AMiner-mini:
A People Search Engine For University

Jingyuan Liu*, Debing Liu*, Xingyu Yan*,
Li Dong#, Ting Zeng#, Yutao Zhang*, and Jie Tang*

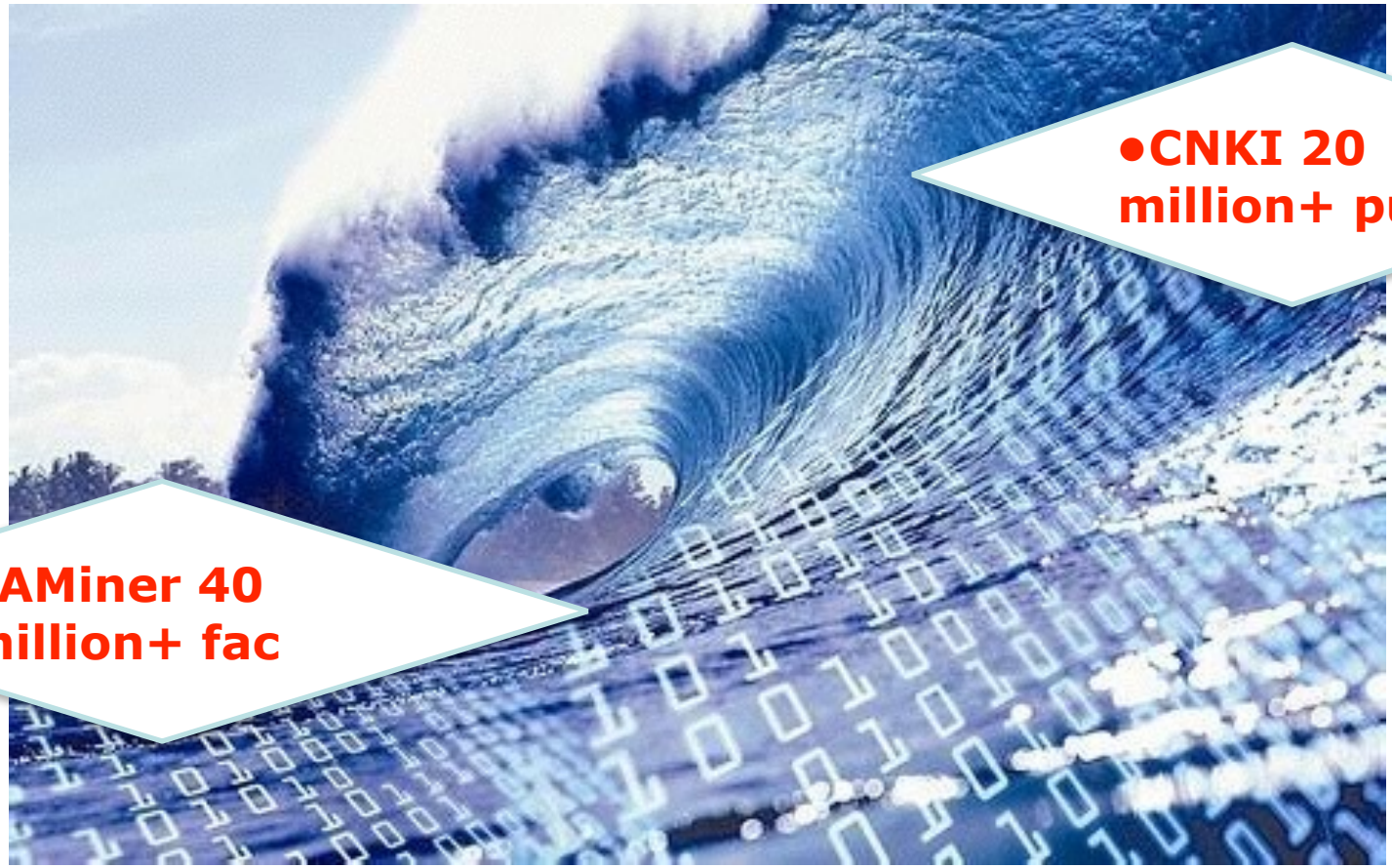
*Dept. of Com. Sci. and Tech. , Tsinghua University
#Tsinghua University Library

System website: http://dlib.lib.tsinghua.edu.cn/

Paper: http://keg.cs.tsinghua.edu.cn/jietang/publications/CIKM14-Liu-et-alAminer-mini.pdf

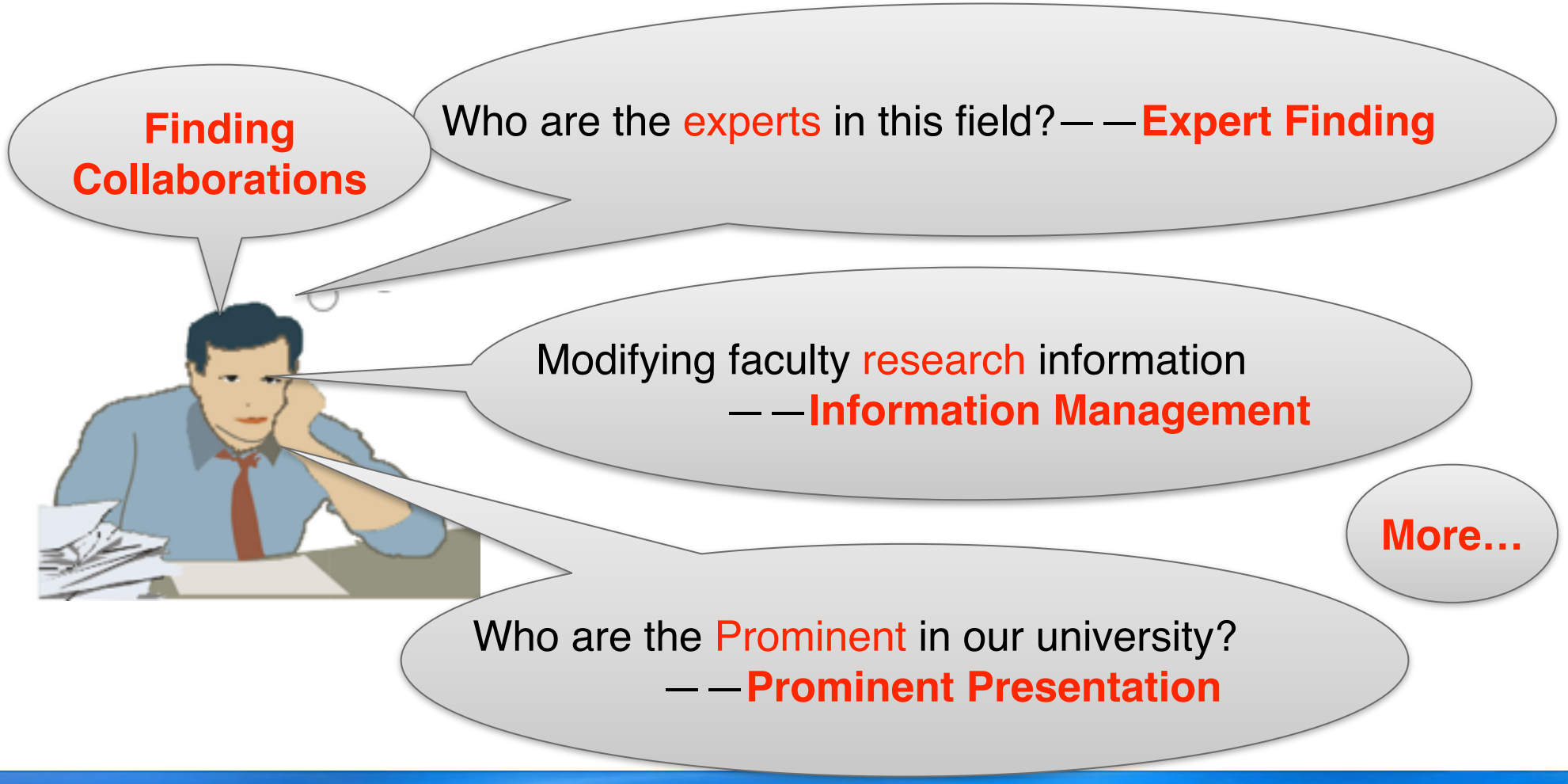
Motivation

- Digital Academic Data Rapid Proliferation



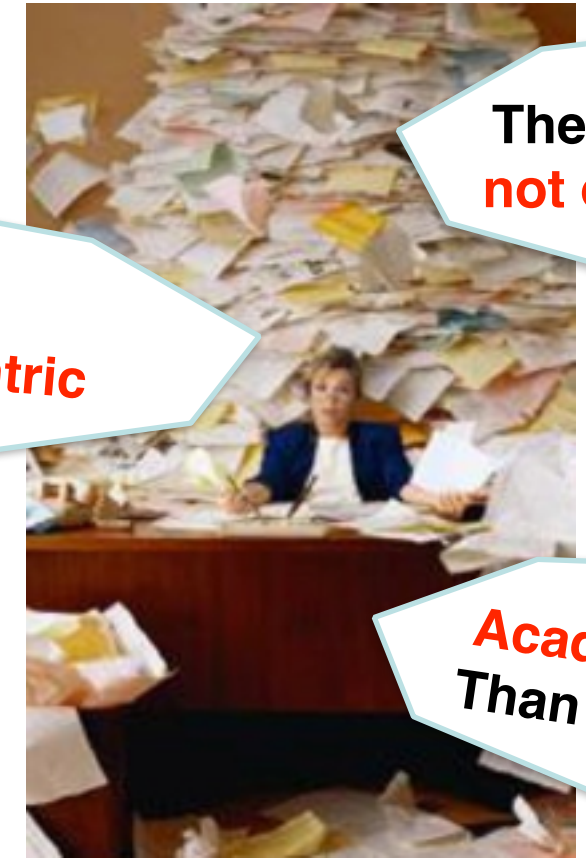
Motivation

- Satisfying Different User Scenarios



Motivation

- People-Centric rather than Data-Centric



Web Search Trend:
Data Centric->People Centric

The Information need is not only about Pub

Academic Search More Than Keywords Matching

What is AMiner-mini?

- A **People Search Engine** for University
  - **Core Techniques:**
    - **Name Disambiguation**
    - **Academic Search**
  - **System Applications:**
    - **Expert Finding**
    - **Prominent Presentation**
    - **Publication Management**
  - **Distributed Structure:**
    - **Distributed Search**

System Statistic

- System mainly contains 3 entities:
  - **Faculty:**

    System contains **10918** faculties from **90** department

  - **Papers:**

    System contains **259465** papers range from 1981 to 2014

  - **Course:**

    System contains **10253** courses range from 2001 to 2013

Academic Search Algorithm

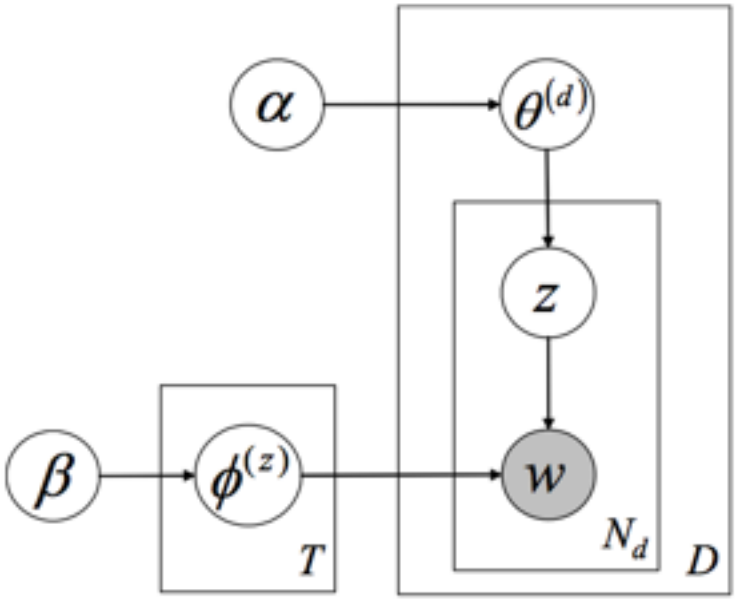
- Modeling Ranking Factors
  - **Relevance**: "relevance" between queries and entities
    - Language Model
    - LDA
  - **Importance**: "important" and "influential"
    - Random Walk
    - Prominent title
  - **Popularity**: "popular" entities
    - User feedback
    - Random Serendipity

Academic Search Algorithm

- Combing Ranking Factors
  - $Score = \omega_R * Relevance + \omega_I * Importance + \omega_p * Popularity$
  - weights are initially manually set
  - weights are 0.6, 0.2, 0.2 separately

Academic Search Algorithm

- Statistic Topic Model
  - Using LDA to extract hidden topics from textural materials



Academic Search Algorithm

- Search Experiment Result

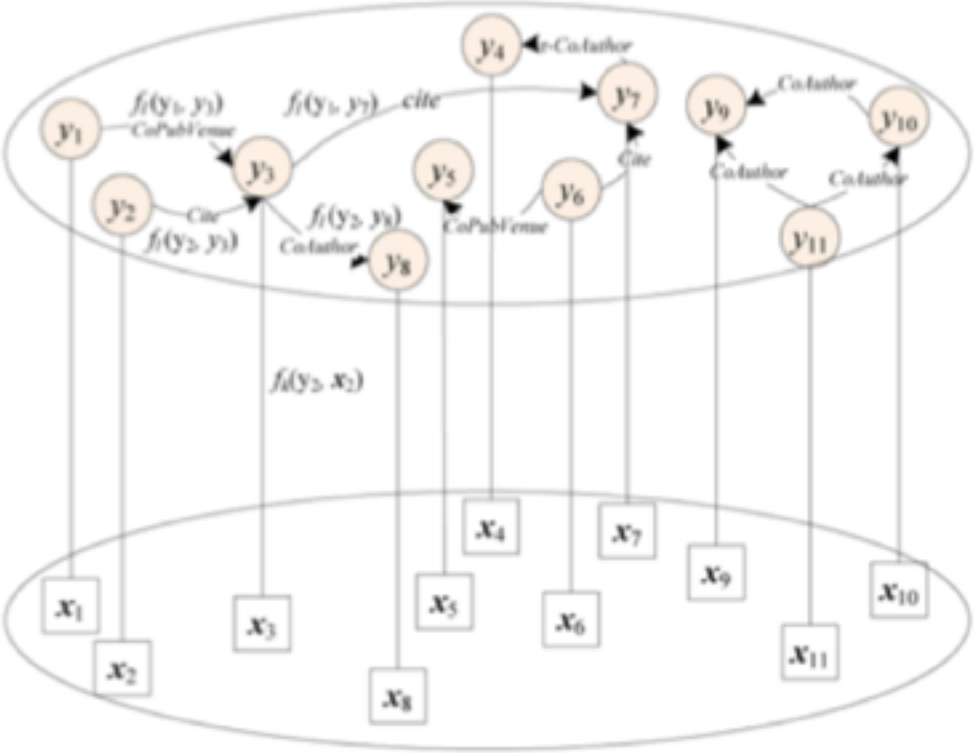
Search Methods	P@5	P@10	MAP
0.3 LDA + 0.7 LM	**0.876**	0.8	**0.912**
0.2 LDA + 0.8 LM	0.864	**0.81**	0.89
0.0 LDA + 1.0 LM	0.872	0.77	0.79
Lucene (TFIDF)	0.773	0.726	0.73

- Obviously outperforms baseline (TF-IDF)
- best combination weights: 0.3 LDA + 0.7 LM

Name Disambiguation Methodology

- Probabilistic HMRF Framework

  - Using a Probabilistic HMRF Framework to cluster ambiguity papers and courses



Name Disambiguation Methodology

- Active Learning Strategy

  - Using active learning strategy to form three-phases disambiguation framework

System Applications

- Expert Finding
  - Implement expert finding via academic search algorithm
  - Search for faculty, pub, course simultaneously

System Applications

- Publication Management
  - Present and Modify faculty information of the personal academic research interest, publication and courses

System Applications

- Prominent Presentation

  - Present prominent faculties with honored title

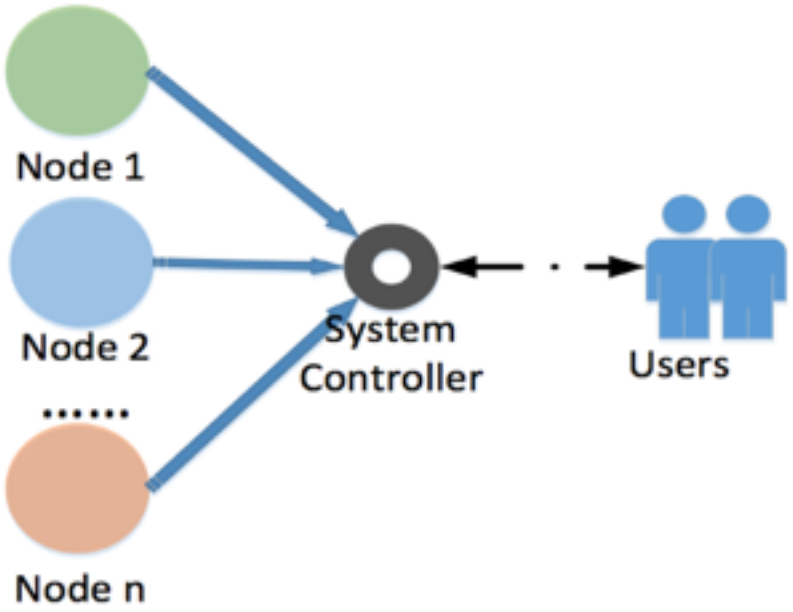
System Applications

- PersonInfo Presentation
  - Research interest
  - Academic social network
  - Research Trend
  - Research Topics

Distributed Structure

- **Intra-** and **Inter-** university level academic services
  - work as single node
  - connect via web server

- **Distributed Search**
  - system controller
  - rerank search result



Deploy your AMiner-mini

- System is cooperated with THU lib
- System is an ongoing project, THU version:
  - http://dlib.lib.tsinghua.edu.cn/
- We plan to build open-source project, find us:
  - git@github.com:toothacher17/AMiner-mini.git
- We are willing to help deploy your own AMiner-mini, contact us:
  - http://keg.cs.tsinghua.edu.cn/jietang/
- The system is developed under J2EE Tapestry Structure

Reference

- J. Tang, A.C.M. Fong, B, Wang, and J. Zhang. A Unified Probabilistic Framework for Name Disambiguation in digital library. In *TKDE*, Volume 24, Issue 6, Pages 975-987, 2012
- K. Balog, Y. Fang, M. de Rijke, P. Serdyukov and L. Si. Expertise Retrieval. In *FTIR,* Volume 6, 2012
- J. Tang, J. Zhang, R. Jin, Z. Yang, K. Cai, L. Zhang, and Z. Su. Topic Level Expertise Search over Heterogeneous Networks. In *Machine Learning Journal*, Volume 82, Issue 2, Pages 211-237, 2011
- R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval (2nd Edition)*. China Machine Press, 2010
- J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang and Z. Su. ArnetMiner: Extraction and Mining of Academic Social Network. In *KDD'08*, pages 990-998, 2008.
- A. Ferrreira, M. Gnocalves, and A. Laender. A Brief Survey of Automatic Methods for Author Name Disambiguation. In *SIGMOD'12*, 2012
- T. Joachims, L. Granka, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search. In *TIS,* Volume 25, 2007
- G. Coulouris, J. Dollimore, and T. Kindberg. *Distributed systems: Concepts and Design (5th Edition)*. China Machine Press, 2011.
- M. Ge, C. Delgado-Battenfeld, and D. Jannach. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *RecSys'10*, 2010

That is all!