

Kevin Bauer
CSCI 7000: Doctoral Level Independent Study
December 5, 2006
Colloquium Review

Theorizing from Data: Avoiding the Capital Mistake

Peter Norvig
Google, Inc.

In this talk, Dr. Norvig shared some insight into the research currently being conducted at Google in the area of language understanding, machine translation, information extraction, and inference. Google has refined its search techniques to such an extent that they now possess an enormous volume of data. A significant problem has been to develop ways to manage this data in such a manner as to first, be useful for information retrieval purposes, and second, to aid in the aforementioned problem domains. For instance, Google has made significant progress in machine translation by using text available on the web that is written in multiple languages concurrently. The main premise of the talk is that the exponential increase in data has had, and is currently having tremendous benefits for problems within the natural languages processing domain. In addition to sharing current and past research activities at Google, Dr. Norvig provided motivation for students to work at Google.