

# *Create rapid business insight with Content Analytics*

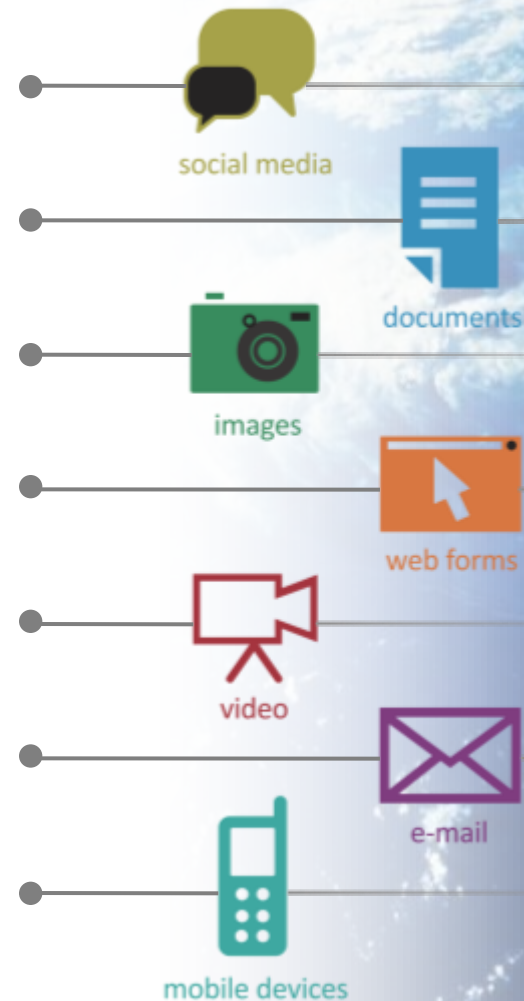
An Enterprise Content Management solution

Rashmi Vittal  
IBM



# Separate the **signal** from the noise

Leveraging content requires the ability to **search, assess** and **analyze** large volumes of **text** in order to understand and determine relevant **insight** quickly ... from multiple information sources **inside and outside the firewall**.



# 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?**

## A Japanese Business Services Provider

# Smart is: **gleaning insight** about customers

*“Insight into customer interaction logs is an information gold mine for us.”*

— General Manager  
Japan Business

*Industry context: computer services*  
*Value driver: improve customer service*  
*Solution onramp: content analytics*

### Business Challenge

A Japanese business services provider operates multiple customer service centers and needed ways to analyze large volumes of information to improve agent training and deliver better customer support.

### What’s Smart?

They implemented content analytics from IBM to understand and process natural language. The solution analyzes customer interactions based on consolidated logs of phone calls, email and Web, identifying keywords.

### Smarter Business Outcomes

Improved agent skills and training, resulting in a 92% reduction in call transfer and 88% improvement in volume. Provides new insights about product issues, resulting in an 88% decrease in product-related calls.

# 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.

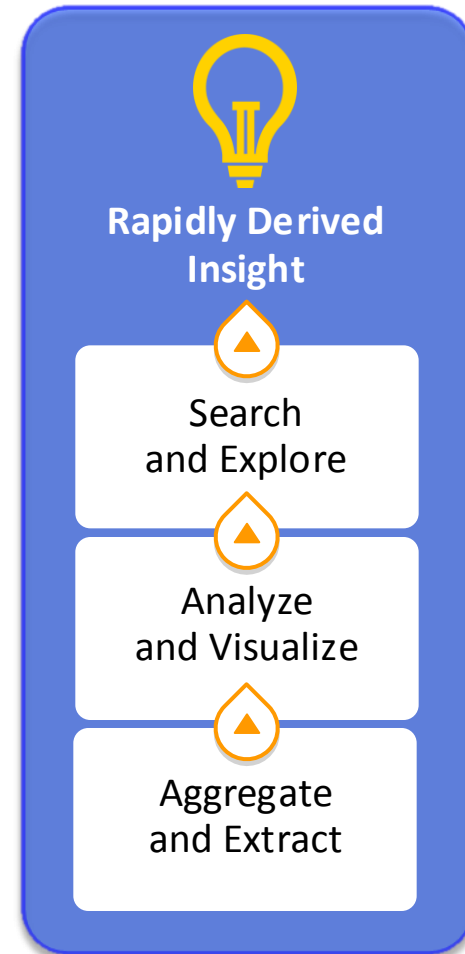
## Search and explore to derive insight

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



# IBM Content Analytics is a platform to derive rapid insight

- Transform raw information into **business insight quickly** without building models or deploying complex systems.
- Derive insight in **hours** or **days** ... not weeks or months.
- **Easy to use** for all knowledge workers to **search** and **explore** content.
- **Flexible** and **extensible** for deeper insights.



External and Internal Content (and Data) Sources  
including Social Media and More



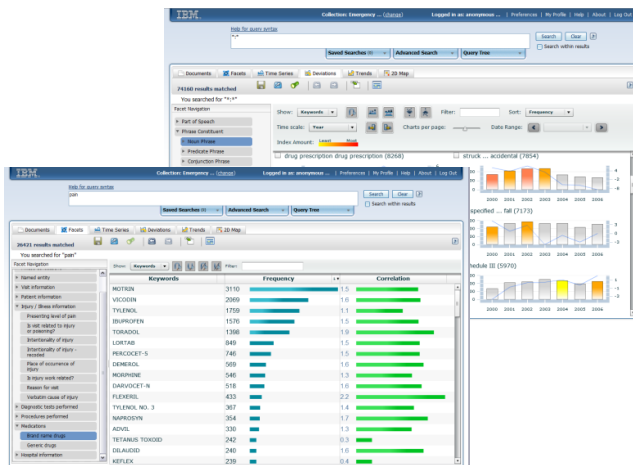
# Text Analytics is the basis for Content Analytics

## 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 **1 oz 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



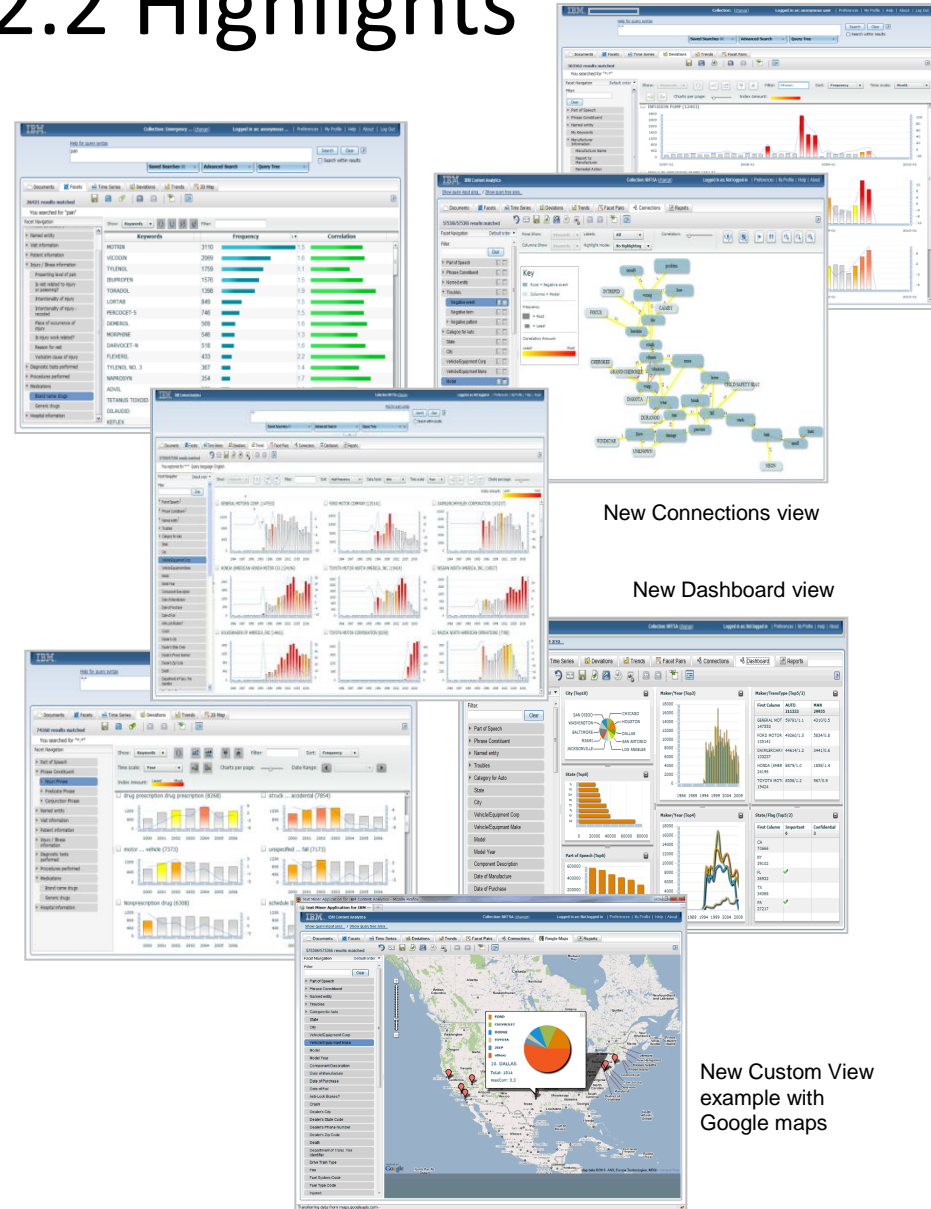
## What is Content Analytics?

*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.

# IBM Content Analytics v2.2 Highlights

*A platform for rapid insight*

- **Dynamically search and explore content for new business insight**
  - New Connections and Dashboard views to easily detect insights; plus add your own custom views
- **Interactively assess for content preservation and decommissioning to reduce storage costs and risk**
- **Powerful solution modeling and support for advanced classification tools for more accurate and deeper insight**
  - Enhanced analytics configuration tools
- **Deliver rapid insight to other systems, users and applications for complete business view**
  - Quickly generate Cognos BI reports, link between Cognos reports and ICA views; deliver analysis to IBM Case Manager solutions



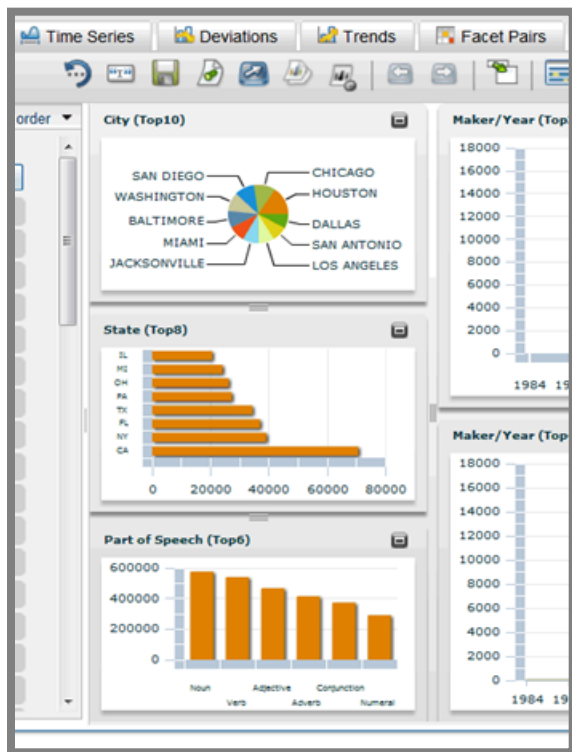
New Connections view

New Dashboard view

New Custom View example with Google maps



# What's new in IBM Content Analytics v2.2?



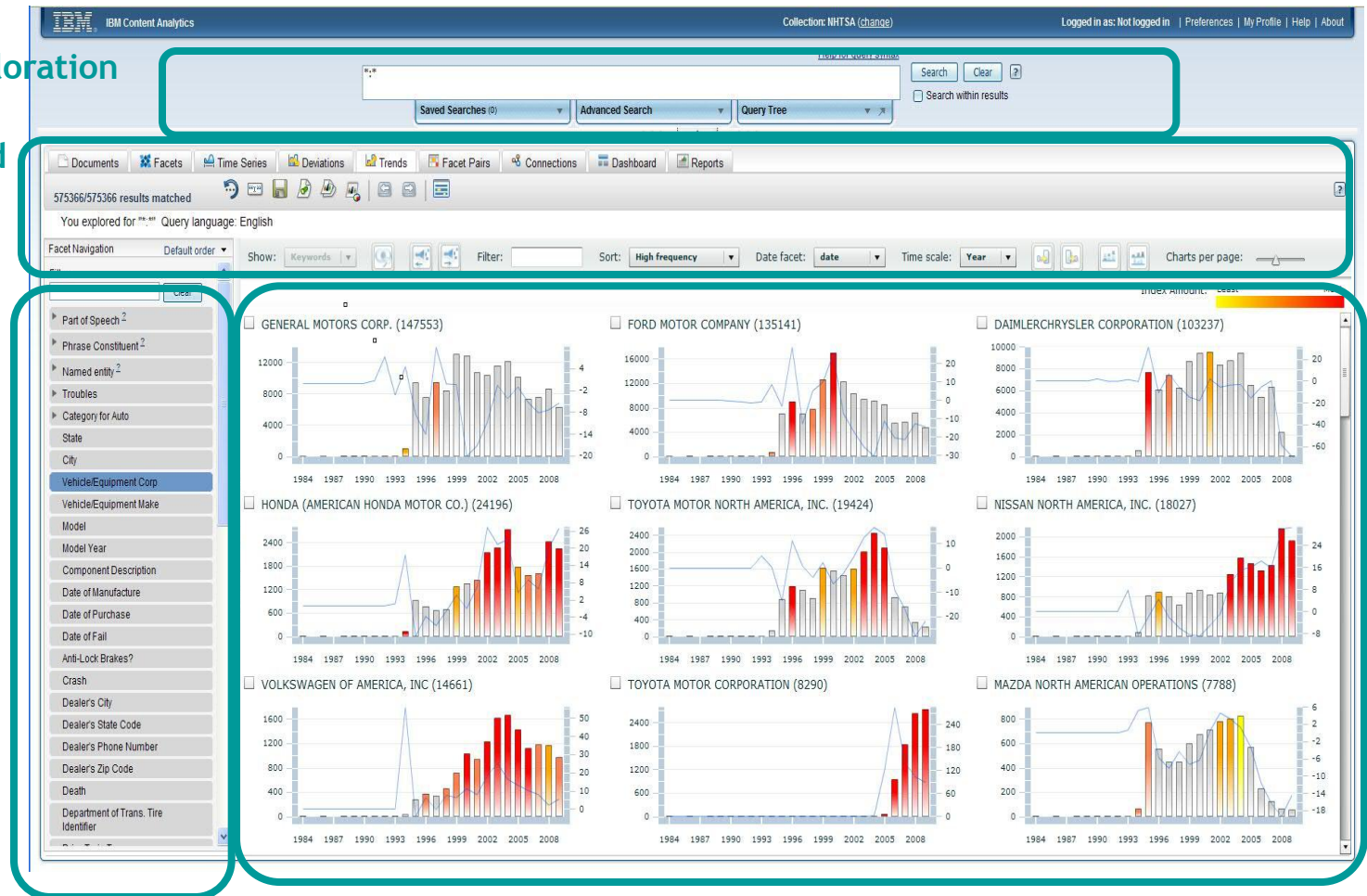
- **New Visualizations in Content Analytics Text Miner**
  - Connections View links highly correlated terms to one another
  - Dashboard view to see 1 or more analytics views in a single window.
  - Query Builder to easily create and save queries.
  - Ability to add custom views
- **Easier integration with Cognos BI reports and models**
  - Quick Cognos® BI report generation
  - Tighter integration with Cognos data models
  - Cognos reports can link from and back to Content Analytics
- **Speed Time to Value: Enhanced analytics configuration tools**
  - Tighter integration with LanguageWare® Resource Workbench (LRW)
  - Parametric dates and numerical range support in Facet Tree Editor
  - Support to auto-detect and add-on new languages
- **Document Analysis Support**
  - Mapping file metadata to auto-generate Facets
  - Documents flagging support
  - Near duplicated document detection
  - Support for Linux® (Redhat) on IBM System z® for file system, databases and web pages
  - Enhanced import/export document analysis to CSV, RDB, etc.
  - Analyze cases files in IBM Case Manager v5.0

# The Rapid Insight User Interface Explained

Search Query Exploration

Views, Filters and Thresholds

Automatically  
Extracted and  
Analyzed  
Concepts, Entities,  
Relationships,  
Meta Data and  
Classifications



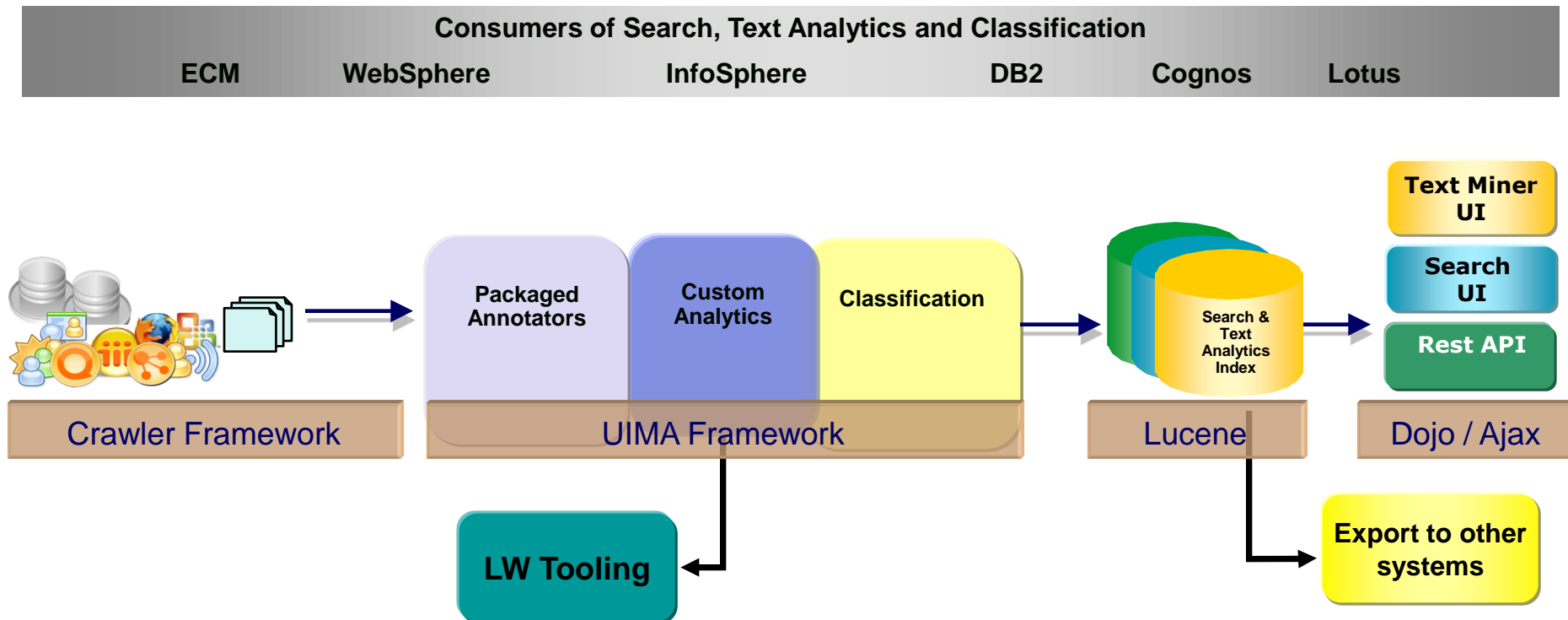
Visualization with Drill Down for Exploration and Assessment

# IBM Content Analytics v2.2 Enterprise-Wide Content Reach

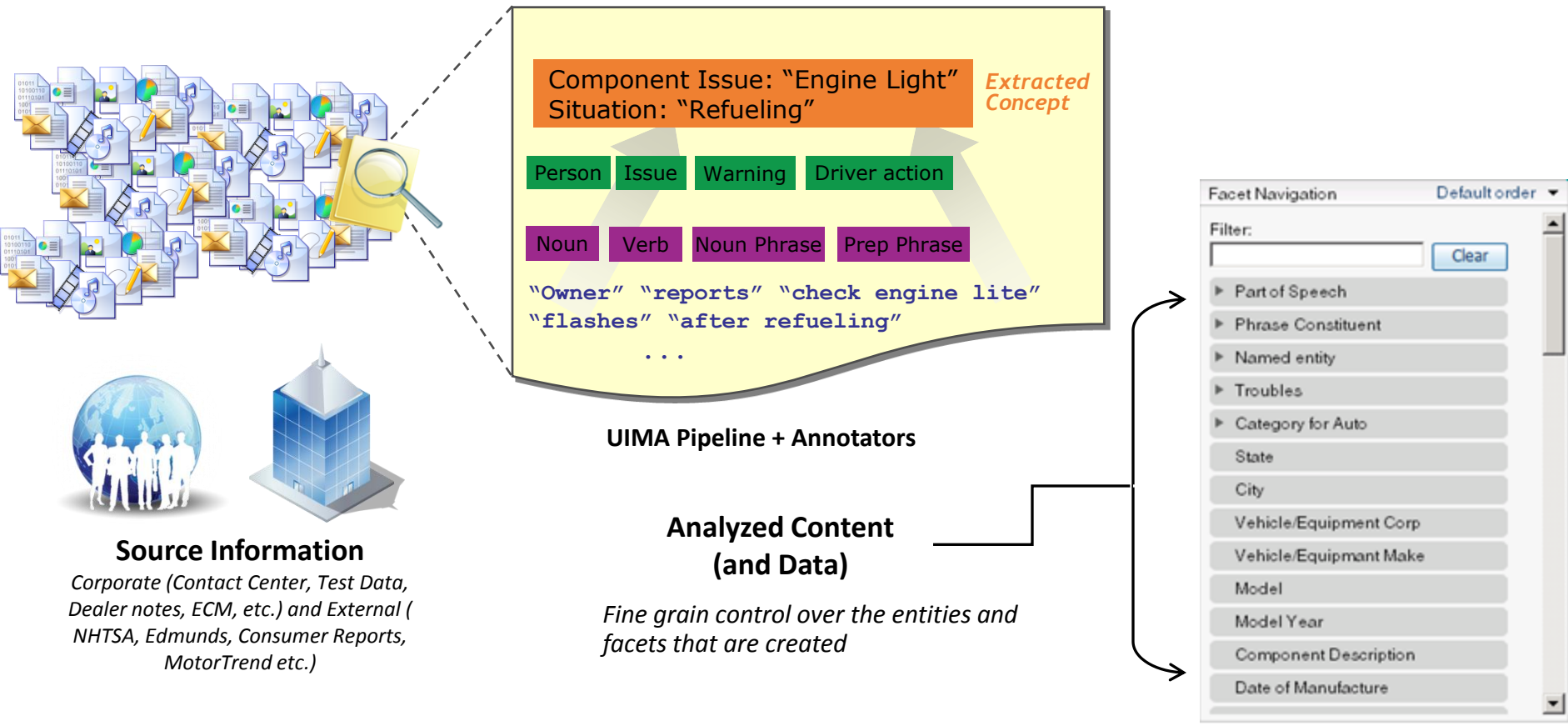
- CA-Datacom 10
- IBM Content Manager Enterprise Edition 8.4, 8.4.1, 8.4.2
- DB2 for iSeries 5.4 and 6.1
- DB2 UDB for Linux, UNIX, Windows 9.1, 9.5 and 9.7
- DB2 for z/OS 8.1, 9.1
- EMC/Documentum 6.0 and 6.5
- FileNet Content Services 5.4, 5.5
- FileNet P8 Content Manager 4.0, 4.5, 4.5.1 and 5.0
- Hummingbird DM 5.1.0.5 with SR6 and 6.0.4
- IMS 10.0 and 11.0.1
- Informix Dynamic Server 11.10 and 11.50
- IBM Case Manager V5.0
- IBM Lotus Connections 2.5.0.1 and 2.5.0.2
- IBM Lotus Domino DM 6.5.1, 7.0,
- IBM Lotus Domino R7, R8 and R8.5
- IBM Lotus Quickr (NSF & J2EE), 8.1, 8.2 and 8.5
- Lotus Web Content Management 6.1, 6.1.5
- Microsoft Exchange Server 2003
- Microsoft Windows SharePoint Services 3.0
  - SharePoint Server 2003 SP2 & 2007
- Microsoft SQL Server 2005, 2008
- MySQL 5.0
- Network News Protocol Newsgroup
- Open Text Livelink Enterprise Server 9.6, 9.7 and 9.7.1
- Oracle 10g and 11g
- Software AG Adabas 7.1
- Sybase 11.9.2, 12.0, 12.5x
- UNIX file systems
- VSAM for z/OS 1.4
- Web (HTTP or HTTPS)
- IBM WebSphere Portal 6.1 and 6.1.5
- Windows file systems: 2003 and 2008 servers

# ECM Discovery and Analytics invest in a common, interoperable technology stack

- Our Search and Text Analytics products (IBM Content Analytics, LanguageWare and eDiscovery) leverage a common architecture and framework.
- Each product interoperates with our IBM Classification Module product.
- Our products integrate with other IBM products and technologies.
- Multiple technologies across IBM in the search and content analytics space that are not compatible.

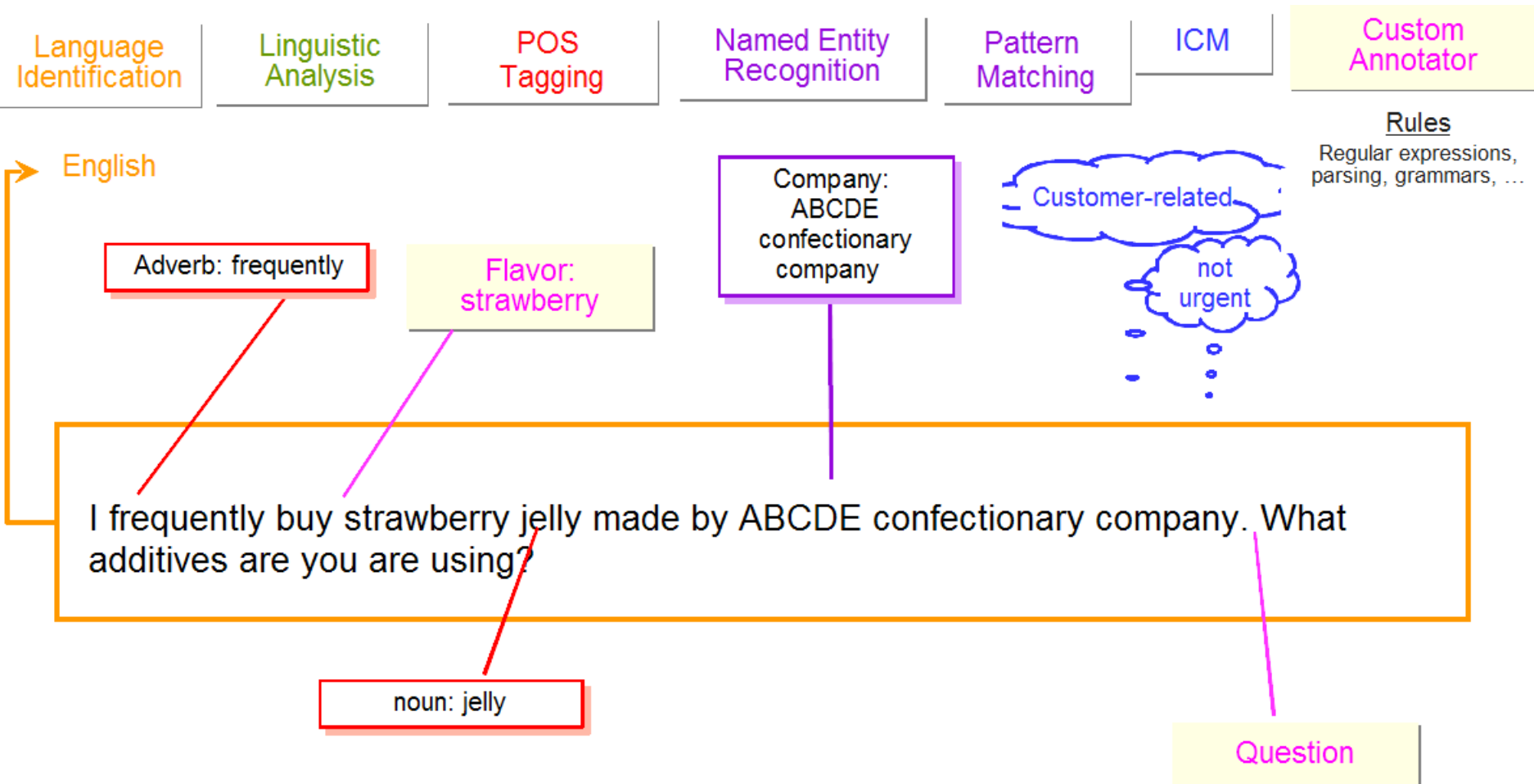


# Search and Content Analytics – How it works



# What do ICA annotators do?

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



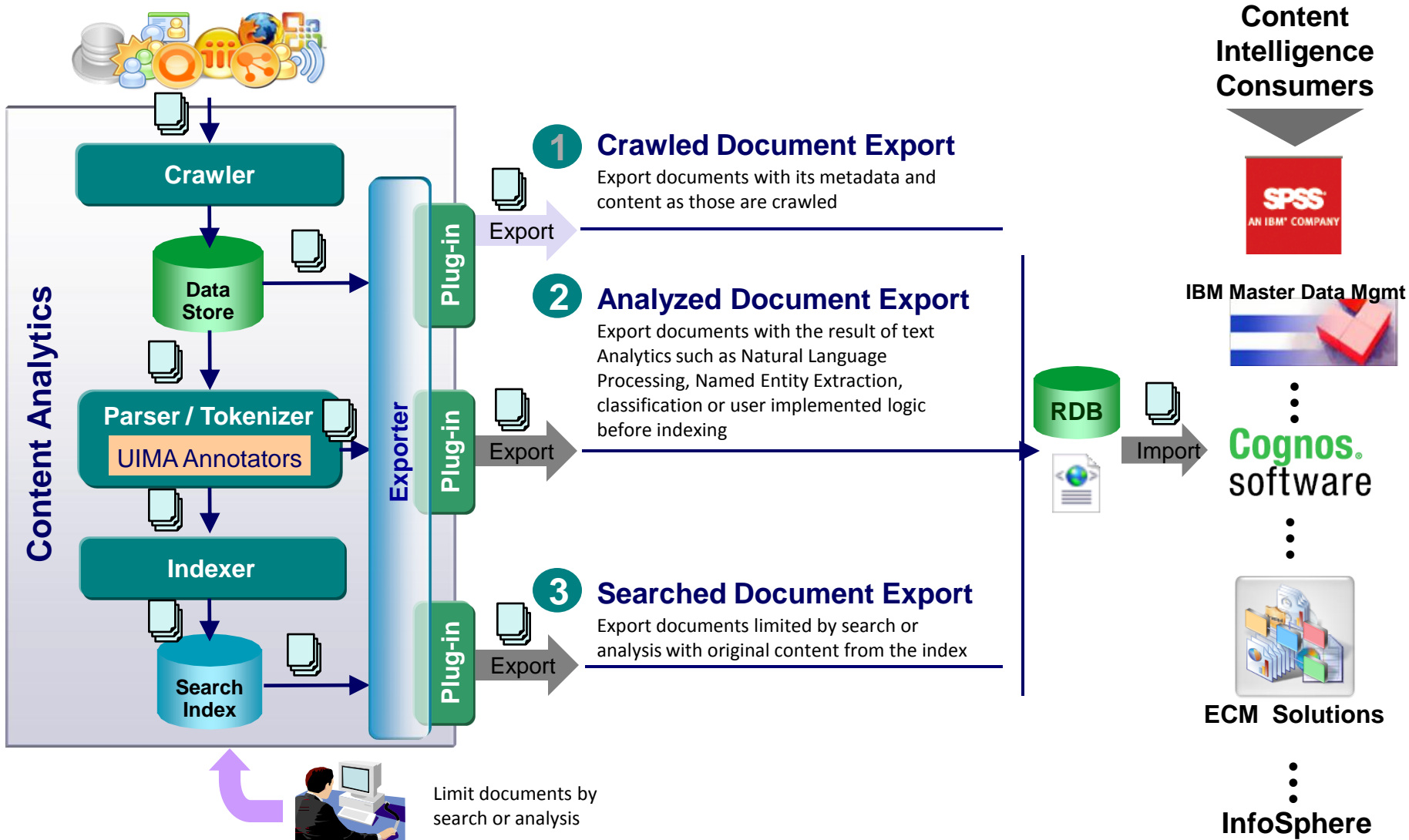


# ICA – Language & Function Support\*

Language	Part of speech	Phrase constituent	Person	Location	Organization	Pattern Matching	ICM Integration
Arabic	X	X				X	X
Chinese	X	X				X	X
Danish	X	X				X	X
Dutch	X	X				X	X
English	X	X	X	X	X	X	X
French	X	X	X	X	X	X	X
German	X	X	X	X	X	X	X
Italian	X	X				X	X
Japanese	X	X	X	X	X	X	X
Portuguese	X	X				X	X
Spanish	X	X	X	X	X	X	X

\*Ability to add new language modules

# IBM Content Analytics: Analysis Export Capability



# 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



## 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



## Crime Analytics

**Analyzing:** Case files, police records, 911 calls...

**For:** Rapid crime solving & crime trend analysis

**Benefits:** Safer communities & optimized force deployment



## Insurance Fraud

**Analyzing:** Insurance claims

**For:** Detecting Fraudulent activity & patterns

**Benefits:** Reduced losses, faster detection, more efficient claims processes



## Automotive Quality Insight

**Analyzing:** Tech notes, call logs, online media

**For:** Warranty Analysis, Quality Assurance

**Benefits:** Reduce warranty costs, improve customer satisfaction, marketing campaigns



## 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



# Partner Solution Snapshot



## ▪ Mindshare

- VOC, sentiment, service quality, product issues, loyalty, reputation management



## ▪ Dayhuff Group

- Content Analytics for Mergers and Acquisition
  - Analysis, assessment, consolidation, decommissioning, mergers and acquisitions, litigation
  - Contest Assessment skills with Records Management



## ▪ Oceanus

- Social networking for securities and investments



## ▪ Syscom IBIG

- Exploit insight lock in images - extract text for fully searchable Reflection Repository



## ▪ Wipro WiQTr (victor)

- Uses text analytics to perform automotive warranty claims analysis



## ▪ HCL America

- iGOVERN Child Care for collaborative management and childcare program execution
- Leverages OmniFind EE to meet information findability requirements



## ▪ JSYMMETRIC

- Multi-industry workplace safety management solution
- Leverages OmniFind EE to meet information findability requirements



## ▪ Prolifics Legal Accelerator

- Tailored solution for legal analysts and business operations
- Leverages OmniFind EE to meet information findability requirements

Oceanus

# Smart is: social media marketing & communications

*“Addressing social networking is a top agenda item within banking and financial services”*

*Industry context: banking and financial services  
Value driver: reach customers thru social media  
Solution onramp: content analytics*



**oceanus**  
Connecting people, process and content

**oceanus SocialView**

**Engage** Search...

Home Active Pending Unclaimed Closed Management Reports Help Log Out

**Welcome, Pam**  
There are currently 60 unclaimed messages, [view now](#)  
You have 2 active conversations, [view now](#)

**Unclaimed Conversations**

**Susan Hilton** 11 days ago

**Terry Ryan** Can you tell me if there are any product changes coming out soon?  
posted 11 days ago

**Susan Hilton** Terry - I think we will have some adjustments to our product portfolio later in Q1  
posted 10 days ago

Close Reply

**MYM Investments** 14 days ago

**Greg McCormick** I am interested in Australian mutual funds ...  
posted 14 days ago

Close Reply

**Pending Messages**  
No Pending Messages

## Business Challenge

Leverage social media for client marketing and communications in the regulated financial services industry. Enable marketing and financial advisors to engage with customers while maintaining compliance with FINRA and SEC regulations.

## What's Smart?

Oceanus SocialView uses IBM Content Analytics to understand the content within messages, enabling a firm to monitor and control social media communications and meets all key industry compliance and marketing requirements.

## Smarter Business Outcomes

Firms can now reach clients through a preferred medium, social media and can access a financially savvy, younger demographic, e.g. 90% of Facebook users are younger than 55 and 69% own mutual funds.

# DEMO

## IBM Content Analytics v2.2

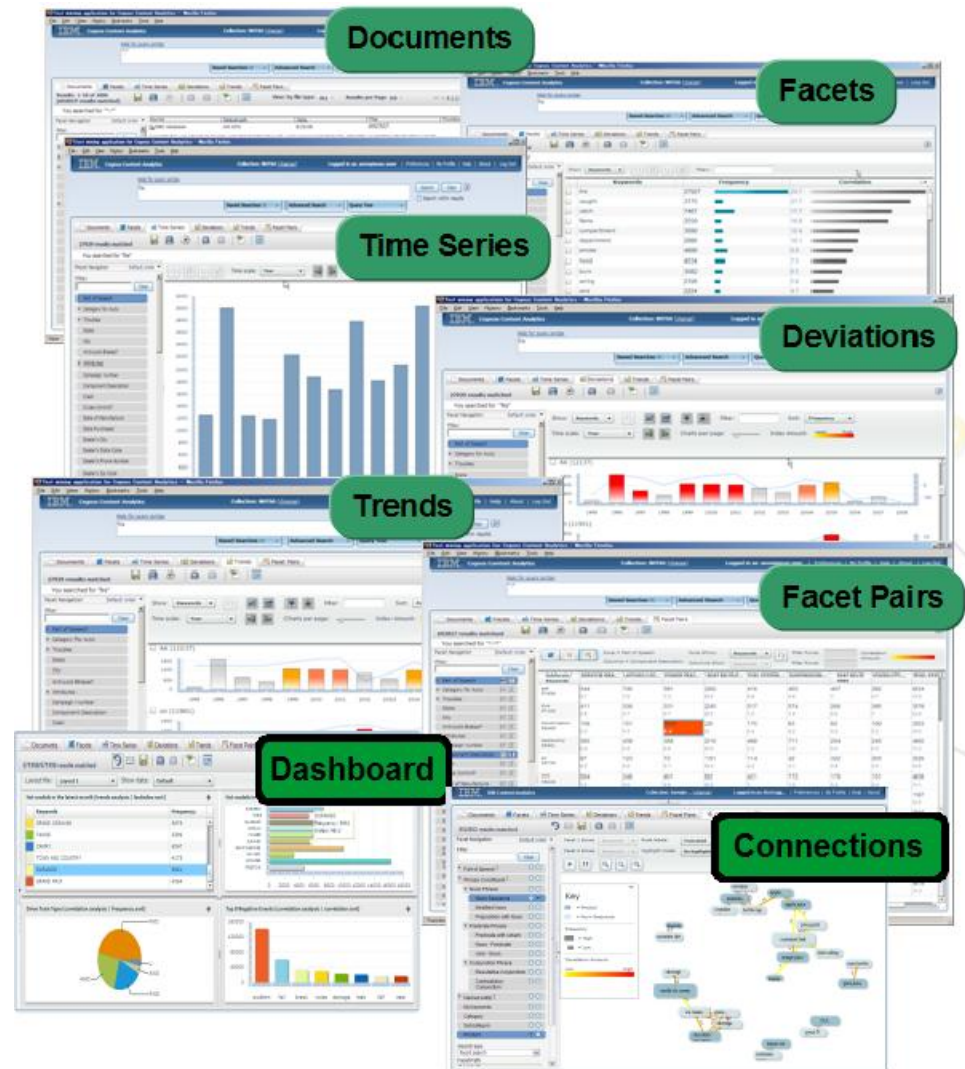


# End to End Scenario (TBD)

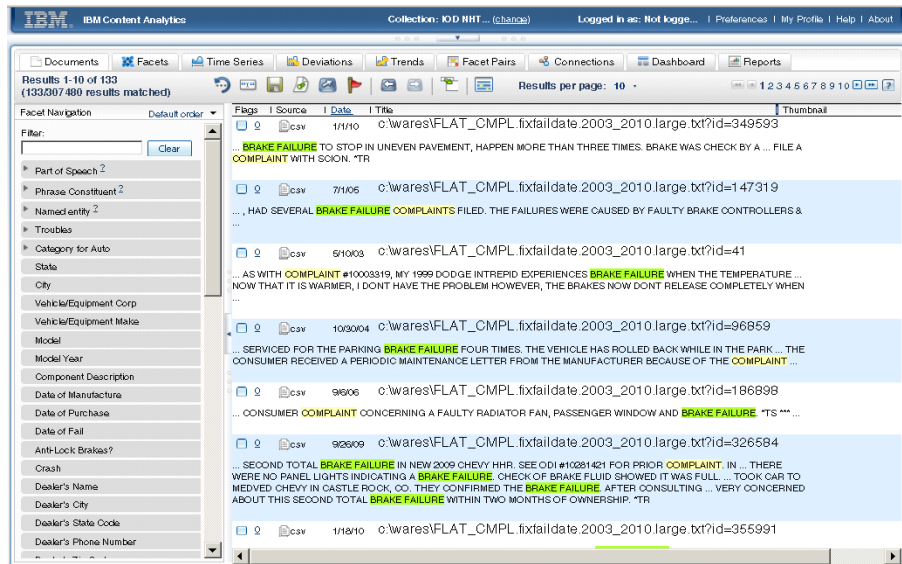
- Load CSV
- Run Analysis
- View

# Rapid Insight Tour

- Walk through each of the 8 views
- Learn about different ways to discover Rapid Insights from content
- Easy to use to search and explore



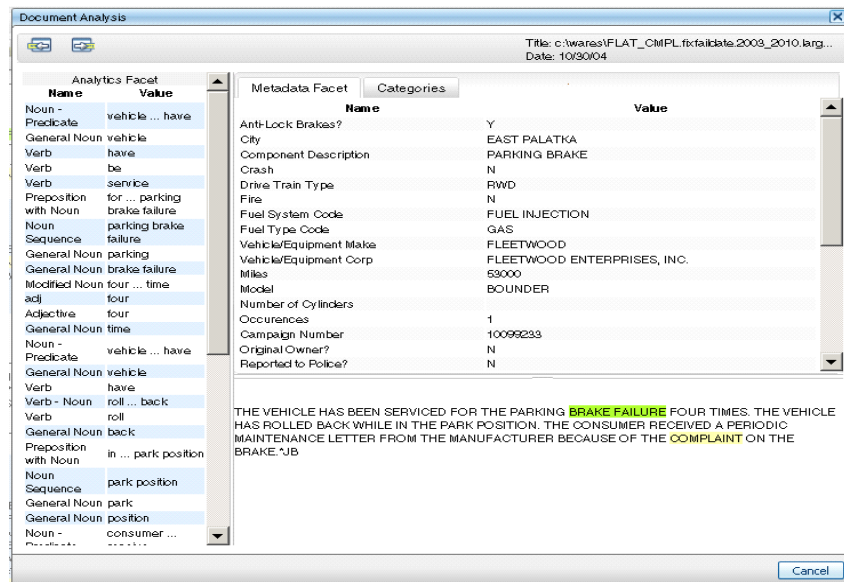
# Documents View



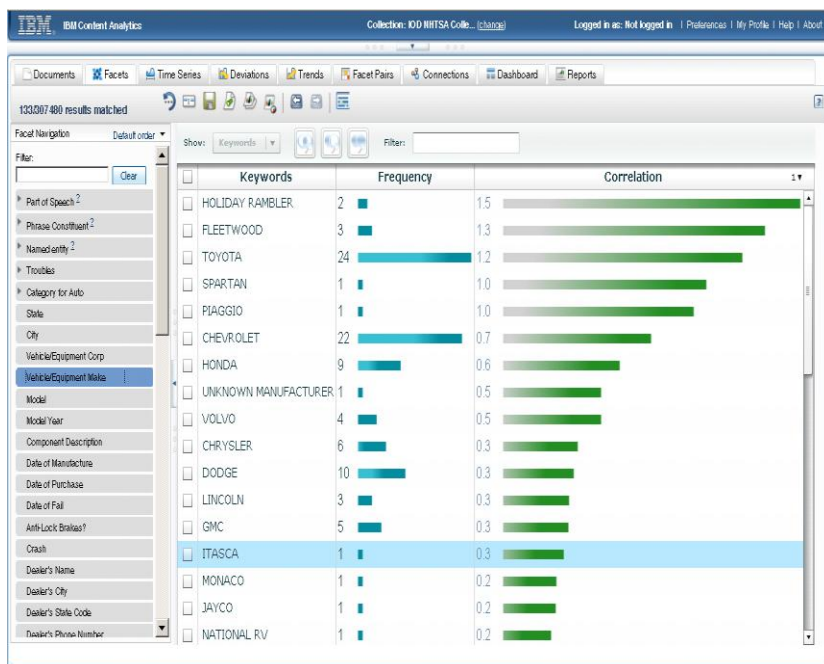
Get the documents matched the current search conditions

- All Search UI capabilities are still ready
- Can export documents to be used in another system

- Document Analysis Dialog
  - Preview the whole document
  - Display document metadata and annotated facets with highlighted texts



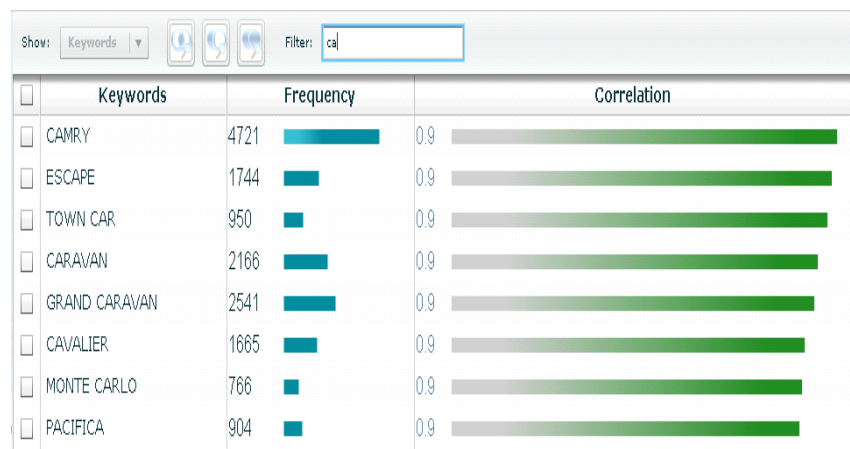
# Facets View



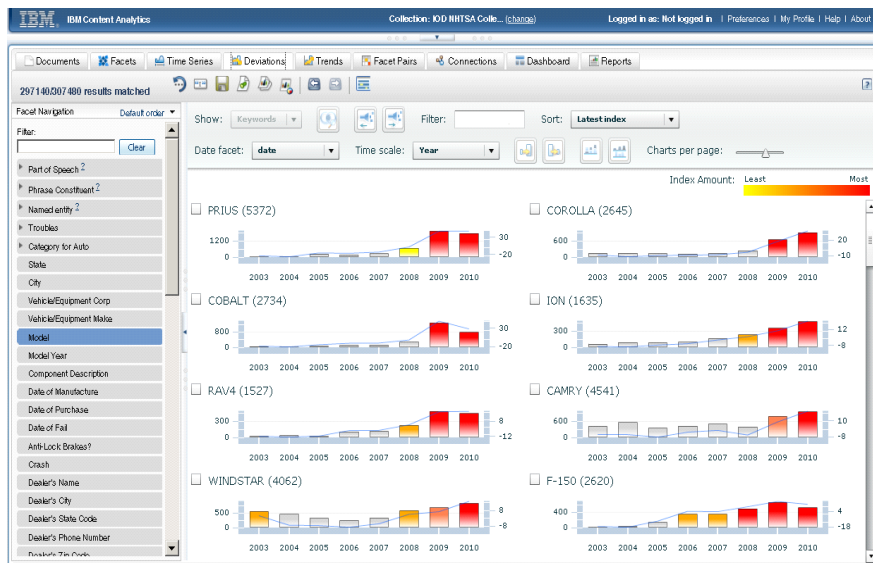
- Quick filter
- Multi-column sort

Show the frequency and correlation indices of keywords belong to the selected facet

- Add selected keyword to current search condition



# Deviations View and Trends View

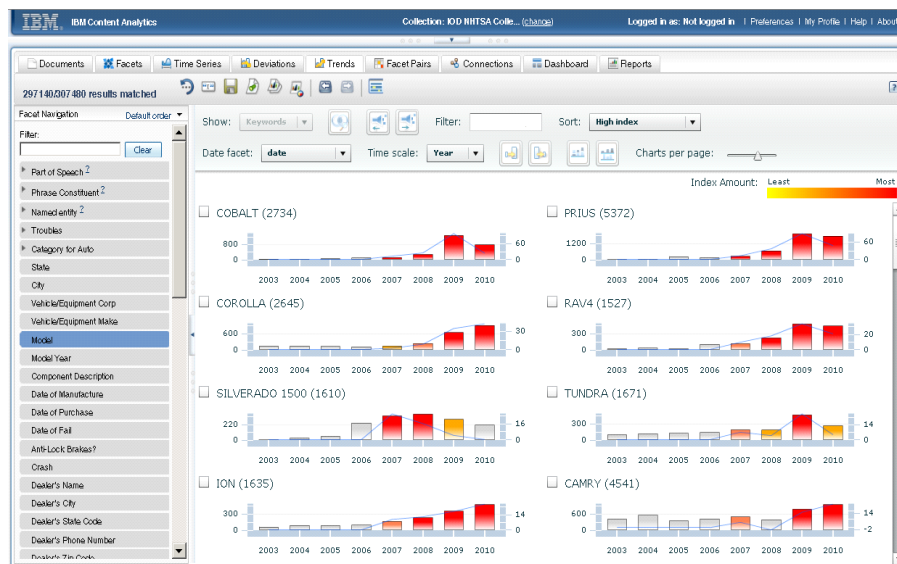


## ■ Trends View

- show how keyword in each time period deviates from average occurrence (against the whole current matched documents)..
- Predict pattern behaviors in future.

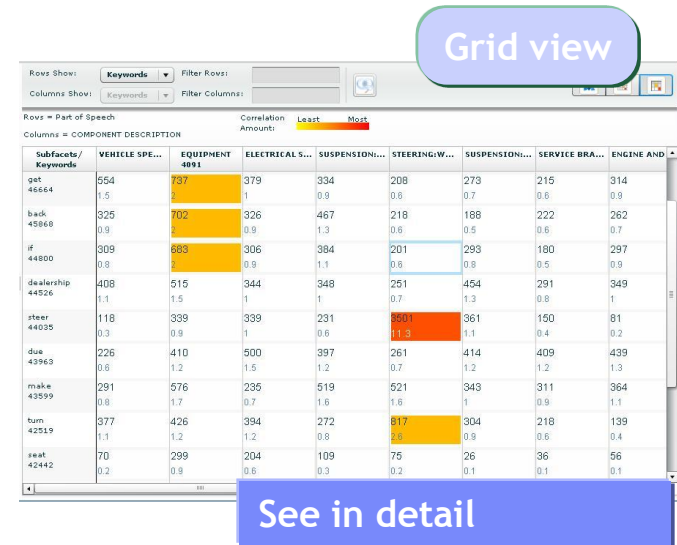
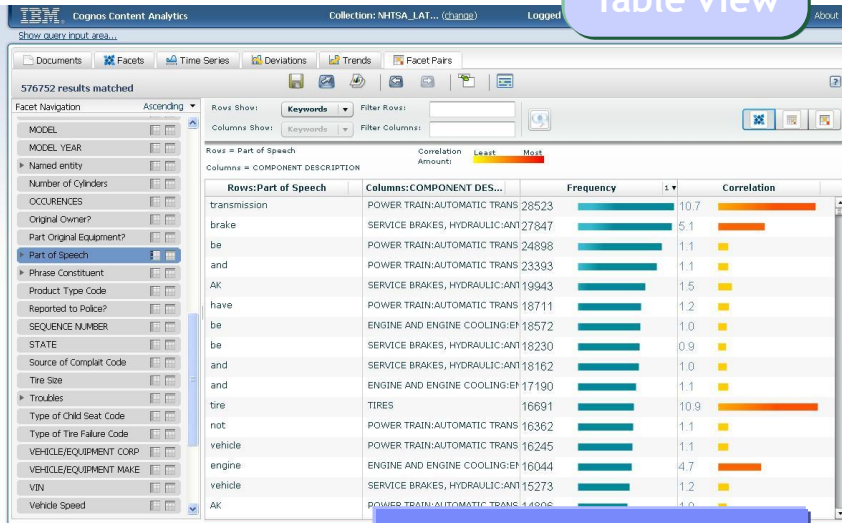
## ■ Deviations View

- show how keyword deviates from average occurrence between other keywords.
- Identify patterns that are cyclic.



# Facet Pairs View

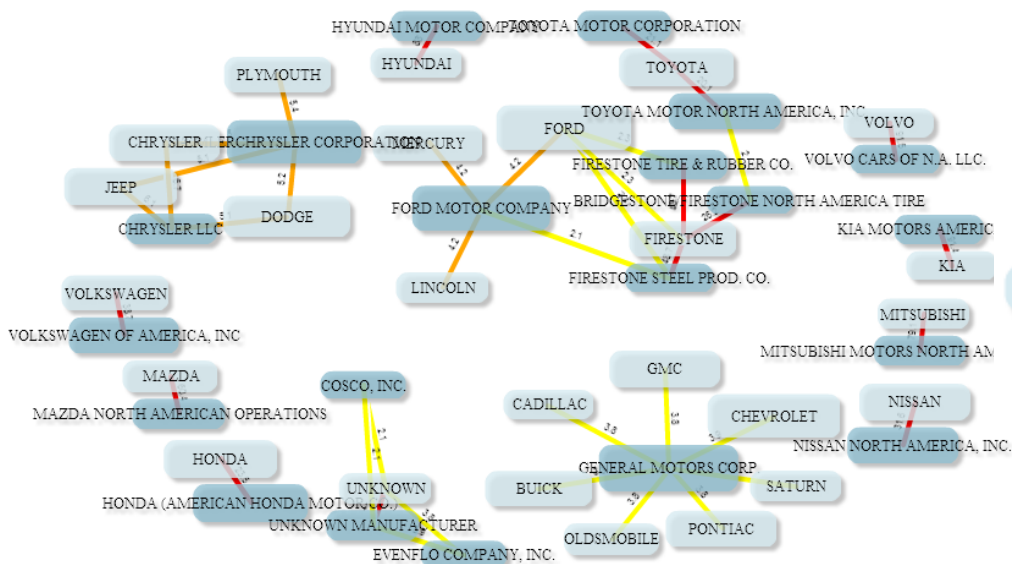
- Show the correlation between keywords belong to two different facets
- 3 view modes



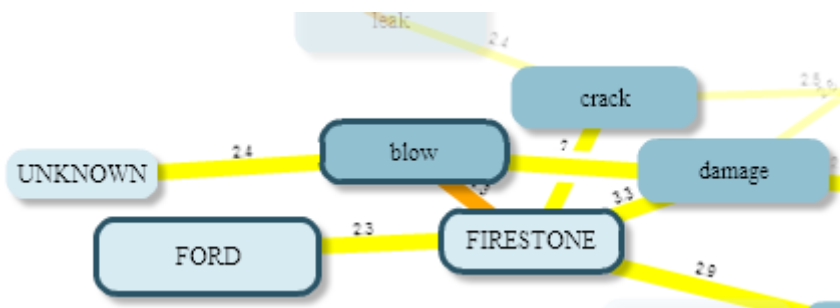


## Connections View links highly correlated terms to one another

- 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”





# Create Dashboard Views for Executive Summaries

IBM Content Analytics
Collection: NHTSA (change)    Logged in as: Not logged in    | Preferences | My Profile | Help | About

Show query input area... / Show query tree area...

Documents
Facets
Time Series
Deviations
Trends
Facet Pairs
Connections
Dashboard
Reports

**575366/575366 results matched**

Facet Navigation    Default order ▼

Filter:

- ▶ Part of Speech
- ▶ Phrase Constituent
- ▶ Named entity
- ▶ Troubles
- ▶ Category for Auto
- State
- City
- Vehicle/Equipment Corp
- Vehicle/Equipmant Make
- Model
- Model Year
- Component Description
- Date of Manufacture
- Date of Purchase
- Date of Fail
- Anti-Lock Brakes?

**City (Top10)**

**State (Top8)**

**Part of Speech (Top6)**

**Maker/Year (Top3)**

**Maker/Year (Top4)**

**Maker/TransType (Top5/2)**

First Column	AUTO	MAN
GENERAL MOT 147553	59781/1.1	4310/0.5
FORD MOTOR 135141	49260/1.0	5834/0.8
DAIMLERCHRY 103237	44614/1.2	3441/0.6
HONDA (AMER) 24196	8879/1.0	1858/1.4
TOYOTA MOT 19424	8558/1.2	967/0.9

**State/Flag (Top5/2)**

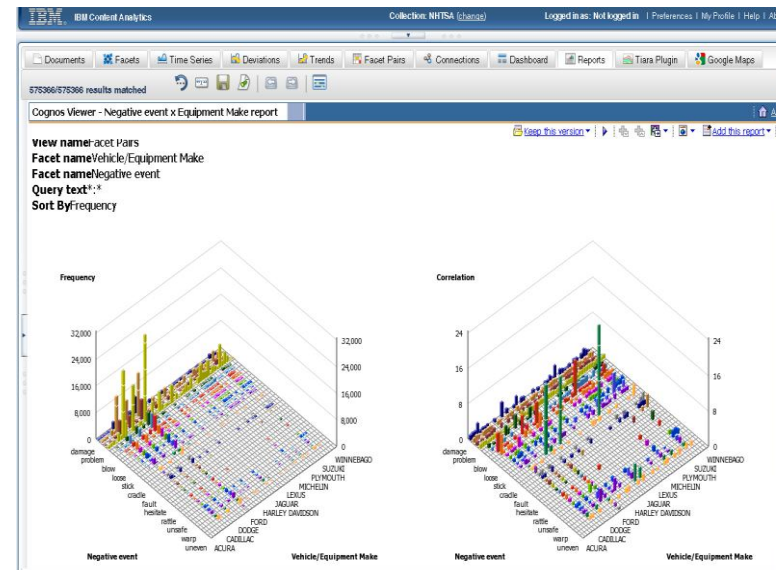
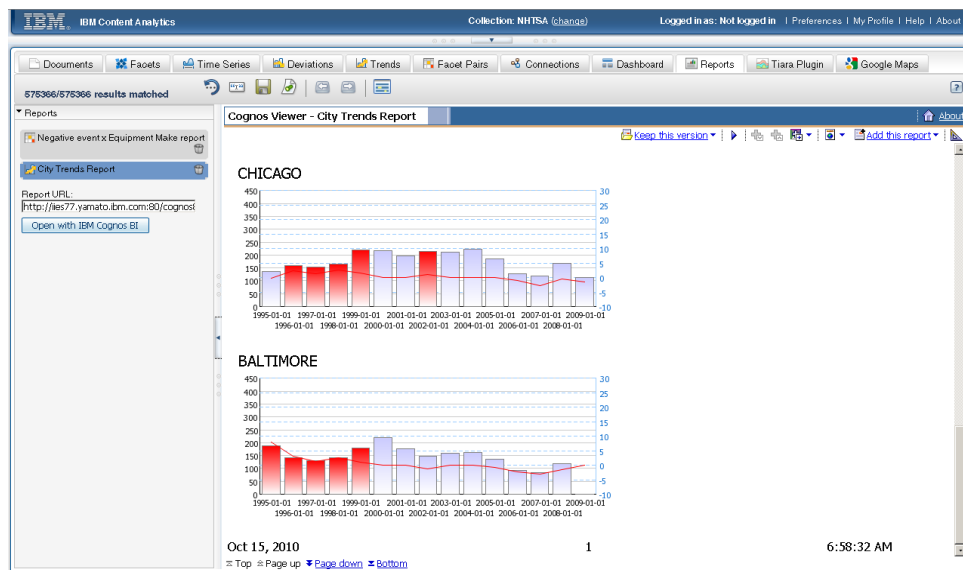
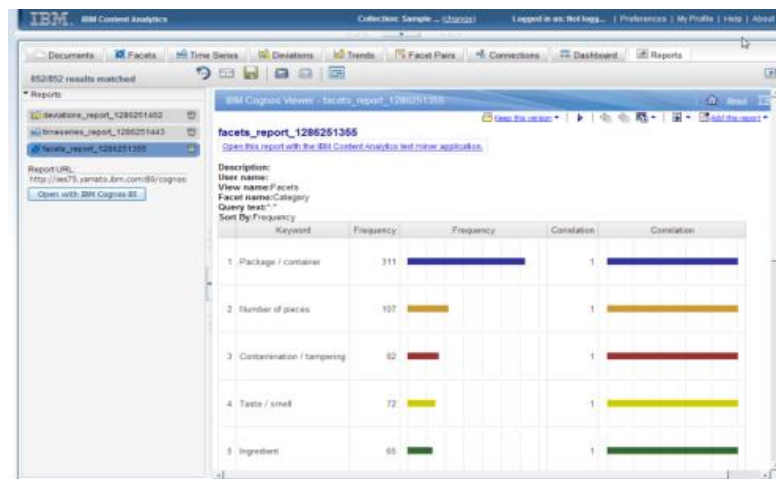
First Column	Important	Confidential
CA 70666		
NY 39102		
FL 36932	✓	
TX 34368		
PA 27217	✓	

© 2011 IBM Corporation

# ICA integrates with Cognos BI reports

From ICA Text Miner, a user can:

- Issue a request to create a report
- List the created reports
- Open the created report
- Delete the created report
- Cognos reports can link to and from Text Miner



# Where to find more information about ICA

IBM Content Analytics ✕ Search

About 13,700,000 results (0.18 seconds)

[Advanced search](#)

**IBM - Content Analytics - Software** ☆ 🔍

**Content analytics** software that helps companies gain new business insights through the analysis of unstructured content. View a demo of **IBM Cognos Content ...**

[www.ibm.com/software/data/content-management/analytics/](http://www.ibm.com/software/data/content-management/analytics/) - [Cached](#)

**You Tube**   | Br

**IBM Content Analytics**  
 IBMECM 6 videos

**IBM Content Analytics**  
 Over 50 years of text analysis and discovery

1:17 / 4:42 480p

**Content Analytics**

Learn what Watson is, how it works, and how the technology can be adapted to solve problems.

→ [Register to listen to the ECM podcast.](#)

**IBM Content Analytics**

Know. Trust. Leverage.

Analyze unstructured content to unlock critical business insight.

→ [Register for new webinars and white papers.](#)

## More Information

- *Solution Briefs*
- *Press Releases*

# Terminology and Definitions

- annotation - Information about a span of text. For example, an annotation could indicate that a span of text represents a company name.
- annotator- a software component that performs linguistic analysis tasks and produces and records annotations.
- character rules - LRW rules to recognize sequences of characters (LRW way to do regular expressions)
- concept Extraction / Entity extraction- A text analysis function that identifies significant vocabulary items (such as people, places, or products) in text documents and produces a list of those items.
- dictionary– a list of words for document processing to use to create annotations
- lexical analysis- The overall process by which LanguageWare segments and normalizes text
- metadata– data about a document, such as size and modified date
- normalization- determining a single string representation for a word or term found in text. This single string representation may also be called lemma, citation form, canonical form. In LRW, since we include Semantic Normalization (IBM=International Business Machines, Big Blue) we use the term *Normal Form*
- parsing rules – LRW rules to recognize patterns of words, they run in the LRW rules engine
- regular expressions - A flexible means of identifying sequences of characters (such as URLs). Written in a formal language that can be interpreted by a regular expression processor.
- tokenization- The simple mechanical process of breaking up white space delimited text into words.