Report of NAACL 2010 Workshop Committee

Richard Sproat and David Traum

We were selected as co-chairs of NAACL workshops on June 24th 2009, and formed into a committee with the workshop chairs for ACL and COLING into a joint committee to select workshops for all three conferences. This committee had to come up with a call for proposals, proposal management process, review procedure, and final decisions.

There was some initial confusion about large SIG-sponsored workshops - whether these should go through the normal review process, or were pre-approved, or needed to be decided on early by the committee but differently than the normal process. The conclusion seemed that this committee should make decisions, though they should be made early. This should be clarified for next year (no large SIG workshops were proposed for NAACL HLT this year).

The committee worked well for generating the call for papers. Following a recommendation of the previous committee, START was used for submission and management of workshop proposals, which went smoothly. With 55 total proposals, the START system was very helpful. The decision procedure we decided on was to have the chairs for each conference review those workshops that selected their conference first, and then pick up other choices if there was more space available. This process was non-optimal, in that it wasn't completely clear exactly how many workshops could be selected for each venue, and thus it was difficult to decide on which workshops were in fact available to second and third choice venues. With this large number of proposals, the overall yield was rather low compared to many recent years, and thus many quality workshop proposals had to be rejected. The committee was not broad enough with enough specialists in all areas to be confident that the proposals that were accepted were better than those that had not.

The results of the joint workshop selection were mixed. On the one hand there was a positive that some workshops were accepted to conferences other than their first choice (including some appearing at NAACL), which would have been more difficult with separate review processes. On the other hand, the committee's workload was much higher than we felt it should have been, due to the need for coordination, and there was much more confusion at several stages in the process. The committee was divided about whether this process was a net positive. Some felt that the old way where each conference is run separately is preferable (perhaps the dates could be staggered to allow resubmission to second choices). Others felt that there was some merit in the joint selection, but an earlier start is needed and better ways of efficiently making the decisions based on good information (with perhaps a broader review committee) is needed.

In total 16 workshops were selected to be colocated with NAACL HLT 2010. For most of these the selection was straightforward, based on the materials submitted by the workshop organizers. In one case however -- Workshop 16, the *Young Investigators in the Americas Workshop* -- we required a couple of iterations with the organizers to get a proposal that was acceptable as a workshop. The original proposal was a little too much like a proposal just to get a bunch of young researchers from Latin America to come to NAACL-HLT, with no specific topic for the submitted papers. We convinced the organizers that they should have a theme, and it was agreed that the topic would be any work on languages spoken in the Americas south of the Rio Grande/Bravo.

For each of the workshops below we list:

1. **Title (with link to website)**

2. Names of organizers
3. Original size estimate (numbers of participants)
4. Number of submissions
5. Number of papers selected for presentation
6. Total number of registrants per the close of online registration

1) Active Learning for NLP (ALNLP)
Organizers: Burr Settles, Kevin Small, and Katrin Tomanek
Size estimate: 35-40
10 submissions
Number of Papers: 6 (including invited talk)
Number of registrants: 20

2) Workshop on Computational Approaches to Analysis and Generation of Emotion in Text
Organizers: Diana Inkpen and Carlo Strapparava
Size estimate: 25
24 submissions
Number of Papers: 19 (including invited talk)
Number of registrants: 31

3) Computational Approaches to Linguistic Creativity (CALC-10)
Organizers: Paul Cook and Anna Feldman
Size estimate: 30-40
9 submissions
Number of Papers: 7 (including invited talk)
Number of registrants: 14

4) Computational Linguistics and Writing: Writing Processes and Authoring Aids (CL&W 2010)
Organizers: Michael Piotrowski, Cerstin Mahlow, and Robert Dale
Size estimate: 40
15 submissions
Number of Papers: 9
Number of registrants: 15

5) Computational Linguistics in a World of Social Media: #SocialMedia
Organizers: Ben Hachey and Miles Osborne
Size estimate: 70
21 submissions
Number of Papers: 18 (including 2 invited talks)
Number of registrants: 34

6) Computational Neurolinguistics
Organizers: Brian Murphy, Kai-Min Chang, and Anna Korhonen
Size estimate: 30-40
10 submissions
Number of Papers: 10 (including invited talk)
Number of registrants: 11

7) Creating Speech and Text Language Data With Amazon's Mechanical Turk
Organizers: Chris Callison-Burch and Mark Dredze

Size estimate: 30

44 submissions

Number of Papers: 35 (including invited talk)

Number of registrants: 60

8) Extracting and Using Constructions in Computational Linguistics
Organizers: Magnus Sahlgren and Ola Knutsson

Size estimate: 30-50

8 submissions

Number of Papers: 6

Number of registrants: 12

9) Formalisms and Methodology for Learning by Reading (FAM-LbR)
Organizers: Rutu Mulkar-Mehta, James Allen, Jerry Hobbs, Eduard Hovy, Bernardo Magnini, and Chris Manning

Size estimate: 50-70

16 submissions

Number of Papers: 15

Number of registrants: 44

10) The 5th Workshop on Innovative Use of NLP for Building Educational Applications
Organizers: Joel Tetreault, Jill Burstein, and Claudia Leacock

Size estimate: 40-50

28 submissions

Number of Papers: 14

Number of registrants: 32

11) Second Louhi Workshop on Text and Data Mining of Health Documents (Louhi '10)
Organizers: Hercules Dalianis, Martin Hassel, and Sumithra Velupillai

Size estimate: 40

16 submissions

Number of Papers:

Number of registrants: 25

12) Semantic Search (SemanticSearch 2010)
Organizers: Donghui Feng, Jamie Callan, Eduard Hovy, and Marius Paşca

Size estimate: 30-40

11 submissions

Number of Papers: 15 (including invited talk)

Number of registrants: 29

13) Speech and Language Processing for Assistive Technologies
Organizers: Melanie Fried-Oken, Kathleen F. McCoy, and Brian Roark

Size estimate: 30-50

16 submissions

Number of Papers: 14

Number of registrants: 23

14) Statistical Parsing of Morphologically Rich Languages (SPMRL 2010)

Organizers: Djamé Seddah, Sandra Kübler, and Reut Tsarfarty

Size estimate: 40-60

13 submissions

Number of Papers: 13 (including preface by organizers, and invited talk)

Number of registrants: 43

15) 6th Web as Corpus Workshop (WAC-6)

Organizers: Adam Kilgarriff, Dekang Lin, and Serge Sharoff

Size estimate: 40

6 submissions

Number of Papers: 6 (including invited talk)

Number of registrants: 28

16) Young Investigators in the Americas Workshop

Organizers: Ted Pedersen and Thamar Solorio

Size estimate: 16

21 submissions

Number of Papers: 8

Number of registrants: 17

Just out of interest we computed the correlation between the predicted sizes and the final online registration numbers: when people gave an estimate range, we used the mean of the range. The R for the correlation is only 0.29, suggesting that one can't make much use of these kinds of estimates.

Workshops were divided into days based on requested preference from workshop organizers for individual days and to avoid overlapping other workshops, as well as fitting to available room sizes. Due to low numbers of acceptable papers, 4 workshops dropped down to 1/2 day, giving a total of 14 workshop days.

The scheduling was a little bit challenging, having to negotiate with the publications chairs midway through the process. It would be best to nail down the date by which workshop organizers must send workshop proceedings to the publications chairs in advance of the call for workshop proposals, but at least by the time of acceptance. From the point of view of the workshops, this should be as late as possible that still allows the publications chairs sufficient time to review the draft proceedings. Another issue that came up late was the possible size for the workshop papers and proceedings. In previous years, with hard-copy publications, there was an overall page limit. This year, there was no overall page limit, but it was decided that no workshop paper should be longer than the main conference papers (8 pages plus one for references).

This was decided somewhat late in the process and led to some required revisions at the last minute. Our recommendation for the future is to give workshop chairs autonomy in this regard, but if there is some limit it should be decided before the call for workshop proposals or at very least at the time of acceptance.