

Welcome!

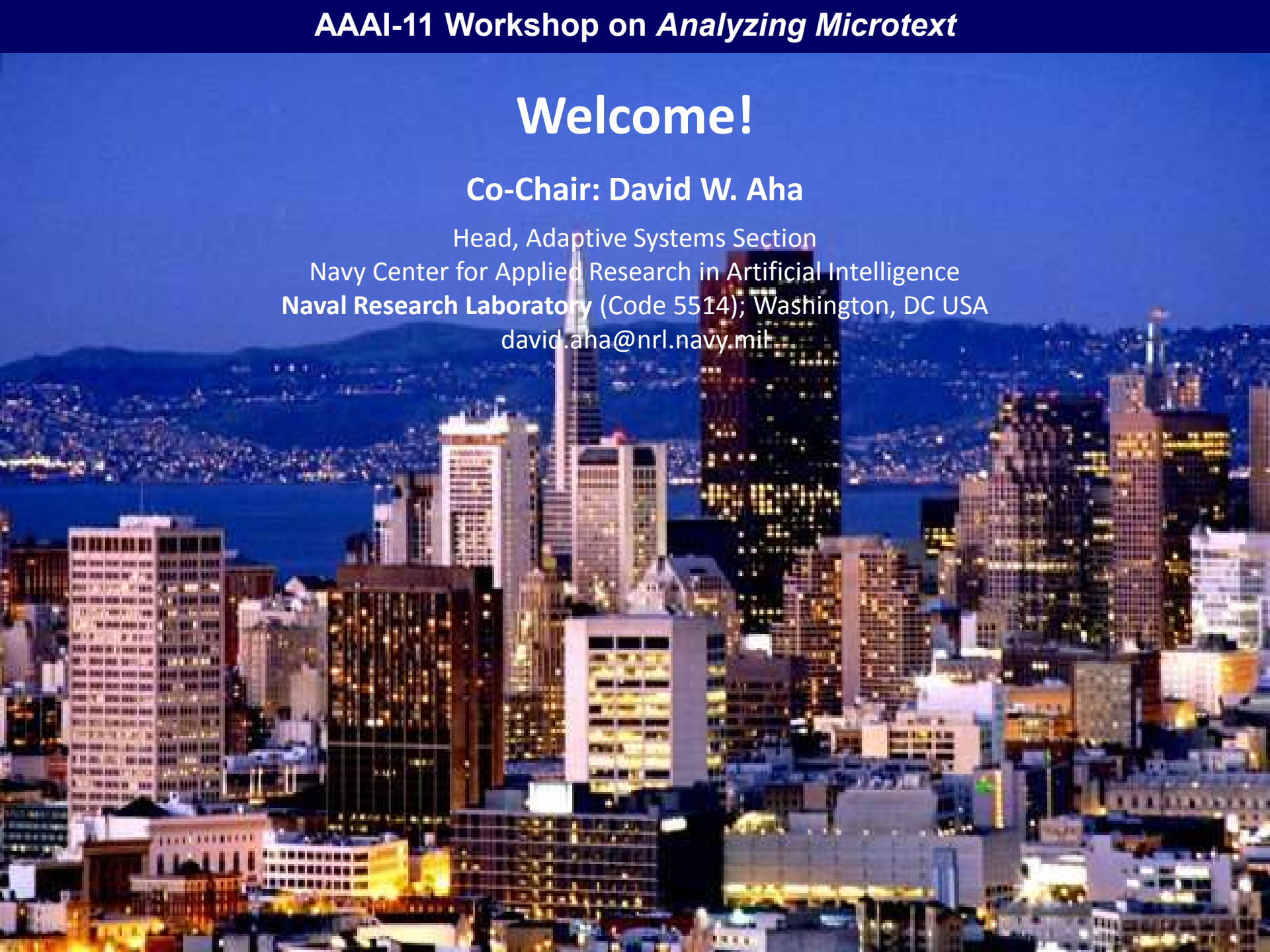
Co-Chair: David W. Aha

Head, Adaptive Systems Section

Navy Center for Applied Research in Artificial Intelligence

Naval Research Laboratory (Code 5514); Washington, DC USA

david.aha@nrl.navy.mil



Why?

NRL project: Chat Attention Management for Enhanced Situation Awareness

- **Situation:** Substantial Navy chat use
- **Problem:** Chat can be distracting!
- **Research Foci**
 1. Chat urgency detection
 2. Chat room summarization
 3. 3-D audio cueing



Our focus: Chat analysis (Uthus & Aha, in submission)

- Not (yet?) popular
- An example of *microtext* (Ellen, ICAART-11)
- Twitter analysis is popular
- We generalized 😊

Community

- Meetings exist re: social media, military chat, etc.
 - e.g., ACL/HLT-11 Workshop: *Language in Social Media*
 - e.g., *Joint Chat Conference*
- Nothing specifically focused on *microtext*
- Our goal here: Bring the communities/interests together
 - **Provide a research forum for cross-fertilization of ideas pertaining to analysis of microtext**

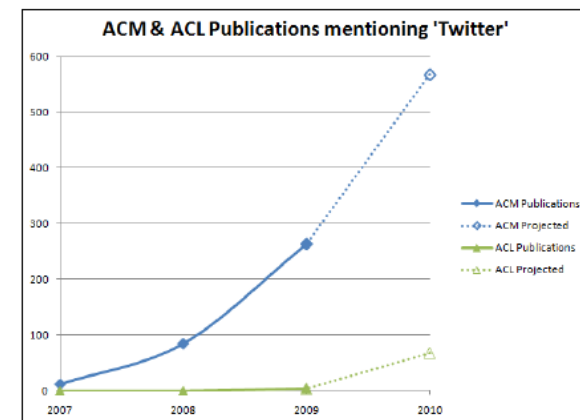


Figure 1: Twitter's growth in the academic community.

Who? (partial list; 40+ expected)

1. Aha, David (Naval Research Laboratory)
2. Bacon, Diana (SHAI)
3. **Carpenter, Tamitha** (SHAI)
4. Chakrabarti, Chayan
5. Cotner, Carl
6. Davison, Brian (Lehigh U.)
7. De la Rosa, Josep Luis (U. Girona)
8. Dent, Kyle (PARC)
9. Duchon, Andrew (Aptima)
10. Duffy, LorRaine (SPAWAR)
11. **Ellen, Jeffrey** (SPAWAR)
12. **Elsner, Micha** (U. Edinburgh)
13. Fleischman, Michael (Bluefin Labs)
14. Foster, Jennifer (Dublin City U.)
15. Quinsulon, Israel
16. Joshi, Ashutosh (Firstrain, Inc.)
17. Liu, Fei
18. Macskassy, Sofus (Fetch Technologies)
19. **Markman, Vita** (Disney IMG)
20. Mustafaraj, Eni (Wellesley College)
21. Oard, Douglas (U. Maryland, College Park)
22. Ortiz, Pedro (USNA)
23. Pottenger, William (Intuidex, Inc.)
24. Ramachandran, Sowmya (SHAI)
25. Sajnani, Hitesh (UCI)
26. Sanders, Geoffery (Lawrence Livermore National Laboratory)
27. Schilder, Frank (Thomson Reuters)
28. Tang, Yi-Jie (National Taiwan U.)
29. Thompson, Cynthia (Deloitte LLP)
30. Uthus, David (NRC/NRL)
31. Weil, Shawn (Aptima)
32. **Young, Joel** (USNA)

Legend

- Underline: Workshop Co-Chair
- **Green**: Invited speaker

What?

1. Introduction: The Broad Coverage of Microtext (Aha)

0830-0845: Workshop Focus, Goals, & Overview (**David W. Aha**, NRL)

0845-0905: All About Microtext: Philosophy and a Survey of Current Results (**Jeffrey Ellen**, SPAWAR San Diego)

0910-0925: Normalizing Microtext (Zhenzhen Xue, Dawei Yin, & **Brian D. Davison**, Lehigh U.)

0930-0945: Analysis of C2 and “C2-Lite” Micro-message Communications (**Andrew Duchon**, Robert McCormack, Brian Riordan, Charlotte Shabarekh, Shawn Weil, & Ian Yohai, Aptima Inc.)

0950-1005: Multi-Label Classification of Short Text: A Study on Wikipedia Barnstars (**Hitesh Sajnani**, Sara Javanmardi (UCI), David McDonald (U. Washington), & Cristina V. Lopes (UCI))

1010-1025: A Microtext Corpus for Persuasion Detection in Dialog (**Joel Young**, Craig Martell (NPS), Parnav Anand (UCSC), Pedro Ortiz (USNA), & Henry Tucker Gilbert (NPS))

1030-1100: Break

2. Chat (Oard)

1100-1120: Chatrooms, IM, Tweets, and Iconic Language: An Analysis of Navy Echelon Differences and the Role of the Written Word in Warfare (**LorRaine Duffy**, SPAWAR San Diego)

1125-1150: Shortcuts and “Cheap” Tricks for Analyzing Microtext: Bootstrapping from What We Already Know (**Craig Martell**, Naval Postgraduate School)

1155-1220: Dialogue Structure in Microtext (**Micha Elsner**, U. Edinburgh)

1225-1330: Lunch (Organized by Panel Topic)

3. Keynote talk and Poster Session (Uthus)

1330-1410: The Flow of Real-Time Information (**Abdur Chowdhury**, Twitter)

1420-1530: Posters (1-minute, 1-slide introductions at the start of this section)

- The Role and Identification of Dialog Acts in Online Chat (**Tamitha Carpenter** & Emi Fujioka, Stottler Henke Associates, Inc.)
- Learning Ontology from the Web for Microtext Processing (**Boris A. Galitsky**, Gábor Dobrocsi, & Josep Lluís de la Rosa, U. Girona)
- Statistical Relational Learning Technologies for Microtext Applications (Christopher A. Kelly, Christopher D. Janneck, & William M. Pottenger, Intuindex Inc.)
- Unsupervised Discovery of Fine-Grained Topic Clusters in Twitter Posts (**Vita Markman**, Disney IMG)
- What Edited Retweets Reveal about Online Political Discourse (**Eni Mustafaraj** & Panagiotis Takis Metaxas, Wellesly College)
- Domain Adaptation in Sentiment Analysis of Twitter (**Viswa Mani Kiran Peddinti** & Prakriti Chintalapoodi, U. Southern California)
- Untangling Topic Threads in Chat-Based Communication: A Case Study (**Sowmya Ramachandran**, Randy Jensen, Oscar Bascara, Tamitha Carpenter (SHAI), Todd Denning (AFRL/RHA), & Shaun Succillon (AFRL))
- Modeling Socio-Cultural Phenomena in Online Multi-Party Discourse (**Tomek Strzalkowski** (SUNY Albany & Polish Academy of Sciences), George Aaron Broadwell, Jennifer Stromer-Galley, Samira Shaikh, Ting Liu (SUNY Albany), & Sarah Taylor (Lockheed Martin))
- A Comparison between Microblog Corpus and Balanced Corpus from Linguistic and Sentimental Perspectives (**Yi-jie Tang**, Chang-Ye Li, & Hsin-Hsi Chen, National Taiwan U.)
- What Are Tweeters Doing: Recognizing Speech Acts in Twitter (Renxian Zhang, Dehong Gao, & Wenjie Li, Hong Kong Polytechnic U.)

1530-1600: Break

4. Microblogs: Twitter (Ramachandran)

1600-1615: Through the Twitter Glass (**Kyle Dent** & Sharoda Paul, Palo Alto Research Center)

1620-1635: #hardtoparse: POS Tagging and Parsing the Twittersverse (**Jennifer Foster**, Ozlem Cetinoglu, Joachim Wagner, Joseph Le Roux, & Stephen Hogan, Dublin City U.)

1640-1745: **Panel Discussion**

- **Panelists:** **Douglas Oard** (U. Maryland), **Vita Markman** (Disney IMG), **David Uthus** (NRC/NRL), and **Sofus Macskassy** (Fetch Technologies and U. Southern California)

Wrapup (Aha)

1745-1800: Lessons Learned, Related Work, Open Problems, & Future Plans (**David W. Aha**, NRL)

Panel

Topics

- Douglas Oard (U. Maryland): **Information Retrieval**
- Vita Markman (Disney IMG): **Social Media Challenges in NLP**
- David Uthus (NRC/NRL): **Summarization**
- Sofus Macskassy (Fetch Technologies & U. Southern California): **Statistical Relational Learning**

Structure

1. Lunch groups:
 - Each panelist will lead a group to some nearby restaurant, and will discuss their topic
 - **Suggested restaurants (tentative):**
 - **The Plant: Café Organic** (reservations for up to 10)
 - **The Market Bar** (reservations for up to 10) // Douglas Oard
 - OSHA (Thai)
 - Pho Hoa Noodle Soup (Vietnam)
 - ***Please join your selected group leader!***
2. Panel discussion (1640-1745)
 - 10min (max) presentations
 - 25min discussion
3. Dinner (all invited) after the Opening Reception

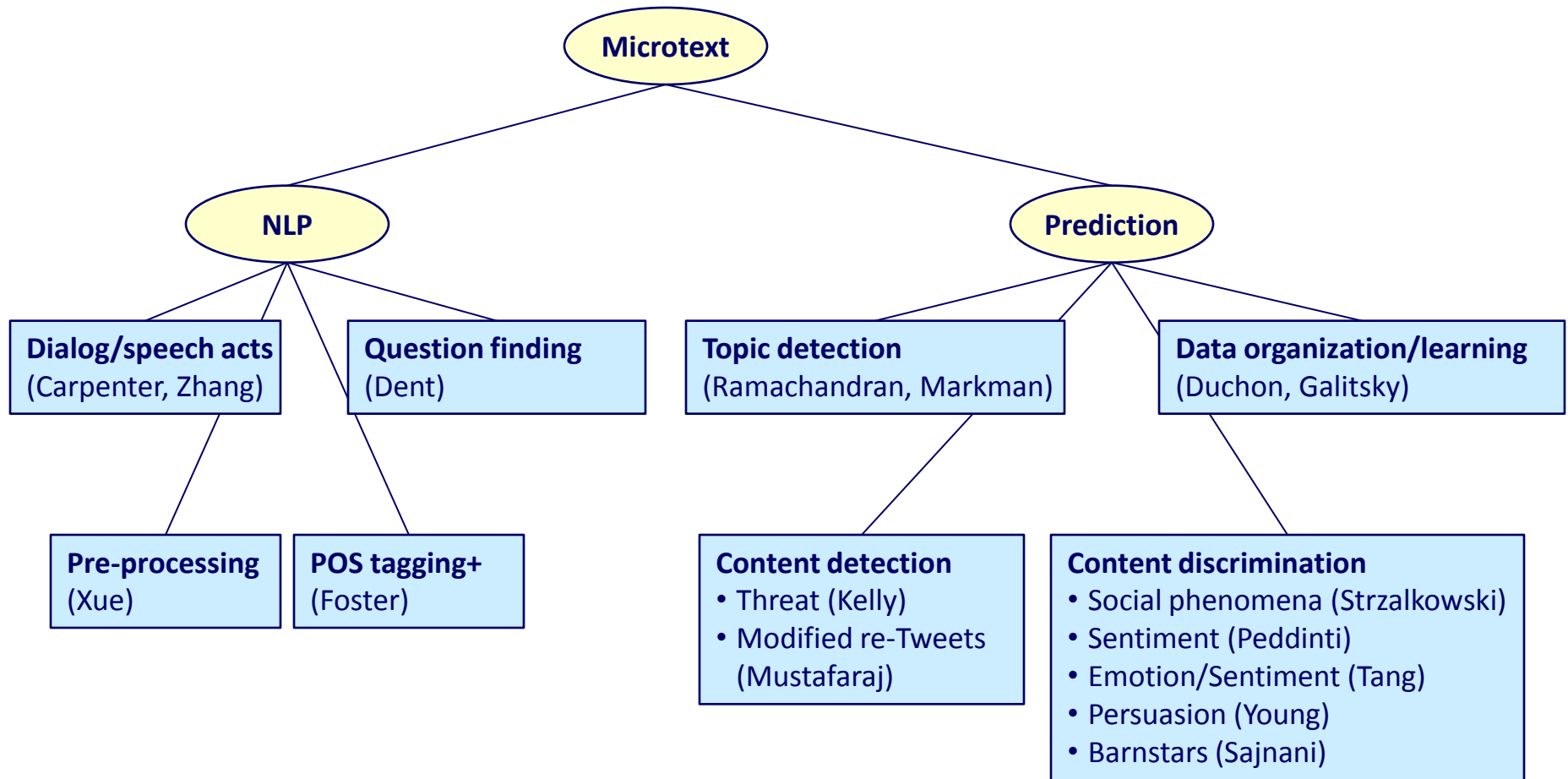
Summary of Presentations: Details

1 st Author	Domain	Task	Contributions
Carpenter	Chat	Identifying dialog acts	Distinction of chat from voice conversations; methods; Chat-IE prototype
Dent	Twitter	Finding Qns in Tweets	NLP tools for Twitter analysis; Tweeting Turker study (supportive)
Duchon	Micro-messages	Data organization	Cross-domain analyses for types of communications; initial data analysis
Foster	Twitter	POS tagging & d.-parsing	Evaluation of WSJ-trained Malt; value of retraining it on dependency trees
Galitsky	Search results, captions	Ontology construction & syntactic generalization	Evaluation of search relevance improvement & text similarity improvement
Kelly	Chat	Threat identification	Data extraction & evaluation showing Higher-Order Naïve Bayes performs well
Markman	Twitter	Topic discovery w/ LDA	Data preprocessing and padding shown to be critical (re: noise, sparsity)
Mustafaraj	Twitter	Discovering re-tweets w/ comments or edits	Argues for the importance of analyzing <i>individual</i> tweets for political discourse; outline of a supervised learner for comment labeling
Peddinti	Twitter	DA for sentiment analysis	Techniques for selecting out-of-domain data; algorithms; domain sim. metrics
Ramachandran	Chat	Topic identification	Comparison of unsupervised with supervised approaches on T-REX data
Sajnani	Wikipedia Barnstars	Classifying Barnstars	Evaluation of multi-class supervised algorithms; utility of contextual features
Strzalkowski	Chat	Social phenomena detection	Methods for detection; evidence that mid-level social phenomena can be reliably detected and can differentiate participants
Tang	Plurk and Sinica	Emotion prediction	Different linguistic and sentiment analysis methods are needed
Xue	Twitter and SMS data	Normalization for NLP	Multi-channel (source theory) model for normalization; argues for using orthographic; phonetic; and contextual info; and acronym expansion
Young	Hostage transcripts	Persuasion detection	Introduction of hostage negotiation corpus; comparison of supervised learners for persuasion detection
Zhang	Twitter	Speech act recognition	Use of word- and character-based features; empirical study

Summary of Presentations: Domain

1 st Author	Domain	Task	Contributions
Carpenter	Chat	Identifying dialog acts	Distinction of chat from voice conversations; methods; Chat-IE prototype
Kelly		Threat identification	Data extraction & evaluation showing Higher-Order Naïve Bayes performs well
Ramachandran		Topic identification	Comparison of unsupervised with supervised approaches on T-REX data
Strzalkowski		Social phenomena detection	Methods for detection; evidence that mid-level social phenomena can be reliably detected and can differentiate participants
Dent	Microblog	Finding Qns in Tweets	NLP tools for Twitter analysis; Tweeting Turker study (supportive)
Foster		POS tagging & d.-parsing	Evaluation of WSJ-trained Malt; value of retraining it on dependency trees
Markman		Topic discovery w/ LDA	Data preprocessing and padding shown to be critical (re: noise, sparsity)
Mustafaraj		Discovering re-tweets w/ comments or edits	Argues for the importance of analyzing <i>individual</i> tweets for political discourse; outline of a supervised learner for comment labeling
Peddinti		DA for sentiment analysis	Techniques for selecting out-of-domain data; algorithms; domain sim. metrics
Tang		Emotion prediction	Different linguistic and sentiment analysis methods are needed
Xue		Normalization for NLP	Multi-channel (source theory) model for normalization; argues for using orthographic; phonetic; and contextual info; and acronym expansion
Zhang		Speech act recognition	Use of word- and character-based features; empirical study
Duchon	Other Microtext Sources	Data organization	Cross-domain analyses for types of communications; initial data analysis
Galitsky		Ontology construction & syntactic generalization	Evaluation of search relevance improvement & text similarity improvement
Sajnani		Classifying Barnstars	Evaluation of multi-class supervised algorithms; utility of contextual features
Young		Persuasion detection	Introduction of hostage negotiation corpus; comparison of supervised learners for persuasion detection

Summary of Presentations: Task



Welcome!



Panel

Topics

- Douglas Oard (U. Maryland): **Information Retrieval**
- Vita Markman (Disney IMG): **Social Media Challenges in NLP**
- David Uthus (NRC/NRL): **Summarization**
- Sofus Macskassy (Fetch Technologies and U. Southern California): **Statistical Relational Learning**

Structure

1. Lunch groups:

- Each panelist will lead a group to some nearby restaurant, and will discuss their topic
- ***Please join your selected group leader!***

2. Panel discussion (1640-1745)

- 10min (max) presentations
 - **Background:** What is this topic and how is this relevant to analyzing microtext?
 - **Status:** What (briefly) is the status of research on this topic?
 - **Problems:** What challenges need to be addressed?
 - **Focus:** What's your specific interest(s) in this topic?
 - **Future:** What advances on this topic do you expect will occur during the next 5 years?
 - What do we need to do to build a research community around these topics? Leverage some existing conference? Build some new venue? Get DARPA/NSF/EU/... to create some large research program?
 - Name one good dissertation topic that is ripe, and that some clever student could start working on now.
 - What was your biggest surprise today? For example, what do you know now that you didn't know at the start of the day that will change your own work in the future?
- 25min discussion

3. Dinner (all invited) after the Opening Reception