第十五章：

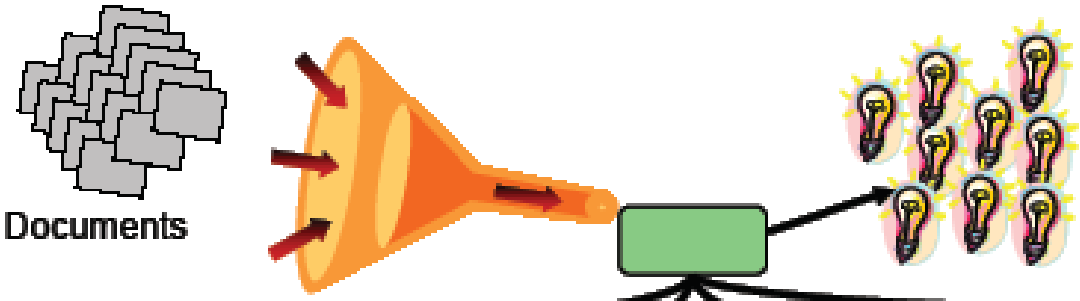
文本挖掘工具与应用

杨建武

北京大学计算机科学技术研究所

Email:yangjianwu@icst.pku.edu.cn

Gartner view of Unstructured Data Management



- Search by Query
  - Keyword-Based
  - Concept-Based
  - NLP-Based
  - Parametric
- Search by Navigation
  - Linear Browsing
  - Ad-Hoc Hyperlinks
  - Taxonomies
  - Visual Maps
  - Ontologies
- Others
  - Personalization
  - Alerting "Agents"
  - What's Related?
  - Skills
  - People
- Text Mining & Visualization
  - Summarization
  - Categorization
  - Clustering
  - Information Extraction
  - Trending & More

Text Mining by Task

- Information retrieval
- Text categorization
- Document clustering
- Information filtering / topic detection
- Text summarization
- Question and answer
- Taxonomy/concept/relationship mining
- Visualization and user interface

Text Mining by Industry

- Biotechnology
- Consumer products
- CRM, Consulting, Marketing
- Education
- Government
- Healthcare
- Insurance
- Other Industry

传统商业方面的应用

Discovering Unexpected Information From A Competitor

- Assume your boss ask you to find out what new information your competitor provides
  - E.g., to learn from the competitor
  - E.g., to design counter measures (对策)
- Text mining techniques that maybe useful
  - novelty detection, text classification, information extraction
- Major problems:
  - How to model what you already know?
    - » Incorporating user's existing knowledge
  - What unexpected information about competitors to find?
  - Algorithms
  - System architecture

Find Unexpected Information About Competitors

- What is unexpected information?
  - Is relevant to the user
  - Is unknown to the user, or contradicts the user's existing beliefs or expectations
- • Examples
  - Unexpected services provided by competitors
  - Unexpected products provided by competitors
- How to measure unexpectedness (novelty)?
  - Between two web sites
  - Between two pages

应用：企业竞争情报

- 为企业收集和分析数据，以识别出现的威胁或问题。
- 跟踪新闻稿、专利公布和合并与收购活动可以帮助确认由于竞争对手、供应商、顾客或合作伙伴的策略变化而导致的潜在威胁。
- 监控和分析新闻组和邮件列表中顾客张贴的内容和对呼叫中心的投诉可以帮助发现市场动态和品牌观念的趋势。

应用： CRM

Interaction data
- Offers
- Results
- Context

10-50% improvement

Attitudinal data
- Opinions
- Preferences
- Needs

10-40% improvement

10-30% improvement

Descriptive data
- Attributes
- Characteristics
- Self-declared info
- (Geo)demographics

Behavioral data
- Orders
- Transactions
- Payment history
- Usage history

应用： CRM


Google
("phone" OR "telephone")
Search Web
548 blocked
AutoFill
Options
cell
cellular
Complaints.com
Search Our Consumer Complaints Database >>
cell phone
Search
Publicize and Read Consumer Complaints
Complaints.com - publicize and read consumer complaints
gain leverage to help solve your complaints
post your complaints for public view
learn from other consumer complaints
read actual consumer experiences
share your own personal consumer experiences
check to see how a business is performing, treating customers
get help with your product / service / customer service problem or complaint
help other consumers avoid the same problems that you have experienced personally
Send/Post Your Complaint
Consumers:
e-mail complaints to:
complaints@complaints.com
Businesses:
e-mail replies to:
manager@complaints.com
Subscribe to our Free, Privacy Protected E-mail Newsletter | Browse Consumer Complaints - by Date
About | Press | Terms | Privacy | How it Works | Business Replies

应用：电子商务网站

- 电子商务最需要
  - 第一是吸引新的用户，增加已有用户的忠实度，
  - 第二是减少系统运行的开销和成本。
- 最有效的方法
  - 记忆及分析用户的浏览兴趣和习惯，为用户提供真正个性化的网上资讯服务。
- 文本挖掘可为电子商务网站提供三个独特功能：
  - 「内容相关推荐」自动监察用户的浏览习惯及内容并随时推送相关资讯及网站；
  - 「协同推荐」自动记忆及分析用户的浏览习惯，让用户可随时进入浏览所推介的内容；
  - 「精确搜寻」会应用户指定的要求，在网上世界搜寻最精确的资料。

应用：BBC公司

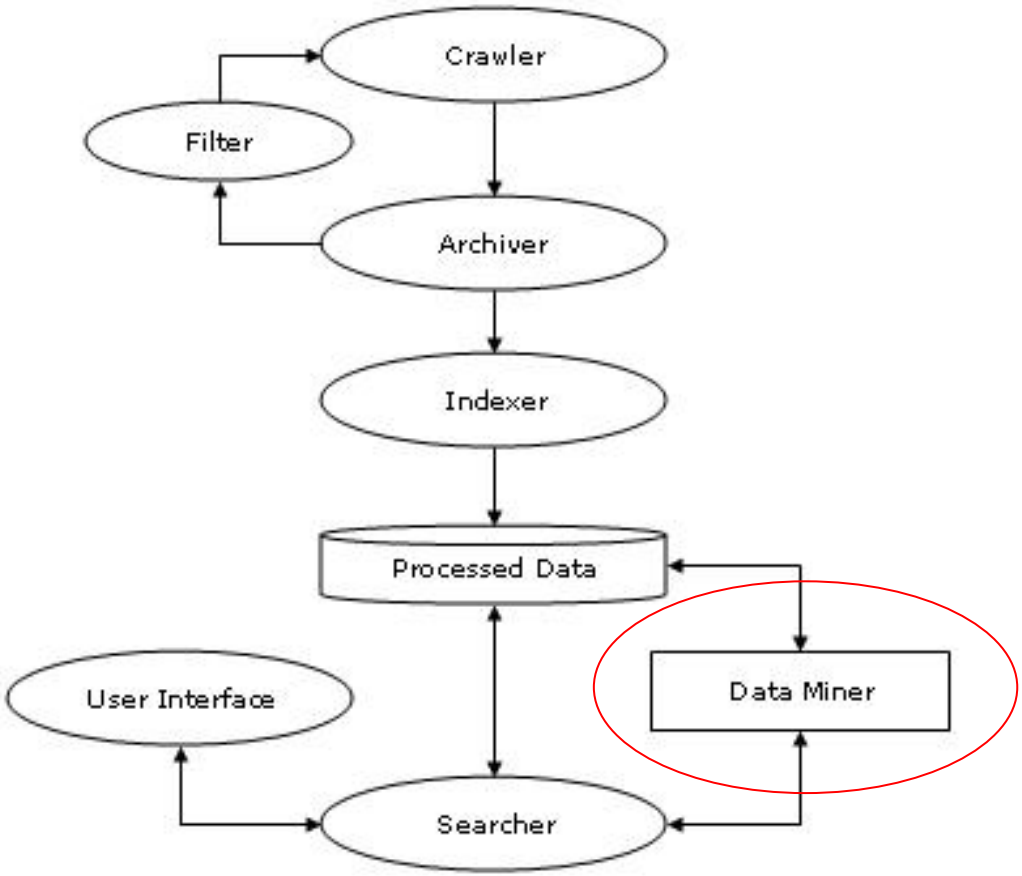
- BBC，英国广播公司每天从世界各地涌进130万份各种格式的新闻消息，每天要对这些信息进行处理，储存，分析，做新闻连接，还有网页新闻发布，要同时支持上百万用户的使用。
- 以前BBC用人力的处理方法，每天需要上百人来阅读，分析，人工贴标签，人工网页连接。耗资巨大，随着信息量的增加越来越不可行。
- 文本挖掘技术使整个过程全部自动化。现在BBC的网页上可以提供20种自然语言的信息检索，即时的信息连接，用户的信息个人化。
- 系统运行的成本却比以前减少了数倍，现在每天只用几个人来管理整个系统就够了。

应用： the Health Industry

- Patients with characteristics X and symptoms Y should get test Z
- Some information is easy to extract from medical forms
  - E.g., patient characteristics such as gender
  - E.g., diagnostic tests assigned
- Some information must be extracted from the text
  - E.g., symptoms such as headache
- Techniques used:
  - Text classification
  - Information extraction (template-filling)

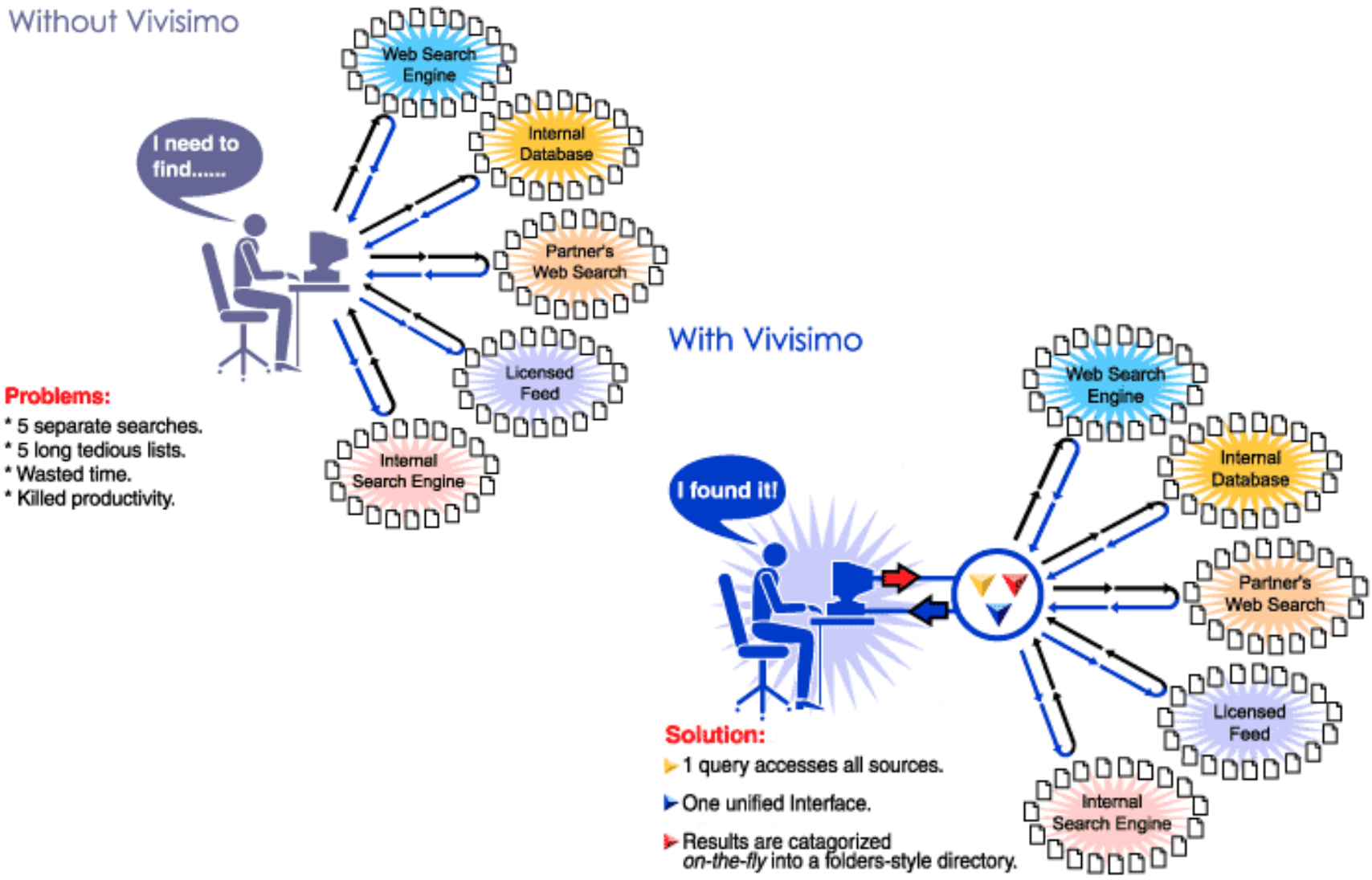
搜索引擎方面的应用

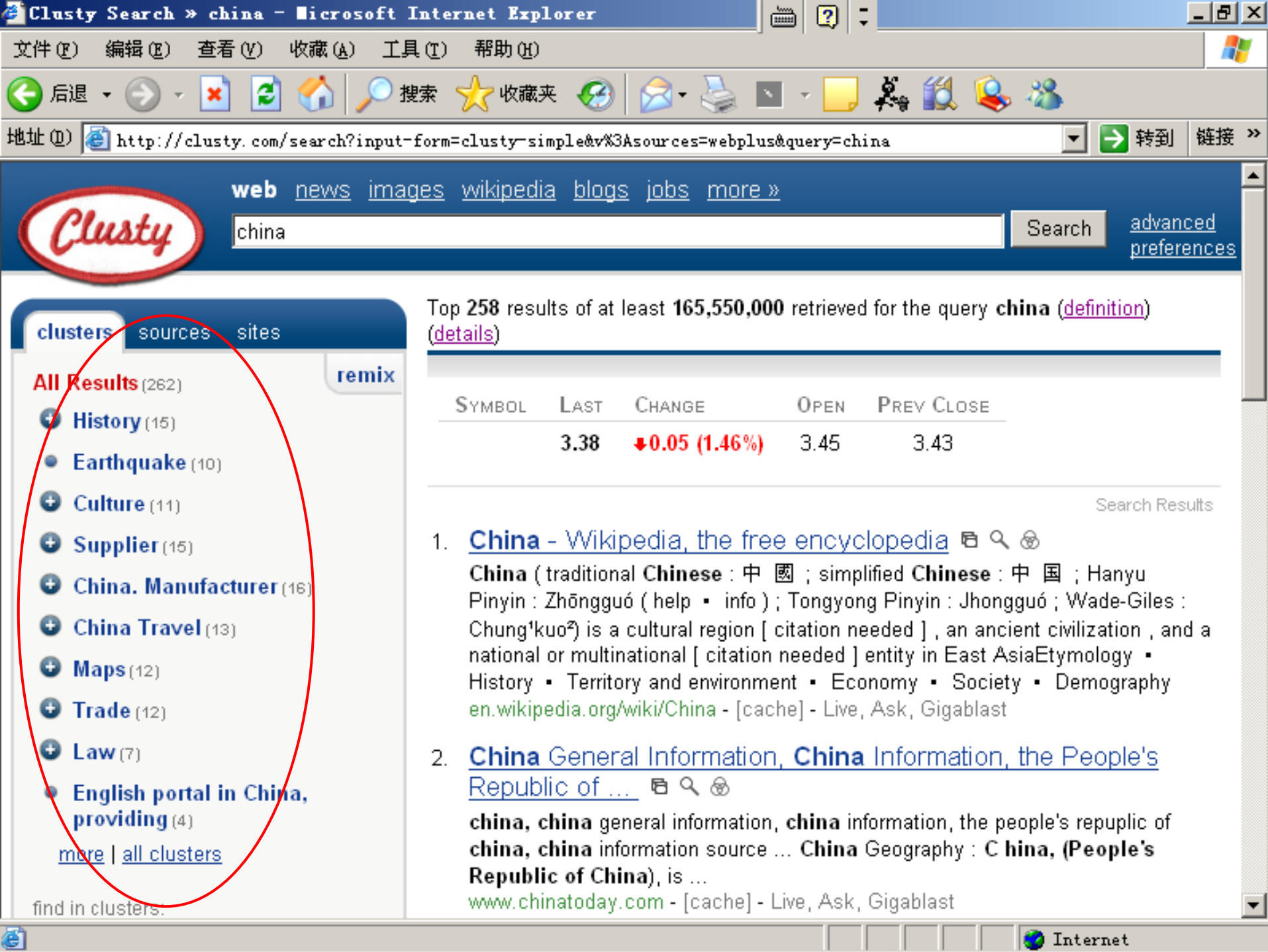
应用： Search Engines



ARCHITECTURE OF A SEARCH ENGINE

Vivisimo Search Engine: (http://clusty.com/)




Clusty Search » china - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹
地址(D) http://clusty.com/search?input-form=clusty-simple&v%3Asources=webplus&query=china
转到 链接
Clusty
web news images wikipedia blogs jobs more »
china
Search
advanced preferences
clusters sources sites
remix
All Results (262)
History (15)
Earthquake (10)
Culture (11)
Supplier (15)
China. Manufacturer (16)
China Travel (13)
Maps (12)
Trade (12)
Law (7)
English portal in China, providing (4)
more | all clusters
find in clusters:
Top 258 results of at least 165,550,000 retrieved for the query china (definition) (details)
Symbol Last Change Open Prev Close
3.38 ↓0.05 (1.46%) 3.45 3.43
Search Results
1. China - Wikipedia, the free encyclopedia
China (traditional Chinese : 中 國 ; simplified Chinese : 中 国 ; Hanyu Pinyin : Zhōngguó (help • info) ; Tongyong Pinyin : Jhongguó ; Wade-Giles : Chung¹kuo²) is a cultural region [citation needed] , an ancient civilization , and a national or multinational [citation needed] entity in East AsiaEtymology • History • Territory and environment • Economy • Society • Demography
en.wikipedia.org/wiki/China - [cache] - Live, Ask, Gigablast
2. China General Information, China Information, the People's Republic of ...
china, china general information, china information, the people's repuplic of china, china information source ... China Geography : China, (People's Republic of China), is ...
www.chinatoday.com - [cache] - Live, Ask, Gigablast
Internet

Google 资讯 - Windows Internet Explorer

http://news.google.cn/nwshp?hl=zh-CN&tab=wn

文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

网页 图片 地图 资讯 视频 音乐 更多

Google 资讯 BETA 谷歌

搜索资讯 搜索所有网页 高级资讯搜索 使用偏好

搜索和浏览 25,000 个不断更新的资讯来源。

- 焦点
- 财经
- 娱乐
- 科技
- 互联网
- 体育
- 社会
- 汽车
- 房产
- 教育
- 热门报道

资讯快讯

文字版本

标准版本

图片集锦

RSS

移动资讯服务

焦点

13分钟前自动生成

加拿大经济已受到流感疫情冲击

国际在线 - 1小时前

新华社消息（记者赵青）：截至4月30日下午，加拿大已确诊34例甲型H1N1流感病例，成为继墨西哥、美国之后确诊感染患者最多的国家。作为与墨西哥有着密切经济联系和人员往来的北美国家， ...

中网资讯中心

新浪网 - 四川在线 - 中国网

所有 41,627 篇资讯文章 »

日本媒体继续关注日本首相麻生太郎访华 新华网 相关(2,006条) »

消防法新规:处理突发应急救援都成消防"分内事" 新华网 相关(269条) »

黑龙江伊南河火场调整作战部署 扑救兵力增至10530人 新华网 相关(835条) »

热点网谈：湖北二级公路取消收费应惠民到底 荆楚网 相关(821条) »

昆明市民可通过邮件推行社区医生上门服务 云南网 相关(18条) »

铁路上海站今迎客流最高峰 凤凰网 相关(102条) »

从通用福特看美国汽车业的破产之路 凤凰网 相关(1,411条) »

"纪念'五三、五四大轰炸'70周年文物资料陈列展"开展 华龙网 相关(21条) »

驾车欲袭击荷兰女王一家的肇事者刚遭解雇(图) 凤凰网 相关(368条) »

个性化谷歌资讯

"阿育王塔"在南京开展 452颗宝珠分7类(图)
中新网 - 2小时前 - 相关(207条) »

微软Xbox战略总监跳槽到苹果
腾讯网 - 1小时前 - 相关(17条) »

重压之下福原爱身心俱疲 女双获胜连声感谢搭档
中国网 - 49分钟前 - 相关(291条) »

银行遭抢开后门助警"诱捕"劫匪
法制晚报 - 26分钟前 - 相关(237条) »

国际汽联颁布明年新规 预算帽上调禁止赛中加油
搜狐 - 49分钟前 - 相关(37条) »

看房的多拍板的也多 长沙房交会第2日人气不减
中国网 - 6小时前 - 相关(202条) »

优先解决本地生源就业问题
中国网 - 1小时前 - 相关(1,442条) »

本地资讯预览：北京

添加北京 添加其他城市

北京高速不停车收费系统今运行 开通ATM机充值

搜狐 - 1小时前

本报讯（记者陈斯） 今日零时起，本市安装的所有不停车收费系统（ETC）

北京一男子抢劫银行被警方当场擒获
搜狐 - 5小时前 - 相关(237条) »

首都经济贸易大学招办主任：二志愿有20%比例

Internet | 保护模式：禁用 100%

ThinkPad - Google Product Search - Windows Internet Explorer

http://www.google.com/products?q=ThinkPad

文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H) 链接

Web Images Maps News Video Gmail more ▾ My Shopping List

Google ThinkPad Search Products Search the Web Advanced Product Search Preferences

Products

Results **1** - **10** of about **443,213** for **ThinkPad**. (**0.16** seconds)

Sponsored Links

ThinkPad - Official Site
www.Lenovo.com Save Up to $500 on **ThinkPad** Laptops w/ Intel Technology. Free Shipping!

Show only: Google Checkout Free shipping Show grid view Sort by: Relevance

Enter location for tax and shipping: ZIP or city, state OK

Lenovo **ThinkPad** T61 7658 - Core 2 Duo 2.2 GHz - 14.1 " - 1 GB Ram ...
Microsoft Windows XP Professional, 5.1 lbs, Lithium ion battery 3.6 hour(s), 13.2 " x 9.3 " x 1.3 "
ThinkPad T Series is the perfect balance of performance and portability. Designed for highly mobile users, these notebooks deliver outstanding functionality and long battery ...
1 review - Add to Shopping List

$1,059 to $1,492
from 15 sellers
Compare prices

Lenovo **ThinkPad** X61 Tablet 7767 - Core 2 Duo 1.6 GHz - 12.1 " - 1 ...
Black, Microsoft Windows XP Tablet PC Edition, 4 lbs, Lithium ion battery 9 hour(s), 10.8 " x 9.6 " x 1.3 "
The **ThinkPad** X61 Tablet delivers hardware-based security and enhanced manageability, along with exceptional performance and outstanding mobility. And like all X Series ...
Add to Shopping List

$1,461 to $1,972
from 14 sellers
Compare prices

Lenovo **ThinkPad** X300 6478 - Core 2 Duo 1.2 GHz - 13.3 " - 2 GB Ram ...
Black, Microsoft Windows XP Professional, 3.4 lbs, Lithium ion battery 6.5 hour(s), 12.5 " x 9.1 " x 1 "

$1,463 to $2,948
from 23 sellers
Compare prices

Sponsored Links

Sony VAIO Laptop Computer
Free Shipping! Intel® Centrino®
2. A new standard in mobility.
www.SonyStyle.com/Vaio

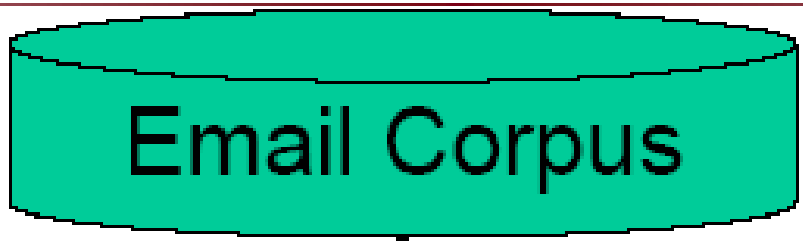
Lenovo **Thinkpad** at Amazon
Big Savings on Lenovo **thinkpad**
Free 2-Day Shipping w/Amazon Prime!
Amazon.com/Electronics-Accessories

Cheap **Thinkpad** T60p
Looking for **Thinkpad** T60p on sale?
Compare Laptop Computers & save!
www.thinkpad-t60p.best-price.com

Cheap Notebook Prices
Save on Notebooks! Shop 1,000's by Processor, Display, Weight & More.
www.NexTag.com/Notebook-Computers

Internet | 保护模式: 禁用 100%

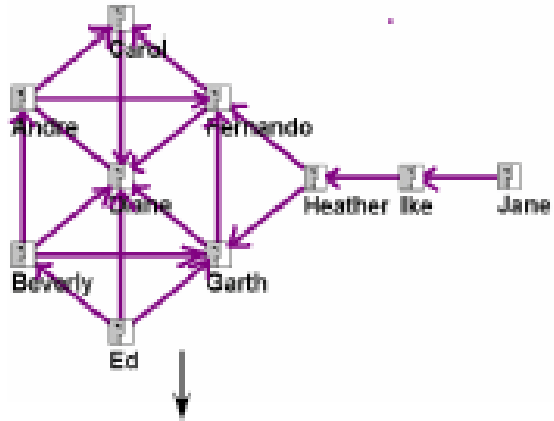
Finding Topic-Specific Experts



Query:

Restaurants near CMU

Social network: who often sends or receives email about restaurants



List of experts:

1. Diane
2. Fernando

Text Mining Tools

IBM DB2 Intelligent Miner

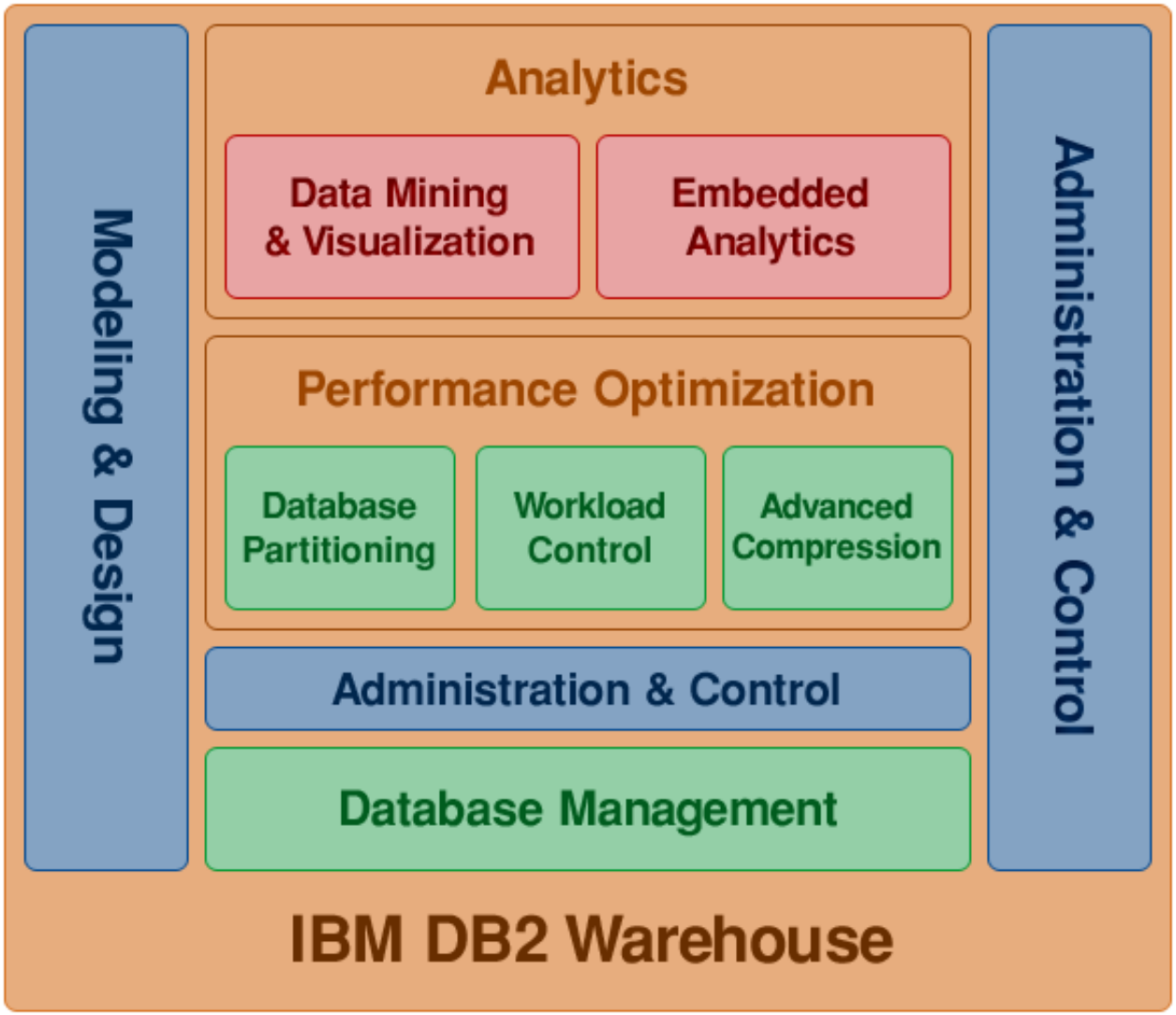
IBM DB2 Intelligent Miner

- IBM DB2 Intelligent Miner：
  - Intelligent Miner for Data
    - 可以寻找包含于传统文件、数据库、数据仓库和数据中心中的隐含信息。
  - IBM Intelligent Miner for Text
    - 允许企业从文本信息中获取有价值的客户信息。
- 1998年 Intelligent Miner for Text V 2.2

DB2 Data Warehouse Editions

- March 14, 2006, IBM announced withdrawal from marketing and end of support for the Intelligent Miner tools.
- DB2 Data Warehouse Editions is the replacement product.

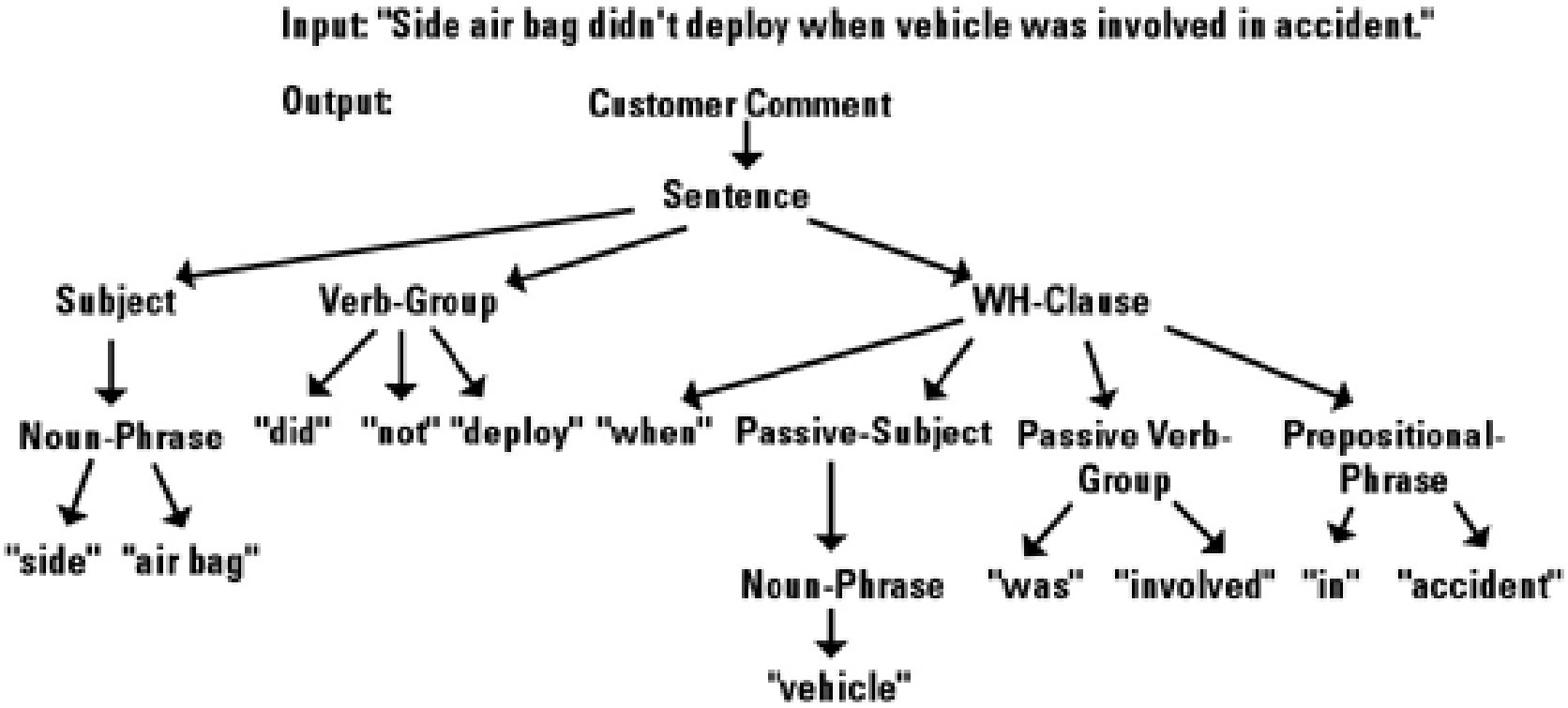
DB2 Data Warehouse Editions



IBM Intelligent Miner for Text

- 扩展了IBM 的数据采集功能，可以从文本文档和数据源获取信息。
- 文本数据源可以是Web 页面、在线服务、传真、电子邮件、Lotus Notes 数据库、协定和专利库。
- 数据源可以包括客户反馈、在线新闻服务、电子邮件和Web 页面。
- 功能包括：
  - 识别文档语言，建立姓名、用语或其它词汇的词典
  - 提取文本的涵义，将类似的文档分组，并根据内容将文档归类。
  - 文本搜索引擎和Web 文本搜索功能

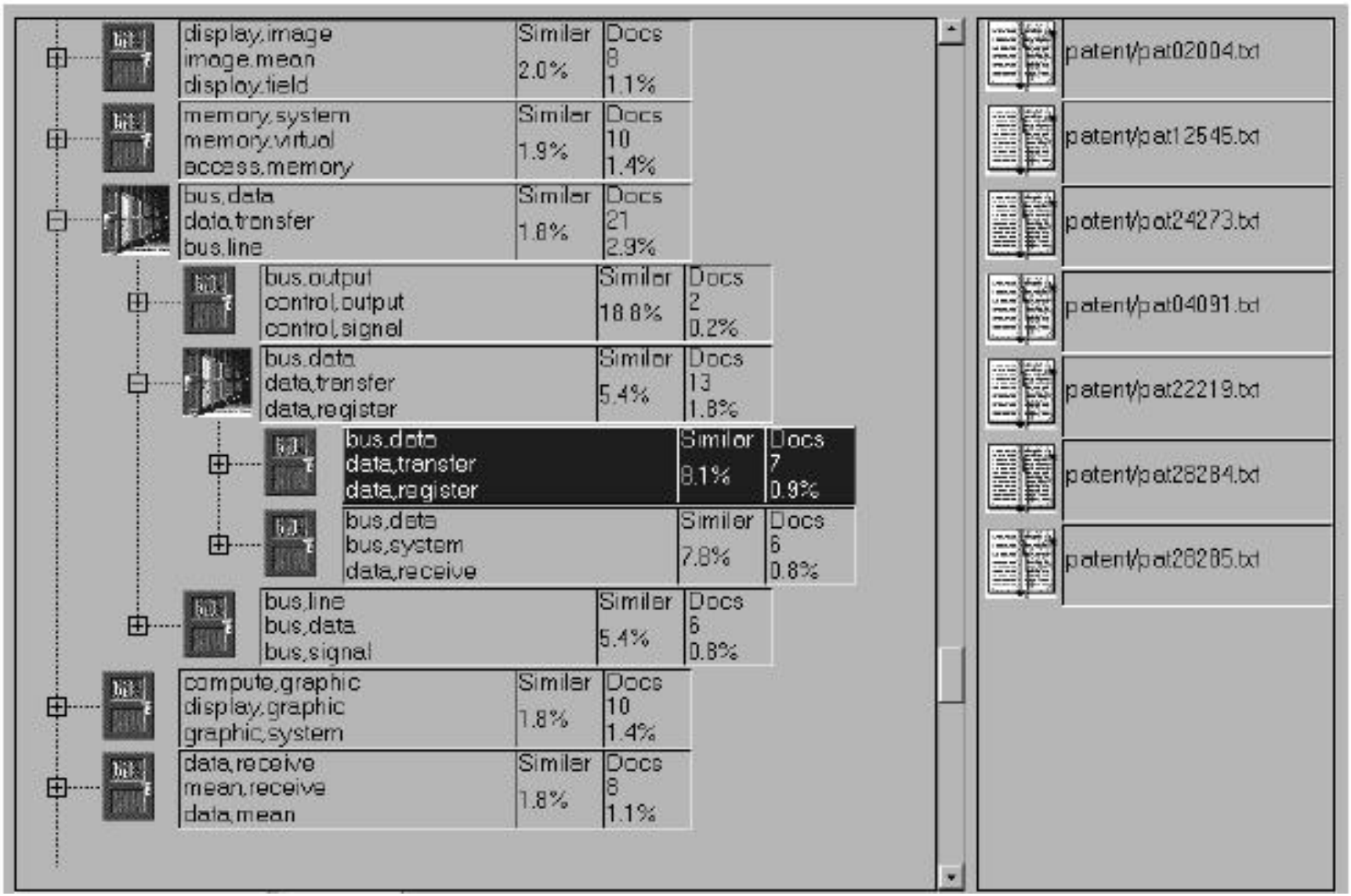
单词角色及其关系的解析树



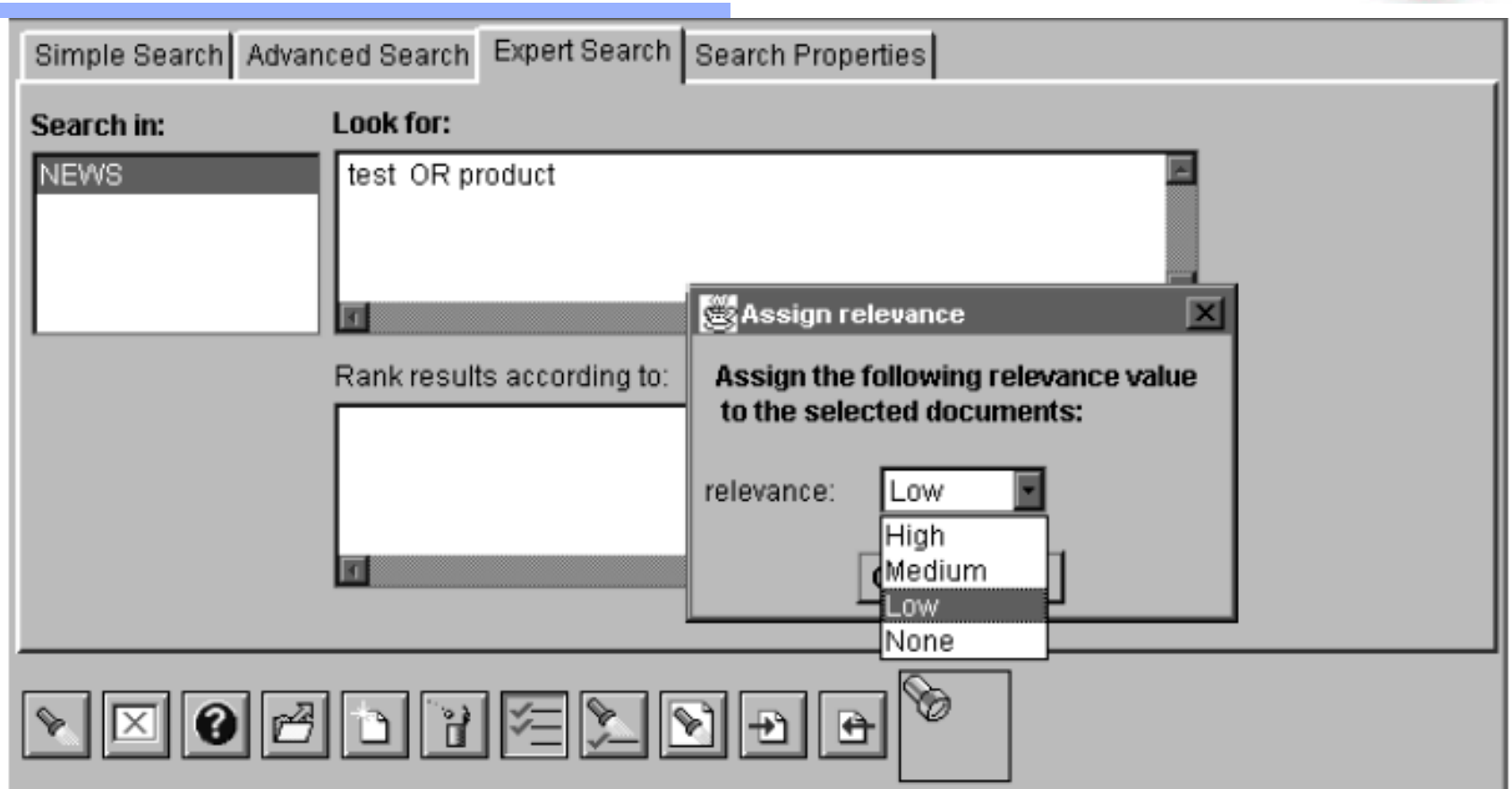
元数据提取与自动分类

类别	关键字 1	关键字 2	关键字 3	最大金额
Intranet applications	employee productivity	web applications	network availability	$20,500
Personnel policy	maternity leave	health benefit	parental leave	$
Lotus Notes information	database replication	e-mail	collaboration applications	$45
Commuter information	bus line	telecommuting options	railroad station	$1.25
Office ergonomics	wrist rest	Spinal curvature	voice recognition	$99

自动聚类


display,image
image,mean
display,field
Similar 2.0%
Docs 8 1.1%
memory,system
memory,virtual
access,memory
Similar 1.9%
Docs 10 1.4%
bus,data
data,transfer
bus,line
Similar 1.8%
Docs 21 2.9%
bus,output
control,output
control,signal
Similar 18.8%
Docs 2 0.2%
bus,data
data,transfer
data,register
Similar 5.4%
Docs 13 1.8%
bus,data
data,transfer
data,register
Similar 8.1%
Docs 7 0.9%
bus,data
bus,system
data,receive
Similar 7.8%
Docs 6 0.8%
bus,line
bus,data
bus,signal
Similar 5.4%
Docs 6 0.8%
compute,graphic
display,graphic
graphic,system
Similar 1.8%
Docs 10 1.4%
data,receive
mean,receive
data,mean
Similar 1.8%
Docs 8 1.1%
patent/pat02004.txt
patent/pat12545.txt
patent/pat24273.txt
patent/pat04091.txt
patent/pat22219.txt
patent/pat28284.txt
patent/pat28285.txt

多种检索



#	Rank	Document ID	Collection Name	Size	Count	Relevance
1	38	NEWS0007.doc	NEWS	180	1	
2	35	NEWS0031.doc	NEWS	1356	9	
3	28	NEWS0041.doc	NEWS	126	3	
4	28	NEWS0052.doc	NEWS	212	3	

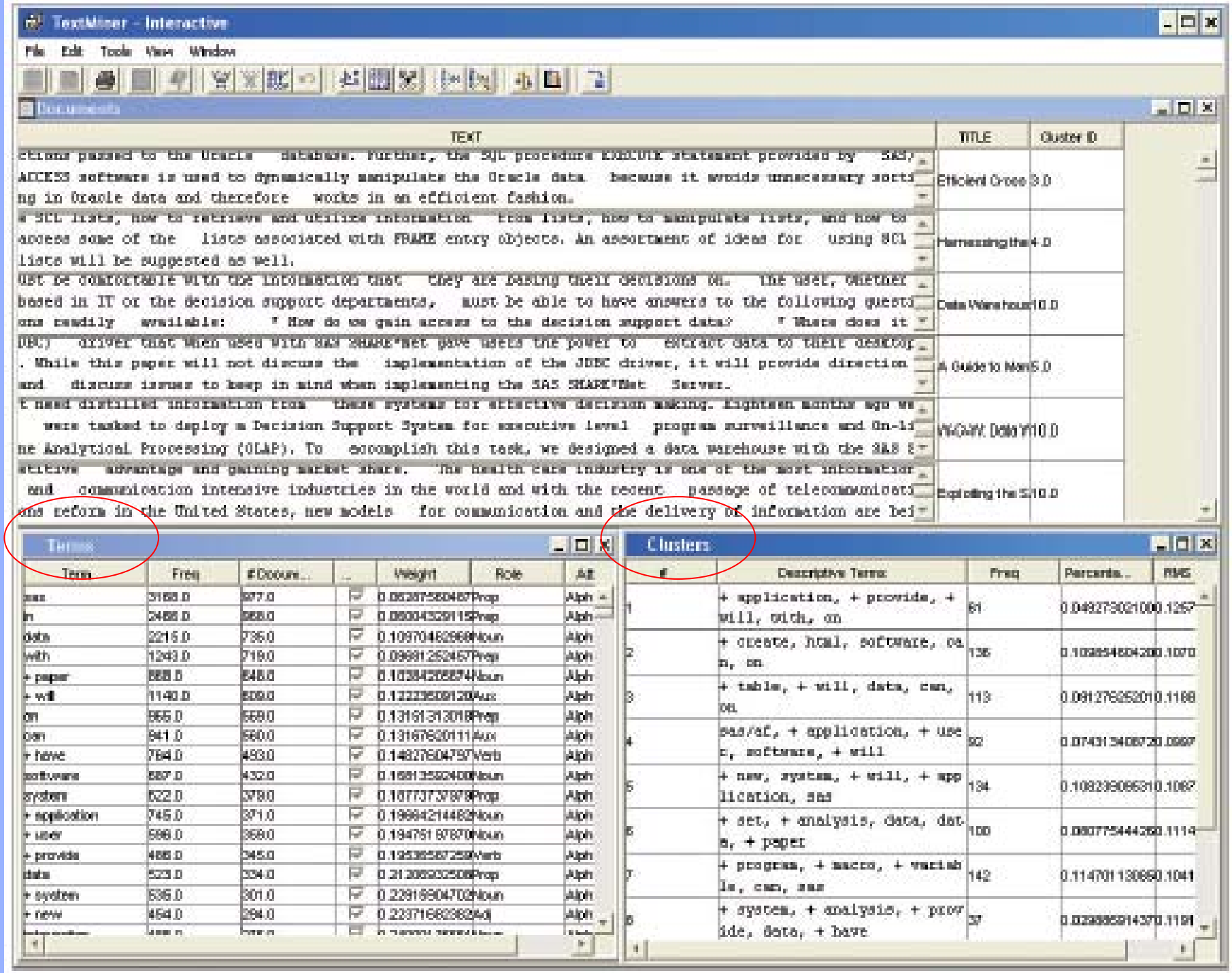
SAS Text Miner

SAS® Text Miner

- SAS Text Miner provides a rich suite of tools for discovering and extracting intelligence from large document collections.
- It helps identify trends and business opportunities and generates meaningful insights to key business issues more efficiently and with less risk.

SAS® Text Miner

- SAS Text Miner provides a rich suite of text processing and analysis tools
  - Text documents can be clustered automatically into groups
  - Text documents can be classified into predefined categories
  - Conjunction with structured data to build predictive models
- Text mining can be described as a three-step process:
  - accessing the unstructured text,
  - parsing the text and turning it into actionable data,
  - and analyzing the newly created data.
- For each step, SAS Text Miner provides state-of-the-art tools that enable organizations to efficiently extract intelligence from large text collections.

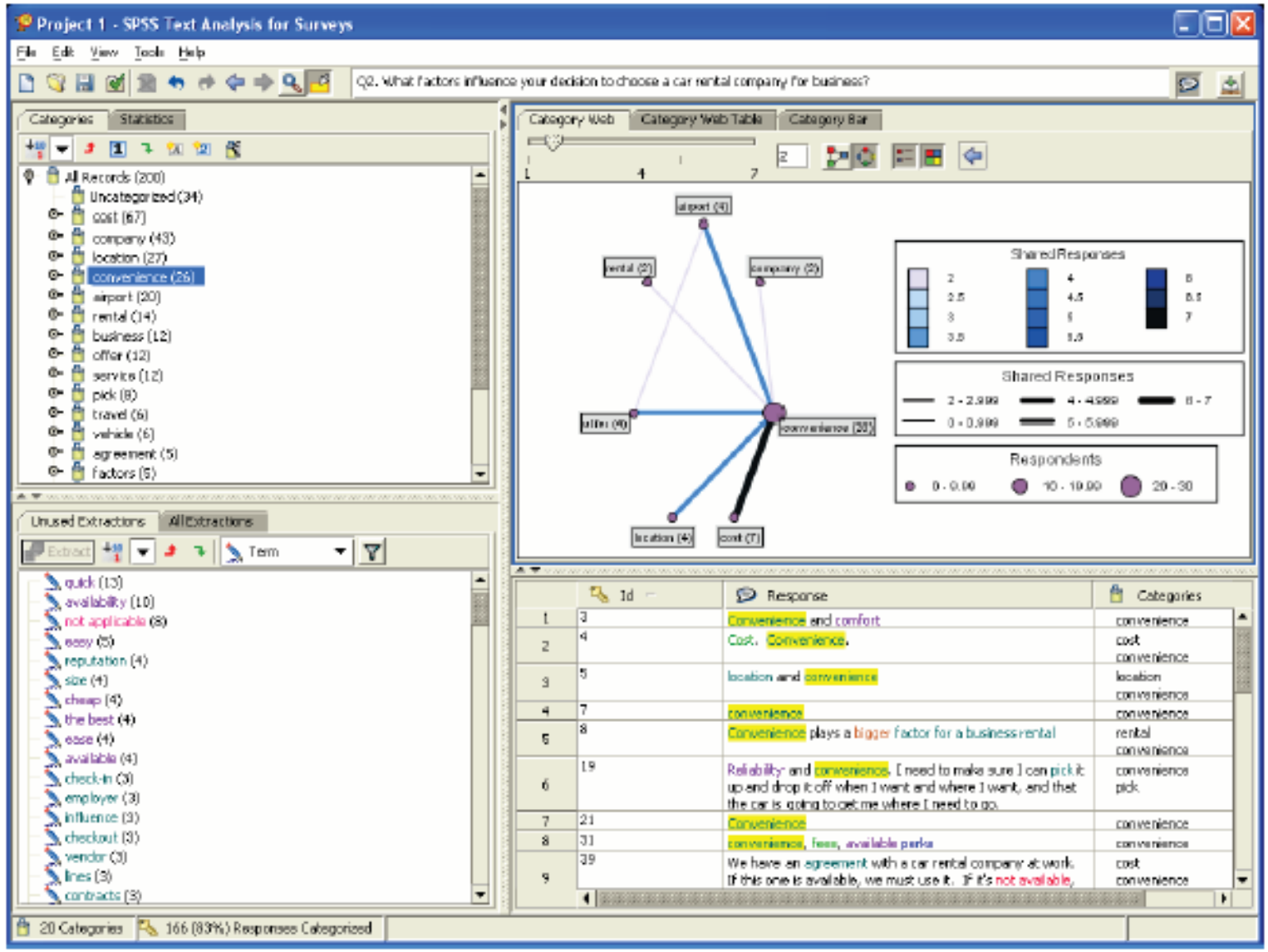

TextMiner - Interactive
File Edit Tools View Window
Documents
TEXT
TITLE
Cluster ID
Efficient Cross 3.0
Harnessing the 4.0
Data Warehousi 10.0
A Guide to Man 5.0
Exploiting the S 10.0
Terms
Term
Freq
Weight
Role
Clusters
#
Descriptive Terms
Freq
Percenta...
RMS
1 + application, + provide, + will, with, on 61 0.049273021000.1257
2 + create, html, software, on, on 135 0.109894604200.1070
3 + table, + will, data, can, on 113 0.091276052010.1168
4 sas/af, + application, + use r, software, + will 92 0.074313408720.0997
5 + new, system, + will, + app lication, sas 134 0.108239086310.1087
6 + set, + analysis, data, dat a, + paper 100 0.080775444260.1114
7 + program, + macro, + variab le, can, sas 142 0.114701130650.1041
8 + system, + analysis, + prov ide, data, + have 37 0.029885914370.1191

SPSS Text Mining

SPSS Text Mining

- Predictive Text Analytics™
- More than 1000 companies use SPSS Text Mining software, including most of the top 500 Fortune Companies

Classification and Categorization

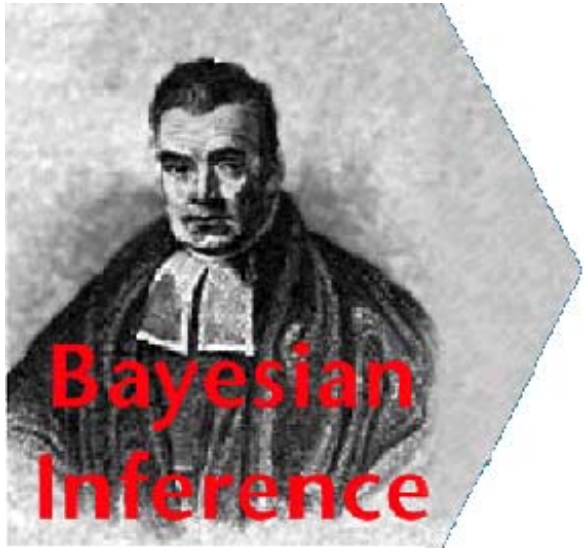


Autonomy IDOL Server

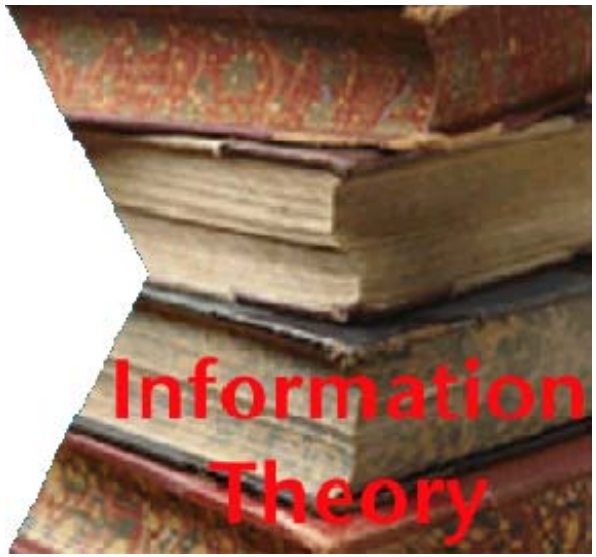
Autonomy

- Autonomy Systems Plc总部设于英国剑桥和美国旧金山，拥有遍布世界各地的分公司。
- Autonomy提供全面完整的智能软件结构，自动化地处理，操作和应用不规整的信息。
- 不规整的信息指的是我们周围越来越多的人们所熟悉的信息，比如电子邮件，因特网网页，电子报表，Word文件, pdf文件,语音文件等等。

The Solution



+



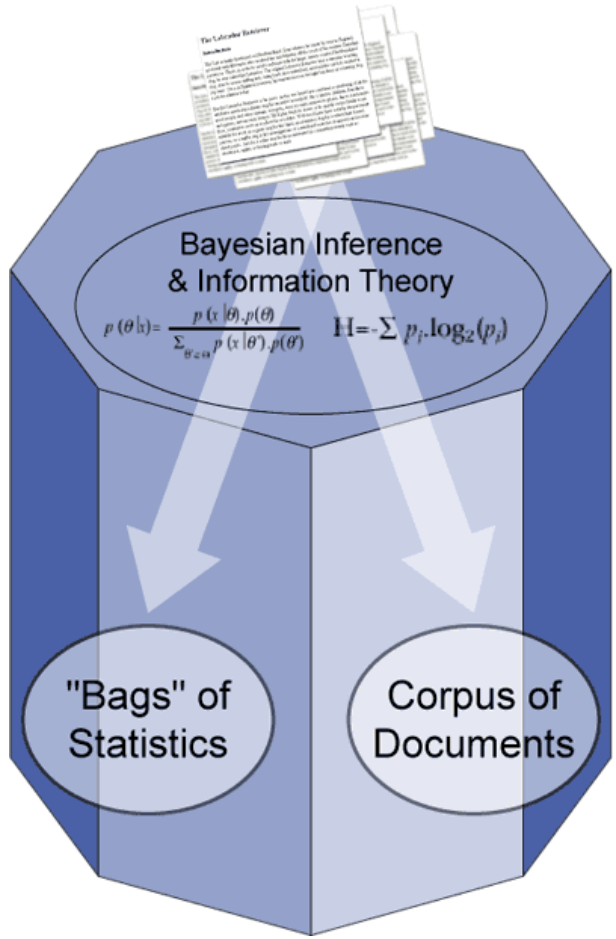
Proprietary Pattern Matching Technology

=

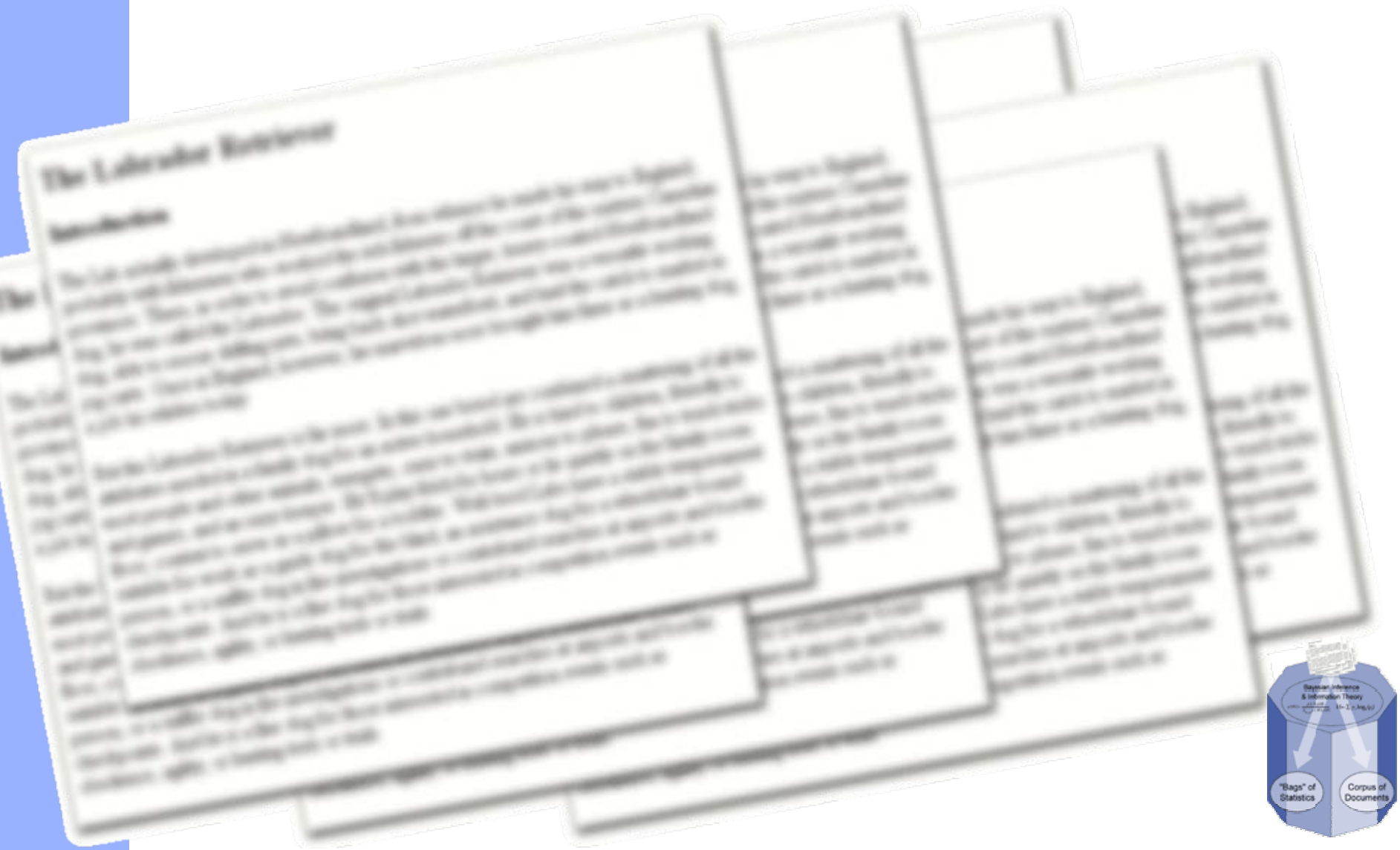
- Based on research from Cambridge University
- Algorithm to extract “concepts” from text and learn
- Language independent
- Significant intellectual property content
- **Data Agnostic!**

Statistics Generation from The Corpus

Using Bayesian Inference and Shannon's Information Theory, Autonomy builds "Bags" of statistics from a corpus of documents



IDOL Server Identifies Key Concepts



IDOL Server Identifies Key Concepts

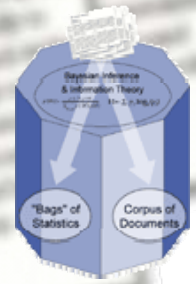
The Labrador Retriever

Introduction

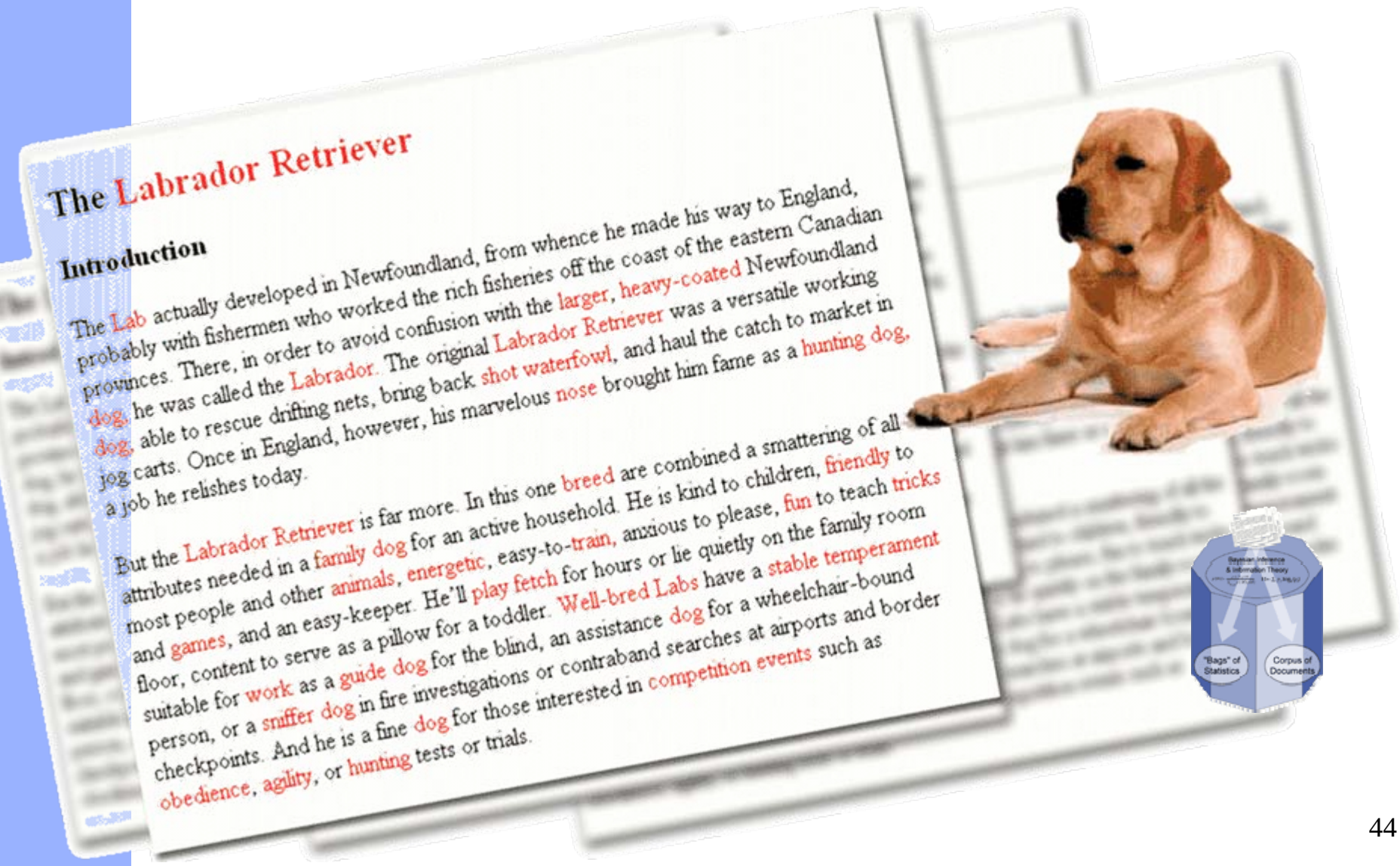
The Lab actually developed in Newfoundland, from whence he made his way to England, probably with fishermen who worked the rich fisheries off the coast of the eastern Canadian provinces. There, in order to avoid confusion with the larger, heavy-coated Newfoundland dog, he was called the Labrador. The original Labrador Retriever was a versatile working dog, able to rescue drifting nets, bring back shot waterfowl, and haul the catch to market in jog carts. Once in England, however, his marvelous nose brought him fame as a hunting dog, a job he relishes today.

But the Labrador Retriever is far more. In this one breed are combined a smattering of all attributes needed in a family dog for an active household. He is kind to children, friendly to most people and other animals, energetic, easy-to-train, anxious to please, fun to teach tricks and games, and an easy-keeper. He'll play fetch for hours or lie quietly on the family room floor, content to serve as a pillow for a toddler. Well-bred Labs have a stable temperament suitable for work as a guide dog for the blind, an assistance dog for a wheelchair-bound person, or a sniffer dog in fire investigations or contraband searches at airports and border checkpoints. And he is a fine dog for those interested in competition events such as obedience, agility, or hunting tests or trials.

拉布拉多猎狗

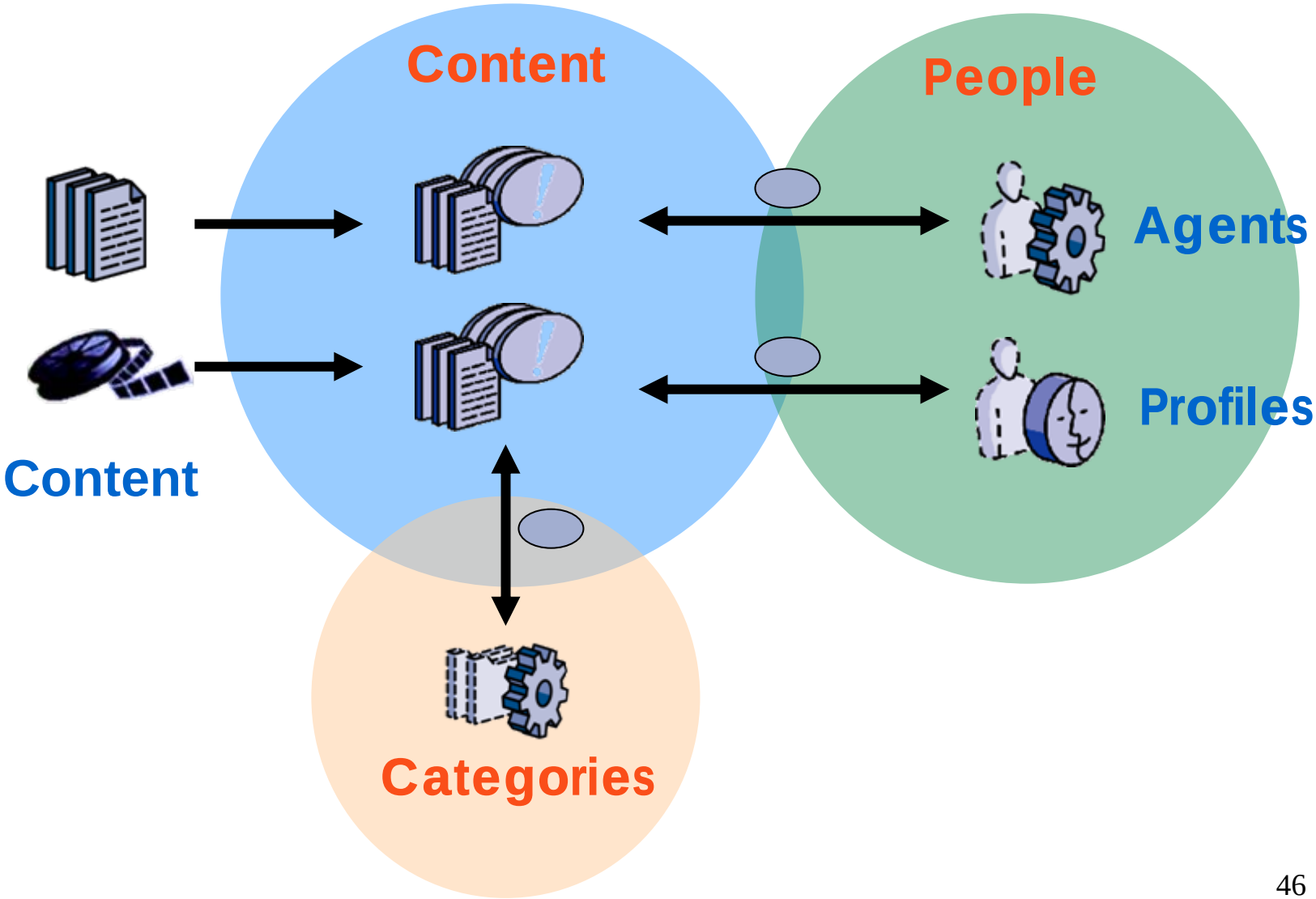


IDOL Server Identifies Key Concepts



And Stores Statistics on Document

IDOL Server Operations



Autonomy Product Overview

Business Solutions

- Video & Audio
- Call Center
- Business Intelligence
- Portal
- Compliance
- CRM-KM-ERP-DM

Application Builder

OEM Integration

Intelligent Data Operating Layer

Administration

Security

Voice & Video

Unstructured

XML & Structured

Audio

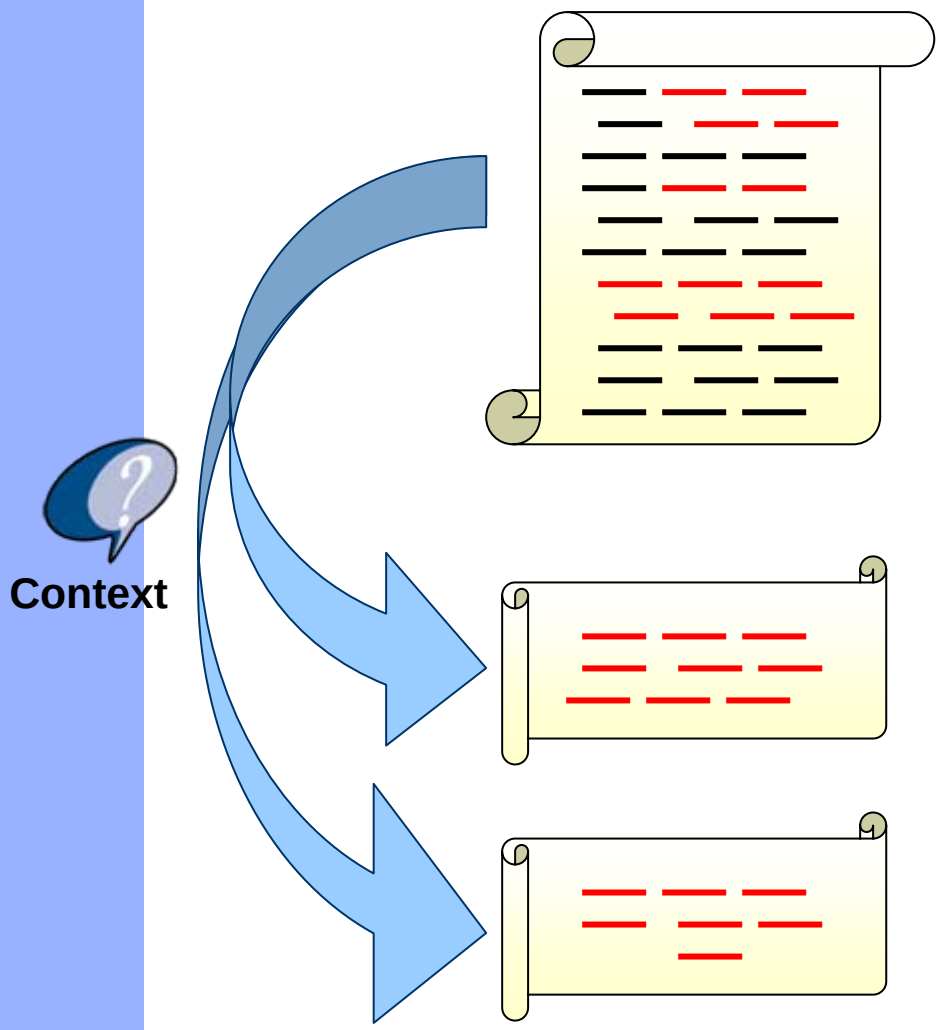
Video

People

Connectors

LCM

Automatic Summarization

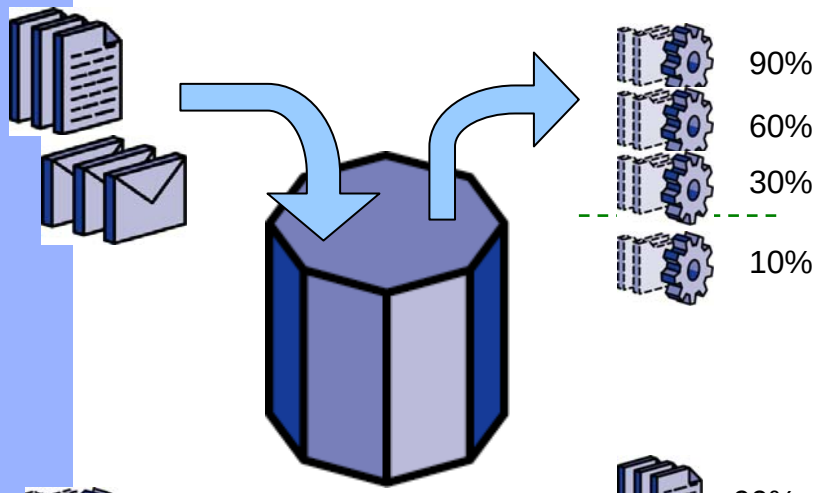


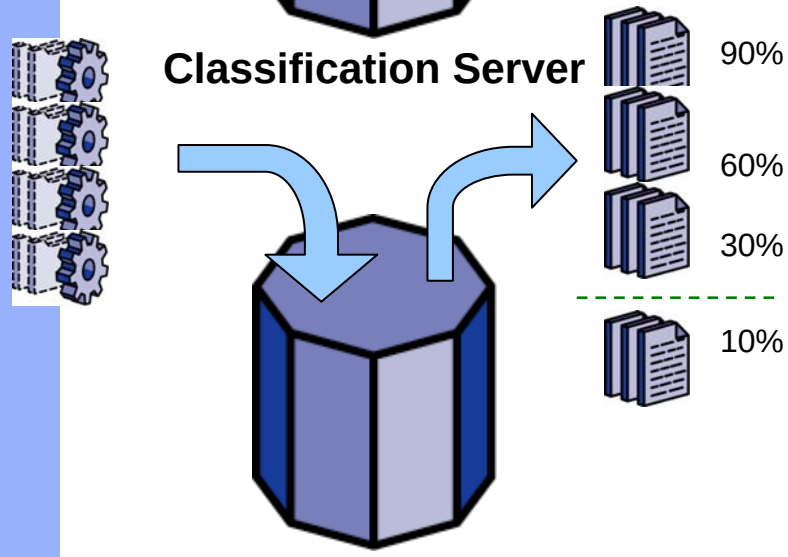
- CONCEPTUAL
- CONTEXTUAL

- AUTOMATIC
- REALTIME
- CROSS-DATA SOURCE
- CROSS-DATA FORMAT

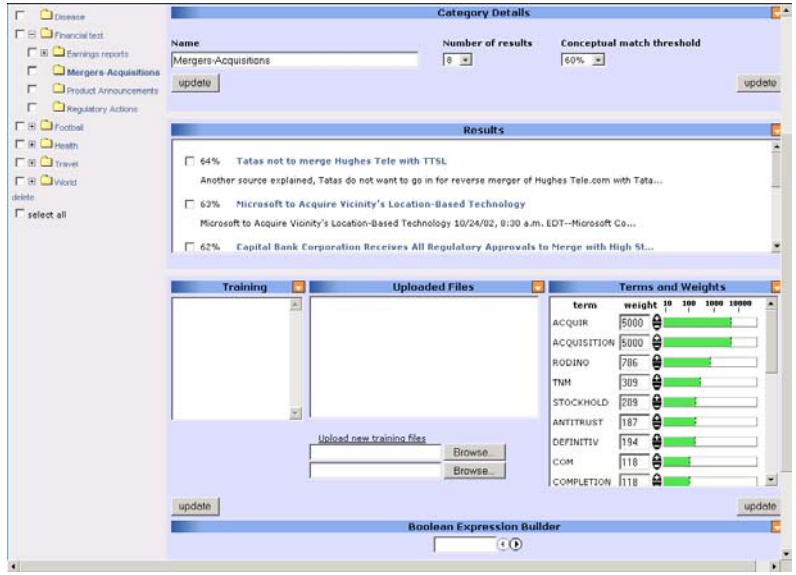
- **Quick Summary 1st N Lines**
- **Concept Summary**
- **Concept Query Summary**

Automatic Categorization

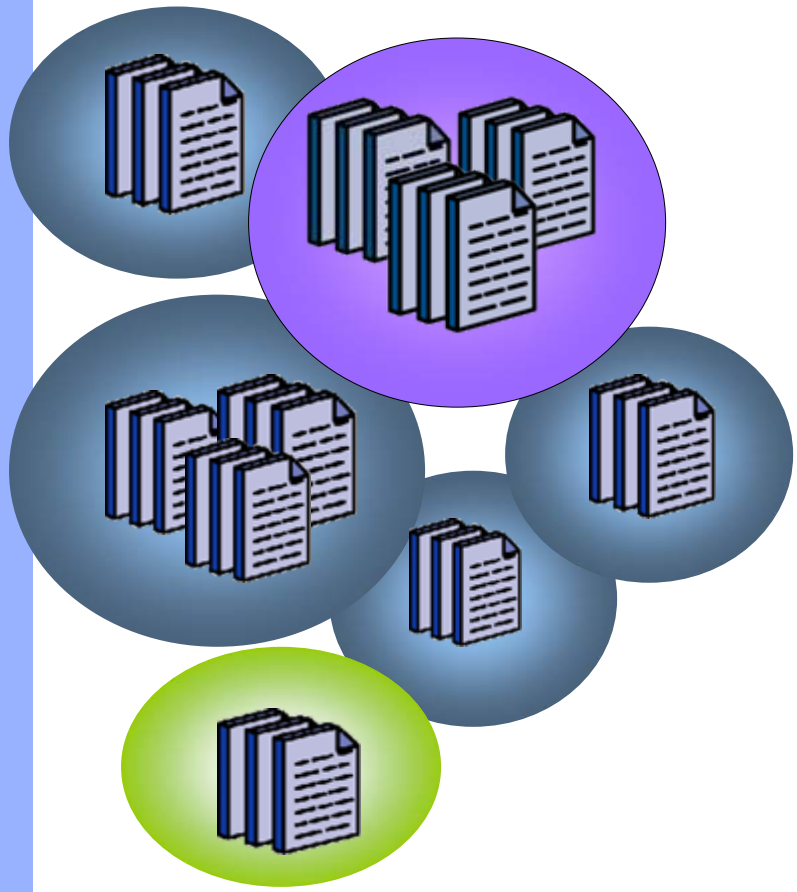




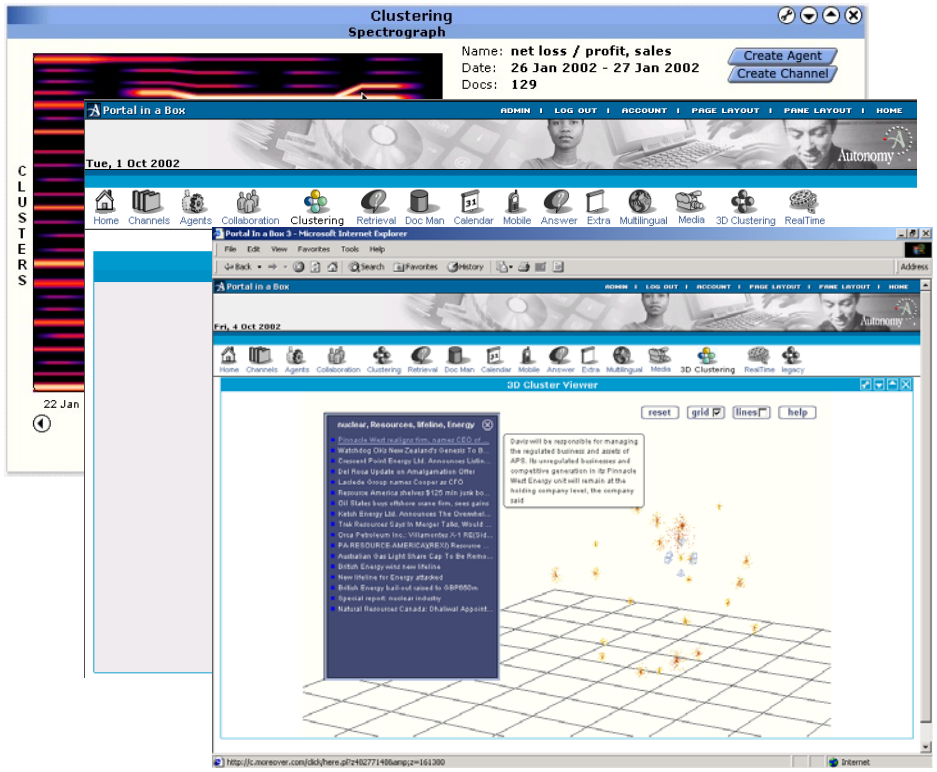
- TRAIN BY EXAMPLE/ BOOLEAN
- 7000 WEB CATEGORIES
- 500 NEWS CATEGORIES
- FTSE World Global CATEGORIES
- CROSS-DATA SOURCE/FORMAT
- LEGACY COMPATIBILTY



Clustering

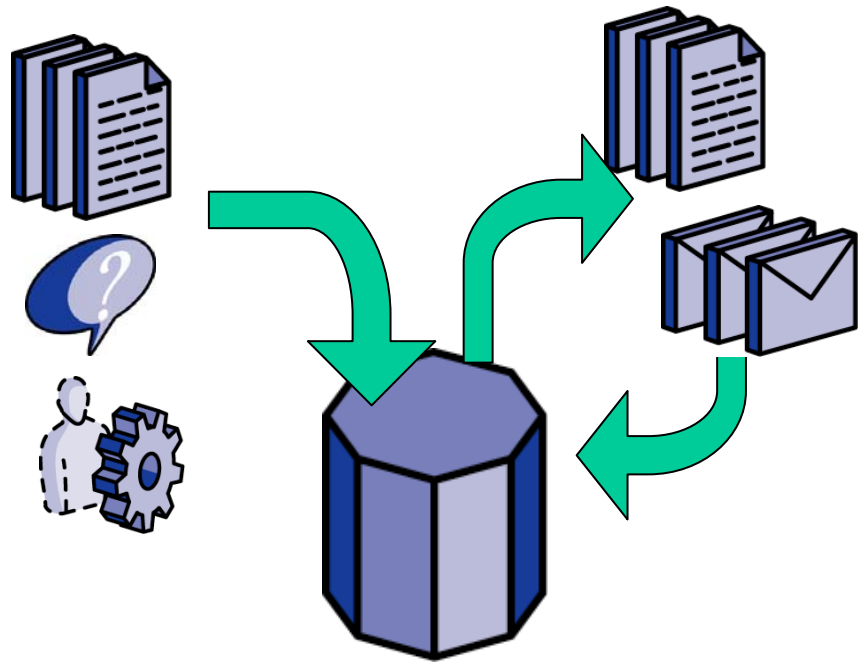


- CLUSTER DATA, PROFILES
- WHAT'S HOT / WHAT'S NEW
- XML OUTPUT
- VISUALIZATION

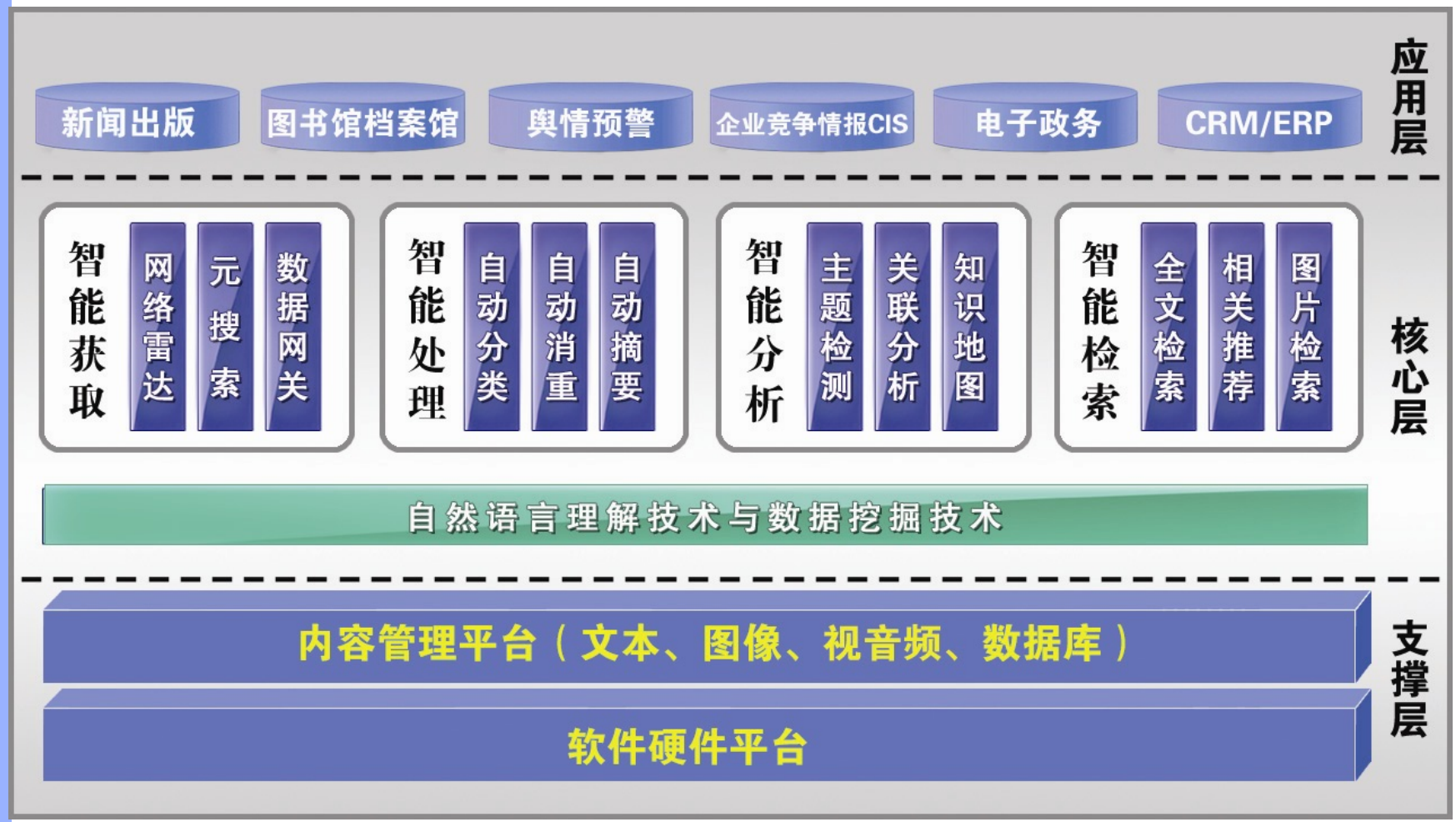


Retrieval

- CONCEPTUAL
- NATURAL LANGUAGE
- KEYWORD
- Boolean including
  - AND, NOT, OR, EOR, XOP, NEARnn, NEARnn, WNEARnn, BEFORE, AFTER, () BRACKETED.
- Parametric
- Federated
- Fuzzy Search
- Proximity Search
- Field Search
- Soundex
- Wildcard
- Cross-lingual / Multi-lingual
- Cross Data format/Source



方正智思


应用层
新闻出版
图书馆档案馆
舆情预警
企业竞争情报CIS
电子政务
CRM/ERP
核心层
智能获取
网络雷达
元搜索
数据网关
智能处理
自动分类
自动消重
自动摘要
智能分析
主题检测
关联分析
知识地图
智能检索
全文检索
相关推荐
图片检索
自然语言理解技术与数据挖掘技术
支撑层
内容管理平台（文本、图像、视音频、数据库）
软件硬件平台

方正智思—研究论文

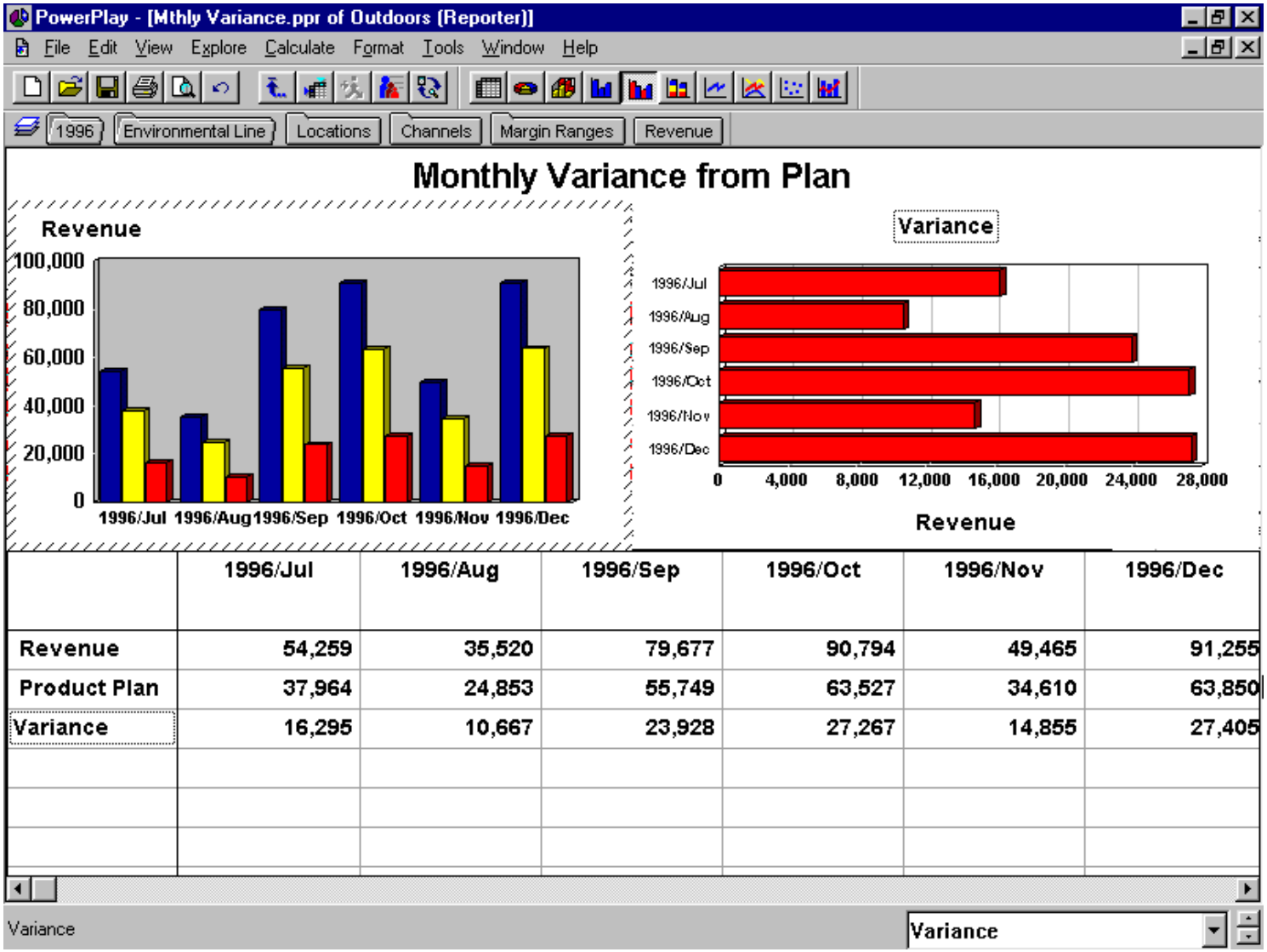
- 顶级会议论文(**Oral Paper**)
  - **IJCAI2007**：Manifold-Ranking Based Topic-Focused Multi-Document Summarization
  - **SIGIR2007**: CollabSum: Exploiting Multiple Document Clustering for Collaborative Single Document Summarizations
  - **ACL2007**: Towards an Iterative Reinforcement Approach for Simultaneous Document Summarization and Keyword Extraction
  - **AAAI2007**: Single Document Summarization with Document Expansion
  - **SIGIR2008**:Multi-Document Summarization Using Cluster-based Link Analysis
  - **AAAI2008**:Single Document Keyphrase Extraction Using Neighborhood Knowledge
- 国际期刊论文(**Regular Paper**)
  - Information Processing & Management
  - Information Retrieval
  - Knowledge and Information Systems
  - Information Sciences

User Interfaces for Text Mining

User Interfaces for Text Mining

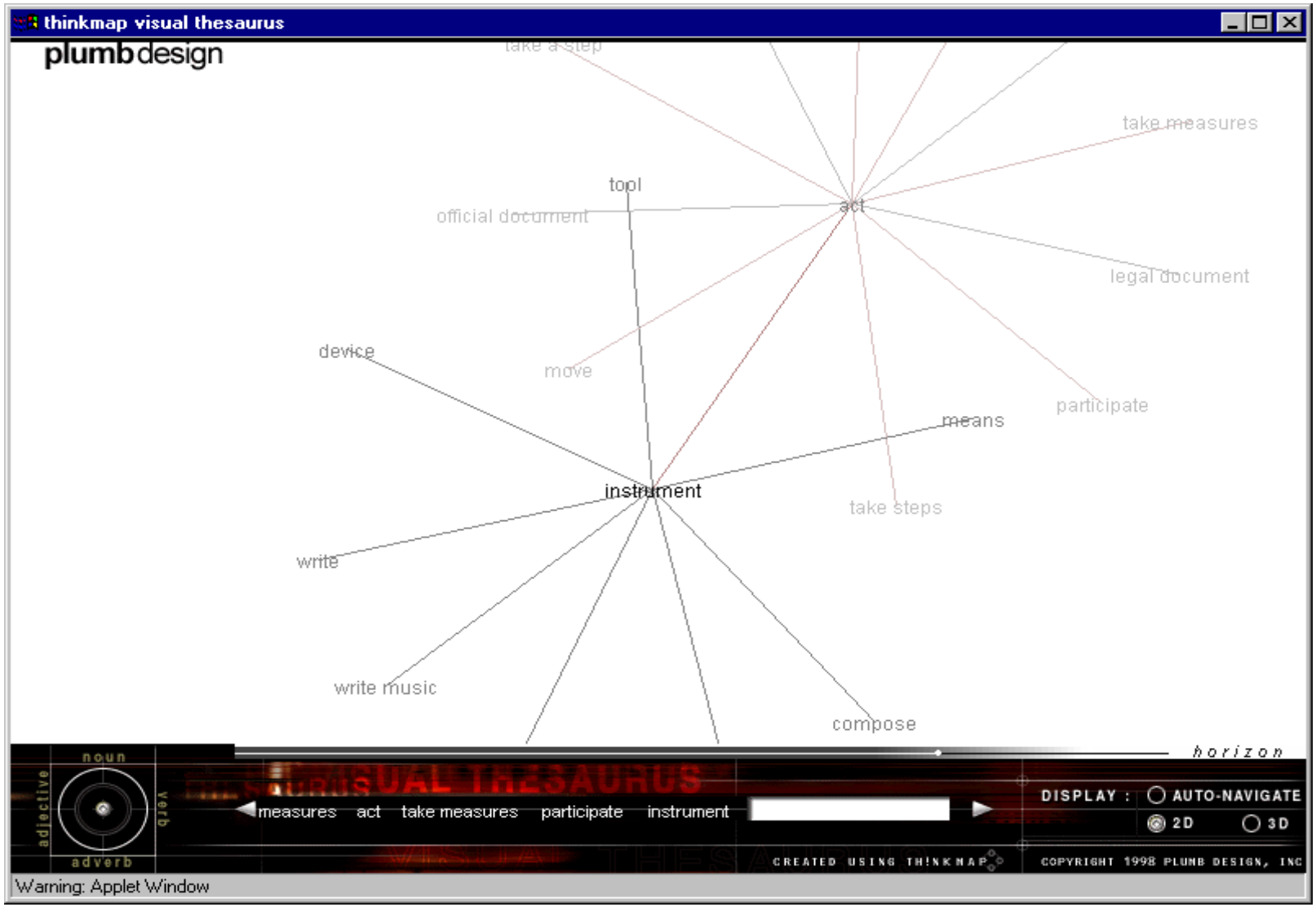
- Need some way to present results of Text Mining in an intuitive, easy to manage form.
- Options:
  - Conventional text “lists” (1D)
  - Charts and graphs (2D)
  - Advanced visualization tools (3D+)
    - Network maps
    - Landscapes
    - 3d “spaces”

Visualization: Charts and Graphs

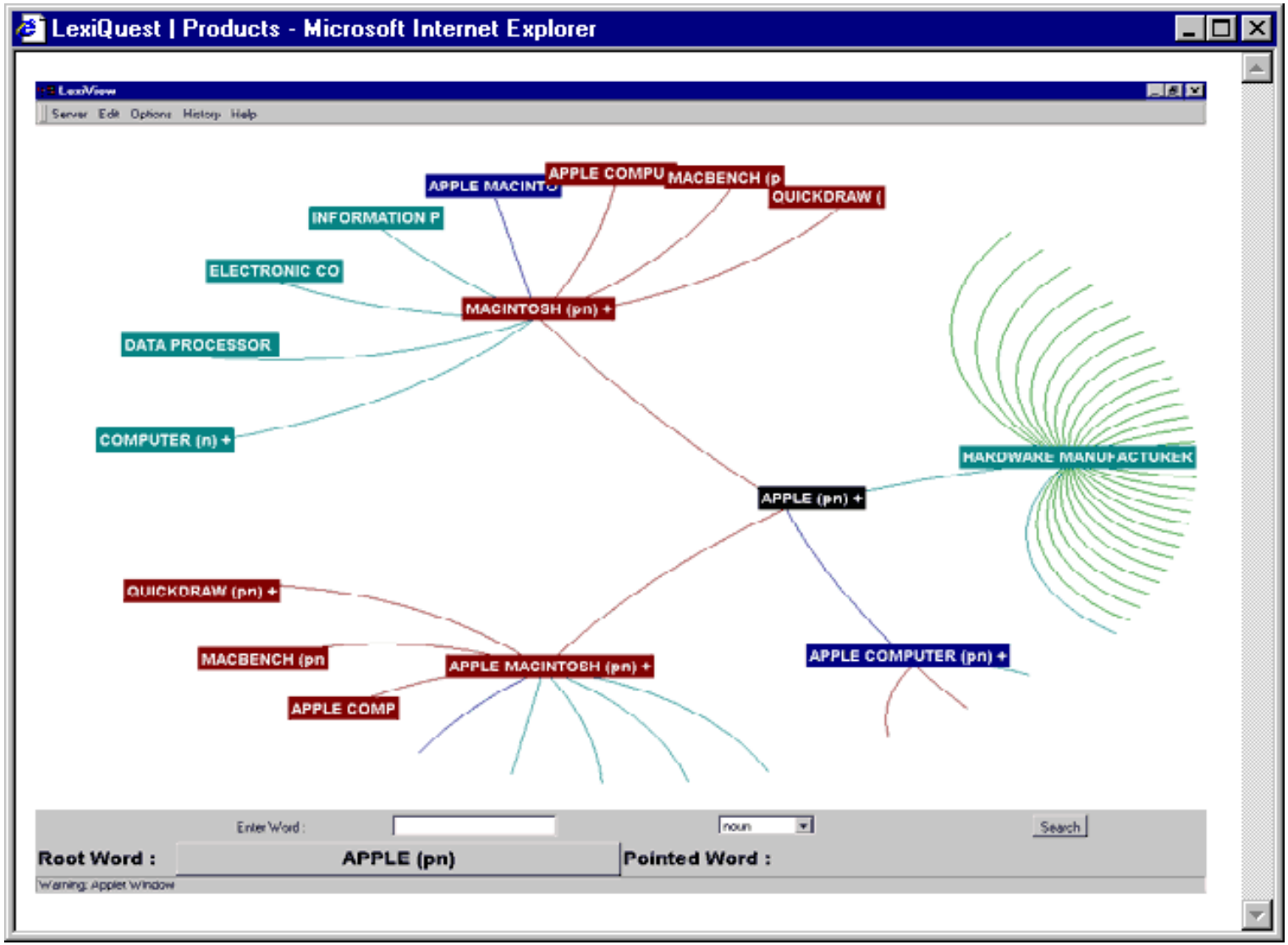


	1996/Jul	1996/Aug	1996/Sep	1996/Oct	1996/Nov	1996/Dec
Revenue	54,259	35,520	79,677	90,794	49,465	91,255
Product Plan	37,964	24,853	55,749	63,527	34,610	63,850
Variance	16,295	10,667	23,928	27,267	14,855	27,405

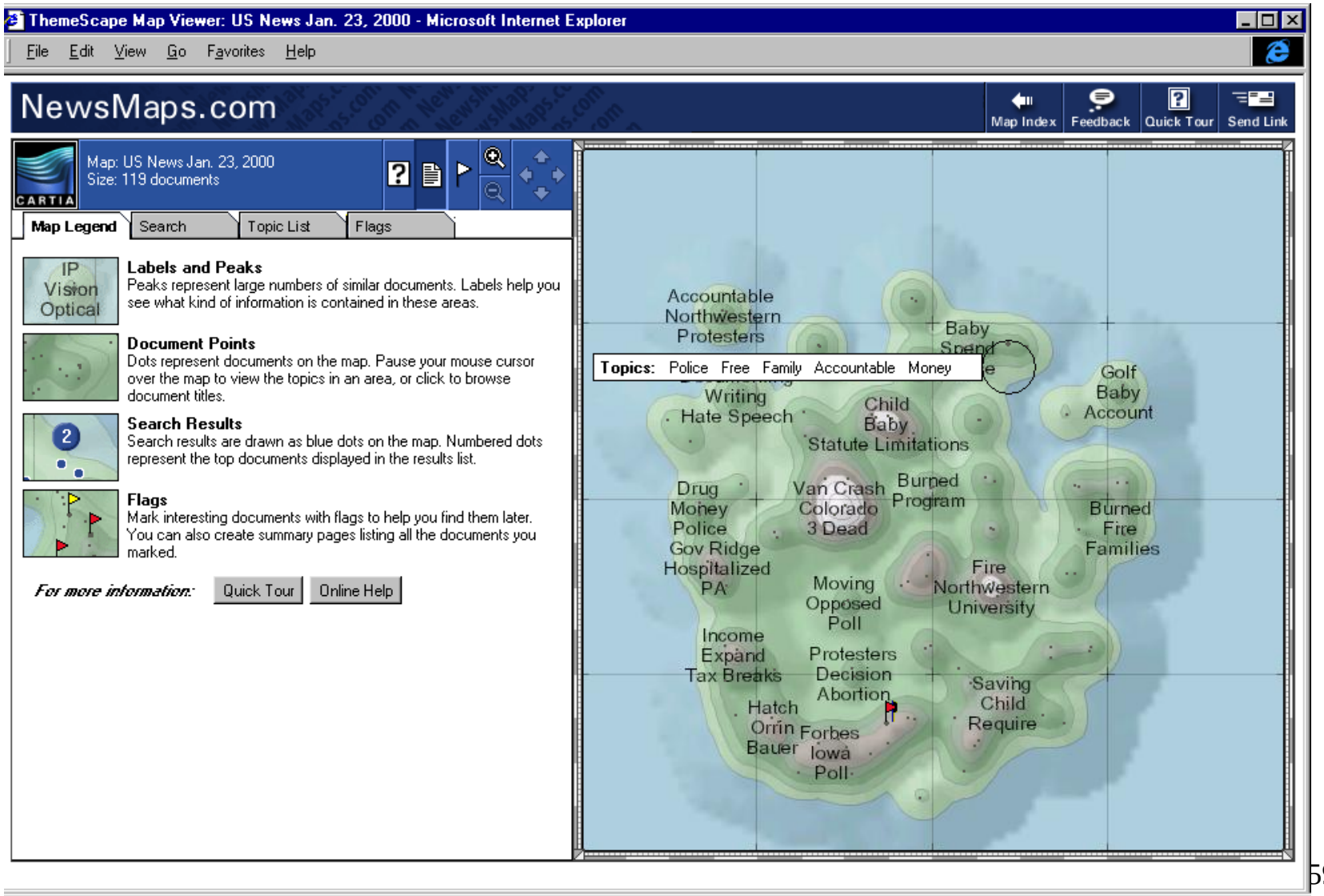
Visualization: Network Maps



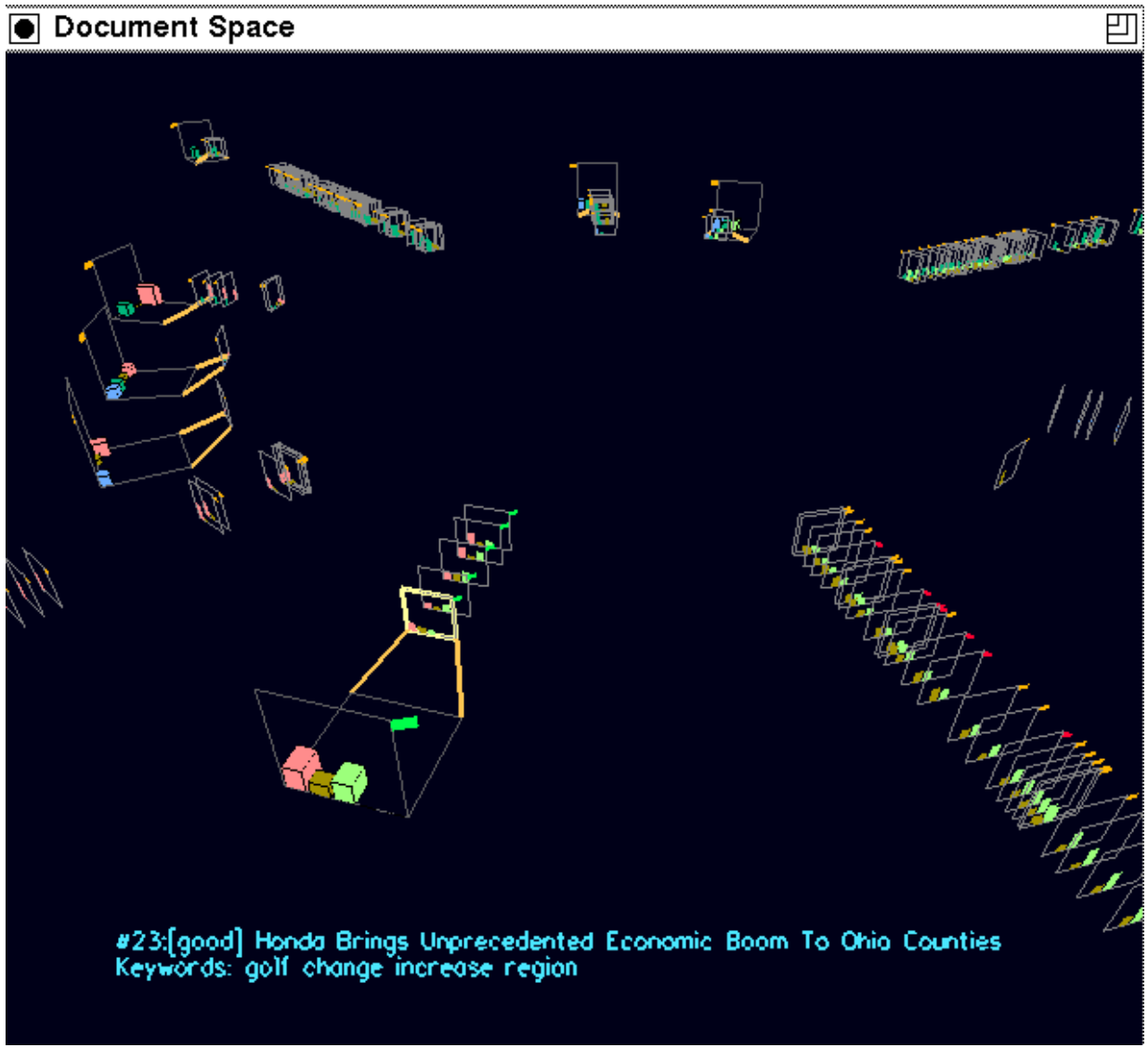
Visualization: Network Maps



Visualization: Landscapes



Visualization: 3D Spaces



互联网采集分析系统 - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 搜索 收藏夹 媒体

地址(D) http://10.1.2.140/Index.action 转到 链接

互联网采集分析系统

首页 | 关注信息 | 热点信息 | 焦点专题 | 超级搜索 | 个人工作平台 | 报告 | 探针 | 系统配置 | 违规转载 | 注销

新闻 跟帖 论坛 博客 Wap 视频

昨日	今日	总计
407354	219049	9206万

预警信息 即时消息

- 国家统计局... [5分钟前]
- 现场目击：... [6分钟前]
- 鲁宁：这一... [8分钟前]
- 专家提醒：... [8分钟前]
- 答谢球童擦... [8分钟前]
- 印度中国争... [9分钟前]
- 俄媒体披露... [10分钟前]
- 普通人见证... [10分钟前]

查看全部

上升最快搜索词

	关键字	排名上升
1	地震捐款	5080
2	1069999301	5010
3	阿坝州政府网	4970
4	中国红十字会总会	4417
5	捐款倡议书	4237
6	壹基金	4030
7	中华慈善总会	3997

全网搜索 搜索

关注信息 >>MORE

- [中央有... 胸有积雷而面如平湖者,可拜上... [05/15 06:25]
- [新疆恐... 日志 [05/15 06:06]
- [物价上... Hello World: ZZ 屁股与脑袋 [05/15 05:30]
- [许霆案] 大猩猩de私生活:从许霆案看到... [05/15 05:27]
- [户籍制度] 日志 [05/15 05:22]
- [粮库空转] 日志 [05/15 05:13]
- [改革开... 记住秋的香:庆北大校庆110周年 [05/15 05:01]
- [阿坝州... 日志 [05/15 05:01]
- [手足口... 汶川地震触动沪深股市 5月12... [05/15 04:04]
- [改革开... 小房说事:zz一篇让人思考的文章 [05/15 03:26]
- [暂住证... DHT-CSRC:著名拳王泰森申请移... [05/15 02:46]
- [涉奥舆情] 中国网民呼吁抵制法货 家乐福... [05/15 02:25]

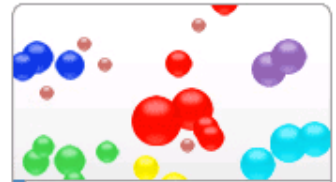
热点信息 >>MORE

- [财经新闻] - 多家公司再报“平安” [95-338]
- [时政新闻] - 灾难无情 河南省政法部门心系灾区 ... [5-132]
- [环球资讯] - 朝鲜总理会见中国大使就四川地震慰问... [4-7]
- [财经新闻] - 国寿开通地震快速理赔通道 [3-21]
- [时政新闻] - 中直机关踊跃开展向地震灾区人民送温... [3-6]
- [时政新闻] - 本网专访中国地震局 谈汶川地震相关情况 [2-58]
- [财经新闻] - 地震凸显社会责任 多家央企捐助四川... [2-107]
- [时政新闻] - 河南省213名抗震救灾医疗队奔赴四川灾区 [2-27]
- [社会新闻] - 救援队伍在北川县城展开拉网式搜索 [2-23]
- [时政新闻] - 1.1万名公安消防特勤官兵奋战灾区 [1-30]
- [时政新闻] - 四川震区降雨短暂停歇 [1-12]
- [环球资讯] - 俄罗斯向中国运送第二批地震救灾物资 [1-9]


今日关注信息地域分布图

一周内关注信息时间分布图


近期主题演化图

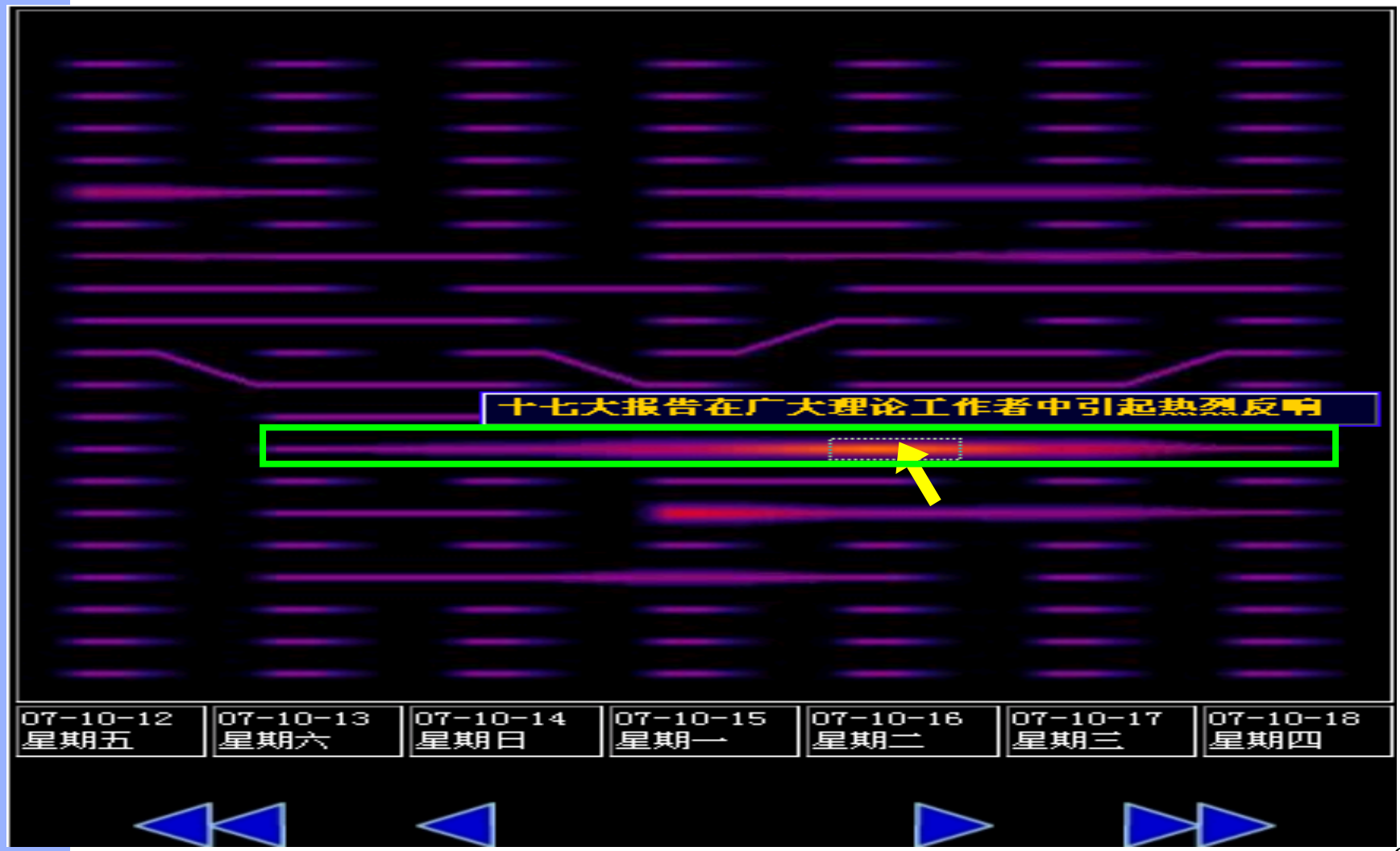

近期主题聚类图

个人定制信息 论坛群发 >> 更多

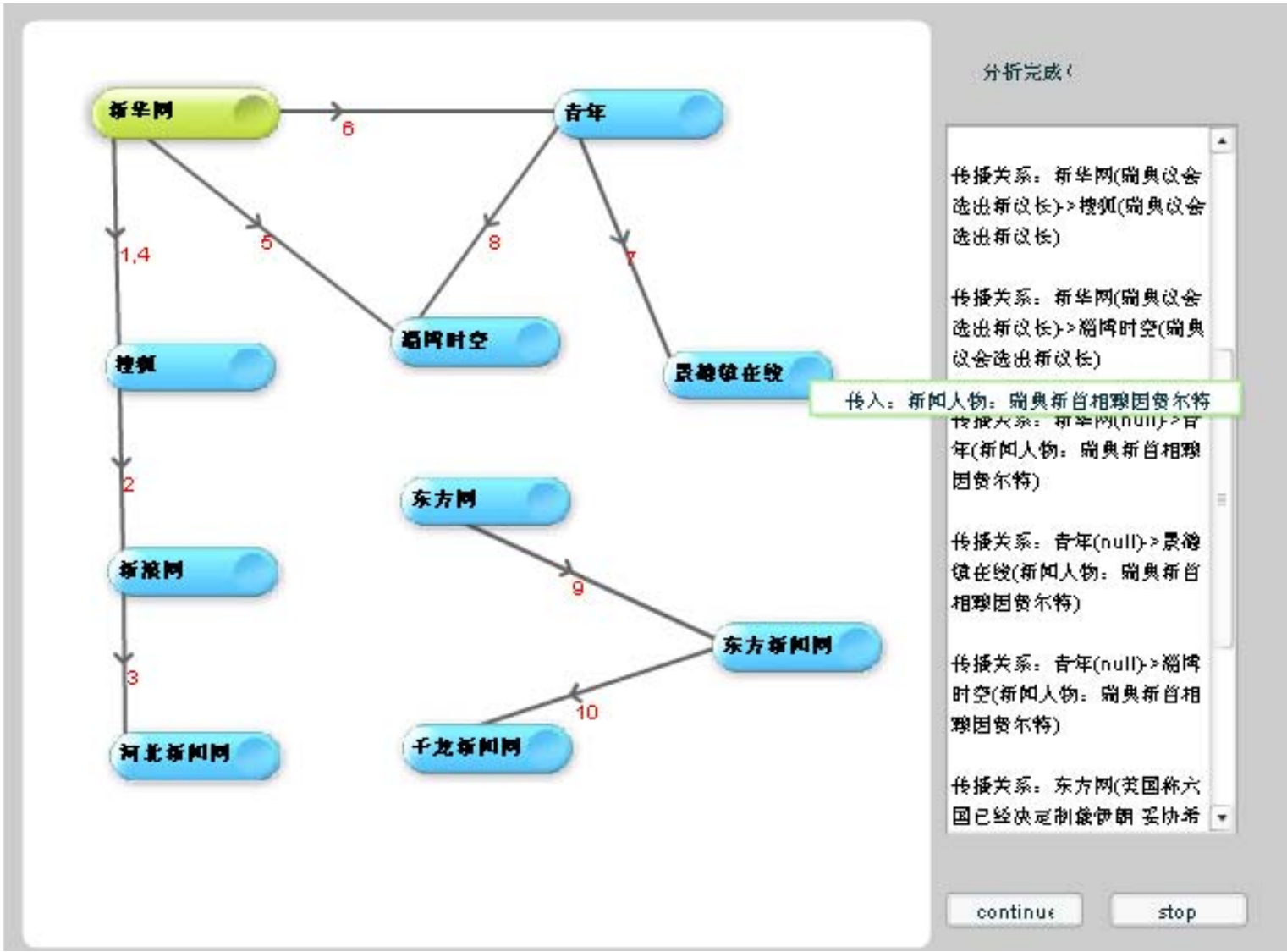
个人收藏夹

http://10.1.2.140/Focus.action?folderId=11&merge=true 可信站点

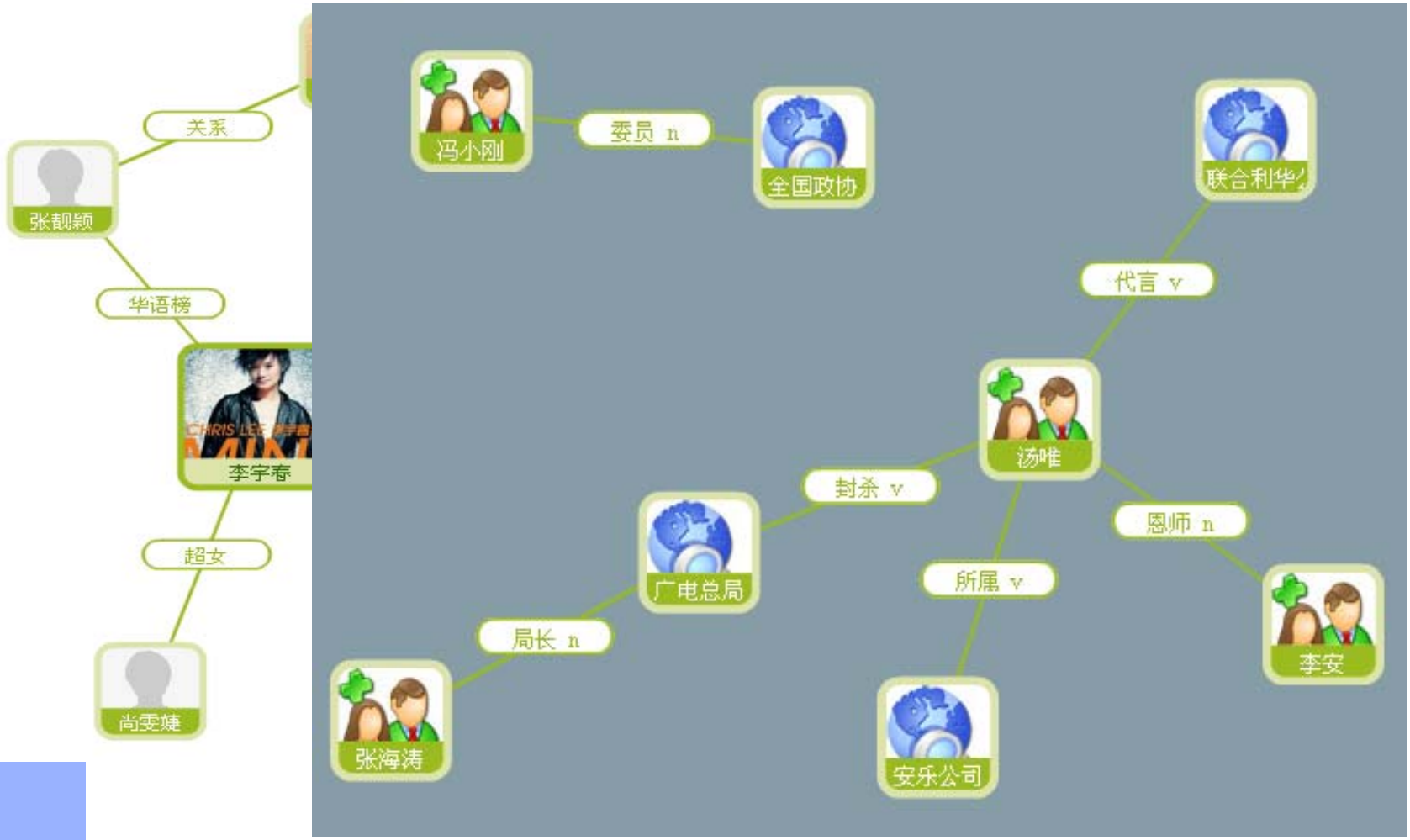
方正智思—话题演化关系分析



方正智思— 话题传播关系分析



方正智思— 实体关系抽取


关系
张靓颖
华语榜
李宇春
超女
尚雯婕
冯小刚
委员 n
全国政协
联合利华
代言 v
汤唯
封杀 v
广电总局
局长 n
张海涛
恩师 n
所属 v
李安
安乐公司

Any Question?