PatentMiner: Topic-level IP Trend Monitoring and Analysis

PI: Jie Tang and Minlie Huang
Key Research Personnel: Bo Wang, Bo Gao, Quan Lin, Fengjiao Wang, Po Hu, Yi Yang
Tsinghua University

Date: August. 2nd, 2010
Motivation

Energy patents

Research papers

e.g., for “oil refinery”,
>179,000 papers
>830 patents...
Motivation (cont.)

• When starting a work in a new research topic;
• Or brainstorming for novel ideas;
• Or analyzing a company.

• What are the “hot” topics recent years?
• What are the most influential research work, researchers, and organizations in a specific topic?
• Which companies are our competitors?
…
Collaborative Research

• **Data source**
  – Patents crawled from USPTO (www.uspto.gov)

• **Research topics**
  – Patent feature extraction
  – Patent similarity and summarization
  – Automatic topic detection
  – Research topic trend tracking

• **System — TrendMiner**
  – Develop software framework for document acquisition, search, and results visualization
TrendMiner Overview

1. Document analysis/IR engine
   - Signature term extraction
   - Summarization
   - Classification
   - Co-Ranking
   - Topic discovery
   - Evaluation analysis

2. Graphical modeling
   - Topic modeling

3. Storage and access
   - Access interface
     - Indexing
     - Storage

4. Heterogeneous sources
   - Energy patents
   - Energy Journals / Conferences
   - Technical reports
   - Social network
   - More...
Patent Search

Summary

Inventor

Company
Patent Similarity and Recommendation

Patent information

Classification

Extracted attributes and features

Similar patents
200 topics have been discovered automatically from the patent data
Competitor Discovery

Mobil Oil Corporation

Discovered competitors:

- UOP (score: 12.3105, # patents: 41)
- UOP LLC (score: 10.3919, # patents: 40)
- ExxonMobil Research and Engineering Company (score: 2.6905, # patents: 8)
- UOP Inc. (score: 2.1600, # patents: 17)
- Stone & Webster Process Technology, Inc. (score: 1.6493, # patents: 9)
- Chevron Research Company (score: 1.1981, # patents: 736)
- Chevron Research and Technology Company (score: 1.1756, # patents: 93)
- W. R. Grace & Co.-Conn. (score: 0.9440, # patents: 2)
- Uop LLC (score: 0.8815, # patents: 1)
- Exxon Research and Engineering Company (score: 0.7391, # patents: 10)
- Amoco Corporation (score: 0.7767, # patents: 5)
- Texaco Inc. (score: 0.5335, # patents: 1523)
- Shell Oil Company (score: 0.3470, # patents: 27)
- Catalytic Distillation Technologies (score: 0.2916, # patents: 81)
- Mobil Oil Corp. (score: 0.2280, # patents: 198)
- Petroleo Brasileiro S.A.-Petrobras (score: 0.2129, # patents: 3)
- The M. W. Kellogg Company (score: 0.1813, # patents: 58)
- Davy Process Technology Limited (score: 0.1535, # patents: 16)
- Elliott Turbomachinery Company, Inc. (score: 0.1512, # patents: 5)
- Elliott Turbomachinery Co., Inc. (score: 0.1512, # patents: 19)
- Engelhard Corporation (score: 0.1494, # patents: 297)
TrendMiner Today

* Patent data:
  > 1.5M patents
  > 1.2M inventors
  > 200K companies
  > 3.0M citation relationships

* Journal data:
  > 2k journal papers
  > 3.7k authors

The crawled data is increasing to >70 Gigabytes.
Introduction of Key Components

• **Deep Document Analysis**
  – Patent classification
  – Key components extraction
  – Multi-patent summarization
  – Summary rephrasing

• **Topic-level Heterogeneous Co-Ranking**
  – Topic detection
  – Co-ranking patents, companies, and inventors
  – Ranking refinement via user feedbacks

• **Evolutionary Topic Analysis**
  – Topic evolution analysis
  – Competitor analysis
  – Name disambiguation
Multi-patent Summarization

1. Search relevant patents
2. Extract concepts to represent the Knowledge Space
3. Score concept’s importance to the query
4. Solve by Integer Linear Programming

Diversified contents in limited length

Knowledge space

Relevant Patents

User Query

Maximum Coverage Model
The invention provides a combination well logging device adapted to be moved inside a borehole to measure the properties of geological formations through which the borehole passes, characterized in that the device comprises a series of elongate logging sections mounted end to end, including a first section adapted to be maintained with a generatrix applied against the wall of the borehole, a second section comprising a body and a measurement pad adapted to be spaced radially from the wall, and a third section adapted to be kept spaced from said wall, first and second hinges connecting the second section to the first and third sections respectively in order to allow respective angles of inclination between the second section and the first and third sections.

Can we automatically generate a simple and easy-to-understand summary?

Solution: rephrasing

The invention provides a well-logging device to be moved inside a borehole to measure the properties of geological formations.
Solution: Modeling with CRFs

- Green nodes are *hidden vars*,
- Purple nodes are *observations*

\[
p(y | x) = \frac{1}{Z(x)} \exp \left( \sum_{e \in E, j} \lambda_j t_j (e, y_{e}, x) + \sum_{v \in V, k} \mu_k s_k (v, y_{v}, x) \right)
\]
Beyond Linear-chain CRFs

\[ p(Y \mid X) = \frac{\exp(\lambda \cdot F(Y, X))}{Z(X)} \]

\[ Z(x) = \sum \exp(\lambda \cdot F(y, x)) \]

\[ F(y, x) = \sum_{e \in \{E^{pc}, E^{ss}, E^{cp}\}} \sum_{j} \lambda_j t_j(e, y|_e, x) + \sum_{v \in V} \sum_{k} \mu_k s_k(v, y|_v, x) \]
Introduction of Key Components

• **Deep Document Analysis**
  – Patent classification
  – Key components extraction
  – Multi-patent summarization
  – Summary rephrasing

• **Topic-level Heterogeneous Co-Ranking**
  – Topic detection
  – Co-ranking patents, companies, and inventors
  – Ranking refinement via user feedbacks

• **Evolutionary Topic Analysis**
  – Topic evolution analysis
  – Competitor analysis
  – Name disambiguation
The Heterogeneous Patent Network

Heterogeneous objects:
- Patent
- Inventor
- Company

Relationships:
- Company own patent
- Patent cite patent
- Inventor write patent
- Inventor belong to company
- Inventor is co-inventors of inventor
- Company co-patentee of company

Challenges:
- How to extract the topics from the heterogeneous network?
- How to track the evolutionary pattern of different topics?
Topic-level Heterogeneous Co-Ranking

Method for interpreting seismic data
Mufti; Irshad R. - 1989, Mobil Oil Corporation (New York, NY)

Seismic data radon dip moveout method
Wang; Cheng-shu - 1998 Vector Seismic Data Processing, ...

Method for converting seismic data from the ...
Robinson; Gary Charles - 2009...

Search with keyword
Search with semantic modeling
Modeling using VSM
Modeling using semantic topics
1d seismic data

Patents
Inventors
Companies

Topics
- Seismic restraint
- Oil refinery
- Exhaust gas
- Drill pipe
- Catalyst system
- Recording system

Query vector
Doc1 vector
Doc3 vector
Doc4 vector
Heterogeneous Co-Ranking

1. **Topic Learning**
   - ExxonMobil
   - Shell
   - Chevron
   - Beasley; Craig J.
   - Hackett; Gary K.
   - Ferber; Ralf
   - Continuous automatic...
   - Method for processing...

2. **Co-Ranking**
   - ExxonMobil
   - Shell
   - Chevron
   - Beasley; Craig J.
   - Hackett; Gary K.
   - Ferber; Ralf
   - Method for enhancing...

**Query**: 1d seismic data

Company

Inventor

Patent
Generative Story of ACT Model

• Generative process

System and Method for Data Mining

Shafiei and Milios

IBM

providing a computing system

clustering documents and terms.

Our model is a four hierarchical

bayesian model. We present efficient

inference techniques based on

Markow Chain Monte Carlo. We

report results in document modeling,

document and terms clustering …

P(c|z) P(w|z)

P(c|z) P(w|z)

Google 0.23
IBM 0.19
model 0.23
learning 0.19
boost 0.17

Microsoft 0.23
IBM 0.19
mining 0.23
clustering 0.19
classification 0.17

Shafiei

NLP

IR

DM

ML

Milios

NLP

IR

DM

ML
ACT Model

Generative process:

1. For each topic $z$, draw $\phi_z$ and $\psi_z$ respectively from Dirichlet priors $\beta_z$ and $\mu_z$;
2. For each word $w_{di}$ in document $d$:
   - draw an author $x_{di}$ from $a_d$ uniformly;
   - draw a topic $z_{di}$ from a multinomial distribution $\theta_{x_{di}}$ specific to author $x_{di}$, where $\theta$ is generated from a Dirichlet prior $\alpha$;
   - draw a word $w_{di}$ from multinomial $\phi_{z_{di}}$;
   - draw a company stamp $c_{di}$ from multinomial $\psi_{z_{di}}$.

$$P(z_{di}, x_{di} | z_{-di}, x_{-di}, w, c, \alpha, \beta, \mu) \propto \frac{m_{x_{di}z_{di}} + \alpha_{z_{di}}}{\sum_z (m_{x_{di}z} + \alpha_z)} \frac{n_{z_{di}w_{di}} + \beta_{w_{di}}}{\sum_{w} (n_{z_{di}w} + \beta_{w})} \frac{n_{z_{di}c_{di}} + \mu_{c_{di}}}{\sum_{c} (n_{z_{di}c} + \mu_{c})}$$
ACT Model (Cont)

• After Gibbs sampling, we can easily estimate the probability of a topic given an author $\theta_{xz}$, the probability of a word given a topic $\phi_{zv}$, and the probability of a company given a topic $\psi_{zc}$:

$$\phi_{zw_{di}} = \frac{n_{zw_{di}} + \beta_{w_{di}}}{\sum_{w_{v}} (n_{zw_{v}} + \beta_{w_{v}})}$$

$$\psi_{z_{cd}} = \frac{n_{z_{cd}} + \mu_{cd}}{\sum_{c} (n_{z_{c}} + \mu_{c})}$$

$$\theta_{xz} = \frac{m_{xz} + \alpha_{z}}{\sum_{z^{'}} (m_{xz^{'}} + \alpha_{z^{'}})}$$
Topic-level Heterogeneous Co-Ranking

Stage 1:
Random walk

Stage 2.
Topic-based relevance

Inventor Graph $G_i$

Company Graph $G_c$

Patent Graph $G_p$

Seismic data

Query

ExxonMobil

top

BP

ExxonMobil

Tom CRF

Tree CRF

Association...

Inventor Graph $G_i$

Company Graph $G_c$

Patent Graph $G_p$

Seismic data

Query

ExxonMobil

top

BP

ExxonMobil

Tom CRF

Tree CRF

Association...

r[x] = \frac{\xi}{|V|} + (1 - \xi) \times \sum_{(x,y) \in E} \lambda_{xy} r[y] P(x|y)

Combination by multiplication

$R[d] = r[d] \times P(q|d)$

$P_{LM}(q|d) = \prod_{w \in q} \frac{N_d}{N_d + \lambda} \cdot \frac{t_f(w, d)}{N_d + \lambda} + (1 - \frac{N_d}{N_d + \lambda}) \cdot \frac{t_f(w, D)}{N_d}$

$P_{ACT}(q|d, \theta, \phi) = \prod_{w \in q} \sum_{z=1}^T P(w|z, \phi_z) \sum_{a \in A_d} P(z|a, \theta_a) P(a|d)$

$P(q|d) = P_{LM}(q|d) \times P_{ACT}(q|d)$
Topic-level Heterogeneous Co-Ranking (2)

Ranking score

\[
r(d_i) = \lambda_{dd} \times \sum_{(d_j, d_i) \in E_{dd}} p(d_i | d_j) \cdot r(d_j) + \lambda_{cd} \times \sum_{(c_i, d_i) \in E_{cd}} P(d_i | e_i) \cdot r(e_i)
+ \lambda_{cd} \times \sum_{(c_i, d_i) \in E_{cd}} P(d_i | c_i) \cdot r(c_i) + \lambda_{td} \times \sum_{(\theta_i, d_i) \in E_{td}} P(d_i | \theta_i) \cdot r(\theta_i)
\]

\[
r(e_i) = \lambda_{de} \times \sum_{(d_j, e_i) \in E_{de}} p(e_i | d_j) \cdot r(d_j)
\]

\[
r(c_i) = \lambda_{dc} \times \sum_{(d_j, c_i) \in E_{dc}} p(c_i | d_j) \cdot r(d_j)
\]

Transition probability

\[
\begin{align*}
P(z_i | a_j) &= \theta_{a_j z_i} \\
P(a_j | z_i) &= \frac{P(z_i | a_j)P(a_j)}{P(z_i)} \\
P(z_i | d_j) &= \frac{1}{A_d} \sum_{x \in A_d} \theta_{x z_i} \\
P(d_j | z_i) &= \prod_{i=1}^{N_d} P(w_{di} | z_i) \\
P(c_j | z_i) &= \psi_{z_i c_j} \\
P(z_i | c_j) &= \frac{P(c_j | z_i)P(z_i)}{P(c_j)} \\
P(q | z_i) &= \prod_{w \in q} P(w | z_i) \\
P(z_i | q) &\propto P(q | z_i)P(z_i)
\end{align*}
\]
User Feedbacks

• User feedbacks can reflect their personal judgment of the relevance between patent and a given query

• The user feedbacks can be exploited from the user click trough data

• Try to refine the ranking performance by utilizing the user feedbacks in click through data
User Feedbacks

Click Through Data

- (patent_1, no click)
- (patent_2, clicked)
- (patent_3, no click)
- (patent_4, clicked)
- (patent_5, no click)

for a given query

Preference Constraints

- (patent_1, patent_2) -1
- (patent_2, patent_3) +1
- (patent_2, patent_5) +1
- (patent_1, patent_4) -1
- (patent_3, patent_4) -1
- (patent_4, patent_5) +1

Data Format: \((x_i^1, x_i^2)\), label: \(c_i = +1/-1\)
Ranking refinement via user feedbacks

1d seismic data

Query-related Patents

TrendMiner

Current Ranking Score $f(q, \text{pat}_i)$

Preference Constraints

Model Learning from feedbacks

Feedback Ranking Score $g(q, \text{pat}_i)$

Final Ranking Score $f(q, \text{pat}_i) \ast g(q, \text{pat}_i)$
Ranking with feedbacks—A Hybrid Approach

- Let the ranking score from the new model between query $q$ and a patent $pat_i$ as $g(q, pat_i)$

$$\min_{\vec{w}} \left\{ \sum_{i=1}^{N} \log[1 + \exp(-c_i \langle \vec{w}, \vec{x}_i^{(1)} - \vec{x}_i^{(2)} \rangle)] + \lambda \| \vec{w} \|^2 \right\}$$

- Let the relevance between query $q$ and a patent $pat_i$ be $f(q, pat_i)$ (via e.g., VSM)

- Finally, the ranking score can be calculated by

$$\text{Score} (q, pat_i) = g(q, pat_i) \ast f(q, pat_i)$$
Introduction of Key Components

• **Deep Document Analysis**
  – Patent classification
  – Key components extraction
  – Multi-patent summarization
  – Summary rephrasing

• **Topic-level Heterogeneous Co-Ranking**
  – Topic detection
  – Co-ranking patents, companies, and inventors
  – Ranking refinement via user feedbacks

• **Evolutionary Topic Analysis**
  – Topic evolution analysis
  – Competitor analysis
  – Name disambiguation
Topic Evolution Analysis I

• Motivation:
  – What are the top companies for a given topic?
  – How does the interest of a company evolve?
**Topic Evolution Analysis II**

- **Motivation:**
  - What are the **top topics** for a company?
  - How does the **topics of a company** **evolve**?

- **Example:**

  ![Mobil Oil Corporation](image)

  **Mobil Oil Corporation** → **TrendMiner**

- **Procedure:**

  ![Diagram](image)

  - Company name
  - ACT
  - Company-related Patents
  - Assign each patent to a topic
  - Topic Ranking
  - Find top 5 topics
  - Patent reassignment
  - Generate evolution pattern

  **Top 5 Topic Trend**
  - Topic 148: Depilation method / Catalyzed vapor phase process
  - Topic 126: Polymerization catalyst / Conformal method
  - Topic 47: Synthesis gas / Acetaldehyde production
  - Topic 69: Efficient process / Production method
  - Topic 71: Sorbic acid process / Catalyst composition
**Competitor Analysis**

**Who are our competitors?**

- **Company name:**
  - Shell Oil Company

**Representative terms**
- Polymerization process
- Hydroformylation process

**TrendMiner**

1. Relevant patents
2. Language model
3. Representative terms extraction
4. Merge

**Patent Database**

**Discovered competitors:**
- Phillips Petroleum Company (score: 0.3672, # patents: 7189)
- Union Oil Company of California (score: 0.3324, # patents: 49)
- Marathon Oil Company (score: 0.2333, # patents: 249)
- Standard Oil Company (score: 0.2364, # patents: 810)
- General Electric Company (score: 0.1213, # patents: 13)
- Minnesota Mining and Manufacturing Company (score: 0.1136, # patents: 19)
- Eastman Kodak Company (score: 0.1369, # patents: 4)
- Chevron Research Company (score: 0.1047, # patents: 766)
- Rohm and Haas Company (score: 0.1029, # patents: 14)
- Atlantic Richfield Company (score: 0.0913, # patents: 777)
- The Standard Oil Company (score: 0.0643, # patents: 203)
- Halliburton Company (score: 0.0624, # patents: 803)
- Exxon Research and Engineering Company (score: 0.0621, # patents: 19)
- Chicago Bridge & Iron Company (score: 0.0602, # patents: 11)
Name Disambiguation

Three different people with "Yue Zhao"

- It is estimated that the 300 most common male names are used by more than 114 million people (taking about 78.74%) in the United States.
Name Disambiguation

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jing Zhang</td>
<td>Shanghai Jiao Tong Univ.</td>
</tr>
<tr>
<td></td>
<td>Yunnan Univ.</td>
</tr>
<tr>
<td></td>
<td>Tsinghua Univ.</td>
</tr>
<tr>
<td></td>
<td>Alabama Univ.</td>
</tr>
<tr>
<td></td>
<td>Univ. of California, Davis</td>
</tr>
<tr>
<td></td>
<td>Carnegie Mellon University</td>
</tr>
</tbody>
</table>

- How to perform the assignment automatically?
- How to estimate the person number?

Proposal of a semi-supervised framework
Name Disambiguation (cont.)

Table 5: Results for 41 names

<table>
<thead>
<tr>
<th>Name</th>
<th>Recall(%)</th>
<th>Precision(%)</th>
<th>F-score(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert Schreiber</td>
<td>72.81</td>
<td>100</td>
<td>84.27</td>
</tr>
<tr>
<td>Michael Smith</td>
<td>70.83</td>
<td>100</td>
<td>82.93</td>
</tr>
<tr>
<td>Hiroshi Tanaka</td>
<td>31.13</td>
<td>100</td>
<td>47.45</td>
</tr>
<tr>
<td>Satoshi Kobayashi</td>
<td>73.41</td>
<td>92.05</td>
<td>81.68</td>
</tr>
<tr>
<td>Philip J. Smith</td>
<td>88.83</td>
<td>100</td>
<td>91.21</td>
</tr>
<tr>
<td>David E. Goldberg</td>
<td>98.26</td>
<td>99.12</td>
<td>98.69</td>
</tr>
<tr>
<td>Yoshio Tanaka</td>
<td>85.48</td>
<td>94.64</td>
<td>89.83</td>
</tr>
<tr>
<td>Hui Yu</td>
<td>94.74</td>
<td>85.71</td>
<td>90.00</td>
</tr>
<tr>
<td>Feng Pan</td>
<td>56.55</td>
<td>84.37</td>
<td>70.72</td>
</tr>
<tr>
<td>Qiang Chen</td>
<td>90.37</td>
<td>100</td>
<td>94.94</td>
</tr>
<tr>
<td>Lei Jin</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Yang Yu</td>
<td>89.53</td>
<td>99.25</td>
<td>94.13</td>
</tr>
<tr>
<td>Ping Zhou</td>
<td>82.00</td>
<td>100</td>
<td>90.11</td>
</tr>
<tr>
<td>Rakesh Kumar</td>
<td>97.49</td>
<td>100</td>
<td>98.73</td>
</tr>
<tr>
<td>David Jensen</td>
<td>87.43</td>
<td>95.56</td>
<td>91.31</td>
</tr>
<tr>
<td>Thomas Wolf</td>
<td>33.33</td>
<td>89.36</td>
<td>48.55</td>
</tr>
<tr>
<td>Michael Lang</td>
<td>53.66</td>
<td>100</td>
<td>69.84</td>
</tr>
<tr>
<td>Thomas D. Taylor</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Manuel Silva</td>
<td>93.55</td>
<td>60.48</td>
<td>91.99</td>
</tr>
<tr>
<td>Charles Smith</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Koichi Furukawa</td>
<td>72.72</td>
<td>86.93</td>
<td>83.10</td>
</tr>
<tr>
<td>Thomas Tran</td>
<td>56.04</td>
<td>100</td>
<td>71.83</td>
</tr>
<tr>
<td>Thomas Hermann</td>
<td>71.30</td>
<td>100</td>
<td>83.25</td>
</tr>
<tr>
<td>Richard Taylor</td>
<td>67.82</td>
<td>100</td>
<td>80.82</td>
</tr>
<tr>
<td>Jim Gray</td>
<td>85.72</td>
<td>93.76</td>
<td>89.24</td>
</tr>
<tr>
<td>Juan Carlos Lopez</td>
<td>89.05</td>
<td>100</td>
<td>94.21</td>
</tr>
<tr>
<td>Sanjay Jain</td>
<td>97.74</td>
<td>100</td>
<td>98.86</td>
</tr>
<tr>
<td>Ajay Gupta</td>
<td>63.03</td>
<td>100</td>
<td>77.32</td>
</tr>
<tr>
<td>Shu lin</td>
<td>79.93</td>
<td>100</td>
<td>88.84</td>
</tr>
<tr>
<td>Michael Siegel</td>
<td>92.87</td>
<td>100</td>
<td>96.30</td>
</tr>
<tr>
<td>Eric Martin</td>
<td>98.31</td>
<td>100</td>
<td>99.15</td>
</tr>
<tr>
<td>Yun Wang</td>
<td>67.65</td>
<td>100</td>
<td>80.70</td>
</tr>
<tr>
<td>Kai Zhang</td>
<td>82.55</td>
<td>100</td>
<td>90.44</td>
</tr>
<tr>
<td>Cheong Chang</td>
<td>83.95</td>
<td>100</td>
<td>91.28</td>
</tr>
<tr>
<td>Daniel Massey</td>
<td>95.24</td>
<td>100</td>
<td>97.56</td>
</tr>
<tr>
<td>Pei Su</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Michael Wagner</td>
<td>84.54</td>
<td>83.97</td>
<td>89.01</td>
</tr>
<tr>
<td>David C. Wilson</td>
<td>89.50</td>
<td>100</td>
<td>94.46</td>
</tr>
<tr>
<td>Gang Luo</td>
<td>1</td>
<td>98.41</td>
<td>99.20</td>
</tr>
<tr>
<td>Bing Liu</td>
<td>93.88</td>
<td>88.99</td>
<td>91.36</td>
</tr>
<tr>
<td>R. Ramesh</td>
<td>68.12</td>
<td>100</td>
<td>81.04</td>
</tr>
</tbody>
</table>
Data source: crawled >1.5 M patents

Research:
- multi-patent summarization
- heterogeneous co-ranking
- topic discovery
- evolution analysis

System:
- patent search
- summarization
- topic trend analysis
- competitor analysis

Conclusion
Our Research Team

Bo Gao
Master Student

Bo Wang
PhD. candidate

Prof. Jie Tang

Fengjiao Wang
Undergraduate

Prof. Minlie Huang

Quan Lin
Master Student

Po Hu
PhD. student

Yi Yang
Undergraduate
Future Work

- **Data source**
  - to crawl energy papers and technique reports

- **Research**
  - Mining hierarchical topic structure
  - Identification of core patents, idea flow…
  - Evolutionary topic map discovery
  - Scalable trend analysis
  - Name disambiguation

- **System**
  - To deploy a real system
Thanks!
Thanks to Adam, Peng, Weichang, Limin, and all team members!!

Demo: http://166.111.134.87
HP: http://keg.cs.tsinghua.edu.cn/persons/tj/