PROPER >
< /EMAIL >