Pragmatic eDiscovery
(Using Enterprise Search to Manage Highly-Involved Corporate Risk, Compliance, and eDiscovery Cases)

March 6, 2008

Framing:
www.arnoldit.com
management consulting and strategy for exceptional projects... white house, bear stevens, trust company of the west, largest web search company, largest software company, us senate police, australian national police, others.

author of:
- beyond search (forthcoming, gilbane group, april 2008)
- google version 2. the calculating predator (now available)
- the google legacy. (available at www.infonotics.com)
- author of enterprise search report (1st, 2nd, and 3rd editions)

www.arnoldit.com -- most of my articles and speeches
www.arnoldit.com/wordpress -- web log

monthly column in kmworld

---

**Beyond Search**
by Stephen E. Arnold

Another "Free" Behind-the-Firewall Search System

It's 6 a.m. in shiny louisville international airport, but the word "international" can be misleading. The news this morning is that microsoft will roll out a "new" sharepoint search service. you can read the breathless infoworld story [here](#). the announcement will be made, i believe, at one, maybe two, separate microsoft conferences this week.

the "free" word is a powerful marketing tool for commercial firms. When it comes to behind-the-firewall search, "free" is a synonym for demonstration product. the set up, configuration, bug process, optimization, and operation of a search or content processing system come with some hidden costs. the license fee is, of course, the cost that the gullible sees upon. when you root around in the financial statements of publicly-traded companies in the search and retrieval business, you find that many are trying to follow in vertiv's pre-sell out footsteps. specifically, vendors want to pump up consulting fees, making them carry the freight for earnings and growth. my recollection is foggy after seven consecutive days of travel, but vertiv was generating more than half its revenues from non-license revenue. the number 65 percent pops in and out of my memory, but
Topics

• Background
• Text processing
• Google’s hiding innovations -- janitors
• Cracking this problem
• Observations

I will post links and a version of this talk on my Web log at www.arnoldit.com/wordpress`
Based on a diagram from SAS in Cary, NC.
Methods and apparatus for determining equivalent descriptions for an ... Jeffrey A. Dean et al

Claims

What is claimed is:

1. A computer-implemented method for determining equivalent descriptions for an information need, comprising:
   identifying a list of queries issued by one or more users;
   identifying a candidate pair of equivalent descriptions by locating two queries that refer to the same information need;
   calculating a score for the candidate pair dependent on the frequency with which the candidate pair occurs in the list of queries identified in step 1; and
   determining whether the pair is equivalent by comparing the score calculated for the candidate pair to a defined threshold value.

2. The computer-implemented method of claim 1, wherein identifying a candidate pair comprises:
   locating two queries that contain at least one term in common, and identifying as a candidate pair the portion of the two queries that are not in common.

3. The computer-implemented method of claim 1, wherein identifying a candidate pair comprises:
   identifying a first description as a term T1 having characters C1, wherein M1 through M3 identify, as a second description, a sequence of terms T21, T22, ..., T2n, and identifying that terms T1 and terms T21, T22, ..., T2n have a candidate pair if each Cn matches the first letter of Tn.

4. The computer-implemented method of claim 1, wherein calculating a score comprises:
   determining a fast frequency with which the candidate pair occurs within the...
Text Processing

A universe of information

Text mining snags other textual content

What search "finds".
Janitors: “Cute” and Obscure

(12) United States Patent
(45) Date of Patent: Jun. 12, 2007
(10) Patent No.: US 7,231,393 B1
(54) METHOD AND APPARATUS FOR LEARNING A PROBABILISTIC GENERATIVE MODEL FOR TEXT

(75) Inventors: Georges Harik, Mountain View, CA (US); Noam M. Shazeer, Mountain View, CA (US)

(73) Assignee: Google, Inc., Mountain View, CA (US)

(21) Appl. No.: 10/788,937
(22) Filed: Feb. 26, 2004

Related U.S. Application Data
(63) Continuation-in-part of application No. 10/976,571, filed on Sep. 30, 2003.

(51) Int. Cl. G06F 17/30 (2006.01)
(52) U.S. Cl. 707/100
(58) Field of Classification Search 707/1, 707/2, 9, 10, 100; 704/251; 700/224

ABSTRACT

One embodiment of the present invention provides a system that learns a generative model for textual documents. During operation, the system receives a current model, which contains terminal nodes representing random variables for words and cluster nodes representing clusters of conceptually related words. Within the current model, nodes are coupled together by weighted links, so that if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node causes the other node to fire with a probability proportionate to the link weight. The system also receives a set of training documents, wherein...
Jonathan Betz is currently a Software Engineer with Google, the Internet’s leading search engine. Prior to joining Google, Jonathan was the Engineering Manager for Analytic Query Services at E.piphany, where he directed development of the analytic core for the industry-leading E.6 marketing automation product. Companies like Expedia, Mutual of Omaha, and H&R Block used products and technologies he developed to produce actionable business insights from terabyte-scale enterprise data stores.

Andrew Hogue is the author of Thresher: automating the unwrapping of semantic content from the World Wide Web. Andrew Hogue, David Karger, WWW ’05: Proceedings of the 14th international conference on World Wide Web, 2005, pp. 89-95. He is also the lead on Google’s January 2008 search views announcement. He’s an inventor of more than eight Google “procedures.”

Import, Normalizer, and Object Agents operate as separate, intelligent “machines” aware of one another’s outputs

Abstract

Import, Normalizer, and Object Agents operate as separate, intelligent “machines” aware of one another’s outputs.
Janitor agents use “appropriate” procedures, access other Google KBs, and interact with the Fact KB rows.

Cracking This Problem
The ISYS 8 Suite

The ISYS Suite of enterprise search, navigation, and discovery is available for site data evaluation and purchase. Comprised of three core applications: ISYS Search, ISYS Search Advanced, and ISYS Search Center, our latest technology addresses several key enterprise search requirements, such as:

- Text mining, discovery, and expert location via ISYS Search Center.
- Power and scalability through content caching, indexing, and federated search support.
- Access to business users, third parties, and external content such as Bedrock.
- Full-text search using natural language through enhanced ISYS Search Center.
- Support for search within SharePoint search, Office 2007 and ability to search information throughout the enterprise.

Learn more about the ISYS Suite and test drive the software today:

- [Download ISYS 8 demo](#) to search on PCs and networks
- [Download ISYS Search Advanced](#) to search across entire enterprise
- [Download ISYS Search Center](#) to add search to your OTM solution

ISYS Highlights

- [ISO Compliance](#)
- [Integrates with Windows SharePoint Services](#)
- [Supports multiple languages](#)
- [Integrates with Microsoft Office 2007](#)
- [Supports federated search](#)
- [Supports web scraping](#)
- [Supports social networking](#)
- [Supports virtualization](#)
- [Supports web 2.0 technologies](#)
- [Supports mobile devices](#)
- [Supports AJAX technologies](#)
- [Supports virtualization](#)
- [Supports cloud computing](#)
- [Supports virtualization](#)
- [Supports cloud computing](#)
- [Supports virtualization](#)
Methods and systems for member-correlated advertisement in a member network - Patent 20140002446

Reference: Most Relevant (10 Items)

1. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

2. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

3. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

4. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

5. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

6. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

7. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

8. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

9. Patent 20130206885
   - Method and system for displaying a member's contact information in the service
   - ID: 5
   - Invention: [Invention Information]
   - Date: [Date]
   - Invention: [Invention Information]

10. Patent 20130206885
    - Method and system for displaying a member's contact information in the service
    - ID: 5
    - Invention: [Invention Information]
    - Date: [Date]
    - Invention: [Invention Information]
Observations
Balance