Julia Hockenmaier, PhD

Assistant Professor

Department of Computer Science

University of Illinois

**Title:** Describing images in natural language: Towards visually grounded semantics

**Abstract:** When we read a descriptive sentence like “People are shopping in a supermarket”, we picture an indoor scene where customers are pushing shopping carts down aisles of produce or other goods, standing to look at the items on the shelves, or waiting in line to pay, etc. That is, if we understand a sentence, we infer what other facts are likely to be true in any situation described by that sentence. These inferences are an integral part of language understanding, but they require a great deal of commonsense world knowledge. In this talk, I will consider two tasks that require systems to draw similar inferences automatically.

First, I will describe our work on developing systems and data sets to associate images with sentences that describe what is depicted in them. I will show that systems that rely on visual and linguistic features that can be obtained with minimal supervision perform surprisingly well at describing new images. I will also define a ranking-based framework to evaluate such systems. In the second part of this talk, I will describe how we can combine ideas from distributional lexical semantics and denotational formal semantics to define novel measures of semantic similarity. We define the 'visual denotation' of linguistic expressions as the set of images they describe, and use our data set of 30K images and 150K descriptive captions to construct a 'denotation graph', i.e. a very large subsumption hierarchy over linguistic expressions and their denotations. This allows us to compute denotational similarities, which we show to yield state-of-the-art performance on tasks that require semantic inference.

References

M. Hodosh, P. Young and J. Hockenmaier (2013) "Framing Image Description as a Ranking Task: Data, Models and Evaluation Metrics", Volume 47, pages 853-899

http://www.jair.org/papers/paper3994.html

P. Young, A. Lai, M. Hodosh, and J. Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions.

Transactions of the Association of Compuational Linguistics (TACL) 2014, 2(Feb), pages 67-78. http://www.transacl.org/wp-content/uploads/2014/02/41.pdf