

## Guest Lecture

The Department of Political Science and the Digital Science Center  
would like to invite you to the following presentation:

**Kenneth Benoit**

London School of Economics and Political Science

“More than Unigrams Can Say: Detecting Meaningful Multi-word  
Expressions from Political Texts”

Almost universal among existing approaches to text mining is the adoption of the bag of words approach, counting each word as a feature without regard to grammar or order. This approach remains extremely useful despite being an obviously inaccurate model of how observed words are generated in natural language. Many substantively meaningful textual features, however, occur not as unigram words but rather as multi-word expressions (MWEs): pairs of words or phrases that together form a single conceptual entity whose meaning is distinct from its individual elements. Here we present a new model for detecting meaningful multi-word expressions, based on the novel application of a statistical method for detecting variable-length term collocations. Combined with frequency and part-of-speech filtering, we show how to detect meaningful MWEs with an application to public policy, political economy, and law. We extract and validate a dictionary of meaningful collocations from three large corpora totalling over 1 billion words, drawn from political manifestos, legislative floor debates, and US federal and Supreme Court briefs. Applying the collocations to replicate published studies using unigrams only applied to each field, we demonstrate that using collocations can improve accuracy and validity over the standard unigram bag of words model.

About the speaker

Kenneth Benoit is Professor of Computational Social Science in the Department of Methodology at the London School of Economics and Political Science. His current research focuses on computational, quantitative methods for processing large amounts of textual data, mainly political texts and social media.

**Date and Time: Monday, 20 January, 2 p.m.**

**Place: Campus Innrain, Innrain 52e, HS 6 (Lecture Hall 6, ground floor)**