



Summary & Future Directions

WWW 2023 Tutorial

Turning Web-Scale Texts to Knowledge: Transferring Pretrained Representations to Text Mining Applications

Yu Meng, Jiaxin Huang, Yu Zhang, Jiawei Han

Computer Science, University of Illinois at Urbana-Champaign

April 30, 2023

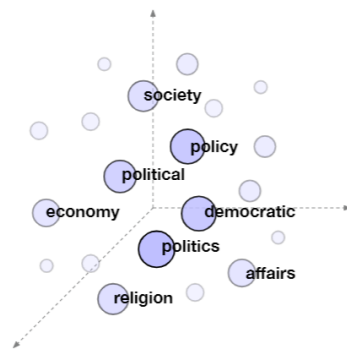
Towards Factual, Automatic, and Efficient Text Mining

Text Corpus

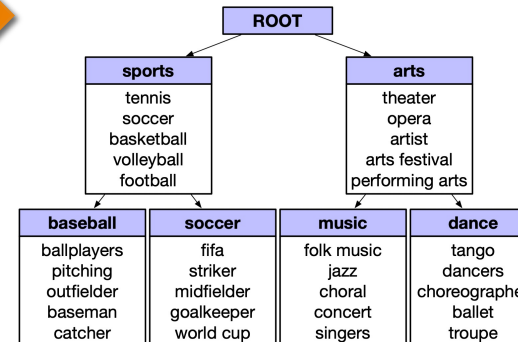


Existing KB

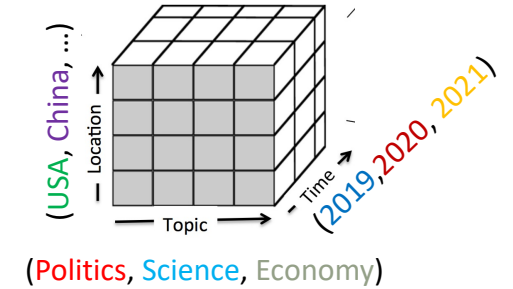
Pretrained Language Model



Topic Discovery



Weakly-Supervised Text Classification



Summary: from Unstructured Text to Knowledge

- ❑ Leverage the Power of Text Embedding and Language Models to Transform Unstructured Text into Structured Knowledge
- ❑ Mining Structures from Massive Unstructured Text (Texts → Structures)
 - ❑ Automated Text Representation Learning
 - ❑ Automated Multi-Faceted Taxonomy Construction
 - ❑ Automated Topic Mining
 - ❑ Automated Text Classification for Document Assignment
 - ❑ Automated Comparative Summarization in Multidimensional Text Cube
- ❑ Still a lot of work to do from unstructured text to structured knowledge

Our Journey: From Big Data to Big Structures & Knowledge



Han, Kamber and Pei,
Data Mining, 3rd ed. 2011

Yu, Han
Link

SYNTHESIS
DATA N
Jiawei Han

*SYNTHESIS LECTURES ON
DATA MINING AND KNOWLEDGE DISCOVERY*

Jiawei Han, Lise Getoor, Wei Wang, Johannes Gehrke, Robert Grossman, Series Editors

Information Networks, 2012

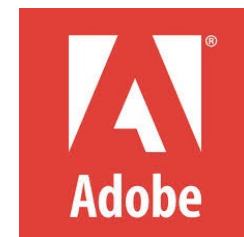
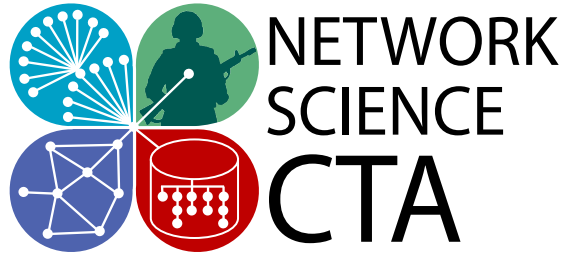
Y. Sun: SIGKDD'13 Dissertation Award

Jiawei Han and Han, Mining Latent Entity
Structures, 2015

Wang: SIGKDD'15 Dissertation Award

Acknowledgement

- Thanks for the research support from: ARL/NSCTA, NIH, NSF, DHS, ARO, DTRA





Thank you !
Q&A

