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…

... functional for the electronic exchange **energy**: Generalized gradient approximation
JP Perdew, W Yue - Physical Review B, 1986 - APS
Accurate and simple density functional for the electronic exchange **energy**: Generalized gradient approximation. ... Accuracy is currently limited2-4 by approximations made for the exchange **energy** as a functional of the density. ...
Cited by 4480 - Related articles - All 7 versions

Accurate spin-dependent electron liquid correlation **energies** for local spin density ...
SH Vosko, L Wilk, M Nusair - Canadian Journal of ..., 1980 - article.pubs.nrc-cnrc.gc.ca
Abstract: We assess various approximate forms for the correlation **energy** per particle of the spin-polarized homogeneous electron gas that have frequently been used in applications of the local spin density approximation to the exchange-correlation **energy** functional. By ...
Cited by 5659 - Related articles - Cached - All 8 versions

Motivation (cont.)

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

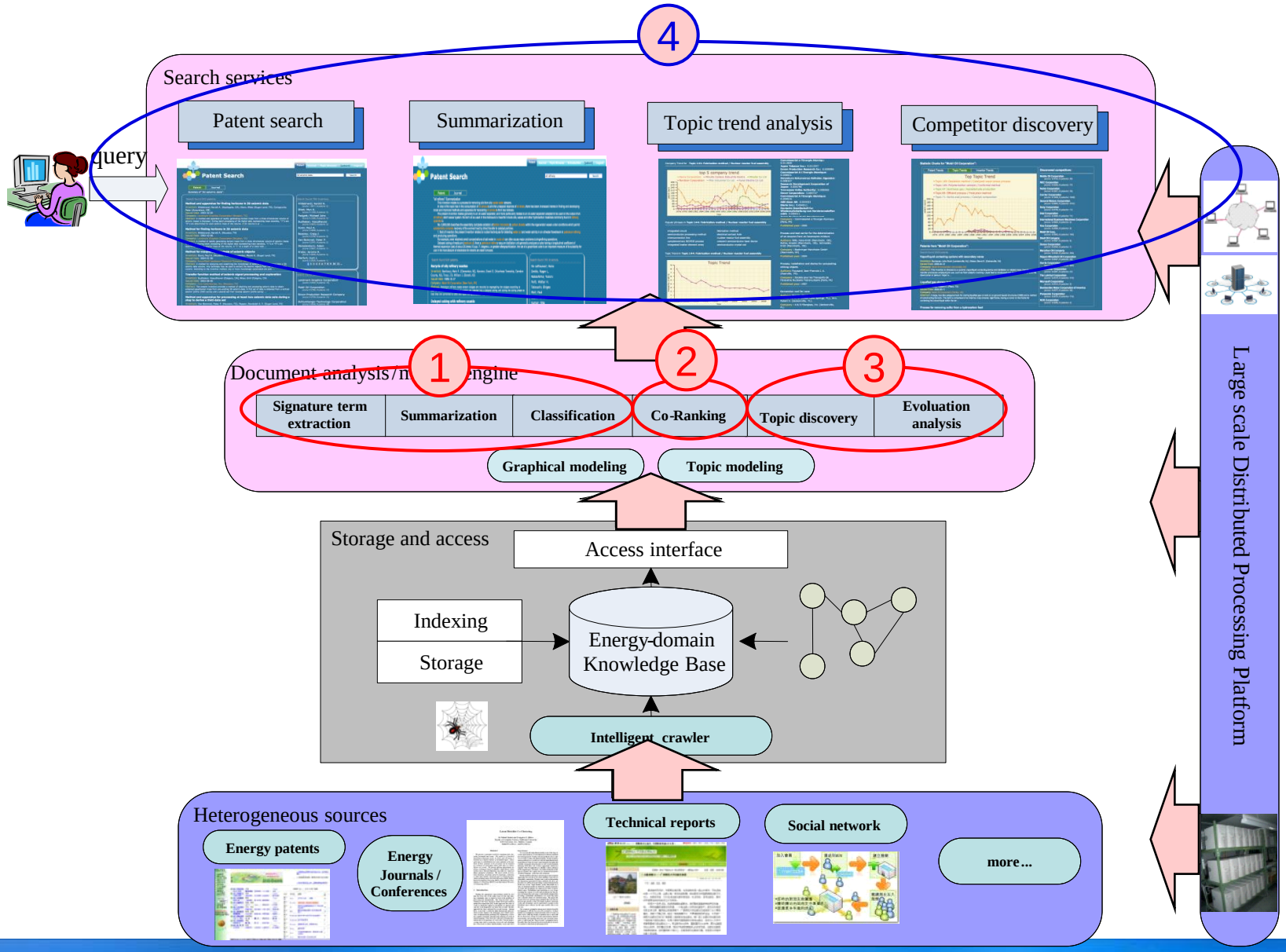
- What are the **"hot" topics** recent years?
- What are the **most influential** research work, researchers, and organizations in a specific topic?
- Which companies are our **competitors**?

…

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



Patent Search

Patent | Journal | Topic Browser | [patent] | Logout

Patent Search

3d seismic data | Search

Patent | Journal

Summary of " 3d seismic data" :

Although there are many advantages (some of which are set forth in the aforementioned published article of dobrin and ward) to recording seismic signals in digital form there still remains the need to make available to the seismic prospector avisible record of the seismic data, or portions of it.

Thetravel time of the wavefront across the array depends upon the seismic wave phase velocity, the wavelength of the acoustic wave, the array length and the depth of the reflector, among other factors.

This invention relates to seismic exploration, and, more particularly, to a seismic array with bidirectional capabilities and a method of seismic exploration.

It is well-known to those skilled in the art that the complex of vibrations received at a given seismic detector array do not consist totally of waves reflected from subsurface boundaries.

In such data processing operations, the digitized seismic data stored on magnetic tape is converted to seismic analog signals.

This information in turn is employed to evaluate the subsurface formations for their oil and gas bearing properties.

In such common depth point exploration it is the principal object of the geophysicist to obtain an accurate representation of the subsurface layering by interpreting between the common depth point or primary reflection signals and the variousnoise signals that are also recorded on the seismic traces.

Such a section is thus a display of wave form data utilizing time along a vertical axis, and distance (locations of the sources) along a horizontal axis.

Search found 952 patents.

Method and apparatus for finding horizons in 3D seismic data
Inventors: Hildebrand; Harold A. (Applegate, CA), Alam; Aftab (Sugar Land, TX), Caragounis; Peter (Amersham, GB)
Issued Date: 1993-10-05
Company: Landmark Graphics Corporation (Houston, TX)
Abstract: A method and apparatus of rapidly generating horizon maps from a three dimensional volume of seismic traces is disclosed. During batch processing of the digital data representing trace wavelets, "1"'s are determined for each seismic trace of the volume. A bit volume is pr ...

Method for finding horizons in 3D seismic data
Inventors: Hildebrand; Harold A. (Houston, TX)
Issued Date: 1992-10-06
Company: Landmark Graphics Corporation (Houston, TX)
Abstract: A method of rapidly generating horizon maps from a three dimensional volume of seismic traces is disclosed. During batch processing of the digital data representing trace wavelets, "1"'s or "0"'s are determined for each seismic trace of the volume. A "1" at a depth of a trace indic ...

Method for morphologic analysis of seismic objects
Inventors: Dunn; Paul A. (Houston, TX), Czernuszenko; Marek K. (Sugar Land, TX)
Issued Date: 2004-01-06
Company: ExxonMobil Upstream Research Company (Houston, TX)
Abstract: A method for analyzing and classifying the morphology of seismic objects extracted from a 3D seismic data volume. Any technique may be used to extract the seismic objects from the 3D seismic data volume. According to the inventive method, one or more morphologic parameters are sele ...

Transfer function method of seismic signal processing and exploration
Inventors: Sudhakar; Vasudhaven (Calgary, CA), Blias; Emil (Calgary, CA)
Issued Date: 2002-08-06
Company: Core Laboratories, Inc. (Houston, TX)
Abstract: The present invention provides a method of obtaining and processing seismic data to obtain enhanced geophysical maps from pre-existing 3D seismic data. A first set of data is obtained from a vertical seismic profile (VSP) survey and a second set from reverse seismic profile survey ...

Method and apparatus for processing at least two seismic data sets during a step to derive a third data set
Inventors: Van Bemmel; Peter P. (Houston, TX), Pepper; Randolph E. F. (Sugar Land, TX)
Issued Date: 1999-02-16
Company: GeoQuest, a division of Schlumberger Technology Corporation (Houston, TX)
Abstract: A processing system stores a special software known as the Cubemath software. The Cubemath software allows the processing system to process, analyze, and interpret a second set of seismic data relative to a first set of seismic data obtained from either a repeat seismic operation o ...

Apparatus and method for evaluation of picking horizons in 3-D seismic data
Inventors: Sitoh; Mun K. (Houston, TX)
Issued Date: 1996-07-16
Company: Landmark Graphics Corporation (Houston, TX)
Abstract: Apparatus and method are disclosed for use with a horizon picking method for 3D seismic data. Variables such as picking method, whether simple or iterative, score required of an iterative method and .DELTA.T between target wavelets and seed wavelet are input into the computer. Duri ...

Apparatus and method for evaluation of score failures in picking of 3-D seismic data
Inventors: Sitoh; Mun K. (Houston, TX)
Issued Date: 1997-10-07

Search found 786 inventors.

- Hildebrand; Harold A. (score: 1.8517, # patents: 2)
- Sitoh; Mun K. (score: 0.1258, # patents: 2)
- Padgett; Michael John (score: 0.1085, # patents: 4)
- Sudhakar; Vasudhaven (score: 0.0975, # patents: 2)
- Dunn; Paul A. (score: 0.0899, # patents: 1)
- Beasley; Craig J. (score: 0.0760, # patents: 6)
- Van Bemmel; Peter P. (score: 0.0741, # patents: 2)
- Gersztenkorn; Adam (score: 0.0601, # patents: 1)
- Krebs; Jerome R. (score: 0.0546, # patents: 4)
- Marfurt; Kurt J. (score: 0.0524, # patents: 1)

1 2 3 4 5 6 7 8 9 10 11 ...

Search found 388 companies.

- Landmark Graphics Corporation (score: 2.0446, # patents: 10)
- Mobil Oil Corporation (score: 0.2766, # patents: 31)
- Exxon Production Research Company (score: 0.2749, # patents: 31)
- Schlumberger Technology Corporation (score: 0.2556, # patents: 30)
- Western Atlas International, Inc. (score: 0.2491, # patents: 27)
- Amoco Corporation (score: 0.2140, # patents: 19)
- ExxonMobil Upstream Research Company (score: 0.1999, # patents: 12)
- Atlantic Richfield Company (score: 0.1384, # patents: 19)
- WesternGeco, L.L.C. (score: 0.1358, # patents: 14)
- WesternGeco L.L.C. (score: 0.1298, # patents: 15)

1 2 3 4 5 6 7 8 9 10 11 ...

Patents on "3d seismic data"

Summary

Inventor

Company

Patent Similarity and Recommendation

Patent | Journal | Topic Browser

Patent Search

Search

Method and system for processing geophysical survey data (7113868)

Issued Date: 2006-09-26
Inventors: Brewster; James (Lincoln Park, NJ)
Abstract: A method and system for processing geophysical survey data is provided. Gravity gradiometry data is processed by approximating density values of terrain signals to remove contributions of the terrain from the gravity gradiometry data. Initially, a predicted gravity gradient data set for a geographical area is calculated using terrain data with unit density. Next, a mathematical transformation is applied to recorded gravity gradiometer data of the geographical area and the predicted gravity gradiometry data. The transformation divides the data into component sub-functions so that each component function can be processed separately. The component sub-functions are processed to estimate densities of the geographical area. Then, using all the estimated densities, gravity gradients due to all of the terrain can be calculated, and subtracted from the recorded gravity gradient signals to reveal gravity gradients due to subsurface masses, for example. [link]

Patent information

Invention Class: Mixed

Classification

Attributes & Features:

Attributes	Features
a processor data storage data storage executable	system gravity gradiometer instrument ggi fulensor gradiometer ftg computer
	fixed geophysical parameter predicted geophysical data gravity gradiometer instruments gravity gradiometer instrument corresponding wavelet coefficients fulensor gradiometer ftg a computer readable medium central processing unit
dividing the geophysical data into components	surface density signal representative fulensor gravity gradient signals gravity gradiometer instrument fulensor gradiometer ftg recorded wavelet coefficients predicted wavelet coefficients a computer readable medium central processing unit

Extracted attributes and features

Similar Patents:
(7236885) Method and system for geophysical data acquisition on an airship Macfarlane; John
(5349527) Method of seismic time migration using a massively parallel computer Pieprzak; Andrew
(6467543) System and process for secondary hydrocarbon recovery Talwani; Manik
(6152226) System and process for secondary hydrocarbon recovery Talwani; Manik
(6954698) System and method for surveying underground density distributions Tryggvason; Bjarni V.
(4060718) Geodetic survey method Huddle; James R.
(6594585) Method of frequency domain seismic attribute generation Gersztenkorn; Adam
(6560537) Prospecting technique Matthews; Larry
(5530679) Method for migrating seismic data Albertin; Uwe K.
(5235554) Method for correcting impulse response differences of hydrophones and geophones as well as geophone coupling to the water-bottom in dual-sensor, bottom-cable seismic operations Barr; Frederick J.

Similar patents

Topic Trend Analysis

Topic 124: Magnetic transfer method / Disk apparatus

Inventors/Authors	Companies/Journals
Gill Hardayal Singh, Smith Gordon James, Lille Jeffrey S., Pinarbasi Mustafa Michael, Lin Tsann, Pinarbasi Mustafa,	Hitachi Global Storage Techn Read-Rite Corporation Hitachi Global Storage Techn Lattice Semiconductor Corpor

Topic 125: Air conditioning system / Fuel cell exhaust gas

Inventors/Authors	Companies/Journals
Breault Richard D., Porte Alain, Weiler Wolfgang, Aubry Jacques A., Liang George, Cronin Michael J.,	United Technologies Corporat Aerospatiale Societe National Thiokol Corporation Lockheed Corporation

Topic 126: Polymerization catalyst / Conformal method

Inventors/Authors	Companies/Journals
Ellis III Frampton E., Chang Main, Lundberg Robert D., Fung Shun C., Speca Anthony N., Soled Stuart L.,	Exxon Research and Engineer Exxon Research & Engineerin Amoco Corporation Standard Oil Company

Topic 127: Optical apparatus / Magnetic tape

Inventors/Authors	Companies/Journals
Sasaki Yoshitaka, Shiba Haruo, Yokoyama Kenji, Coon Warren, Deutsch Ralph, Hoshino Yoshihiro,	TDK Corporation Yamaha Corporation Yamaha Corporatio Nippon Gakki Seizo Kabushiki

200 topics have been discovered automatically from the patent data

Patent | Journal | Topic Browser | [patent] | Logout

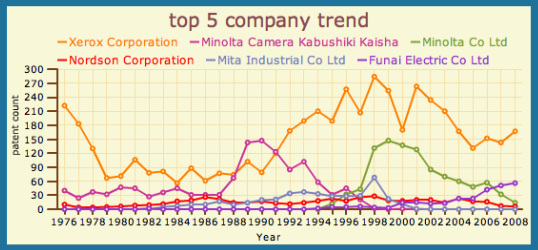
Patent Search

Search

Hot words in **Topic 144: Fabrication method / Nuclear reactor fuel assembly**

method:	0.041002	**apparatus:**	0.022601
device:	0.018508	**process:**	0.016586
nuclear:	0.014949	**system:**	0.014201
fuel:	0.011924	**reactor:**	0.010500
liquid:	0.005802	**assembly:**	0.005553
production:	0.005268	**measuring:**	0.004983
radioactive:	0.004841	**gas:**	0.004841
material:	0.004770	**high:**	0.004698
heat:	0.004022	**producing:**	0.003880
storage:	0.003809	**metal:**	0.003773

Company Trend for **Topic 144: Fabrication method / Nuclear reactor fuel assembly**



Active authors in **Topic 144: Fabrication method / Nuclear reactor fuel assembly**

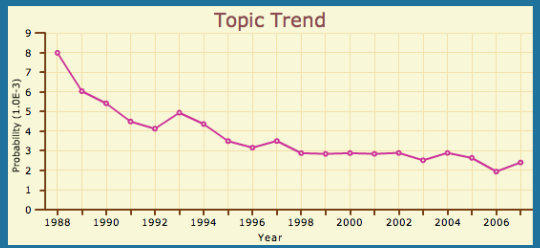
Show Top Sites

Lazzari Jean-Pierre	Schoening Josef	Wagner Anthony S.	Nylund Olov	Lang Fred D
Deleonibus Simon	Clerc Jean-Frederic	Galomb David E.	Jeuch Pierre	Bruel Michel
Lang Fred D.	Nagaoka Tadayoshi	Steiger Ronald P.	Perotto Christian	Erbstoesser Steven R.
Steingroever Erich	Leib Kenneth G.	McIntyre Frederic S.	Aronne Armand J.	Spohr Reimar

Popular phrases in **Topic 144: Fabrication method / Nuclear reactor fuel assembly**

integrated circuit	fabrication method
semiconductor processing method	electrical contact hole
interconnection line	nuclear reactor fuel assembly
complementary BiCMOS process	crescent semiconductor laser device
integrated heater element array	semiconductor crystal rod

Topic Trend in **Topic 144: Fabrication method / Nuclear reactor fuel assembly**



Hot conferences/jounals in **Topic 144: Fabrication method / Nuclear reactor fuel assembly**

Commissariat a l'Energie Atomique: 0.115066
Exxon Production Research Company: 0.030535
Grumman Aerospace Corporation: 0.027410
ExxonMobil Upstream Research Company: 0.014639
United Kingdom Atomic Energy Authority: 0.012207
Commissariat a l'Energie Atomiqu: 0.011808
Japan Tobacco Inc.: 0.011557
Exxon Production Research Co.: 0.009586
Commissariat A L'Energie Atomique: 0.008831
Doryokuro Kakunenryo Kaihatsu Jigyodan: 0.008327
Research Development Corporation of Japan: 0.006796
Tennessee Valley Authority: 0.006566
Dover Corporation: 0.005727
Commissariat A l'Energie Atomique: 0.004616
ABB Atom AB: 0.004532
Steris Inc.: 0.004511
Deutsche Gesellschaft fur Wiederaufarbeitung von Kernbrennstoffen mbH: 0.004511
Japan Nuclear Cycle Development Institute: 0.004280
Hochtemperatur-Reaktorbau Gmb: 0.004196
Siemens Power Corporation: 0.003924

Publications in **Topic 144: Fabrication method / Nuclear reactor fuel assembly**

Nuclear reactor fuel assembly
Authors: Feutrel; Claude (Vauhallan, FR).
Company : Commissariat a l'Energie Atomique (Paris, FR)
Published year: 1986

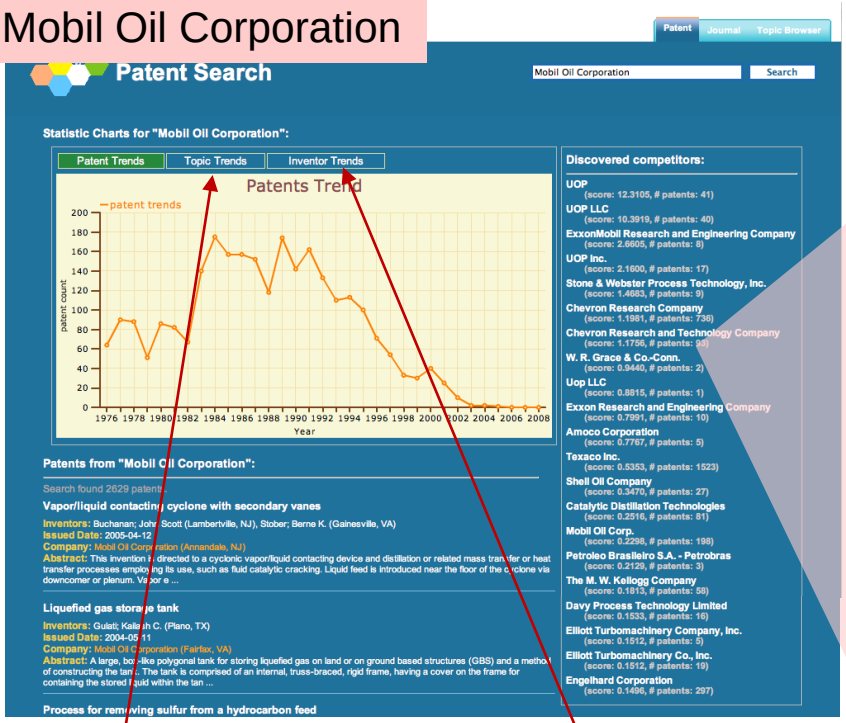
Process and test carrier for the determination of an enzyme from an isoenzyme mixture
Authors: Wilk; Hans-Erich (Mannheim, DE), Rothe; Anselm (Mannheim, DE), Schneider; Erich (Mannheim, DE).
Company : Boehringer Mannheim GmbH (Mannheim, DE)
Published year: 1994

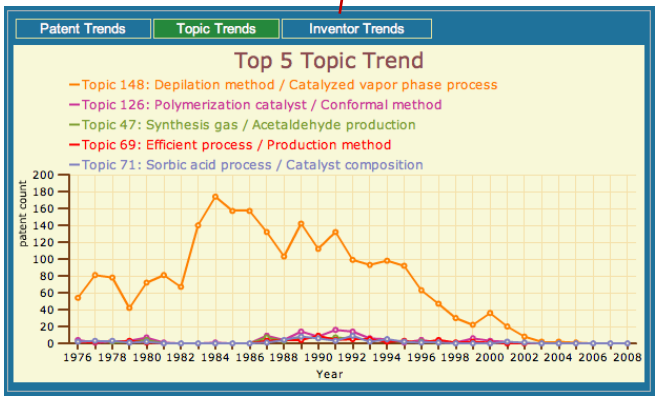
Process, installation and device for compacting oblong objects
Authors: Foussard; Jean-Francois J. A. (Ezanville, FR).
Company : Societe pour les Transports de l'Industrie Nucleaire Transnucleaire (Paris, FR)
Published year: 1987

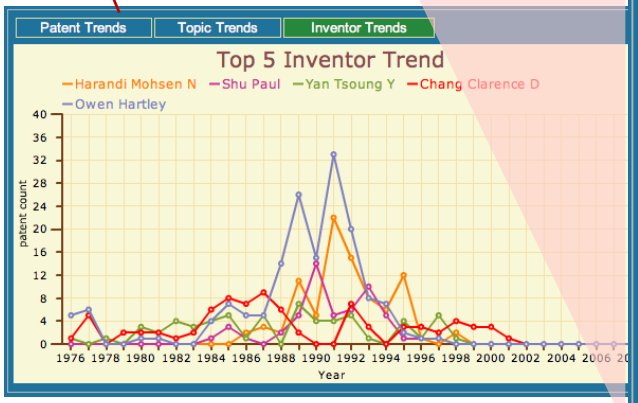
Conversion roof for vans
Authors: Borum; Otis C. (Silver Springs, FL), Borum; Martin O. (Green Cove Springs, FL), Arn; Robert H. (Jacksonville, FL).
Company : A & S Fiberglass, Inc. (Jacksonville, FL)

Competitor Discovery

Mobil Oil Corporation







Discovered competitors:

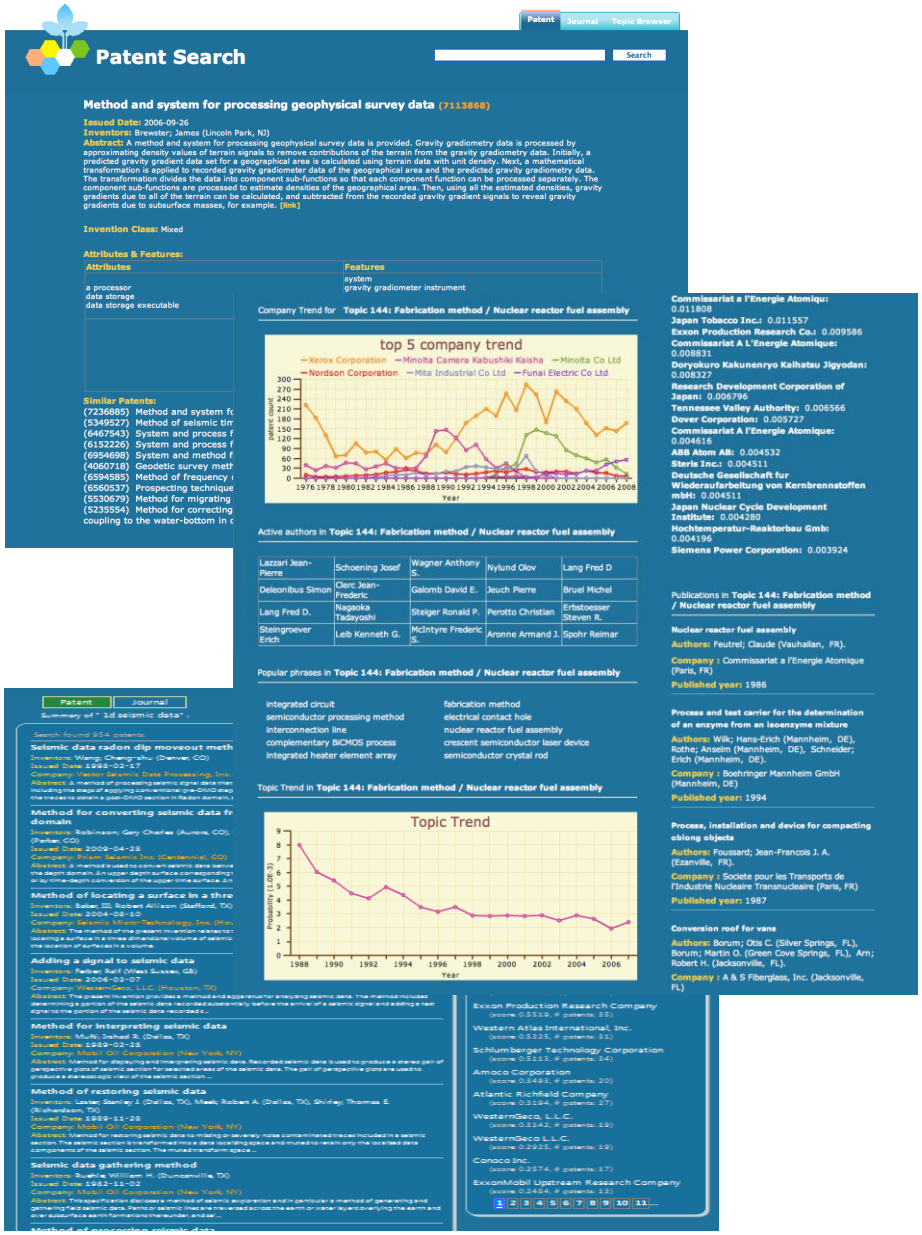
- **UOP** (score: 12.3105, # patents: 41)
- **UOP LLC** (score: 10.3919, # patents: 40)
- **ExxonMobil Research and Engineering Company** (score: 2.6605, # patents: 8)
- **UOP Inc.** (score: 2.1600, # patents: 17)
- **Stone & Webster Process Technology, Inc.** (score: 1.4683, # patents: 9)
- **Chevron 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: 0.7991, # patents: 10)
- **Amoco Corporation** (score: 0.7767, # patents: 5)
- **Texaco Inc.** (score: 0.5353, # patents: 1523)
- **Shell Oil Company** (score: 0.3470, # patents: 27)
- **Catalytic Distillation Technologies** (score: 0.2516, # patents: 81)
- **Mobil Oil Corp.** (score: 0.2298, # 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: 0.1533, # patents: 16)
- **Elliott Turbomachinery Company, Inc.** (score: 0.1512, # patents: 5)
- **Elliott Turbomachinery Co., Inc.** (score: 0.1512, # patents: 19)
- **Engelhard Corporation** (score: 0.1496, # 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

User Query

Search relevant patents

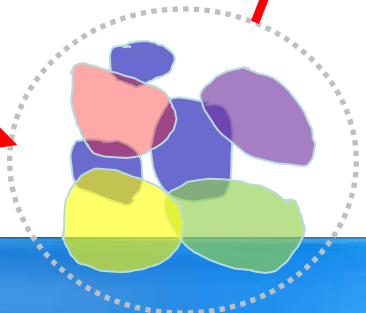
1

Relevant Patents

2

Knowledge space

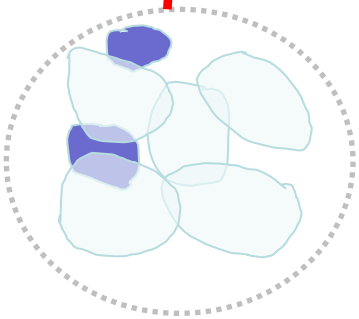
Extract concepts to represent the Knowledge Space



3

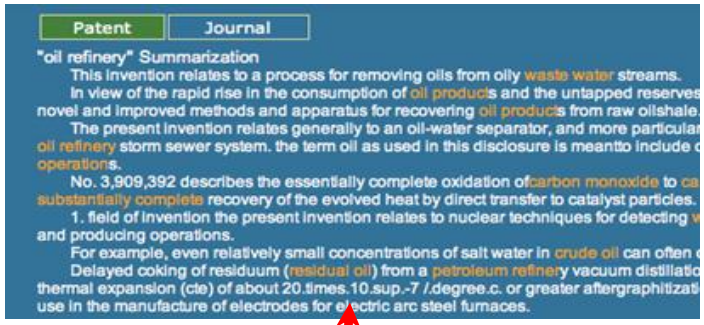
Score concept's importance to the query

Maximum Coverage Model



4

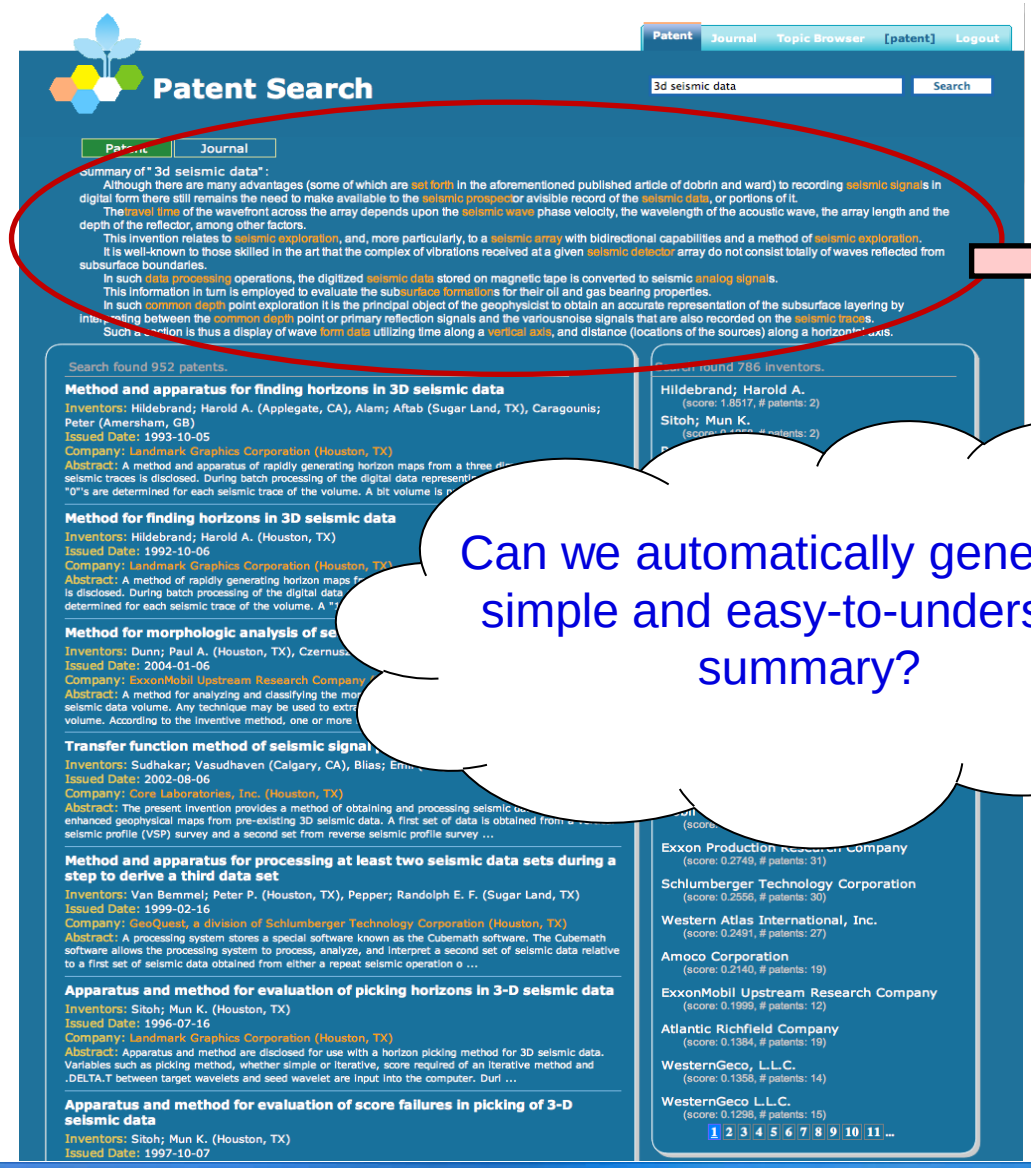
Solve by Integer Linear Programming



Diversified contents in limited length

12

Summary Rephrasing



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

Too long and too complex to understand!!

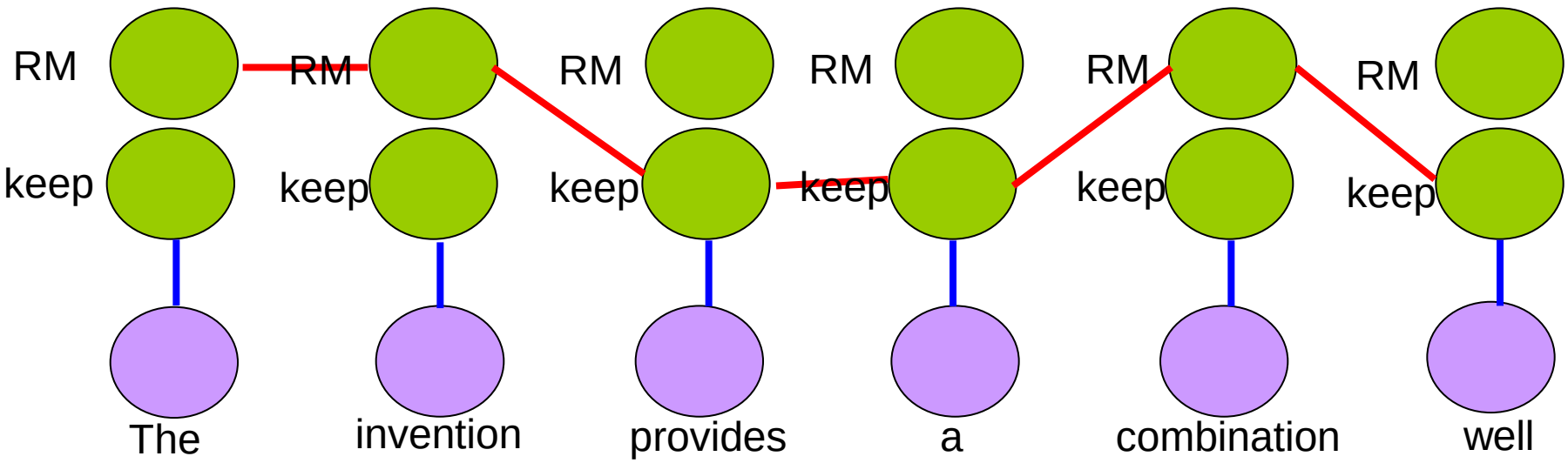
The invention provides a combination well logging device adapted to be moved inside a borehole to measure the properties of geological formations through which the borehole passes, characterized in that the device comprises a series of elongate logging sections mounted end to end, including a first section adapted to be maintained with a generatrix applied against the wall of borehole, a second section comprising a body and a measurement pad adapted to be spaced radially from the application against said wall, and a third section to be kept spaced from said wall, first and second ges 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.

Solution: rephrasing

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

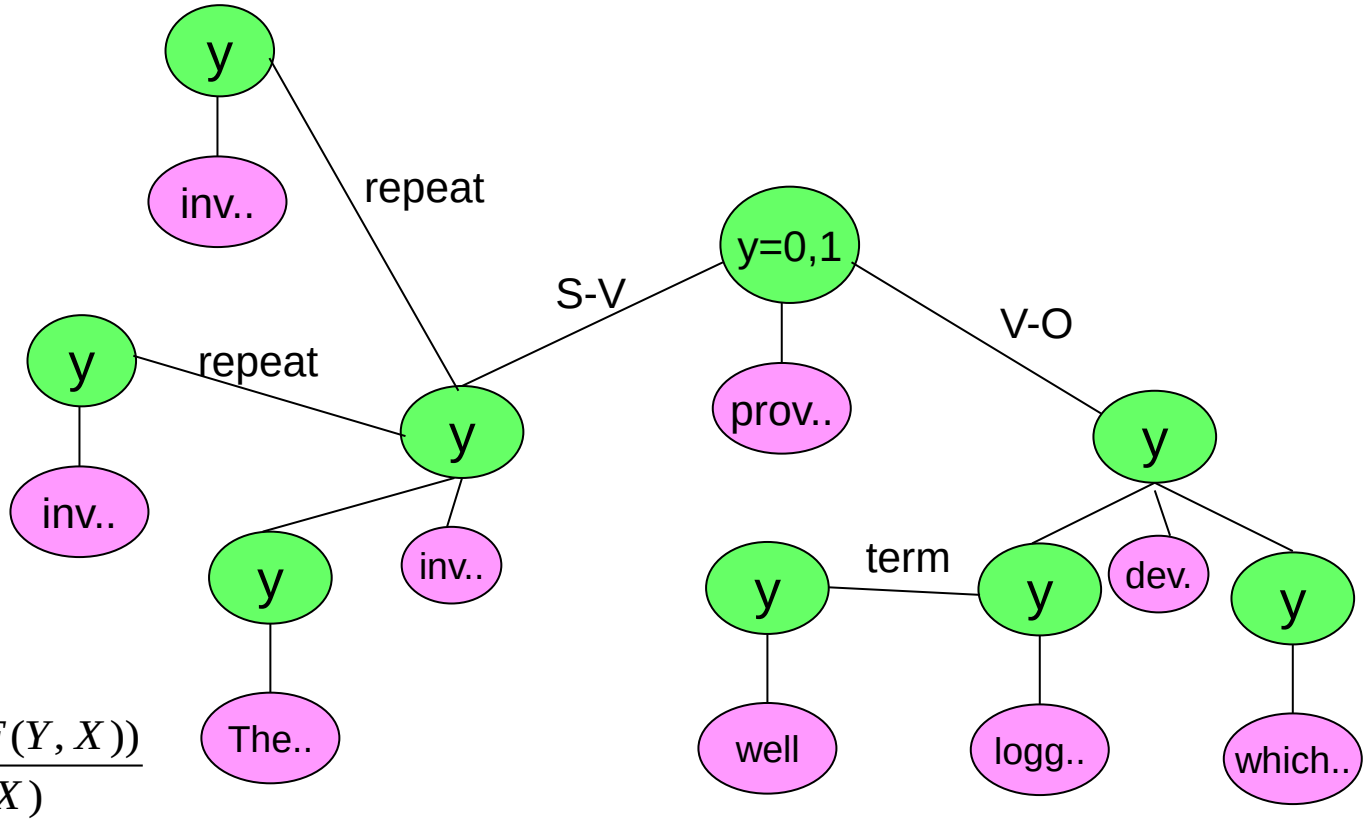
Solution: Modeling with CRFs

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



$$p(y \mid x) = \frac{1}{Z(x)} \exp\left(\sum_{e \in E, j} \lambda_j t_j(e, y|_e, x) + \sum_{v \in V, k} \mu_k s_k(v, y|_v, x) \right)$$

Beyond Linear-chain CRFs



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

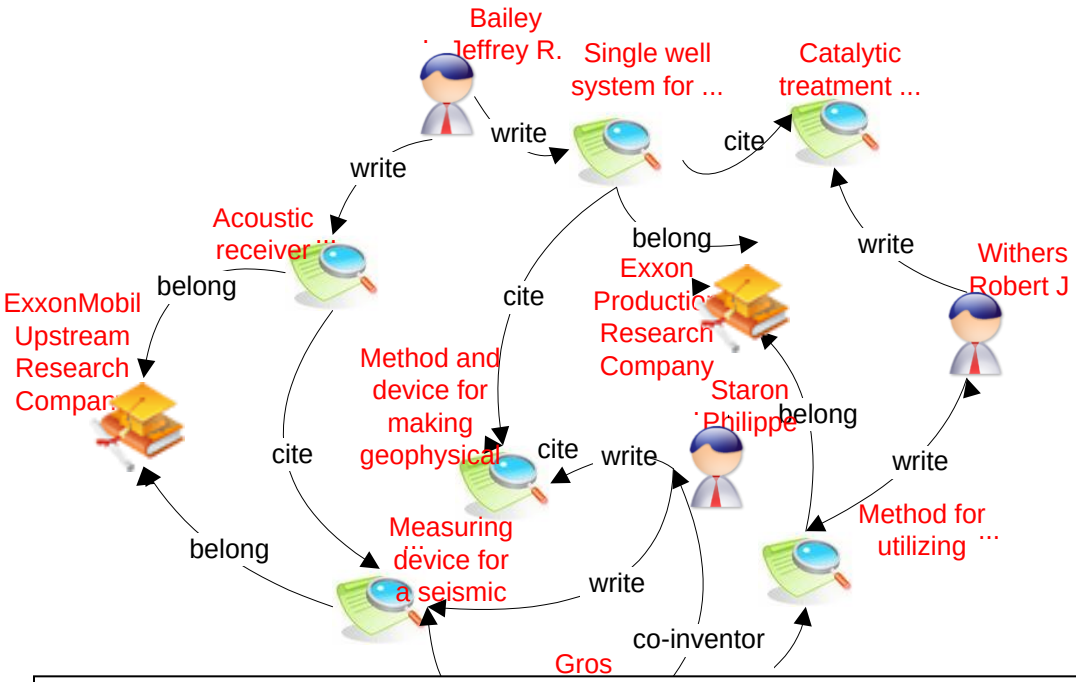
$$Z(x) = \sum_{y} \exp(\lambda \cdot F(y, x))$$

$$F(y,x) = \boxed{\textbf{Complex} \sum_{e \in \{E^{pc}, E^{ss}, E^{cp}\}}} \sum_j \lambda_j t_j(e, y_{|e}, x) + \sum_{v \in V} \sum_k \mu_k s_k(v, y_{|v}, x)$$

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



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

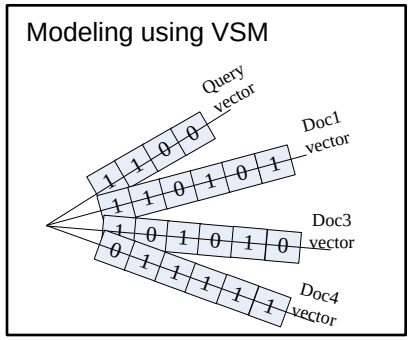
Challenges:

- How to extract the topics from the heterogeneous network?

- How to track the evolutionary pattern of different topics?

Topic-level Heterogeneous Co-Ranking

Search with keyword



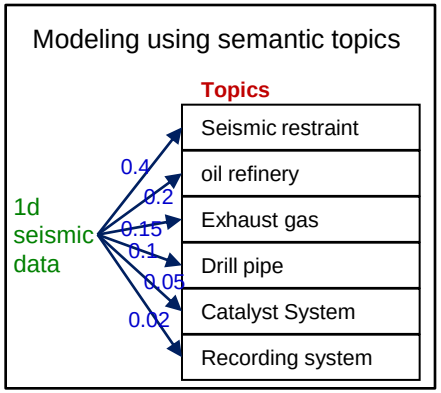
Return

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

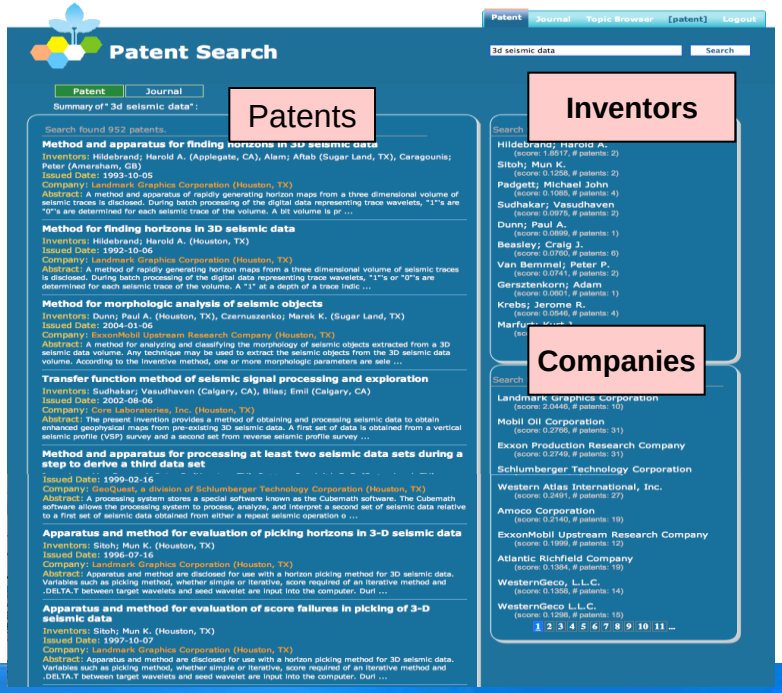
[Seismic data radon dip moveout method](#)
Wang; Cheng-shu - 1998 Vector Seismic Data Processing, …

[Method for converting seismic data from the …](#)
Robinson; Gary Charles - 2009…

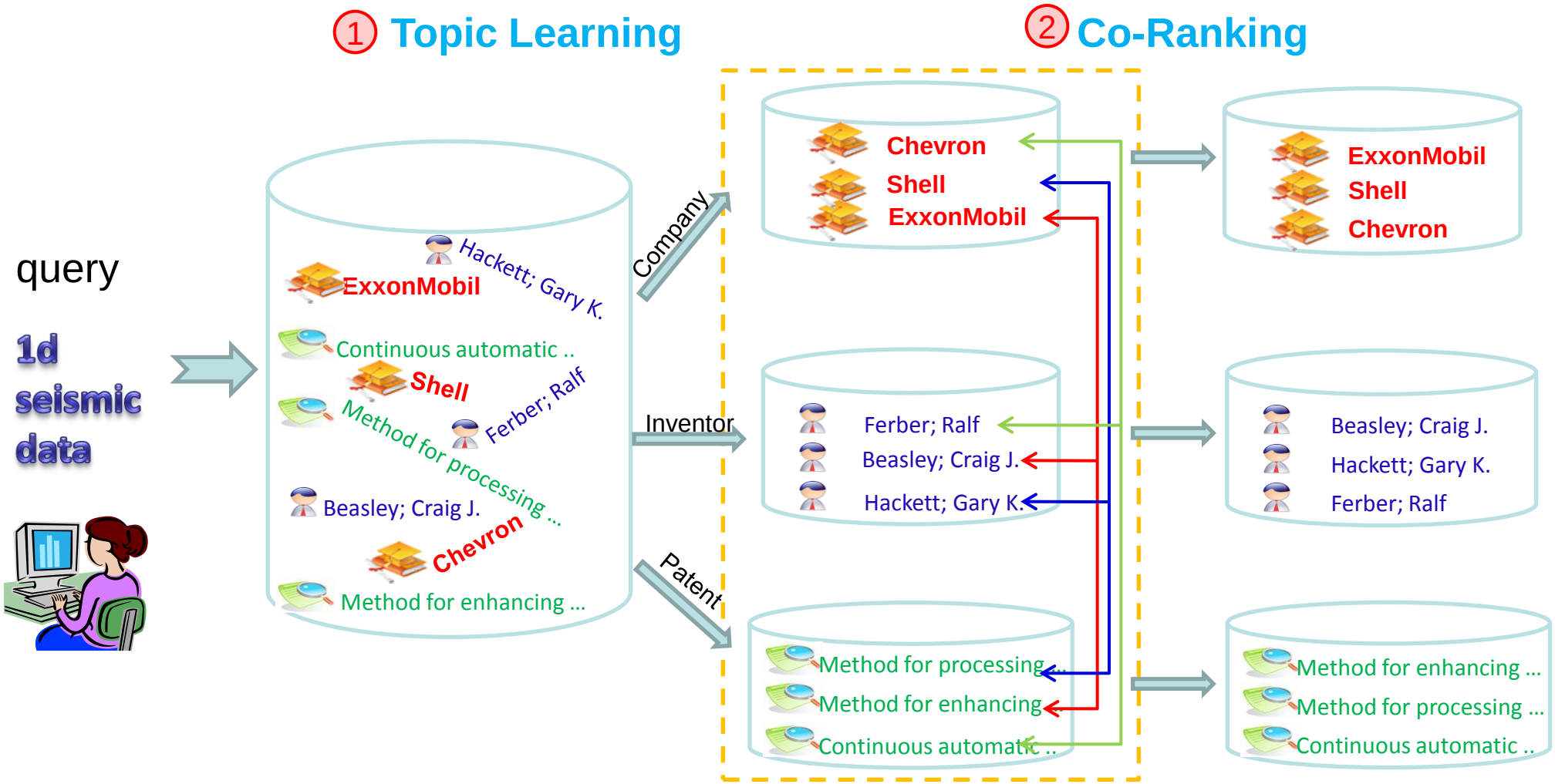
Search with semantic modeling



Return

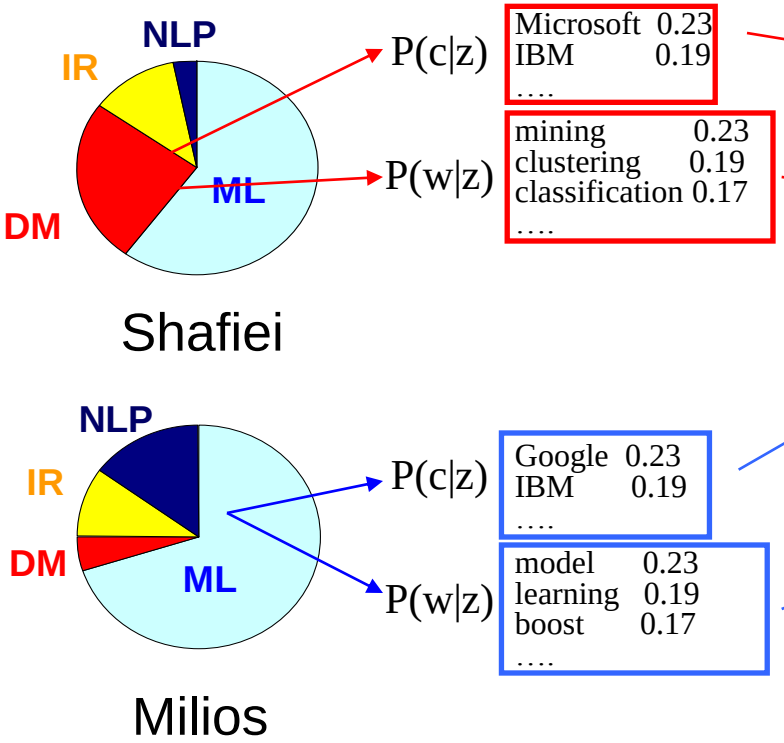


Heterogeneous Co-Ranking



Generative Story of ACT Model

- Generative process



Patent

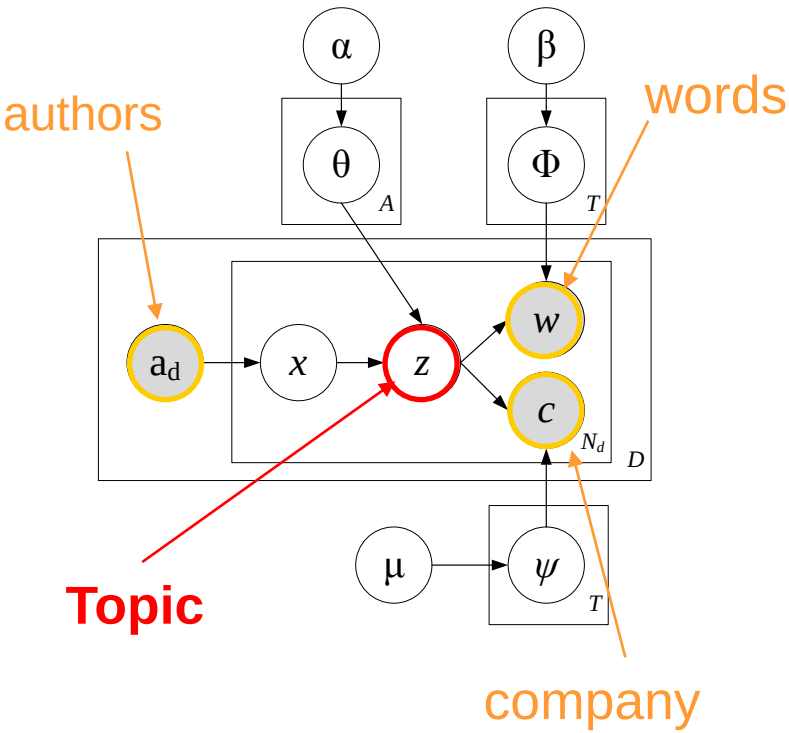
System and Method for Data Mining

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 …

ACT Model



Generative process:

1. For each topic z, draw ϕ_z and ψ_z respectively from Dirichlet priors β_z and μ_z;
2. For each word w_{di} in document d:
   - draw an author x_{di} from $\mathbf{a}_d$ uniformly;
   - draw a topic z_{di} from a multinomial distribution $\theta_{x_{di}}$ specific to author x_{di}, where θ is generated from a Dirichlet prior α;
   - draw a word w_{di} from multinomial $\phi_{z_{di}}$;
   - draw a company stamp c_{di} from multinomial $\psi_{z_{di}}$.

$$P(z_{di}, x_{di} | \mathbf{z}_{-di}, \mathbf{x}_{-di}, \mathbf{w}, \mathbf{c}, \alpha, \beta, \mu) \propto$$

$$\frac{m^{-di}_{x_{di} z_{di}} + \alpha_{z_{di}}}{\sum_z (m^{-di}_{x_{di} z} + \alpha_z)} \frac{n^{-di}_{z_{di} w_{di}} + \beta_{w_{di}}}{\sum_{w_v} (n^{-di}_{z_{di} w_v} + \beta_{w_v})} \frac{n^{-d}_{z_{di} c_d} + \mu_{c_d}}{\sum_c (n^{-d}_{z_{di} c} + \mu_c)}$$

Author-Company-Topic

ACT Model (Cont)

- After Gibbs sampling, we can easily estimate the probability of a topic given an author θ_{xz}, the probability of a word given a topic φ_{zv}, *and the probability* of a company given a topic ψ_{zc}:

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

$$\psi_{zc_d} = \frac{n_{zc_d} + \mu_{c_d}}{\sum_c(n_{zc} + \mu_c)}$$

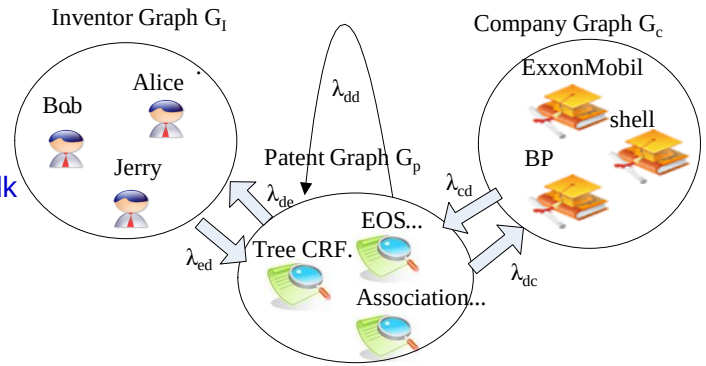
$$\theta_{xz} = \frac{m_{xz} + \alpha_z}{\sum_{z'}(m_{xz'} + \alpha_{z'})}$$

Topic-level Heterogeneous Co-Ranking

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

Stage 1:

Random walk



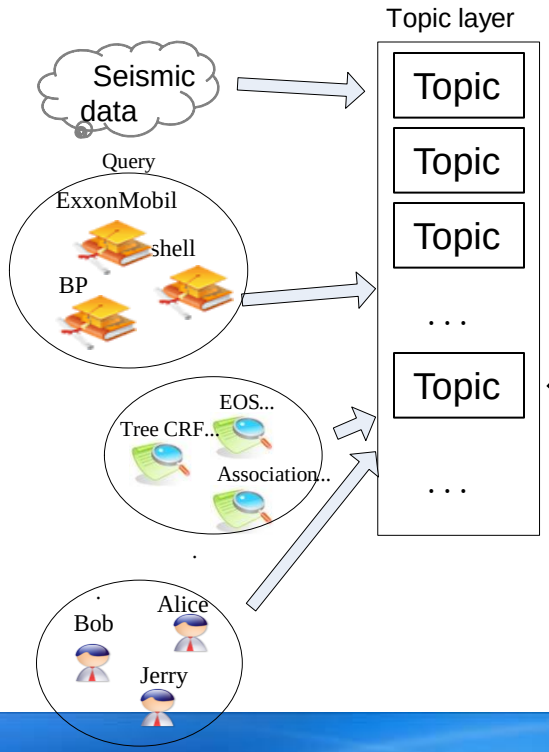
Ranking score

Combination by multiplication

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

Stage 2.

Topic-based relevance



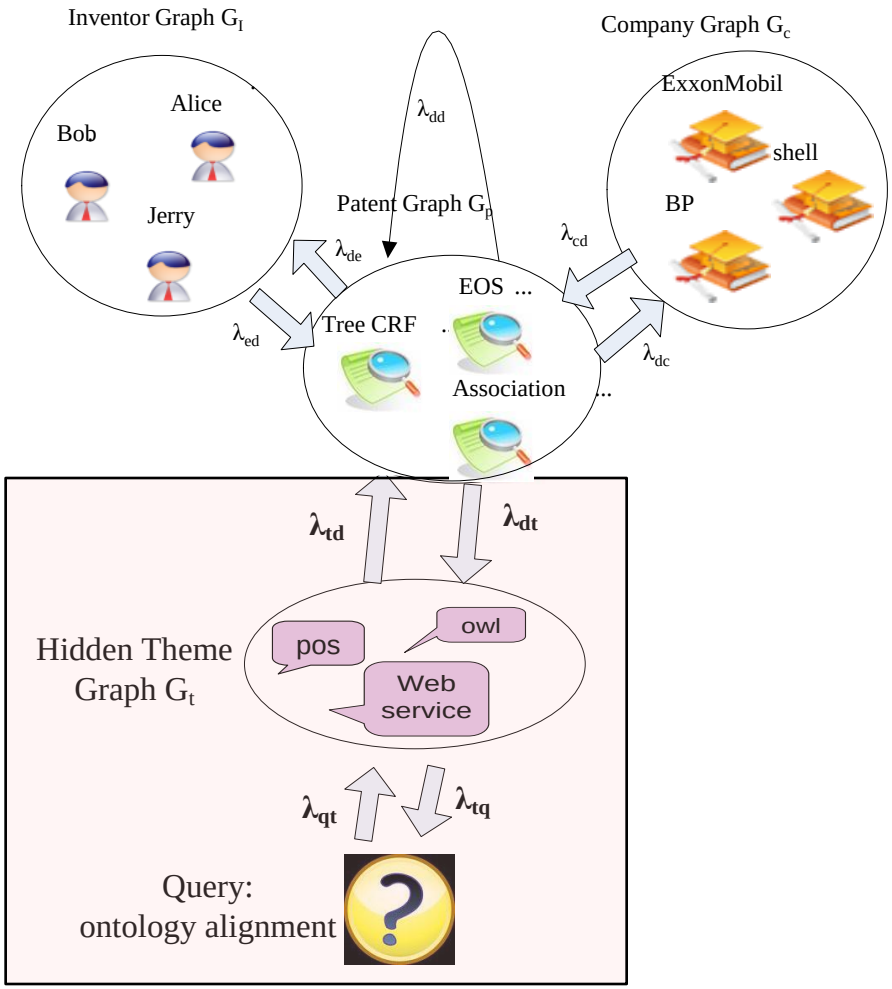
Topic-based relevance score

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

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

$$P(q|d) = P_{LM}(q|d) * P_{ACT}(q|d)$$

Topic-level Heterogeneous Co-Ranking (2)



Ranking score

$$r(d_i) = \lambda_{dd} \times \sum_{(d_j,d_i)\in E_{dd}} \boxed{p(d_i \mid d_j)} \cdot r(d_j) + \lambda_{ed} \times \sum_{(e_i,d_i)\in E_{de}} \boxed{P(d_i \mid e_i)} \cdot r(e_i)$$

$$+ \lambda_{cd} \times \sum_{(c_i,d_i)\in E_{dc}} \boxed{P(d_i \mid c_i)} \cdot r(c_i) + \lambda_{td} \times \sum_{(\theta_i,d_i)\in E_{dt}} \boxed{P(d_i \mid \theta_i)} \cdot r(\theta_i)$$

$$r(e_i) = \lambda_{de} \times \sum_{(d_j,e_i)\in E_{de}} \boxed{p(e_i \mid d_j)} \cdot r(d_j)$$

$$r(c_i) = \lambda_{dc} \times \sum_{(d_j,c_i)\in E_{dc}} \boxed{p(c_i \mid d_j)} \cdot r(d_j)$$

Transition probability

$$P(z_i|a_j) = \theta_{a_j z_i}$$

$$P(a_j|z_i) = \frac{P(z_i|a_j)P(a_j)}{P(z_i)}$$

$$P(z_i|d_j) = \frac{1}{A_d} \sum_{x \in \mathbf{a}_d} \theta_{x z_i}$$

$$P(d_j|z_i) = P(\mathbf{w}_d|z_i) = \prod_{i=1}^{N_d} P(w_{di}|z_i)$$

$$P(c_j|z_i) = \psi_{z_i c_j}$$

$$P(z_i|c_j) = \frac{P(c_j|z_i)P(z_i)}{P(c_j)}$$

$$P(q|z_i) = \prod_{w \in q} P(w|z_i)$$

$$P(z_i|q) \propto P(q|z_i)P(z_i)$$

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)

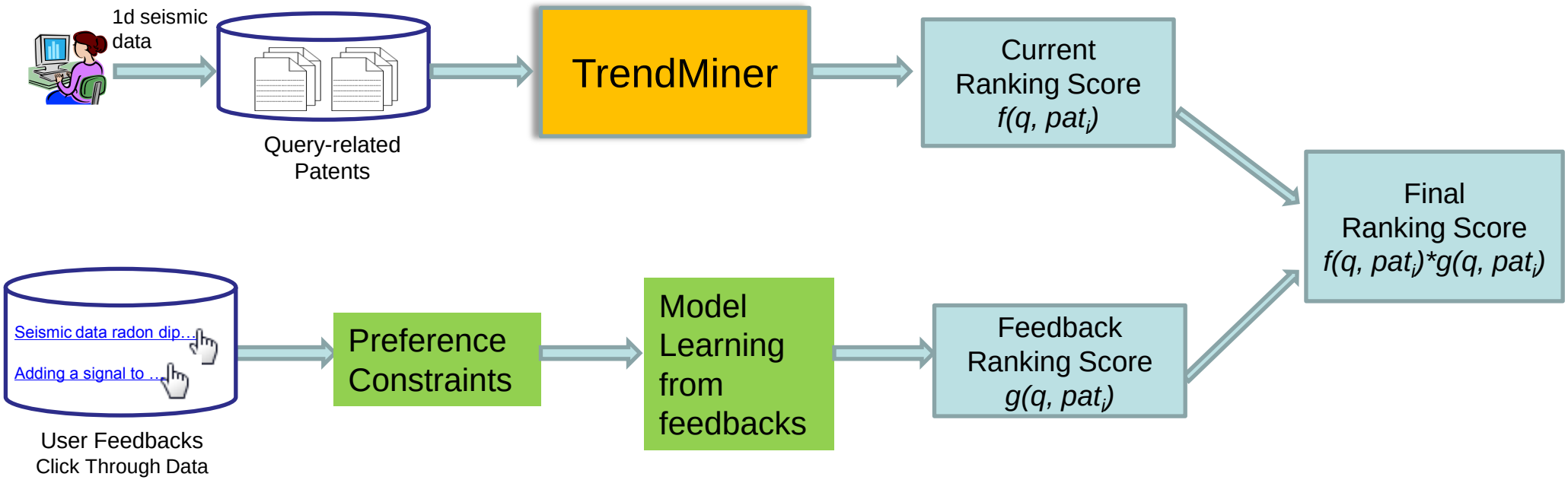
for a given query

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



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)

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

- 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

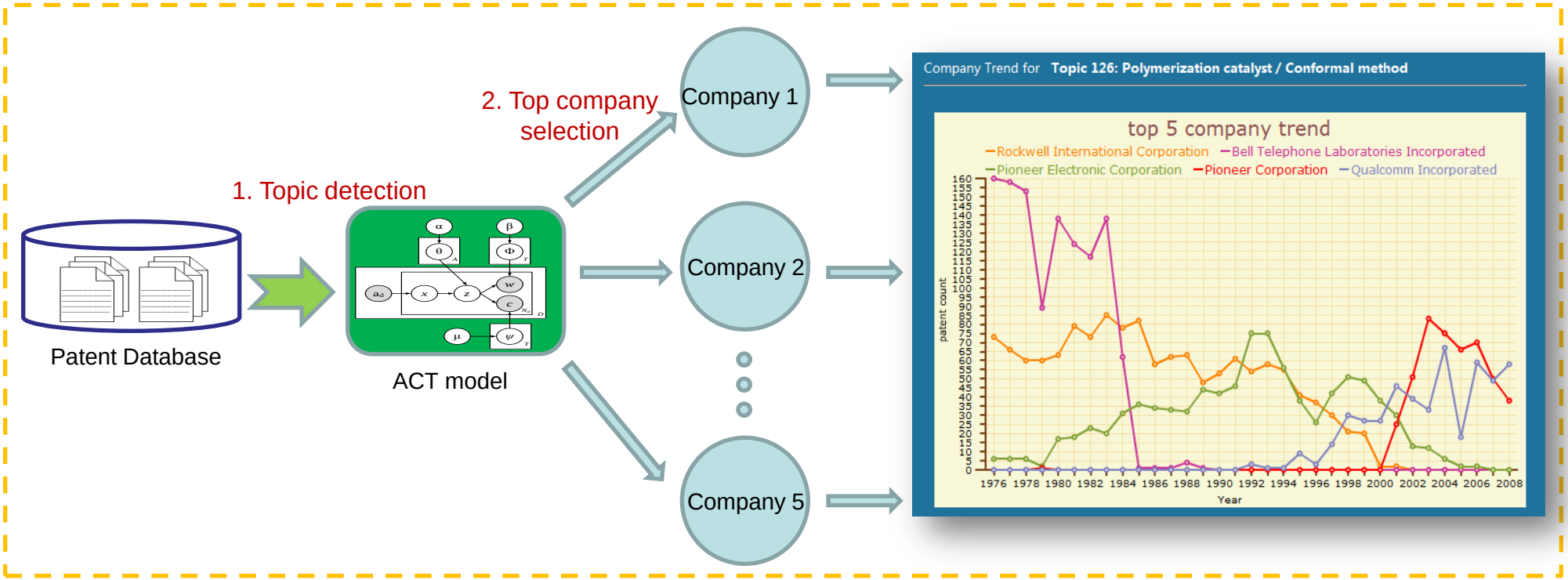
Score (*q*, pat_i) = g(*q*, pat_i) * f(*q*, pat_i)

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?

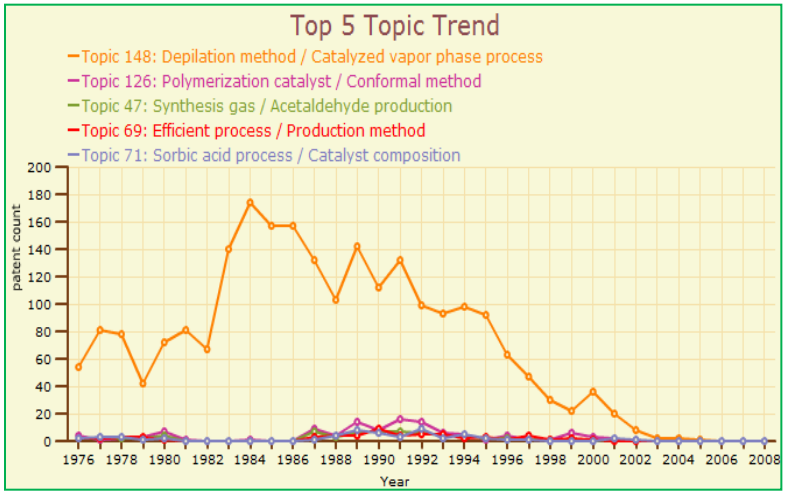


Topic Evolution Analysis II

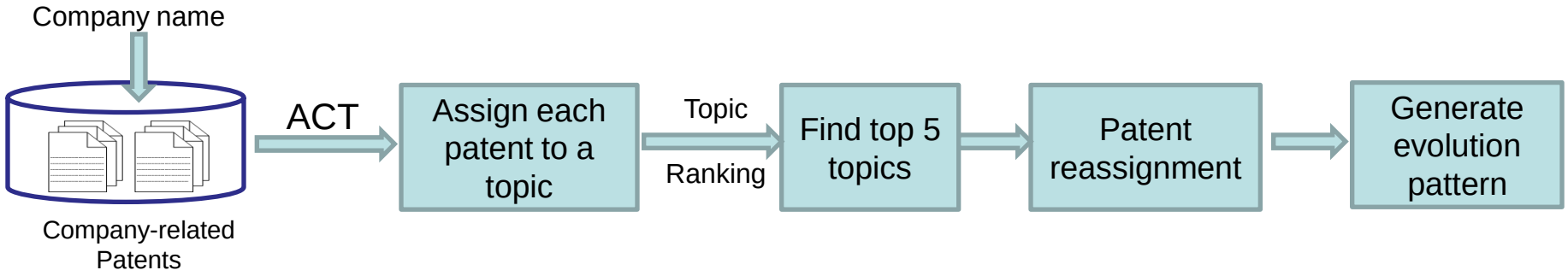
- Motivation:
  - What are the top topics for a company?
  - How does the topics of a company evolve?

- Example:

Mobil Oil Corporation → TrendMiner →



- Procedure:



Competitor Analysis

Shell Oil Company

Representative terms

Polymerization process
Hydroformylation process

Discovered competitors:

- Phillips Petroleum Company (score: 0.3672, # patents: 2189)
- Union Oil Company of California (score: 0.3524, # patents: 49)
- Marathon Oil Company (score: 0.2633, # patents: 249)
- Standard Oil Company (score: 0.2366, # patents: 580)
- General Electric Company (score: 0.1212, # patents: 15)
- Minnesota Mining and Manufacturing Company (score: 0.1136, # patents: 16)
- Eastman Kodak Company (score: 0.1129, # patents: 14)
- Chevron Research Company (score: 0.1047, # patents: 736)
- Rohm and Haas Company (score: 0.1029, # patents: 14)
- Atlantic Richfield Company (score: 0.0913, # patents: 977)
- The Standard Oil Company (score: 0.0832, # patents: 381)
- Halliburton Company (score: 0.0824, # patents: 801)
- Exxon Research and Engineering Company (score: 0.0821, # patents: 13)
- Chicago Bridge & Iron Company (score: 0.0803, # patents: 11)

Company name

TrendMiner

Patent Database

1 Relevant patents

Language model

2 Representative terms extraction

3 ACT

3 ACT

4 Merge

Name Disambiguation

Search "Yue Zhao"

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

Name	Affiliation
Jing Zhang (26)	Shanghai Jiao Tong Univ.
	Yunnan Univ.
	Tsinghua Univ.
	Alabama Univ.
	Univ. of California, Davis
	Carnegie Mellon University

Jing Zhang

List of publications from the DBLP Bibliography Server - FAQ

Coauthor Index - Ask others: ACM DL - ACM Guide - CiteSeer - CSB - Google

2007

83 EE Jing Zhang, Guizhong Liu: Hyperspectral images lossless compression by a novel three-dimensional wavelet coding
82 EE Jing Peng, Dongqing Yang, Changjie Tang, Jing Zhang, Jianjun Hu: CACS: A Novel Classification Algorithm Base
81 EE Jing Zhang, Xi Chen, Ming Li: Computing Exact p-Value for Structured Motif. CPM 2007: 162-172
80 EE Jing Zhang, Jie Tang, Juan-Zi Li: Expert Finding in a Social Network. DASFAA 2007: 1066-1069
79 EE Guojun Chen, Jing Zhang, Xiaoli Xu, Yuan Yin: Real-Time Visualization of Tire Tracks in Dynamic Terrain with LO
78 EE Jing Zhang, Hai Huang: Federate Job Mapping Strategy in Grid-Based Virtual Wargame Collaborative Environmer
77 EE Maria Wimmer, Michael Goul, Jing Zhang: Minitrack: E-Government Information and Knowledge Management. H
76 EE Kai Kang, Jing Zhang, Baoshan Xu: Optimizing the Selection of Partners in Collaborative Operation Networks. IC
75 EE Lingshuang Shao, Jing Zhang, Yong Wei, Junfeng Zhao, Bing Xie, Hong Mei: Personalized QoS Prediction forWe
Training of a Kernel-Based Representative and Discriminative
n for Peer-to-Peer Network Management. International Confer
3D Motion Segmentation from Straight-Line Optical Flow. M

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

$$p(y \mid x) = \frac{1}{Z(x)} \exp\left(\sum_{i,k} D(x_i, y_k) + \sum_{i,j \neq i} \{ D(x_i, x_j) \sum_{c_k \in C} [w_k c_k(x_i, x_j)] \} \right)$$

R	W	Relation Name	Description
r_1	w_1	Co-Conference	$p_i.pubvenue = p_j.pubvenue$
r_2	w_2	Co-Author	$\exists\, r, s>0,\ a_i^{(r)}=a_j^{(s)}$
r_3	w_3	Citation	p_i cites p_j or p_j cites p_i
r_4	w_4	Constraints	Feedbacks supplied by users
r_5	w_5	τ-CoAuthor	τ-extension co-authorship (τ>1)

Proposal of a semi-supervised framework

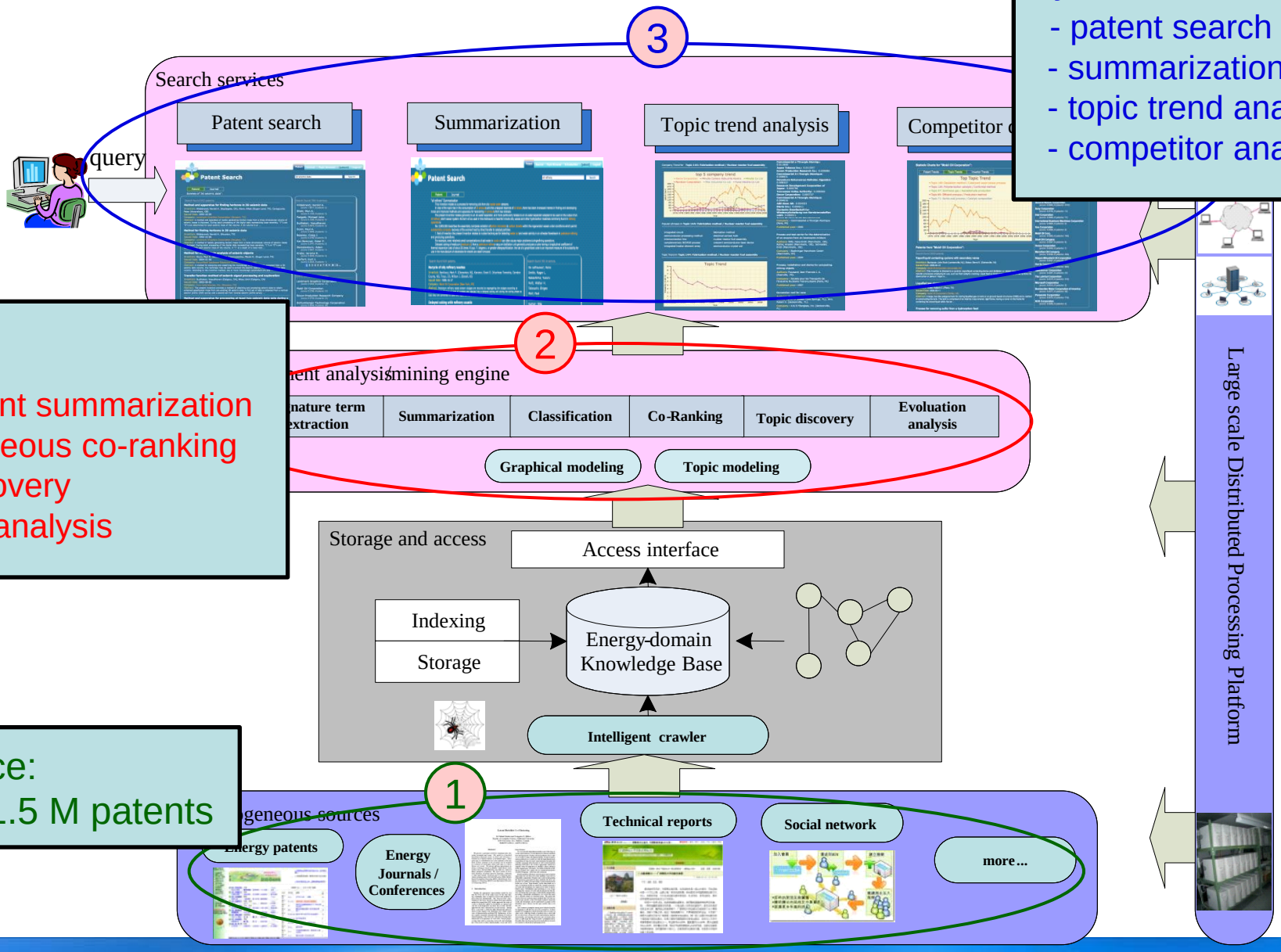
Name Disambiguation (cont.)

Table 5: Results for 41 names

Name	Recall(%)	Precision(%)	F-score(%)
Robert Schreiber	72.81	100	84.27
Michael Smith	70.83	100	82.93
Hiroshi Tanaka	31.13	100	47.48
Satoshi Kobayashi	73.41	92.05	81.68
Philip J. Smith	83.83	100	91.21
David E. Goldberg	98.26	99.12	98.69
Yoshio Tanaka	85.48	94.64	89.83
Hui Yu	94.74	85.71	90.00
Feng Pan	56.55	94.37	70.72
Qiang shen	90.37	100	94.94
Lei Jin	100	100	100
Yang Yu	89.53	99.25	94.13
Ping Zhou	82.00	100	90.11
Rakesh Kumar	97.49	100	98.73
David Jensen	87.43	95.56	91.31
Thomas Wolf	33.33	89.36	48.55
Michael Lang	53.66	100	69.84
Thomas D. Taylor	100	100	100
Manuel Silva	93.55	90.48	91.99
Charles Smith	100	100	100
Koichi Furukawa	72.72	96.93	83.10
Thomas Tran	56.04	100	71.83
Thomas Hermann	71.30	100	83.25
Richard Taylor	67.82	100	80.82
Jim Gray	85.72	93.76	89.24
Juan Carlos Lopez	89.05	100	94.21
Sanjay Jain	97.74	100	98.86
Ajay Gupta	63.03	100	77.32
Shu lin	79.93	100	88.84
Michael Siegel	92.87	100	96.30
Eric Martin	98.31	100	99.15
Yun Wang	67.65	100	80.70
Kai Zhang	82.55	100	90.44
Cheng Chang	83.95	100	91.28
Daniel Massey	95.24	100	97.56
Fei Su	100	100	100
Michael Wagner	84.54	93.97	89.01
David C. Wilson	89.50	100	94.46
Gang Luo	1	98.41	99.20
Bing Liu	93.88	88.99	91.36
R. Ramesh	68.12	100	81.04

Conclusion

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



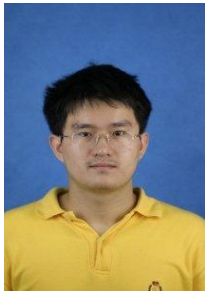
Research:
- multi-patent summarization
- heterogeneous co-ranking
- topic discovery
- evolution analysis

Data source:
crawled >1.5 M patents

Our Research Team

Bo Wang
PhD. candidate

Bo Gao
Master Student



Po Hu
PhD. student

Prof. Jie Tang

Prof. Minlie Huang

Fengjiao Wang
Undergraduate

Quan Lin
Master 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/