IBM Content Analytics Overview

IBM China / Hong Kong Limited

4 December 2012

Unlock valuable insight from content

What our clients are doing with Content Analytics

Understand what customers want before they ask.

Detect fraudulent claims before they are paid.

Dynamically deploy resources to the areas of greatest threat.

Save lives by quickly identifying critical safety defects.

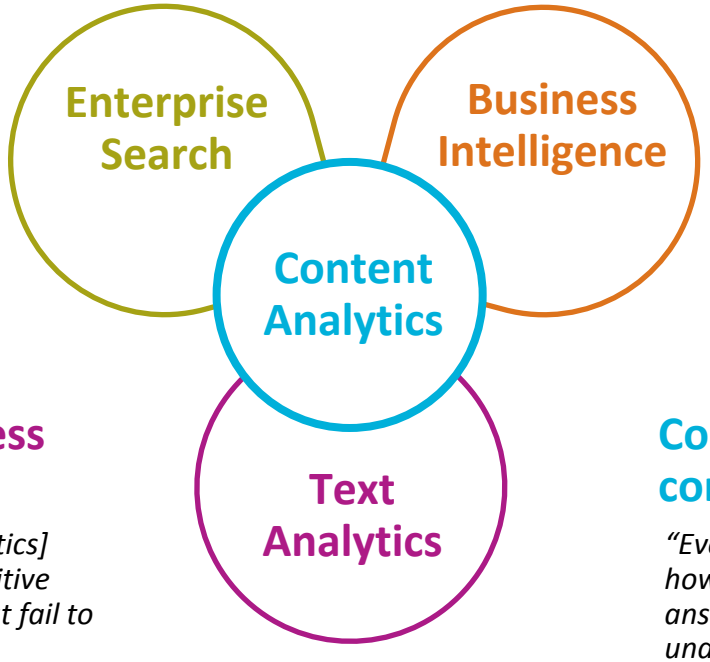
Are you unlocking the value of your unstructured content?

Traditional approaches are converging

More than keyword search is needed

"Making unstructured data searchable is now a presumed primary interface for applications of all kinds, as well as for intranets and content repositories."

– Whit Andrews, Rita Knox Gartner



Analyzing unstructured content no longer optional

"For many business process professionals, access to structured data, even when supported by BI or predictive analytics, lacks sufficient context for customer service, finance, and other areas where communications with customers involves many channels"

– Craig Le Clair Forrester

Increasing in business importance

"Early adopters of [text analytics] are already gaining a competitive advantage. Organizations that fail to do so will be at risk."

– Sue Feldman IDC

Converging toward content analytics

"Every enterprise should understand how content analytics can produce answers to its critical questions; understanding this now will make it possible to exploit these tools as their availability proliferates."

– Rita Knox Gartner

Going from raw information to rapid insight

Uncover business insight through unique visual-based approach

Aggregate and extract from multiple sources

… to form large **text**-based collections from multiple internal and external sources (and types), including ECM repositories, structured data, social media and more.

Organize, analyze and visualize

… enterprise **content** (and data) by identifying trends, patterns, correlations, anomalies and business context from collections.

Search and explore to derive insight

… from collections to confirm what is suspected or uncover something new without being forced to build models or deploy complex systems.

IBM Content Analytics adds value to …

Healthcare Analytics

- ***Analyzing:*** *E-Medical records, hospital reports*
- ***For:*** *Clinical analysis; treatment protocol optimization*
- ***Benefits:*** *Better management of chronic diseases; optimized drug formularies; improved patient outcomes*

Crime Analytics

- ***Analyzing:*** *Case files, police records, 911 calls…*
- ***For:*** *Rapid crime solving & crime trend analysis*
- ***Benefits:*** *Safer communities & optimized force deployment*

Automotive Quality Insight

- ***Analyzing:*** *Tech notes, call logs, online media*
- ***For:*** *Warranty Analysis, Quality Assurance*
- ***Benefits:*** *Reduce warranty costs, improve customer satisfaction, marketing campaigns*

Customer Care

- ***Analyzing:*** *Call center logs, emails, online media*
- ***For:*** *Buyer Behavior, Churn prediction*
- ***Benefits:*** *Improve Customer satisfaction and retention, marketing campaigns, find new revenue opportunities*

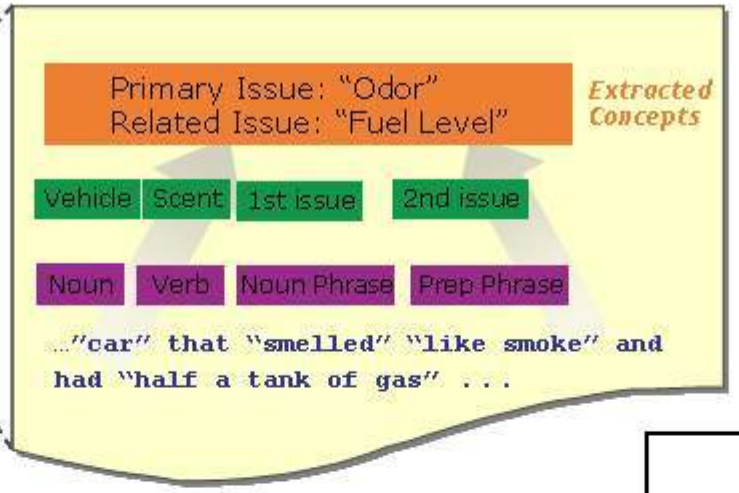
Insurance Fraud

- ***Analyzing:*** *Insurance claims*
- ***For:*** *Detecting Fraudulent activity & patterns*
- ***Benefits:*** *Reduced losses, faster detection, more efficient claims processes*

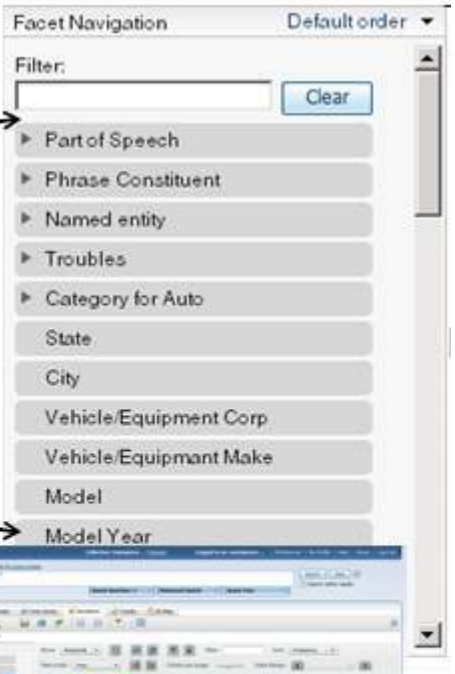
Social Media for Marketing

- ***Analyzing:*** *Call center notes, SharePoint, multiple content repositories*
- ***For:*** *churn prediction, product/brand quality*
- ***Benefits:*** *Improve consumer satisfaction, marketing campaigns, find new revenue opportunities or product/brand quality issues*

Overview



UIMA Pipeline + Annotators



Source Information

Corporate (Contact Center, Test Data, Dealer notes, ECM, etc.) and External (NHTSA, Edmunds, Consumer Reports, MotorTrend etc.)

Analyzed Content (and Data)

Fine grain control over the entities and facets that are created

ECM Solutions

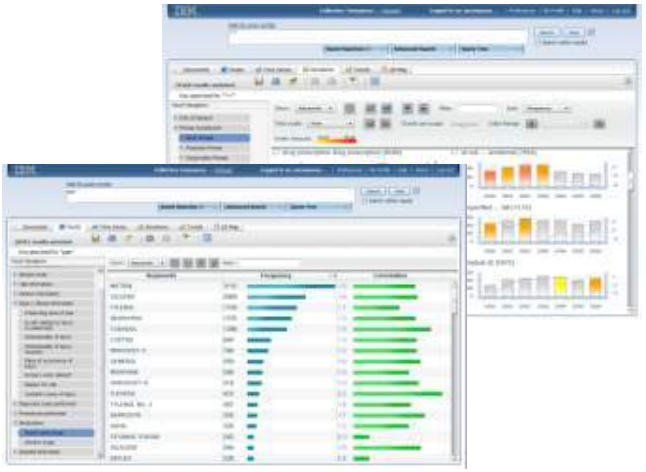
Text Analytics is the basis for Content Analytics

What is Text Analytics?

Text Analytics (NLP*) describes a set of linguistic, statistical, and machine learning techniques that allow text to be analyzed and key information extraction for business integration.

PC 143 (Hunter)
15 June 2006 23:47
Suspect identified himself as John Setsuko. Matched description given by night club doorman (IC1, Male, Ag 22-24 yrs, blue Everton shirt). Stopped whilst driving White Ford Mondeo, W563 WDL. Address given as 22 East Dene Ridge, Copdock, Ipswich. Searched at scene and found in possession of 1oz Cannabis Resin and lockable pocket knife.

Arresting_Officer	PC 143
Arrest_Date_Time	15/06/2006 : 23:47
Suspect_Forename	John
Suspect_Surname	Setsuko
Suspect_VRN	W563WDL
Suspect_Vehicle_Color	White
Suspect_Vehicle_Make	Ford Mondeo
Suspect_Addr_Street	22 East Dene Ridge
Suspect_Addr_Town	Ipswich
Evidence_1_Description	1 oz Cannabis Resin
Classification	Drug possession

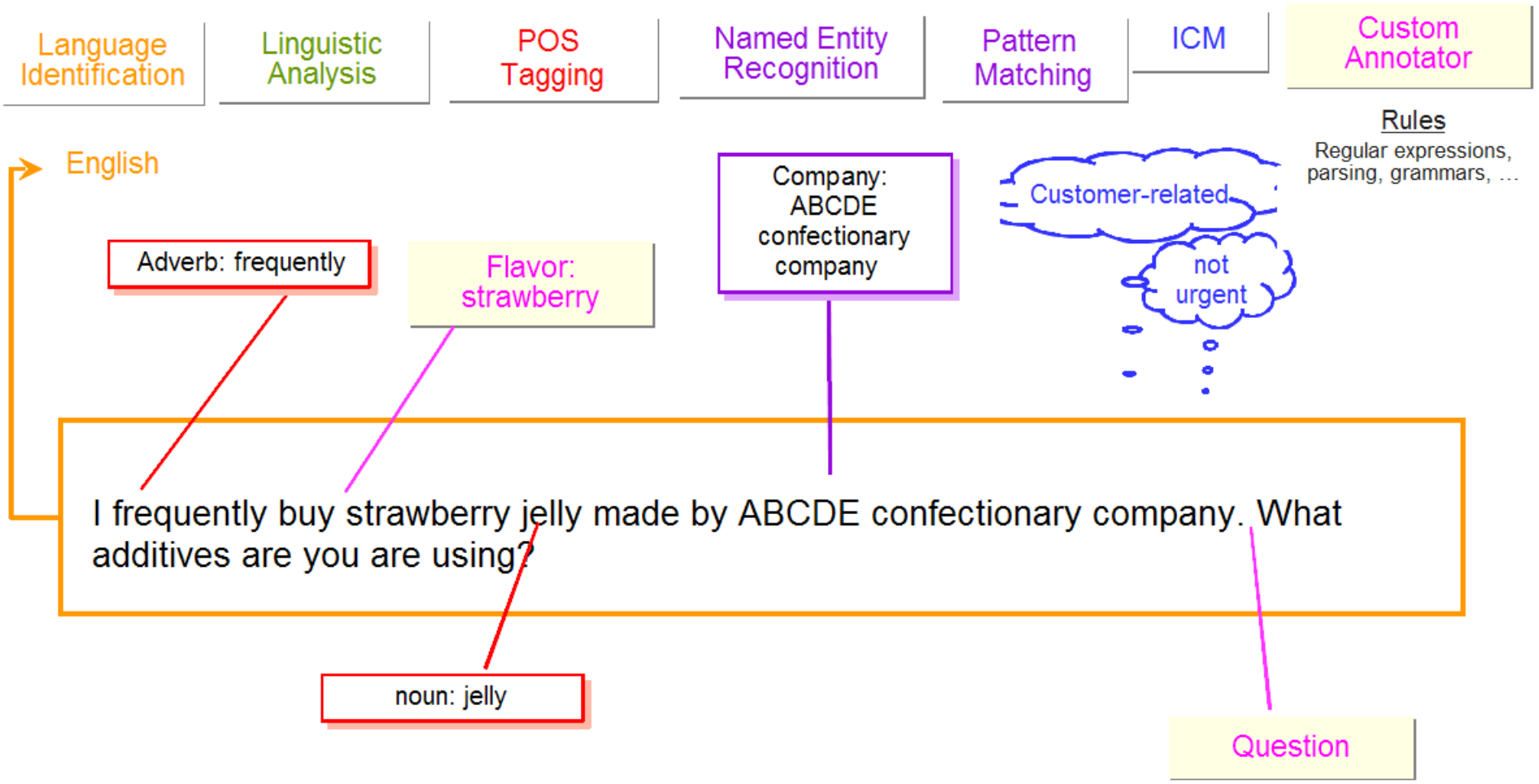


What is Content Analytics?

Content Analytics (Text Analytics + Mining) refers to the text analytics process plus the ability to visually identify and explore trends, patterns, and statistically relevant facts found in various types of content spread across internal and external content sources.

What do ICA annotators do?

annotator- a software component that performs linguistic analysis tasks and produces and records annotations



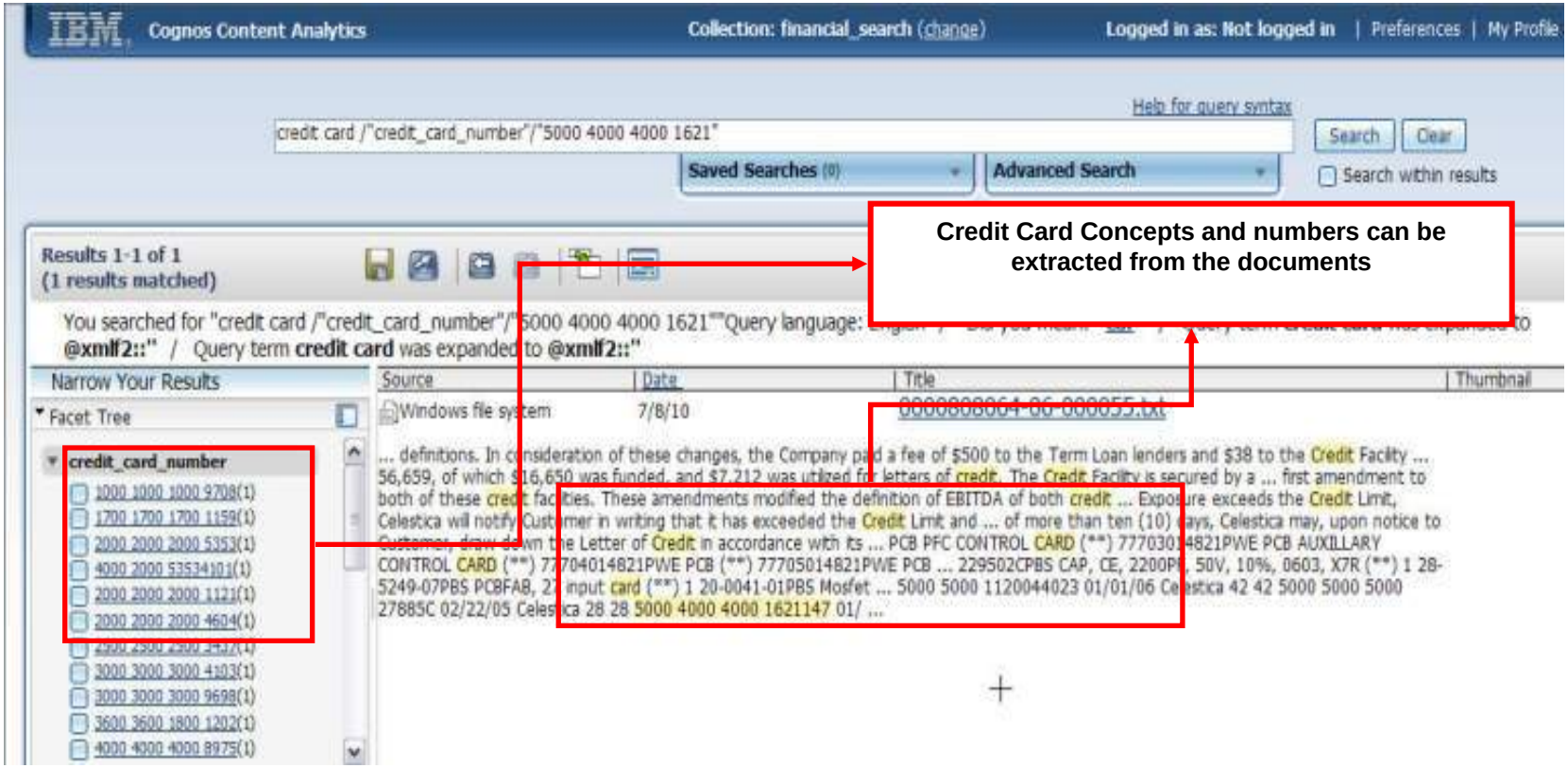
Text Miner applications

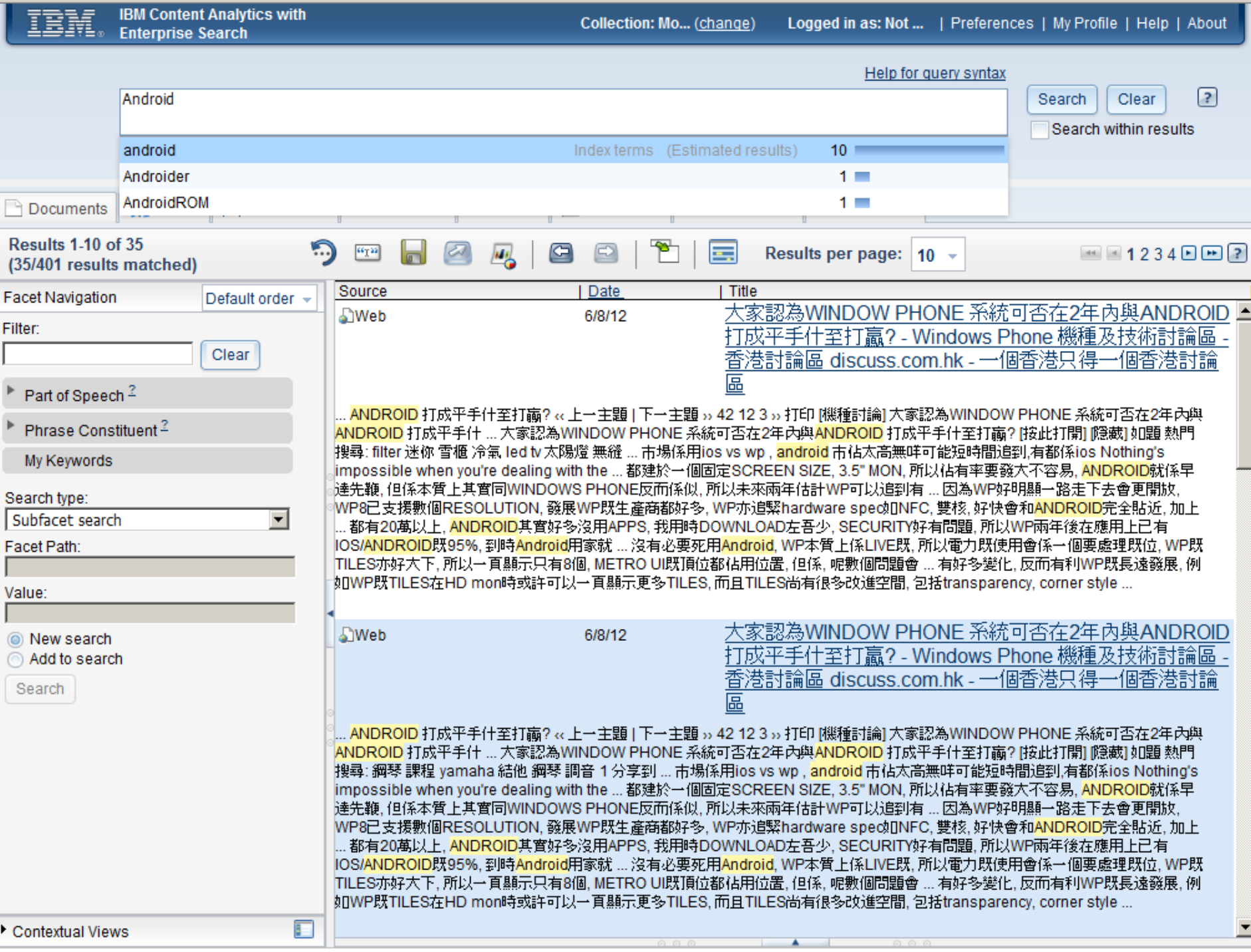
8 views for analysis, exploration and investigation
Dynamically search and explore content for new business insight
Powerful solution modeling and support for advanced classification tools for more accurate and deeper insight
Deliver rapid insight to other systems, users and applications, like Cognos BI or Case Manager, for complete business view

Content Modeling Example

extract regular expressions

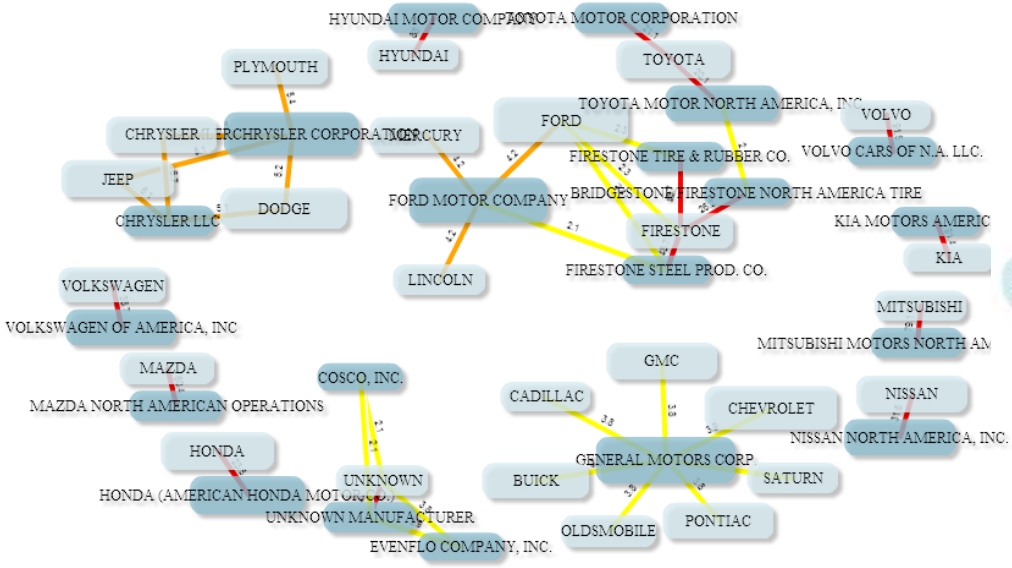
- Identify documents that contain identifiable information like credit cards numbers, social security numbers, etc.



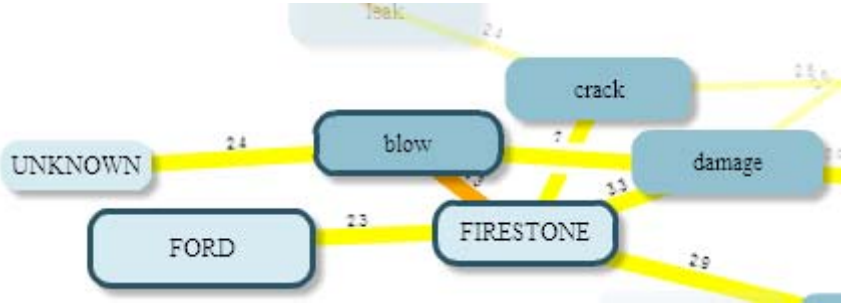

IBM Content Analytics with Enterprise Search
Collection: Mo... (change)
Logged in as: Not ...
Preferences | My Profile | Help | About
Help for query syntax
Android
Search
Clear
Search within results
android Index terms (Estimated results) 10
Androider 1
AndroidROM 1
Documents
Results 1-10 of 35
(35/401 results matched)
Results per page: 10
1 2 3 4
Facet Navigation
Default order
Filter:
Clear
Part of Speech 2
Phrase Constituent 2
My Keywords
Search type:
Subfacet search
Facet Path:
Value:
New search
Add to search
Search
Contextual Views
Source | Date | Title
Web 6/8/12
大家認為WINDOW PHONE 系統可否在2年內與ANDROID打成平手什至打贏? - Windows Phone 機種及技術討論區 - 香港討論區 discuss.com.hk - 一個香港只得一個香港討論區
... ANDROID 打成平手什至打贏? « 上一主題 | 下一主題 » 42 12 3 » 打印 [機種討論] 大家認為WINDOW PHONE 系統可否在2年內與ANDROID 打成平手什 ... 大家認為WINDOW PHONE 系統可否在2年內與ANDROID 打成平手什至打贏? [按此打開] [隱藏] 如題 熱門搜尋: filter 迷你 雪櫃 冷氣 led tv 太陽燈 無縫 ... 市場係用ios vs wp , android 市佔太高無咩可能短時間追到,有都係ios Nothing's impossible when you're dealing with the ... 都建於一個固定SCREEN SIZE, 3.5" MON, 所以佔有率要發大不容易, ANDROID就係早達先鞭, 但係本質上其實同WINDOWS PHONE反而係似, 所以未來兩年估計WP可以追到有 ... 因為WP好明顯一路走下去會更開放, WP8已支援數個RESOLUTION, 發展WP既生產商都好多, WP亦追緊hardware spec如NFC, 雙核, 好快會和ANDROID完全貼近, 加上 ... 都有20萬以上, ANDROID其實好多沒用APPS, 我用時DOWNLOAD左吾少, SECURITY好有問題, 所以WP兩年後在應用上已有IOS/ANDROID既95%, 到時Android用家就 ... 沒有必要死用Android, WP本質上係LIVE既, 所以電力既使用會係一個要處理既位, WP既TILES亦好大下, 所以一頁顯示只有8個, METRO UI既頂位都佔用位置, 但係, 呢數個問題會 ... 有好多變化, 反而有利WP既長遠發展, 例如WP既TILES在HD mon時或許可以一頁顯示更多TILES, 而且TILES尚有很多改進空間, 包括transparency, corner style ...
Web 6/8/12
大家認為WINDOW PHONE 系統可否在2年內與ANDROID打成平手什至打贏? - Windows Phone 機種及技術討論區 - 香港討論區 discuss.com.hk - 一個香港只得一個香港討論區
... ANDROID 打成平手什至打贏? « 上一主題 | 下一主題 » 42 12 3 » 打印 [機種討論] 大家認為WINDOW PHONE 系統可否在2年內與ANDROID 打成平手什 ... 大家認為WINDOW PHONE 系統可否在2年內與ANDROID 打成平手什至打贏? [按此打開] [隱藏] 如題 熱門搜尋: 鋼琴 課程 yamaha 結他 鋼琴 調音 1 分享到 ... 市場係用ios vs wp , android 市佔太高無咩可能短時間追到,有都係ios Nothing's impossible when you're dealing with the ... 都建於一個固定SCREEN SIZE, 3.5" MON, 所以佔有率要發大不容易, ANDROID就係早達先鞭, 但係本質上其實同WINDOWS PHONE反而係似, 所以未來兩年估計WP可以追到有 ... 因為WP好明顯一路走下去會更開放, WP8已支援數個RESOLUTION, 發展WP既生產商都好多, WP亦追緊hardware spec如NFC, 雙核, 好快會和ANDROID完全貼近, 加上 ... 都有20萬以上, ANDROID其實好多沒用APPS, 我用時DOWNLOAD左吾少, SECURITY好有問題, 所以WP兩年後在應用上已有IOS/ANDROID既95%, 到時Android用家就 ... 沒有必要死用Android, WP本質上係LIVE既, 所以電力既使用會係一個要處理既位, WP既TILES亦好大下, 所以一頁顯示只有8個, METRO UI既頂位都佔用位置, 但係, 呢數個問題會 ... 有好多變化, 反而有利WP既長遠發展, 例如WP既TILES在HD mon時或許可以一頁顯示更多TILES, 而且TILES尚有很多改進空間, 包括transparency, corner style ...

Connections View links highly correlated terms to one another

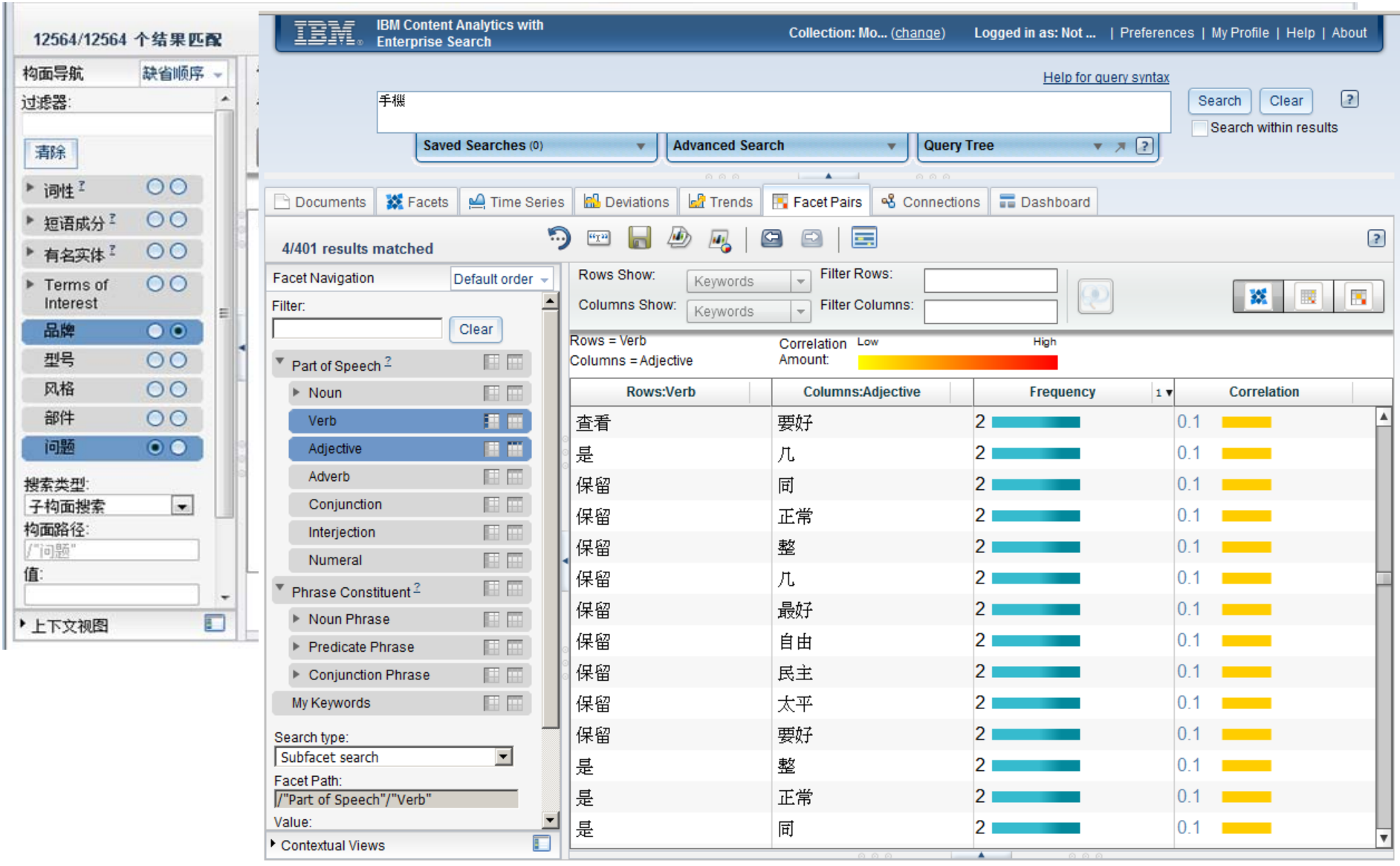
- Show relationship between multiple facet values
- Connections between nodes represents correlation between two facet values
- Color of line represents the importance of correlation index (red is the highest)



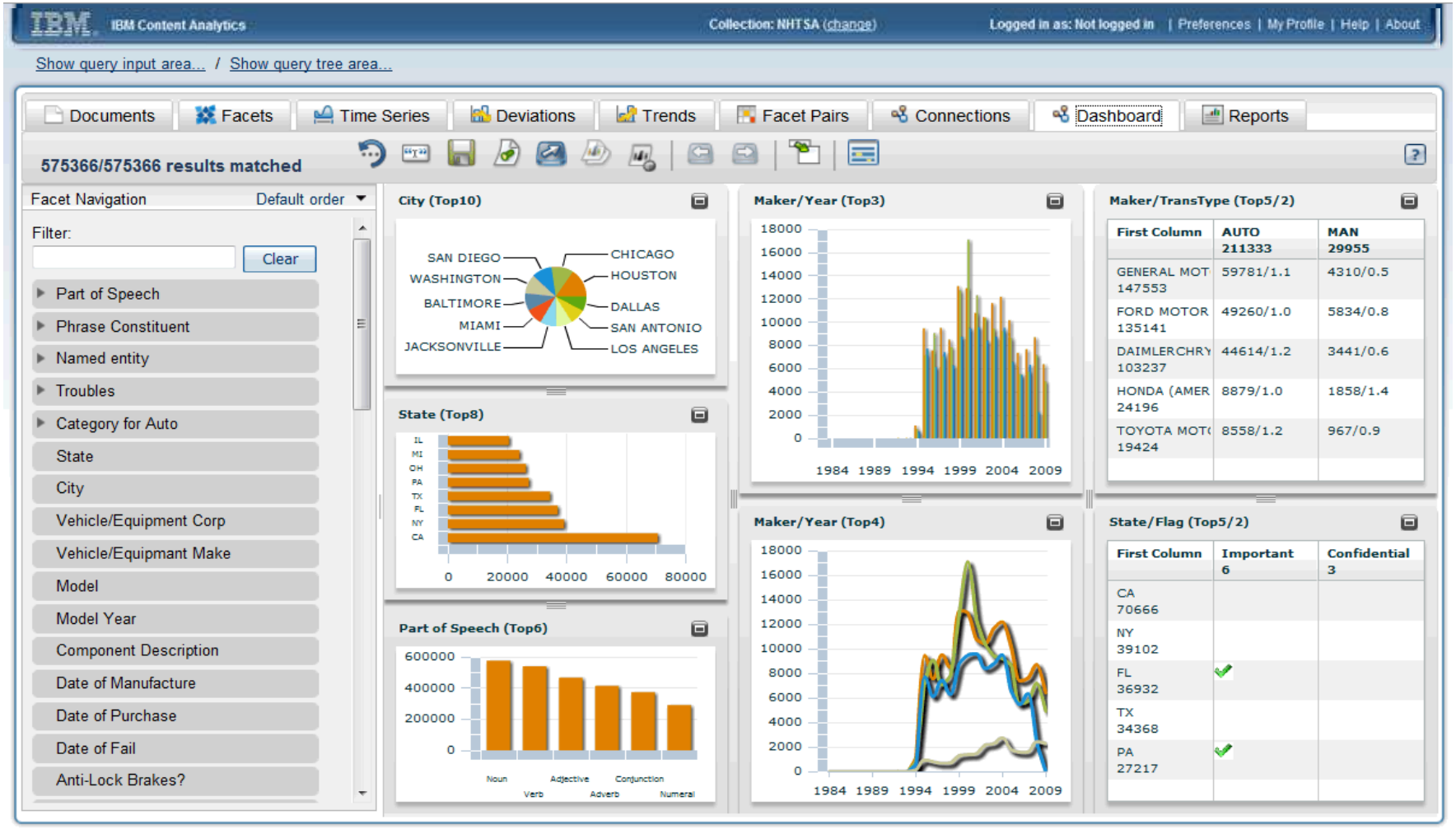
Identify relations between “FORD”, “blow” and “FIRESTONE”



Traditional / Simplified Chinese supported



Create Dashboard Views for Executive Summaries



Smart is: breakthrough content analysis

IBM Watson (Jeopardy)

Industry context: broad industry value
Value driver: improve business decisions
Solution onramp: content analytics

" … an information seeking tool that's capable of understanding your question to make sure you get what you want and then deliver's that content through a naturally flowing dialog"

Dr. David Ferrucci
Principal Investigator
Watson project

Business Challenge

Advance the state of the art in broad domain Question Answer (QA) systems to enable breakthrough applications in many different industries.

What's Smart?

Uses IBM Content Analytics (LanguageWare) in conjunction with other technologies to read, analyze and understand vast sources of unstructured content. Runs many algorithms in parallel to create, compare and determine confidence in candidate answers. Presents answers with a confidence level attached.

Smarter Business Outcomes

Coming to your industry soon! Will deliver value in limitless applications starting with clinical healthcare, customer care, government intelligence and beyond.

Global Financial Services organization specializing in Insurance

Smart is: Slashing risk exposure with Analytics

The need

- Reduce the loss ratio on claims
- Attack fraud
- Maintain optimal level of reserves

The solution

The company contracted with IBM to implement IBM Content Analytics software

- Initially configured to automate the search of 15 different internal data sources going back 15 years for greater insight into claim losses and insured policy lifecycle changes
- Designed with Natural Language Processing (NLP) technologies to enable knowledge-driven searches of both structured and unstructured information
- Aimed at providing one version of the truth by validating policy data across applications and databases
- Built for rapid addition of internal and external data sources as analysis needs expand
- Planned for future integration with IBM SPSS software to enhance predictive analysis on trend data

Projected benefits

- Improve risk assessment models by uncovering unexpected patterns and associations among existing data sources
- Set adequate reserves with a better understanding of the factors contributing to claims losses
- Pinpoint fraud with data mining to identify triggers that may signal bogus claims
- Save millions of dollars in staff time and get results more quickly by automating the risk assessment process

The solution is targeted to reduce losses by correlating information from a range of data sources to gain a more complete picture of insured policy risk exposure

Telecommunications Company listens to the Voice of their Customers

Smart is: reducing customer churn

The need

This telco wanted to improve customer satisfaction levels as its first priority to secure and maintain market share. The client wanted to strategically utilize the "Voice of the Customer" (VoC) to identify new customer opportunities while preventing contract cancellation of existing customers (churn) through rapid responses to incidents or planning of new services.

The solution

IBM Content Analytics processes call center notes, surveys, and customer emails to address the following use cases:

- **Customer Churn**: Detect likely candidates for customer churn. An alerting engine then automatically sends reports to a department that deals specifically with customer churn situations.
- **FAQ Generation**: IBM Content Analytics analyzes customer issues and suggests FAQ candidates for posting to a self-service Web site.
- **Root Cause Analysis:** perform exploratory root cause analysis of customer issues by mining for trends, patterns and unusual product and services associations with customer experiences.

Benefits

Improved accuracy to detect likely churn candidates by 50%.
Improved rates for model and service upgrades to loyal customers.
Started new Premium Club points program based on VoC.
Improved self-service FAQ system
Continually monitor voice of customer for new offerings and services.
Opened kiosks in international airports

"As a result (of ICA), we can easily identify trends and patterns from customer voices across our organization and provide better customer service."

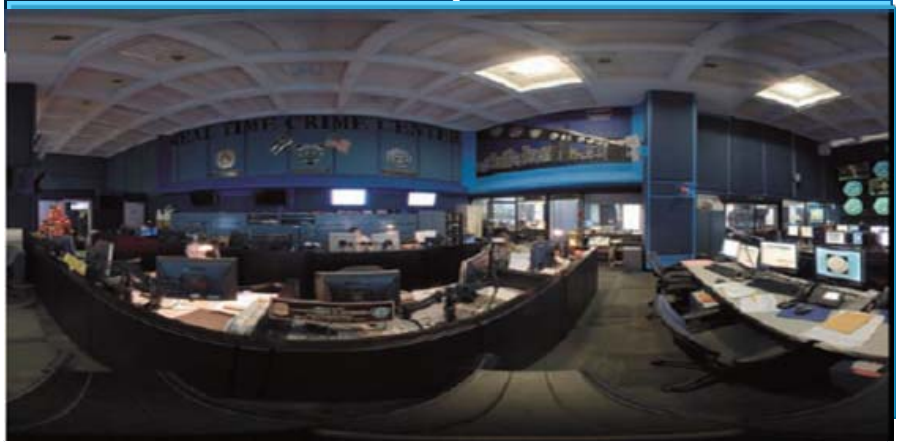
— Manager of Information System Department Group

New York Police Department's Real Time Crime Center uses IBM Content Analytics to crack cases

- Search and analyze complaints, police reports, 911 records, arrest records, and data marts
- All of these forms of text suffer from the common problems of call center text i.e. abbreviations, misspellings, synonyms (Police-specific i.e. perp, ML, FM, MO, pistol, gun, etc...)
- Content Analytics can analyze concepts and find similar situations described in different ways
- In the first week of deployment 2 old murder cases were solved

BUSINESS BENEFITS

- **Find events that keyword search can never find because they are all described differently – what keyword to use?**
- **Content Analytics can describe events, categorize them and allow for concept searches across often unstructured and at times inaccurate descriptions**



The PoliceChief — The Professional Voice of Law Enforcement

http://policechiefmagazine.org/magazine/index.cfm?fuseaction=display&article_id=995&issue_id=92006

Help your customers to start unlocking the insight trapped in their unstructured content

Uncover business insight *quickly to gain customer insight, improve product quality and customer service, detect fraud, optimize decision making and more …*

IBM Content Analytics

Find relevant enterprise content quickly and securely

Analyze enterprise content to unlock the meaning buried in unstructured information

Customize rapid insight to industry and customer specific needs

Enable deeper insights through integration to other systems and solutions