KDD 2012 BEIJING
THE 18TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING
Beijing, China August 12-16, 2012

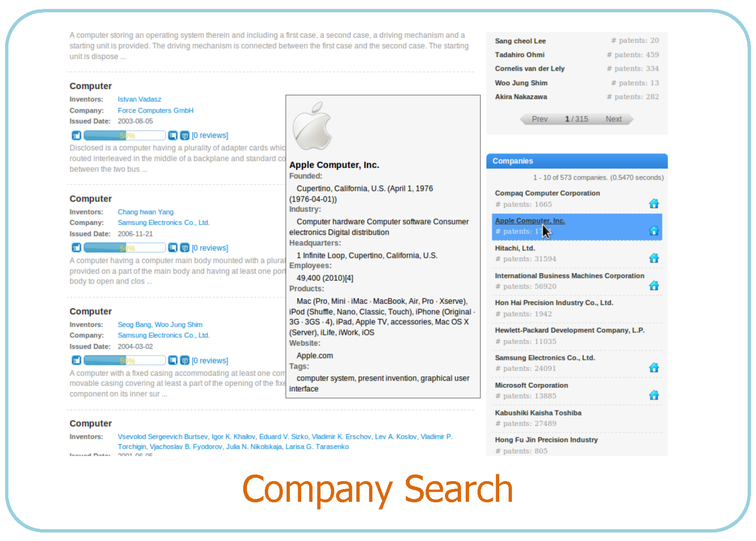
PatentMiner: Topic-dirven Patent Analysis and Mining

**Jie Tang, BoWang, Yang Yang, Po Hu, Yanting Zhao, Xinyu Yan, Bo Gao,
Minlie Huang, Peng Xu, Weichang Li, and Adam K. Usadi
Department of Computer Science, Tsinghua University, China
ExxonMobil Research and Engineering Company, USA**

Introduction

PatentMiner is a free online service used for analyzing and mining patent networking data. By now, we collected 8,000,000+ patents, 400,000+ companies, 2,000,000+ inventors. PatentMiner provides in-depth topic-level analysis functions.

System Overview



Company Search

Major functions in PatentMiner:

❑Patent Search

Find prolific inventors, top company, and best patent

❑Company Analysis

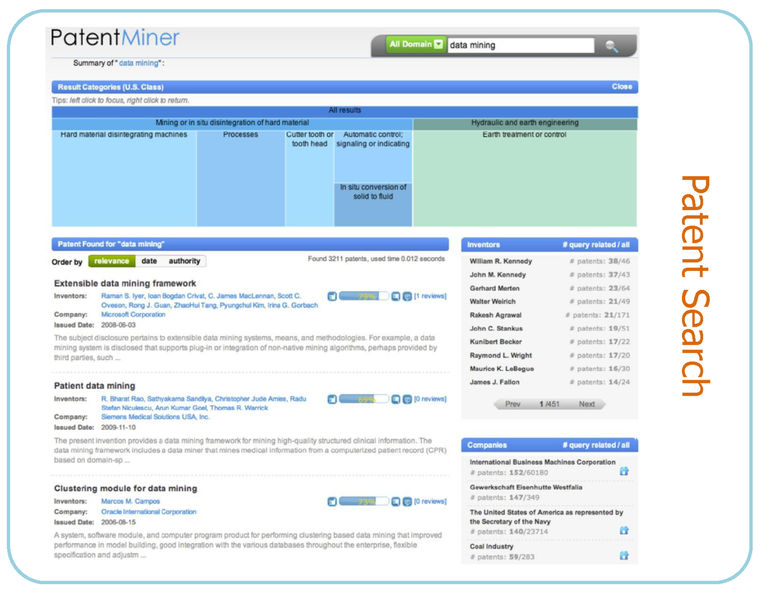
Extract company basic information, and discover topic evolution

❑Topic Cataloging

Model patents and companies with mixture topic distribution

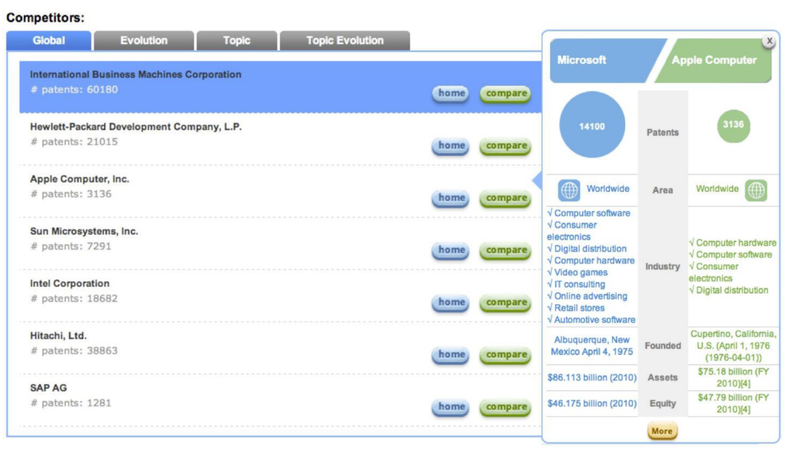
❑Competitive Analysis

Identify your competitors and analyze their technology evolution

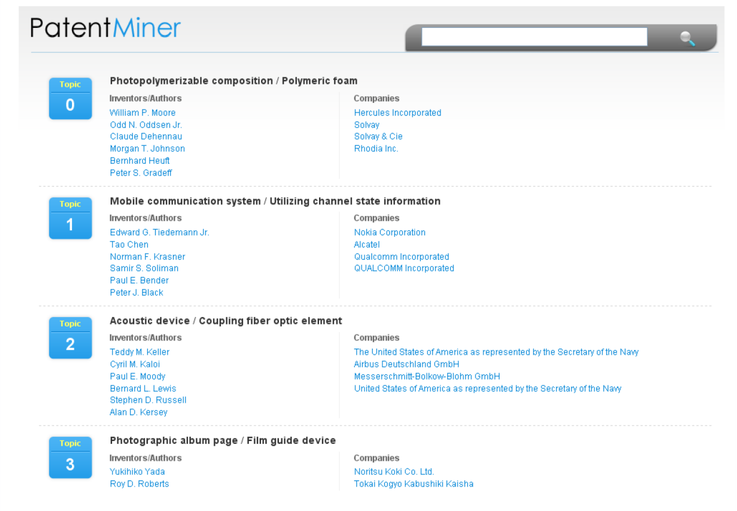


Patent Search

Company Trend



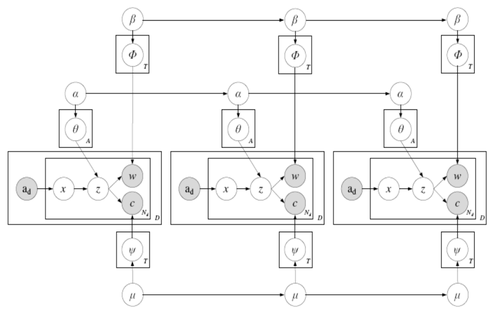
Competitor Analysis



Topic Browser

Technique Issue

Modeling Patent Network



Dynamic Inventor-Company-Topic Model

Inventor-topic smoothing $\Omega_1 = \sum_z (\theta_{az}^t - \theta_{az}^{t-1})^2$

Company-topic smoothing $\Omega_2 = \sum_z (\psi_{cz}^t - \psi_{cz}^{t-1})^2$

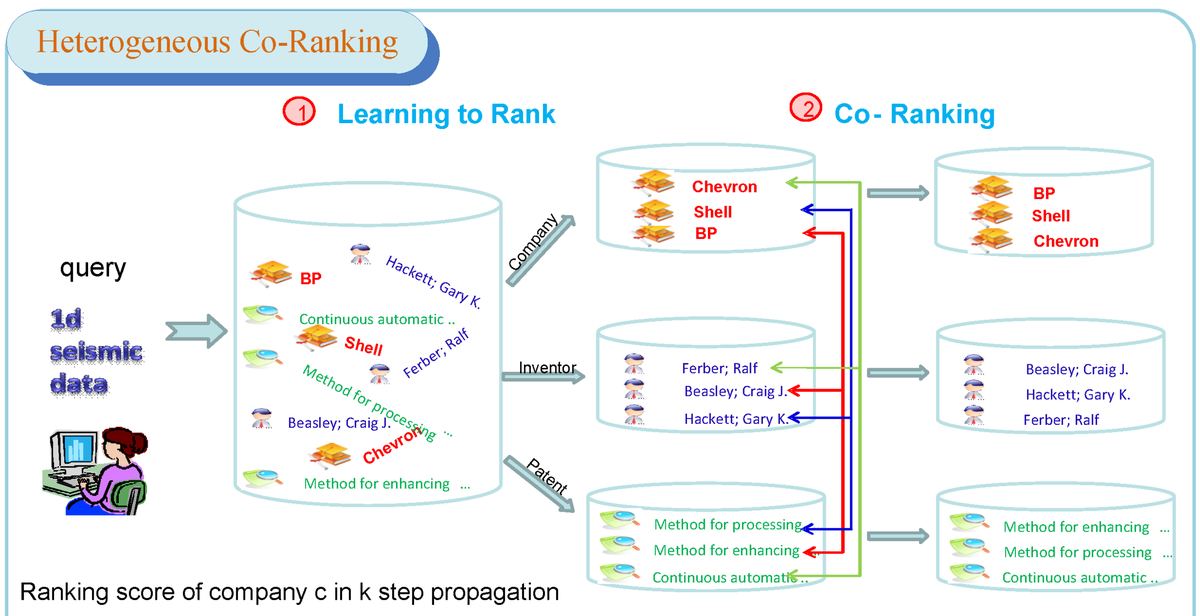
Topic smoothing $\Omega_3 = \sum_z (P(z)^t - P(z)^{t-1})^2$

Objective function

$$\mathcal{O}(\mathbf{D}) = -\mathcal{L}(\mathbf{D}) + \gamma_1\Omega_1 + \gamma_2\Omega_2 + \gamma_3\Omega_3$$

$$\mathcal{L}(\mathbf{D}) = P(\mathbf{x},\mathbf{z},\mathbf{w},\mathbf{c}|\Theta,\Phi,\Psi,\mathbf{a}) = \prod_{d=1}^{M}\prod_{i=1}^{N_d}\frac{1}{A_d} \times \prod_{z=1}^{K}\left(\prod_{x=1}^{A}\theta_{xz}^{m_{xz}}\prod_{j=1}^{W}\phi_{zw_j}^{n_{zw_j}}\prod_{c=1}^{C}\psi_{zc}^{n_{zc}}\right)$$

Heterogeneous Co-Ranking



Ranking score of company c in k step propagation

$$r^k[c] = (1-\xi_1-\xi_2)r^{k-1}[c] + \frac{\xi_1}{|V_a^c|}\sum_{d\in V_a^c} r^{k-1}[a] + \frac{\xi_2}{|V_d^c|}\sum_{d\in V_d^c} r^{k-1}[d]$$

Experimental results

Object	Method	P@1	P@5	MAP	N@1	N@5
Patent	LM	.7001	.6900	.6991	.7021	.6833
	HCR-1	.7592	.7102	.7359	.7592	.7310
	HCR-2	.7598	.7201	.7361	.7600	.7300
	HCR-5	.7600	.7298	.7400	.7678	.7367
Company	LM	.6931	.6790	.6654	.6888	.6532
	HCR-1	.7167	.6833	.7058	.7167	.6934
	HCR-2	.7189	.6900	.7100	.7200	.7000
	HCR-5	.7201	.6999	.7210	.7201	.7031

Competitor Analysis

Method 1: topic comparison based on DICT

Global Competitor Analysis

$$S(c,c') = (\sum_{i=1}^{n} p(z_i|c)p(c'|z_i))^2 + \eta(|\mathcal{D}_c| - |\mathcal{D}_{c'}|)^2$$

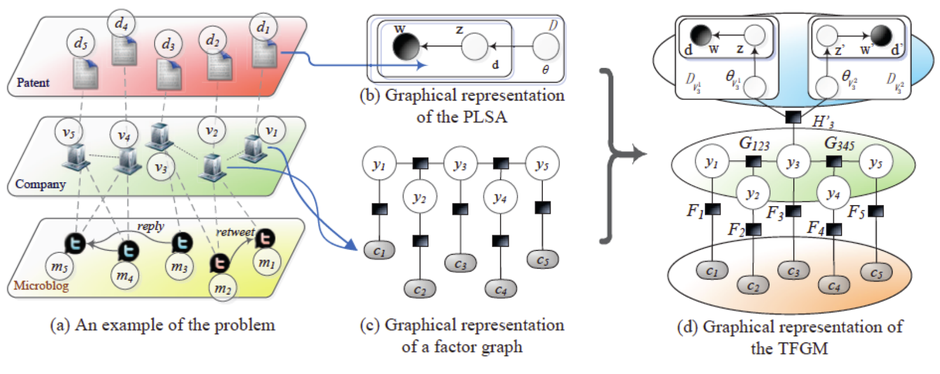
Topic-level Competitor Analysis

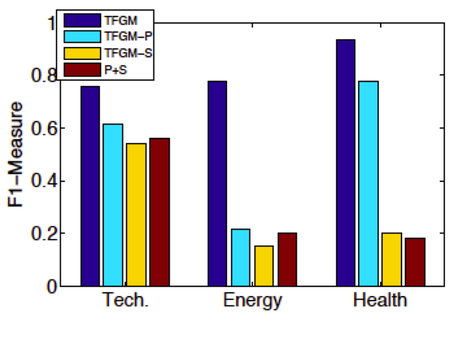
$$S(c,c',z) = (p(c|z) - p(c'|z))^2 + \eta(\sum_{d_i\in\mathcal{D}_{cz}} p(d_i|z) - \sum_{d_j\in\mathcal{D}_{c'z}} p(d_j|z))^2$$

	Methods	P@1	P@5	MAP	N@1	N@5
Global	WBS	.2009	.1087	.2904	.2009	.2841
	TopCom+TBD	.1731	.0846	**.3078**	.1731	.2871
	TopCom+PBC	**.2098**	**.1161**	.2920	**.2098**	**.3085**
Topic	LM+LDA	.1536	.1221	.2643	.1536	.2524
	TopCom+DBC	.1369	.1270	.2388	.1469	.2446
	TopCom+HBC	**.1620**	**.1366**	**.2781**	**.1620**	**.2874**

Experimental results

Method 2: combing patent records and social medias





Patent and social media contribution

Topic	Words	Competitors
Topic #4	image graphics pixel 3d	NVIDIA Vs. Autodesk
		Adobe Vs. VMware
		VMware Vs. Autodesk
		Microsoft Vs. NVIDIA
Topic #31	database distributed query domain	Oracle Vs. Jabil Circuit
		Yahoo! Vs. Jabil Circuit
		Google Vs. Jabil Circuit
		Microsoft Vs. Google
Topic #76	semiconductor toner compositions chamber	Novellus Systems Vs. Intel
		First Solar Vs. CREE
		Applied Materials Vs. IBM
		Motorola Vs. CREE

Examples of topic-level competitors

Paper ID: 315