

TOPIC MODELING THE PAST

Panelists:

Nelson, Robert K., University of Richmond, USA

Mimno, David, Princeton University, USA

Brown, Travis, University of Maryland, College Park, USA

Chair:

Mitsu Inaba. Ritsumeikan Univeristy

Robert K. Nelson

- Dr. Robert K. Nelson is the director of Digital Scholarship Lab at University of Richmond.
- He is an historian of nineteenth-century America.
- He holds a PhD in American studies from the College of William and Mary, and his work has appeared in the Journal of Social History and American Literature.

David Mimno

- Dr. David Mimno is currently CRA Computing Innovation Fellow at Princeton University.
- He worked for an internet auction startup, the NLP group at the University of Sheffield, and a cultural heritage digital library, the Perseus Project.
- He has a particular interest in historical texts and languages.

Travis Brown

- Dr. Travis Brown is Assistant Director of Research and Development at the Maryland Institute for the Technology in Humanities.
- He worked as an editor for the Walt Whitman Archive and was the lead developer of eComma, a web application for collaborative textual annotation.
- He is particularly interested in using computational linguistics techniques for visualizing large collections of literary and historical texts.

Introduction

- Topic modeling is a probabilistic, statistical technique that uncovers themes and topics and can reveal patterns in otherwise unwieldy amounts of text. This panel will, first and foremost, illustrate the interpretative potential of topic modeling for research in the humanities. Robert K. Nelson will analyze the similarities and differences between Confederate and Union nationalism and patriotism during the American Civil War using topic models of two historic newspapers.
- Travis Brown will explore techniques to tailor topic model generation using historical data external to a corpus to produce more nuanced topics directly relevant to particular research questions.
- David Mimno (chief maintainer of the most widely used topic modeling software, MALLET) will describe his work using topic modeling to generate – while respecting copyright – a new scholarly resource in the field of Classics that derives from and organizes a substantial amount of the twentieth century scholarly literature.
- The panel will also address methodological issues and demonstrate new applications of topic modeling, including the challenge of topic modeling across multi-lingual corpora, the integration of spatial analysis with topic modeling (revealing the constructedness of space, on the one hand, and the spatiality of culture, on the other), and the generation of visualizations using topic modeling useful for ‘distant reading.’ The panel thus addresses issues of multilingualism, spatial history, data mining, and humanistic research through computation.

Questions

- What is the key to successful "topic modeling" in the humanities?
- What should we consider to integrate computer-generated "objective" results and "subjective" perspectives of humanities scholars?