

# Discovery in Text: Visualisation, Topics and Statistics

**Wray Buntine**

NICTA

Canberra, Australia

wray.buntine@nicta.com.au

## Abstract

Discovery in or understanding of a text collection can be viewed from many angles: the text aspect of the data mining paradigm, the discover aspect of the information seeking paradigm, or the text content aspect of visualisation. This talk will view topic models as a technique within these paradigms. Some visualisations will be reviewed, as well as a variety of different topic models, and some of the natural language processing issues involved in working with the models. Finally, some of the non-parametric statistical methods underlying the analysis will be reviewed because they are fascinating as well.

## Short Biography

Dr. Wray Buntine joined NICTA in Canberra Australia in April 2007 and is a Principal Researcher working on applying machine learning and probabilistic methods to tasks such as information access and text analysis. In 2009 he was co-chair of ECMLPKDD in Bled, Slovenia and in 2011 he co-organised a PASCAL2 Summer School on Machine Learning in Singapore. He reviews for conferences such as ECIR, CIKM, ECMLPKDD, ICML, KDD, SIGIR, UAI and WWW and is on the editorial board of Data Mining and Knowledge Discovery. He was previously at University of Helsinki, Helsinki Institute for Information Technology, NASA Ames Research Center, UC Berkeley, and Google.