



# Big Data Research as Interaction between Topic Models and Expert Data

A New Approach to Capturing National Online  
Debates

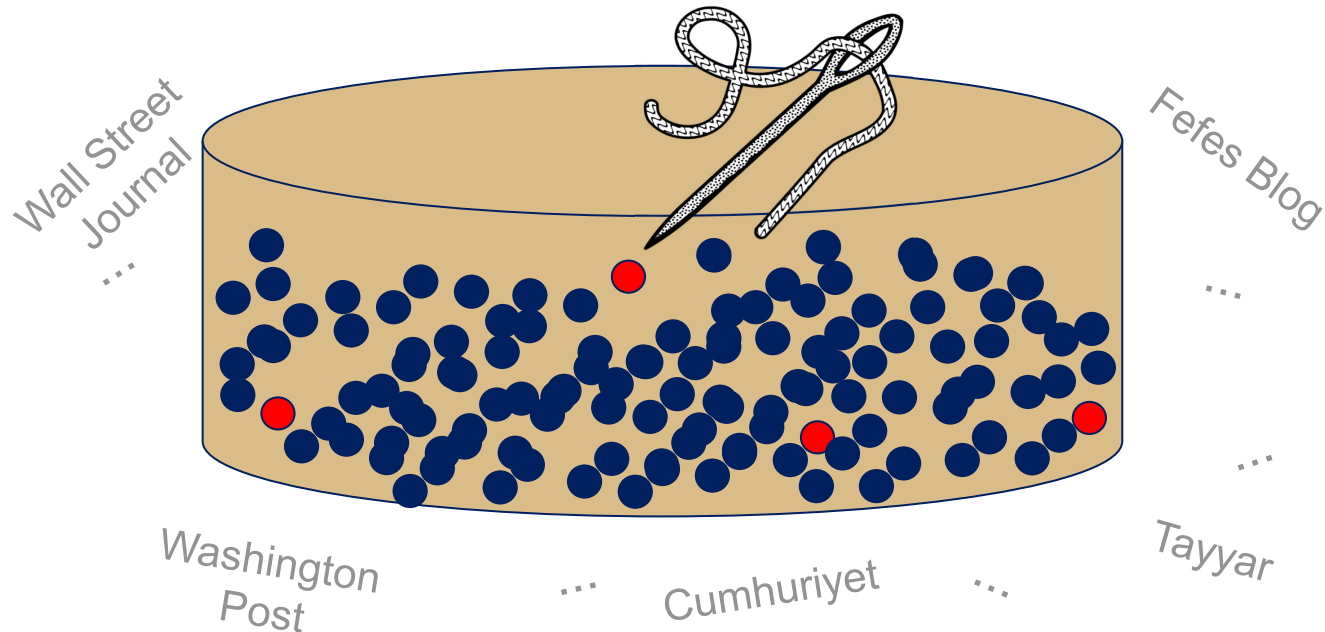
Cäcilia Zirn, Eike Mark Rinke, Charlotte Löb, Hartmut  
Wessler

3<sup>rd</sup> GESIS Computational Social Science Winter  
Symposium 2016

30 November 2016

# Analytical Problem: Thematic Document Classification

Finding the needle in the haystack!



---

# Application background

- six countries (Australia, Germany, Lebanon, Switzerland, Turkey, USA)
- Text data:
  - complete crawl of 101 leading national news websites & blogs
  - one calendar year (1 August 2015-31 July 2016)
  - $\approx$  1,6 million documents (main articles)
- Expert data:
  - media, journalism, and religion scholars

# Analytical Approach: Expert Survey Data Integration with Topic Models

Thematic classification  
based on topic model  
(LDA) process

