Zhihui Yang

zhyangcs@gmail.com
+86 13262778217
No. 1159, Cailun Road, Pudong District, Shanghai, China, (201203)

Research Interests

Text Mining; Exploratory Analysis of Text Data; Database System

Education

2014 - present **Fudan University (FDU), Shanghai, China**
Ph.D. Candidate in Computer Science
Advisor: Prof. X. Sean Wang; **GPA**: 3.566/4

2010 - 2014 **Lanzhou University (LZU), Lanzhou, Gansu Province, China**
B.S. in Information Science and Engineering; **(Outstanding Graduates)**
Advisor: Prof. Wei Su; **GPA**: 4.77/5
Undergraduate Thesis: *Chem2Dot: a CML to Chemical Braille Translation Software*
(Excellent Undergraduate Thesis)

Publications

1. **Yang, Zhi hui**, Ma, H. x., He, Z. y., & Wang, X. S. (2017). Finding maximal ranges with unique topics in a text database. ***World Wide Web Journal (WWWJ)***, 1–22.
2. Ma, H. x., **Yang, Zhi hui**, Jing, Y. n., He, Z. y., & Wang, X. S. (2018). Answering unique topic queries with dynamic threshold. ***World Wide Web Journal (WWWJ)***, 1–20.
3. **Yang, Zhi hui**, Gong, J. y., Liu, C. y., Jing, Y. n., He, Z. y., Zhang, K., & Wang, X. S. (2018). Iexplore: accelerating exploratory data analysis by predicting user intention. In *International conference on database systems for advanced applications* ***(dasfaa)***.
4. Zhou, K. w., **Yang, Zhi hui**, Ma, H. x., He, Z. y., Jing, Y. n., & Wang, X. S. (2017). Design and development of partitional topic model. *Journal of Frontiers of Computer Science and Technology*. doi:10.3778/j.issn.1673-9418
5. Liu, L. h., **Yang, Zhi hui**, He, Z. y., Jing, Y. n., & Wang, X. S. (2017). Unique topic query system based on relational information extraction. In *The 34th national database conference*.
6. Liu, L. h., **Yang, Zhihui**, He, Z. y., Jing, Y. n., & Wang, X. S. (2018). Unique topic query processing on cloud (under review).

Selected Research Experience

- **Text Data Analysis (2014-present), Laboratory for Data Analytics and Security (DAS Lab)** (Main Contributor)
  - We introduced the concept of *unique topics* to discover topics that appear frequently within a small range of documents in contrast to the whole range.
  - We also proposed a pruning-based optimization (PBO) algorithm to find the maximal ranges of the specified unique topic. The PBO algorithm reduced the time complexity from $O(n^3)$ to $O(n^2)$. Additionally, we further reduced the time complexity to $O(n)$.
  - Based on LDA, we developed a new topic model DbLDA to utilize the commonalities inside each subset in a text database.
  - These works was published on **WWWJ2017 and WWWJ2018** .

Selected Research Experience (continued)

- **Exploratory Data Analysis (2016-present), DAS Lab**
  - `Hubble: A Smart System for Data Exploration in Big Data Era, bridge the gap between analysts and data (Participant)`
  - `iExplore: Accelerating Exploratory Data Analysis by Predicting User Intention (Main Contributor)` (i)We introduced an intention model to help the iExplore system have a comprehensive understanding of user's intention. (ii) We also studied the convergence of the intention model to figure out the characteristic of the exploratory process.
  - This work has been accepted by **DASFAA2018** .
- **Chemical Markup Language (2013-2014), WME Lab** (Main Contributor)
  - We designed a method to translate Chemical Markup Language (CML) to Braille to facilitate information accessibility.
  - This work was funded by **Hui-Chun Chin and Tsung-Dao Lee Chinese Undergraduate Research Endowment, CURE**.
  - My undergraduate thesis about this work was awarded **Excellent Undergraduate Thesis.**

Selected Development Experience

- **iExplore**: It is a smart system to assist users in doing the time-consuming data exploration task through predicting user intention. (Independent Project)
- **Inverted Index on Hadoop**: We builded inverted index of 1.9 million webpages (12.27GB) on a Hadoop cluster with 5 machine with 1.5 hours to construct the index. (Cooperative Project)
- **Chem2Dot**: A Java project focuses on translating CML to Braille to facilitate the process of information accessibility. (Independent Project)
- **LocationSender**: An Android App aims at automatic location and sharing. (Cooperative Project)

Selected Skills

Technologies — Hadoop, Spark, Pytorch.

Coding — Java, Python.

Selected Honors and Awards

- 2016 **Outstanding Ph.D. Student at Fudan University.**
- 2015 **Outstanding Students of Master's Degrees at Fudan University.**
- 2014 **the 15^{th} Chun-Tsung Scholar**
- 2014 **Outstanding Graduate at Lanzhou University.**
- 2014 **Excellent Undergraduate Thesis of Lanzhou University.**
- 2013 **IBM University Program Academic Qualification.**
- 2012 **National scholarship.**
- 2011 **National inspirational scholarship.**