

The SUMO Speaker Series for Undergraduates

meets on Wednesdays, 4:30-5:30pm in Room 380-380C (in the basement)

All talks will be accessible to Freshman taking the 50-series
food from Pizza Chicago

Wednesday, October 15th:

Harnessing the World Wide Web to Measure Textual Similarity

Professor Mehran Sahami (C.S. department)

ABSTRACT:

The World Wide Web provides a wealth of data that can be harnessed to help improve information retrieval (i.e., web search) and increase understanding of the relationships between different entities. In many cases, we are often interested in determining how similar two entities may be to each other, such as two pieces of text. In this talk, we examine how this problem can be addressed by making use of mathematical techniques to model and measure textual similarity (specifically examining the difficult case of short text snippets -- even those that may not share any terms in common). We also present a direct application of our similarity measurement technique to improving information access for users of web search engines.

contact Nathan Pflueger (pflueger@stanford.edu) for more information.