

Computational Linguistics in the UK (CLUK '06) March 2006

Final Report for EPSRC Grant EP/D078679/1

August 2006

Summary

The Computational Linguistics in the UK (CLUK) conference in 2006 continued the series of successful annual workshops that maintain and stimulate research in Computational Linguistics (CL) and Natural Language Processing (NLP) in the United Kingdom. The EPSRC grant allowed the organisers to invite a prominent UK academic as keynote speaker, which contributed to the discussion. The grant also allowed the organisers to make and archive a webcast of the proceedings. The live webcast was accessed while the workshop was taking place, and the archived webcast has been frequently accessed since.

The stated objectives of the workshop were met:

1. *to enable UK PhD students working NLP/CL to develop presentation and research skills.* 9 presentations were made from PhD students from 8 UK institutions. As part of the workshop, the Open University provided a doctoral training session to discuss the nature and use of linguistic data in NLP research, and to expose the students to alternative methodologies in the field. As is traditional, the 2006 CLUK workshop was organised principally by PhD students at the host institution, with support as required from the academic staff. Paper reviewing was carried out by staff and students at various academic institutions across the UK. The participating students have therefore experienced all aspects of workshop involvement.
2. *to provide students and researchers with a perspective on current NLP/CL research in the UK.* The workshop attracted submissions and visiting academics from eleven institutions in the UK, with papers presented from eight of those institutions. The topics covered a broad range of problems and approaches, including traditional symbolic approaches to Natural Language Understanding, statistical learning, Natural Language Generation and Question Answering.
3. *to give PhD students the opportunity to establish contacts with other students and academics in the NLP/CL community.* The discussions in the presentation sessions and during the breaks provided opportunities for presenting students to receive feedback from academics working in different areas of NLP/CL.

An additional result of broadcasting the workshop has been a proposed collaboration to investigate the possible uses of the stored archive as a multimedia corpus.

Background

Computational Linguistics UK (CLUK) provides a forum for Computational Linguistics

and Natural Language Processing researchers in the UK. In order to maintain long term excellence in NLP/CL research in the UK, CLUK holds an annual workshop to provide PhD students working in these disciplines with an opportunity to present their work in a typical conference environment at an early stage in their academic careers. As such, the workshop provides essential research training for UK PhD students by providing early feedback from established academics, and ensures that senior researchers can see first hand the directions in which NLP/CL research is moving. Computational linguistics is an import area of international computing research, with the major conferences, ACL and COLING attracting very high quality submissions from international participants. Students who have presented work at past CLUK workshops have often gone on to present at the main international workshops.

This year's workshop revealed several different strands of research within the UK. Particular work presented covered machine learning of grammars, classification of documents by similarity and sentiment, Question Answering and Natural Language Generation. As several institutions were working in each of these areas, the workshop provided the environment for the students and researchers to interact with others working on common problems.

The workshop was held on 8th and 9th March 2006 at the Open University in Milton Keynes. The senior programme committee (see appendix) are of international standing, and the student reviewing panel was drawn from 8 academic institutions across the UK. The reviewing panel contained a roughly equal split between academic staff and students. The workshop was attended by 33 participants altogether, around two thirds of whom were students. An additional 5 attended each day via the webcast.

Key advances and supporting methodology

Webcast

In order to make this year's workshop as accessible as possible to the NLP/CL community, the speaker presentations were made fully available via a webcast, funded via the EPSRC award. The webcast was subsequently archived and has been available since at the CLUK website. The archive of presentations appears to have been a significant success.

The number of hits registered during the workshop itself were:

Number of Distinct IPs for Live Hits = 14
(Open University = 4 | Internet = 10)
Live Event Hits: 17

That is, the webcast registered 17 hits in total from 10 distinct IP addresses from outside the Open University over the course of the workshop.

The number of hits registered by the site at the archive (as at 1st August 2006, approximately five months after the event itself) were:

Total Number of Distinct IPs = 161
(Open University = 26 | Internet = 135)
Post-Event Hits: 674

For the post-event hits, the site has been accessed by 161 distinct IPs, of which 135 were external to the Open University.

We conclude that the funding received from EPSRC has been instrumental in providing researchers in Computational Linguistics with access to the UK workshop. The high access rate also suggests that the proposed objective of the webcast, that students should be able to review their own performance at later dates, has been achieved.

It has also been proposed that the webcast itself could form the basis of future research (see "Further research" section).

Keynote Speakers

The EPSRC grant allowed us to invite the keynote speaker Professor Jon Oberlander from the School of Informatics University of Edinburgh, without an excessive increase in the registration costs.

Professor Oberlander's talk was entitled "The computational linguistics of affect - a personal view." The grant enabled the organisers to invite a prominent member of the community to speak, from one of the main centres of NLP research in the UK. The talk itself also stimulated discussion on possible future directions of NLP, as it represented an individual's view on a new area of NLP research.

The introductory talk was given by the keynote speaker, Professor Anne De Roeck, of the Computing Department at the Open University, whose talk was entitled "Dataset Profiles - investigating the role of data in experimental Natural Language Processing."

Doctoral Event

To support the doctoral training, this year's CLUK also held a half day doctoral event, which was led by Dr. Marian Petre of the Open University. 8 students attended, who were primarily those who were visiting to present at the main conference. The activities involved identifying the use of evidence in NLP projects, and skills for establishing collaborative research networks.

The feedback provided by the students was positive and has been reproduced in the appendix. The session was also favourably commented on by some academic staff, with the suggestion that a similar workshop be used in future CLUK meetings.

Project plan review

By tradition, CLUK is organised and executed primarily by research students with advice from research active academic staff. As described in the original proposal, the planning and execution of the workshop was carried out by Gaston Burek and Dileep Damle, two

PhD students in the Knowledge Media Institute, with support from Dr. Alistair Willis and Dr. Paul Piwek, both lecturers in the department of Computing and members of the NLP group. Proceedings were compiled at the Open University and distributed to participants. Refereeing of papers was carried out by students working in NLP/CL, and senior members of CLUK. Each paper was reviewed by one senior member and two students. The programme committee is given in the appendix.

Research impact and benefit to society

The impact of the workshop will be measured by the success of future publications by those participants who were able to present at CLUK. Not enough time has passed since the workshop took place to be able to determine how much impact the conference has had, and whether the participants found that their experience has helped with their research training. The impact will become apparent once the participants have been able to submit to the major international Natural Language Processing conferences.

It is certainly the case that the funding provided by EPSRC resulted in reinforcement of cross-UK collaboration by being able to fund a speaker from Edinburgh. This ensured that one of the major centres for NLP was represented at the meeting, and so ensured that UK-wide engagement was achieved.

Explanation of expenditure

The funds provided were used as described in the original proposal, for the invited speaker expenses and the staff costs for making the webcast. The grant provided the travel costs for Professor Oberlander to travel to the workshop from Edinburgh, his accommodation in Milton Keynes and his other travelling expenses. The staff costs were for a technician to be available for one day in preparation for the workshop, the two days of the workshop itself, and a further two days afterwards for preparation and archiving of the recorded material. Under full economic costing, the award covered 80% of the costs of the speaker and 80% of the technician's cost.

No further contribution was requested from EPSRC; the remaining costs were covered by the University and registration fee. In accordance with the original intention, the EPSRC grant allowed the organisers to maintain a registration fee at a more acceptable rate for students than might otherwise have been possible.

Further research and dissemination activities

As discussed in a previous section, most of the further research will be carried out by those students who were able to attend and/or present at the workshop.

An additional and unanticipated success of the webcast has been a proposal that the (now archived) webcast could be used as a multimodal corpus. The webcast constitutes a collection of papers, with associated verbal presentations and slides. No such publicly available corpus currently exists. Eric Atwell of the University of Leeds was unable to attend CLUK in person and viewed the proceedings via the webcast. He has since

proposed a collaborative activity to exploit this resource. We are currently in the early stages of investigating this possibility.

As outlined in the original proposal, the webcast has been archived until at least 7th March 2007, and can be viewed via the CLUK website at:

<http://kmi.open.ac.uk/events/cluk06/>

where an electronic version of the proceedings is also available.

CLUK 2006 formed part of the ongoing workshop series; the committee meeting held on the first day of the workshop decided that CLUK 2007 will be held at the University of Cambridge.

Appendix

Programme committee and reviewers

Senior members

Ted Briscoe, Cambridge
John Carroll, Sussex
Alison Cawsey, Heriot Watt
Stephen Clark, Oxford
Kees van Deemter, Aberdeen
Anne De Roeck, The Open University
Roger Evans, Brighton
Rob Gaizauskas, Sheffield
Jonathan Ginzburg, Queen Mary London
Rodger Kibble, Goldsmith College
Alex Lascarides, Edinburgh
Mark Lee, Birmingham
Diana Maynard, Sheffield
Chris Mellish, Aberdeen
Ruslan Mitkov, Wolverhampton
Johanna Moore, Edinburgh
Paul Piwek, The Open University
Donia Scott, The Open University
John Tait, Sunderland
Marilyn Walker, Sheffield
Yorick Wilks, Sheffield
Alistair Willis, The Open University

Student reviewing panel

Gaston Burek, the Open University
Francis Chantree, the Open University
Dileep Damle, the Open University
Simon Dobnik, Oxford
Eva Esteve Ferrer, Sussex
Caroline Gasperin, Cambridge
Debra Haley, the Open University
Shahzad Khan, Cambridge
Georgiana Puscasu, Wolverhampton
David Reitter, Edinburgh
Avik Sarkar, the Open University
Shenghui Wang, Manchester
Jon Herring, Brighton

Feedback from students at doctoral training session

Questions

I found the event useful for networking with other researchers in the field
 My knowledge of the field was broadened beyond my immediate research topic
 My knowledge of key UK researchers in the field increased
 I gained new insights at the colloquium which will be useful in my research

Responses	Average
5 5 5 5 5 4 5 5 4 4 5	4.7
3 5 5 4 4 2 4 4 3 4 5	3.9
4 5 3 5 2 5 4 4 4 3 4	3.9
4 5 4 4 4 5 4 3 4 3 4	4.0

If you presented a paper:

The colloquium helped me gain paper writing skills
 The colloquium gave me useful experience of submitting a paper to a conference/workshop
 I gained useful experience of academic presentations

5 4 4 2 4 5	4.0
5 5 5 3 5 3	4.3
5 5 5 4 5 4	4.7
4 5 5 5 5 3	4.5

If you were a reviewer:

I gained useful experience of reviewing

4 5	4.5
-----	------------