



Towards an Integrated Information Processing Paradigm

EDBT 2023 Tutorial: Mining Structures from Massive Texts by Exploring the Power of Pre-trained Language Models

Yu Zhang, Yunyi Zhang, Jiawei Han

Department of Computer Science, University of Illinois at Urbana-Champaign

Mar 29, 2023

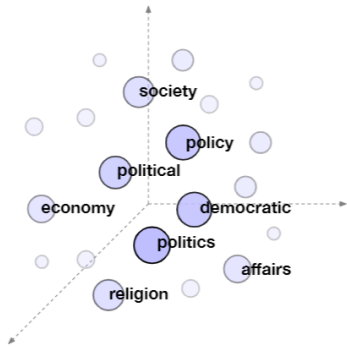
Our Roadmap of This Tutorial

Text Corpus

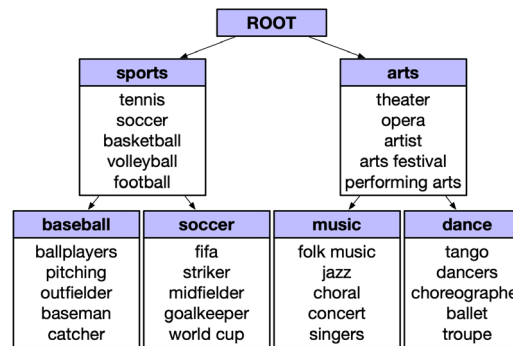


Existing KB

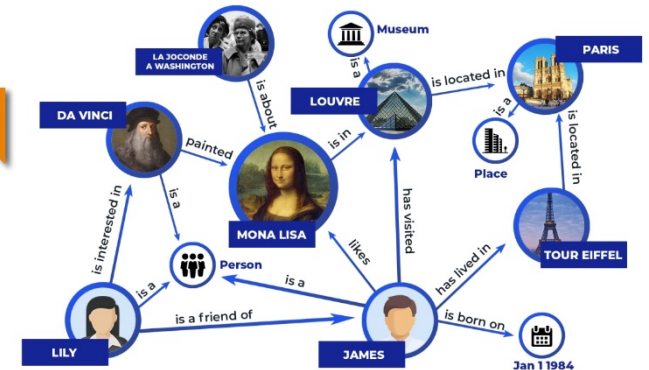
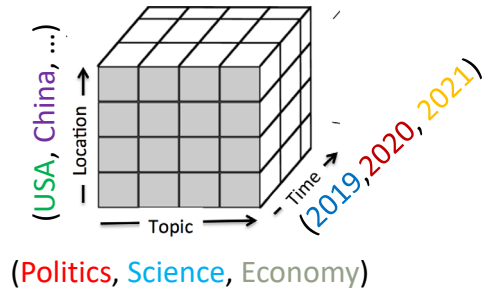
Part I: Pretrained Language Models



Part II: Mining Topic Structures



Part III: Mining Document Structures



Part IV: Mining Entity Structures

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 Topic Discovery
 - ❑ Automated Text Classification for Document Assignment
 - ❑ Automated Multi-Faceted Taxonomy Construction
 - ❑ 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

The image displays a collage of book covers from Morgan & Claypool Publishers. The central focus is the book 'Multidimensional Mining of Massive Text Data' by Chao Zhang and Jiawei Han, which is part of the 'Synthesis Lectures on Data Mining and Knowledge Discovery' series. Other visible covers include 'Mining Structures from Text: An Effort-Light Approach' by Xiang Ren and Jiawei Han, and 'Mining Latent Entity Structures' by Jiawei Han. The background features a colorful mosaic pattern and the text 'DATA Concepts'.

Morgan & Claypool Publishers

Multidimensional Mining of Massive Text Data
Chao Zhang
Jiawei Han
SYNTHESIS LECTURES ON DATA MINING AND KNOWLEDGE DISCOVERY
Jiawei Han, Lise Getoor, Wei Wang, Johannes Gehrke, Robert Grossman, Series Editors

Mining Structures from Text
An Effort-Light Approach
Xiang Ren
Jiawei Han
SYNTHESIS LECTURES ON DATA MINING AND KNOWLEDGE DISCOVERY
Jiawei Han

Mining Latent Entity Structures
Jiawei Han
SYNTHESIS LECTURES ON DATA MINING AND KNOWLEDGE DISCOVERY
Jiawei Han, Lise Getoor, Wei Wang, Johannes Gehrke, Robert Grossman, Series Editors

DATA
Concepts

Morgan & Claypool Publishers

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

Yu, Han, Link

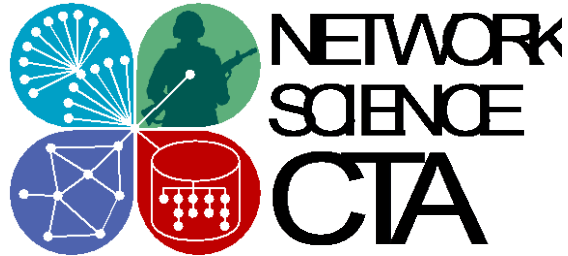
Information Networks, 2012

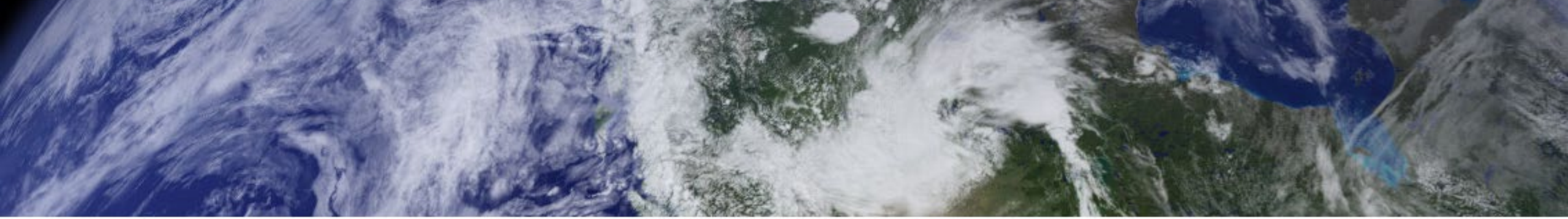
Y. Sun: SIGKDD'13 Dissertation Award

Wang: SIGKDD'15 Dissertation Award

Acknowledgement

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





Thank you !
Q&A

