



Topic
Modeling
Visualization

Interactive
Text Mining
Suite

Demonstration

Topic Modeling of Scholarly Articles:
Interactive Text Mining Suite
Интерактивный инструмент глубокого
анализа текста

Olga Scrivner and Jefferson Davis

Indiana University



Dialogue 2016



Topic Modeling Visualization - Lacunae

Topic
Modeling
Visualization

Interactive
Text Mining
Suite

Demonstration

- Many state-of-the-art topic modeling tools require programming skills
- There is a need for interactive user-friendly tools for humanities researchers
- Tools should incorporate comparison of different methods for topic modeling
- Most tools are platform-dependent and require installation





Interactive Text Mining Suite - ITMS

Topic
Modeling
Visualization

Interactive
Text Mining
Suite

Demonstration

ITMS:

- Web application - **Shiny App** framework for R
- Not constrained by platform dependency, memory limitation or programming skills
- User-friendly
- State-of-the-art statistical and graphical tools (R programming language)
- Large audience (social science, digital humanities, linguistics)

Interactive Text Mining Suite ITMS

About Demo Data Preparation **Data Visualization** References

Word Frequency Topic Modelling Cluster Analysis Punctuation Analysis

ITMS



ITMS Web Application

Topic
Modeling
Visualization

Interactive
Text Mining
Suite

Demonstration



- 1 Use QR Scanner or go to <https://languagevariationsuite.shinyapps.io/TextMining/>
- 2 Smart Phone, iPad or computer
- 3 For phone and iPad - use Dropbox to upload files



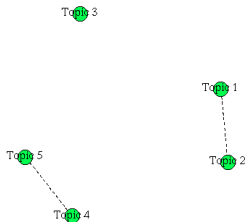
ITMS Functions

Topic
Modeling
Visualization

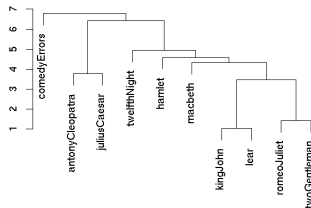
Interactive
Text Mining
Suite

Demonstration

- 1 Input: text and pdf files
- 2 Interactive data pre-processing
- 3 Data visualization (frequency, clustering and topics)



Cluster Dendrogram



- 4 Topic modeling comparison:
 - Collapsed Gibbs Sampling Method (*package lda*)
 - Latent Dirichlet Allocation (*package topicmodels*)
 - Structural Topic Modeling (*package stm*)



Pilot Study Demonstration - Articles

Topic
Modeling
Visualization

Interactive
Text Mining
Suite

Demonstration

Data: articles between 2010 and 2015 with keywords *twitter*, *tweet*, *tweets* in titles

Source: the ACL anthology website
<http://www.aclweb.org/anthology/>

level
topics
event
irony
users
sarcasm
sentiment

	Topic 1	Topic 2	Topic 3	Topic 4
1	event	irony	sentiment	level
2	events	sarcasm	features	disclosure
3	topic	speaker	analysis	model



Future Directions

Topic
Modeling
Visualization

Interactive
Text Mining
Suite

Demonstration

- Support for html and xml files
- Dynamic graphs
- Collocation and Name entity analyses
- Deployment with Hadoop to overcome processing limitations



We will greatly appreciate suggestions, feedback and collaboration initiatives to continue the development of open source interactive text-mining tools.