

# TM 2015 – Topic Models: Post-Processing and Applications Workshop

Nikolaos Aletras, Jey Han Lau, Timothy Baldwin, Mark Stevenson

University College London; King's College London; The University of Melbourne; University of Sheffield  
nikos.aletras@gmail.com; jeyhan.lau@gmail.com; tb@ldwin.net; mark.stevenson@sheffield.ac.uk

## ABSTRACT

The main objective of the workshop is to bring together researchers who are interested in applications of topic models and improving their output. Our goal is to create a broad platform for researchers to share ideas that could improve the usability and interpretation of topic models. We expect this will promote topic model applications in other research areas, making their use more effective.

## Categories and Subject Descriptors

I.7 [Computing Methodologies]: Document And Text Processing

## Keywords

Topic Model, Topic Modeling, Topic Coherence, Topic Evaluation, Topic Similarity, Topic Model Applications

## 1. INTRODUCTION

Topic modelling is a popular tool for modelling document collections and has seen applications in a variety of domains, from medical science to digital humanities. Topic models provide an effective means of extracting topics from large document collections and have the advantage that they require no supervision or document labels. While topic models can be trained automatically, judicious processing of their output is often required in applications where human users interact directly with the topics generated by a model. For example, the raw output from topic models can be difficult to interpret and consequently labels (manually assigned or automatically generated) are often added to increase readability; visualisation of the relationship between documents and topics is not intuitively obvious and its design requires thoughtful consideration. In addition, topics can contain unrelated words which may make the navigation of the document collection difficult; a typical approach to tackle this is to filter out low quality topics by computing topic coherence. This workshop aims to bring together a wide range of

tools and methodologies used in post-processing the output of topic models, with the goal of improving their usability and interpretation within user applications.

## 2. TOPICS OF INTEREST

- Topic coherence
- Extrinsic topic models evaluation
- Topic representation (generation of topic labels)
- Topic similarity
- Applications of topic models
- Visualisation of document collections with topic models (topic browsers)
- Multimodal topic models
- Data sets for evaluating topic qualities, including coherence, labels and similarity

## 3. ORGANISERS

- Nikolaos Aletras (University College London)
- Jey Han Lau (King's College London)
- Timothy Baldwin (The University of Melbourne)
- Mark Stevenson (University of Sheffield)

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author(s). Copyright is held by the owner/author(s).

CIKM'15, October 19–23, 2015, Melbourne, Australia.

ACM 978-1-4503-3794-6/15/10.

DOI: <http://dx.doi.org/10.1145/2806416.2806875>.