

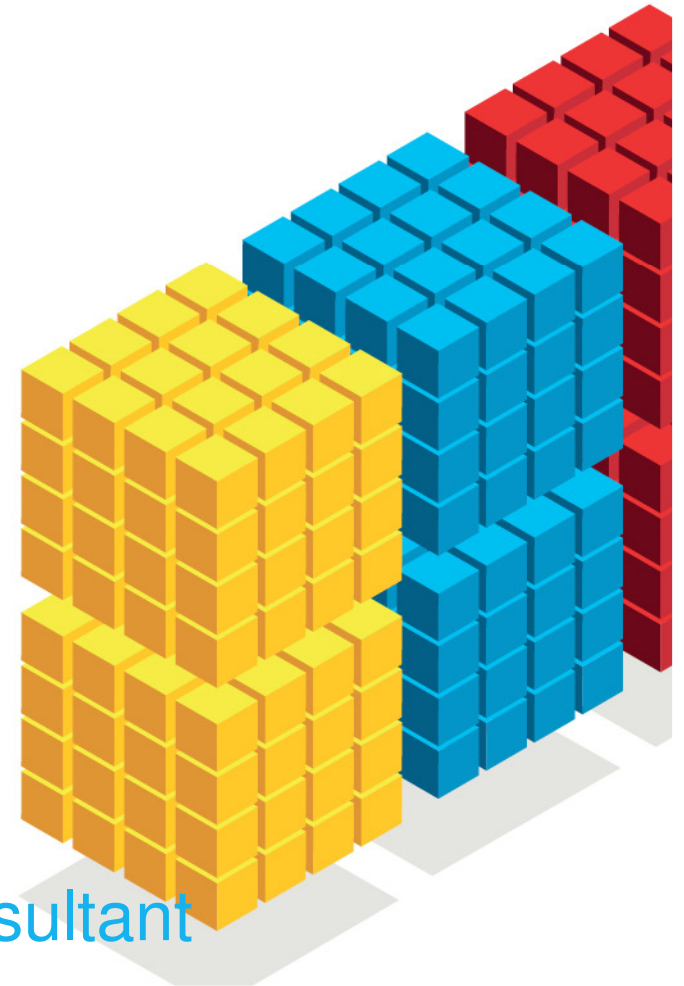
Smarter systems for a smarter planet:

# IBM Work Smarter Seminar Series



## IBM Content Analytics Demonstration

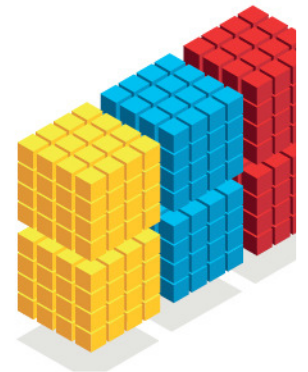
Adrian Barfield – Technical Sales Consultant






## Agenda

- Content Analytics Principles
- Configuring Analytics to your business needs
- Data Extraction Demonstration
- Visualisation and Data mining Demonstration






## Information Explosion



**Volume**  
Every day, **15 petabytes** of new information are being generated. By 2010, the amount of digital information will grow to 988 exabytes



**Variety**  
**80%** of new data growth is unstructured content, generated largely by email, with increasing contribution by documents, images, and video and audio.



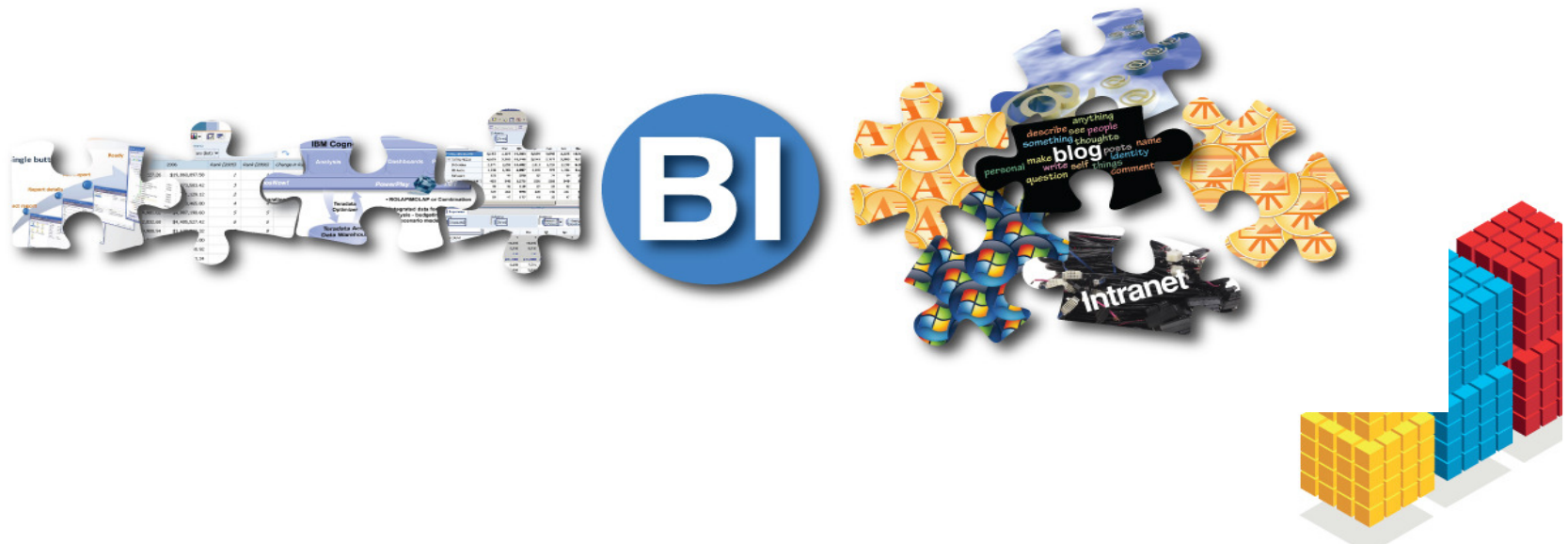
**Velocity**  
An average company with 1,000 employees spends **\$5.3 million** a year to find information stored on its servers. **42%** of managers say they use the wrong information at least once per week.

Information Week, "State Of Enterprise Storage Changing Priorities, Changing Practices", Jan 2009. Delahunty, Steve  
 IDC: Expanding Digital Universe, 2007  
 IDC: "The Hidden Costs of Information Work" April 25, 2006 Accenture: Managers Say the Majority of Information Obtained for Their Work is useless 2007  
[http://newsroom.accenture.com/article\\_display.cfm?article\\_id=4484](http://newsroom.accenture.com/article_display.cfm?article_id=4484)



## Content Analytics: *The Missing Link in BI*

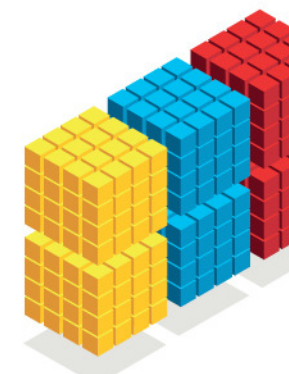
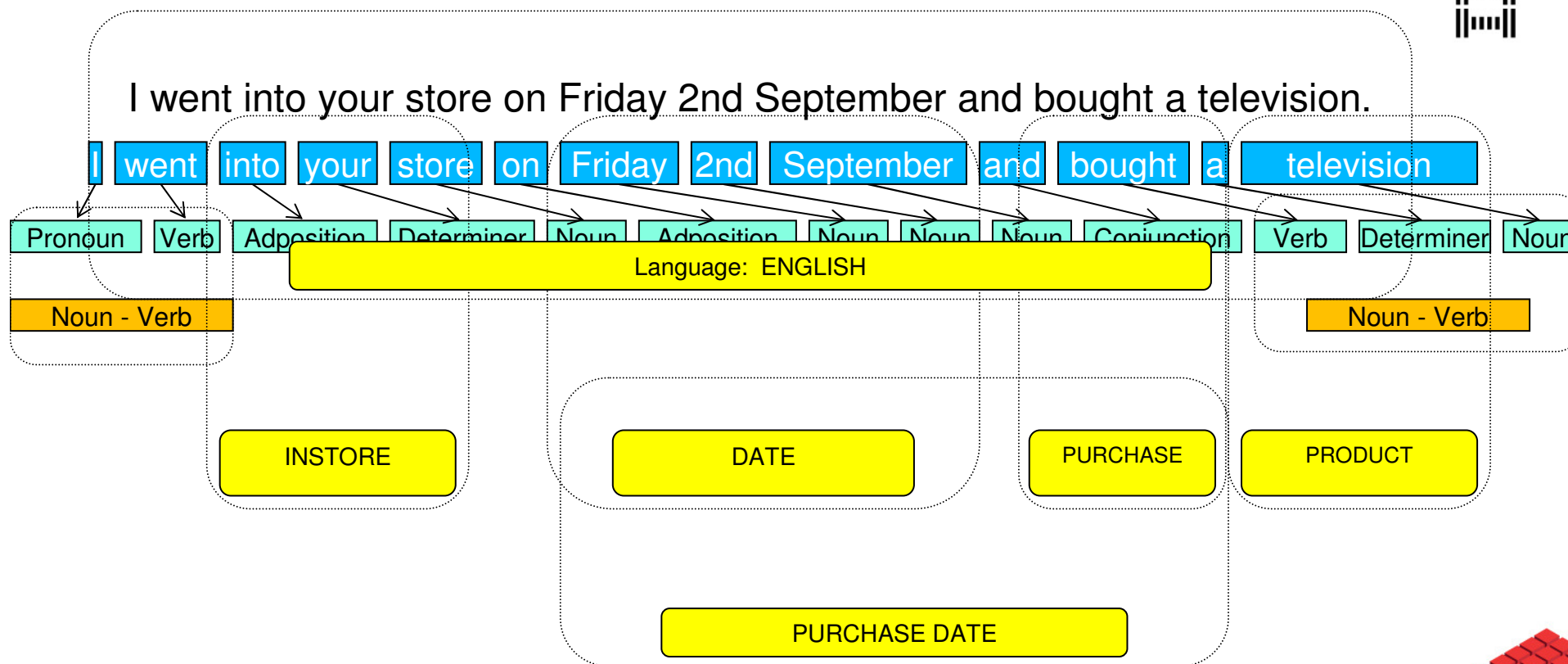
- Content Analytics delivers new insights from the content and context of structured and **unstructured information** ...
  - Internet, desktops, portals, call center notes, repositories, etc.





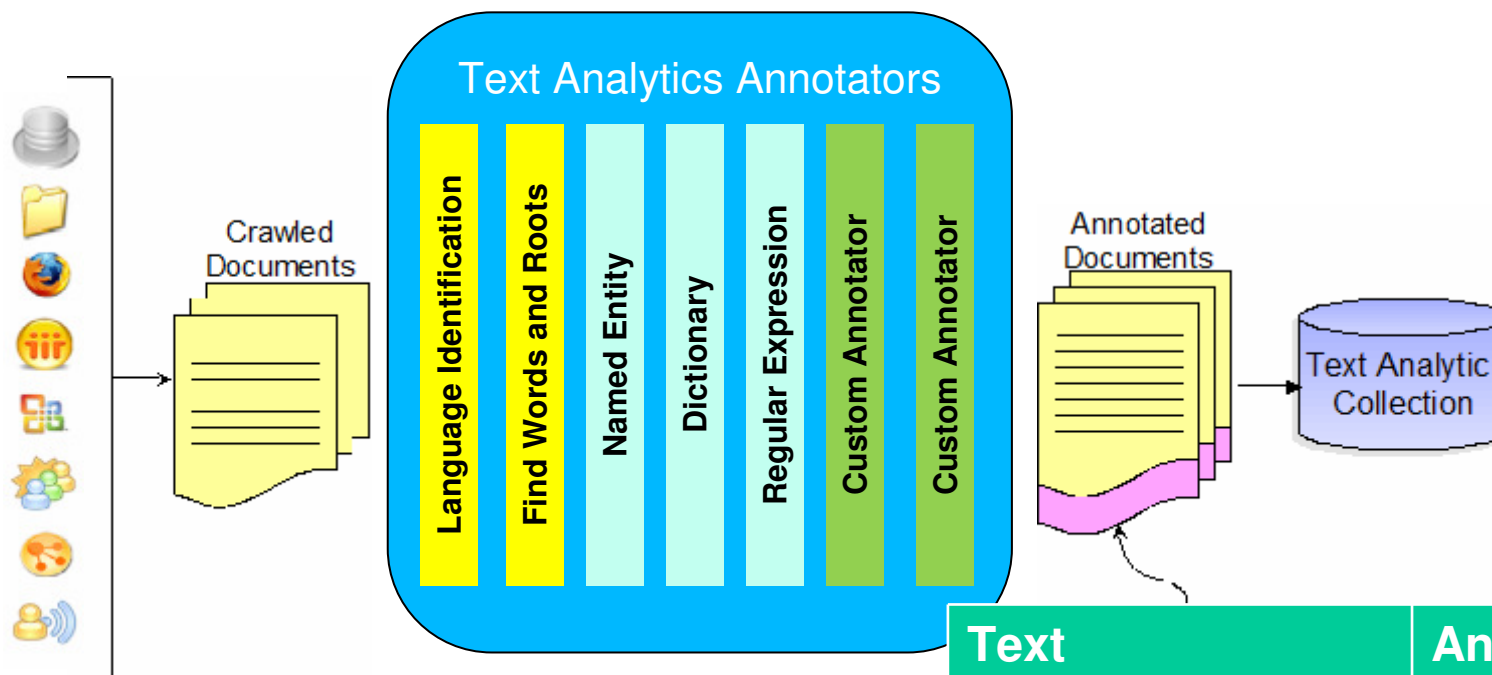


# Content Analytics Principles



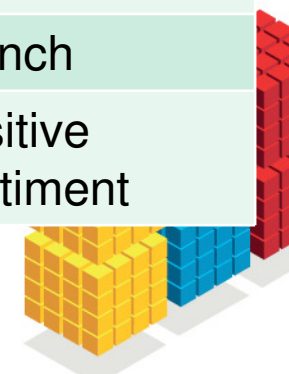


# IBM Content Analytics Pipeline

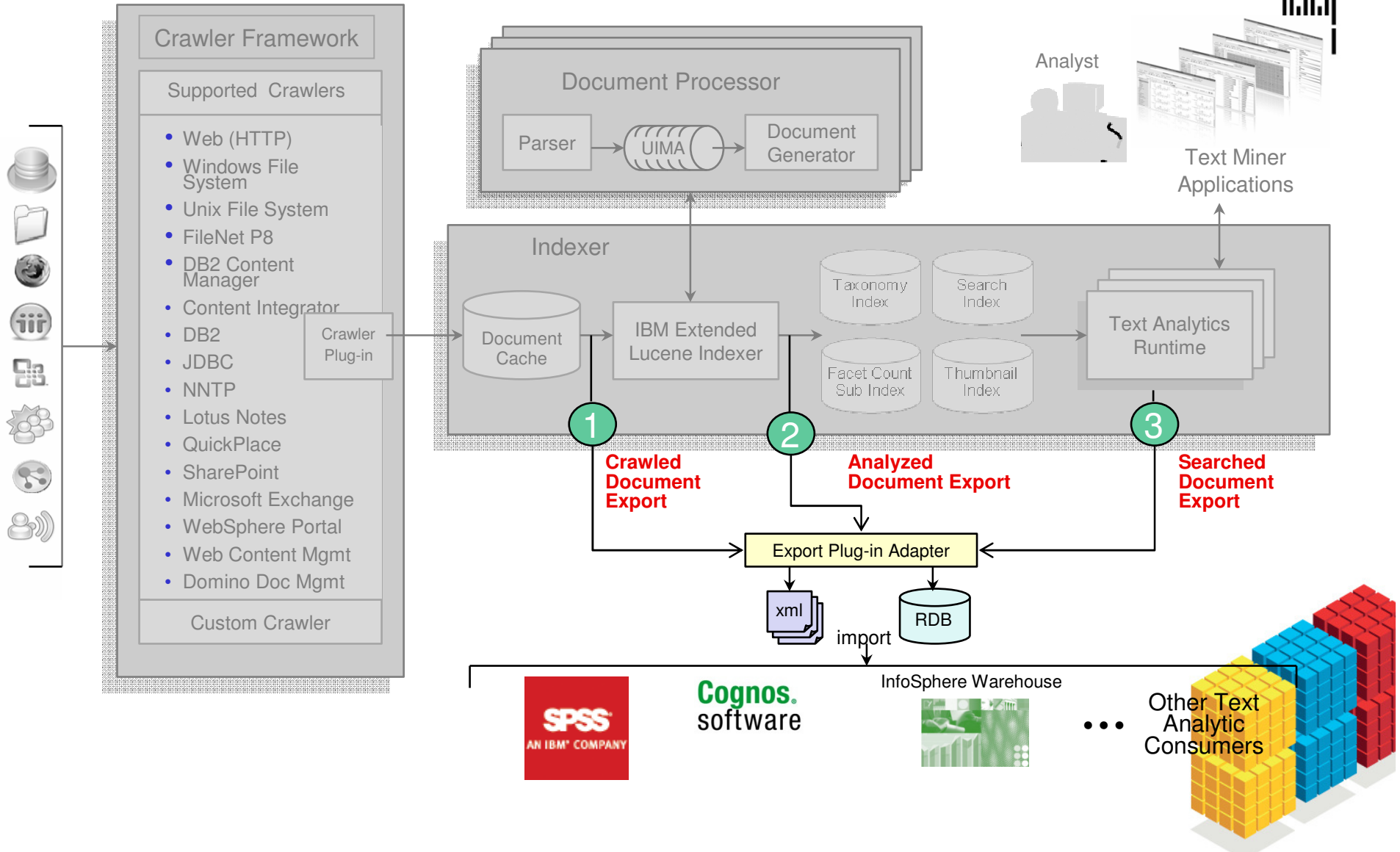


I walked into your Sydney store on Friday 2<sup>nd</sup> September 2011 and purchased a 3D Plasma TV. I wanted to commend the sales team at this branch on their helpfulness.

Text	Annotation
02/09/2011	Purchase date
3D Plasma TV	Product
Sydney Store	Branch
Commend the sales team	Positive sentiment



# Document and Analytics Export Capability (Alerting)



# Configuring Analytics to meet your business needs



- Decide on areas of interest (Voice of customer, Fraud reduction, Product Issues, etc)
- Identify Content Sources
- Determine Data of Interest using:
  - Dictionaries
  - Regular Expressions
  - Parsing Rules
  - Normalisation
- Decide on search terms
- Create standard reports

The screenshot shows the LanguageWare Resource Workbench interface. The main window displays a text document titled "Complaint.txt" with the following content:

Emily Swan  
345, Railway Road  
Erakenville, NSW, 2061  
The Manager,  
Focus Bank  
601a Pacific Highway  
St Leonards  
NSW 2065  
24th July, 2009.  
Ref: Account Number 12345678

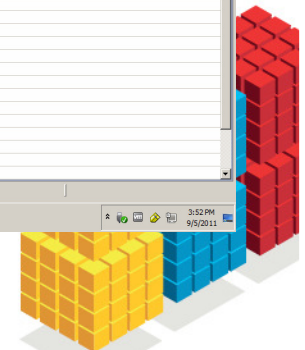
I wish to complain about the charges made to my account

I walked into your Mollongong Branch the other day and was greeted by one of the rudest people I have met. Richard Collins, your branch manager at this location treated me as if I I've was an idiot.

Mark Robinson

The Properties table at the bottom right shows the following data:

complaint	Part of Speech
complaint 1 - 18. awful - unhappy	
awful	Adjective
break	Verb
complain	Verb
delay	Verb
disgust	Verb
disappoint	Verb
hit	Verb
horrible	Adjective
inaccessible	Adjective
malfunction	Verb
out of order	Verb
queue	Noun
queue	Verb
regret	Verb







# Configuring Analytics to meet your business needs

After you add a facet to the facet tree, edit the facet to map search fields to it and to organize how the subfacets are displayed.

**Facet Tree:**

- Root
  - Part of Speech
  - Phrase Constituent
  - Named entity
  - Report Information
  - Device Information
  - Manufacturer Information
  - Patient Information
    - Patient Treatment
    - Patient Outcome**
  - Hazards

**Add a facet:**

\*Facet path:   
Facet name:   
Generate patterns:   
Visible in the text miner:   
Add counts to parent facet:   
**Add**

**Edit a facet:**

\*Facet path:   
Facet name:   
Generate patterns:   
Visible in the text miner:

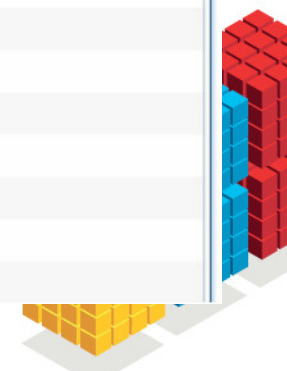
Facet Navigation: Default order

Filter:  **Clear**

- Part of Speech
- Phrase Constituent
- Named entity
- My Keywords
- Voice of Customers (sample)
  - Unfavorable
  - Favorable
    - Favorable predicate
    - Favorable noun**
    - Favorable pattern
    - Favorable phrase
  - Gratitude
  - Requirement
  - Question

Show: **Keywords**    Filter:

	Keywords	Frequency	1	Correlation
<input type="checkbox"/>	good photo	1	<div style="width: 100%;"></div>	1.0
<input type="checkbox"/>	good thing	1	<div style="width: 100%;"></div>	1.0
<input type="checkbox"/>	good discussion	1	<div style="width: 100%;"></div>	1.0
<input type="checkbox"/>	good experience	1	<div style="width: 100%;"></div>	1.0



# Content Analytics Views

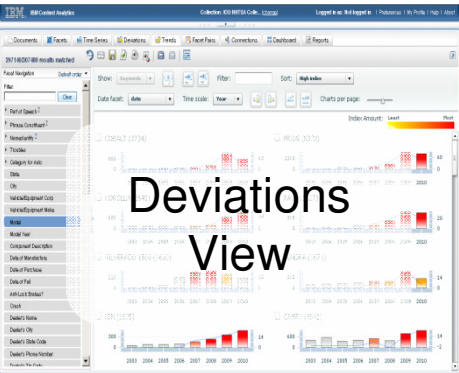
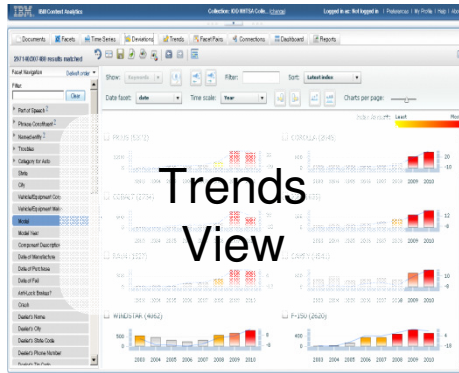
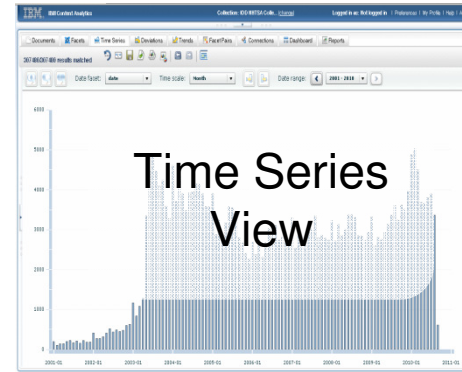


**Document View**

Document snippet: ...MILWAUKEE TO STOP INDEPENDENT, WIPON MORE THAN THREE TIMES HAVE BEEN ON BY A... FILE A COMPLAINT WITH SCORE "75"

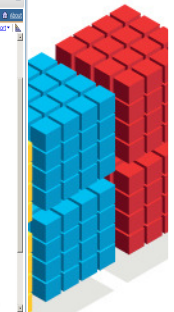
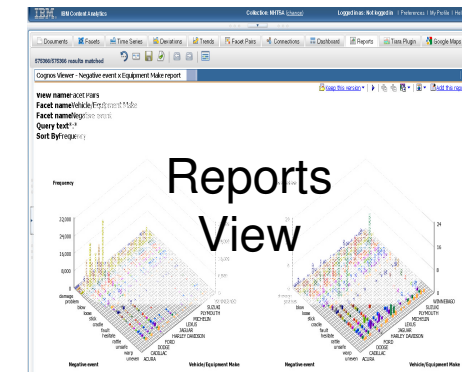
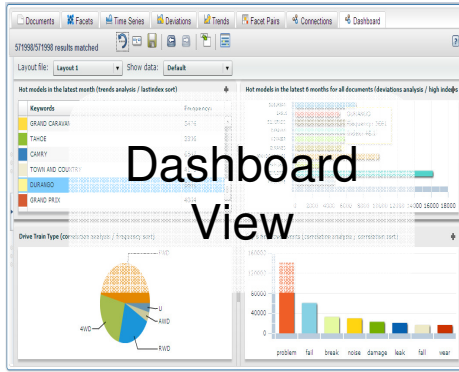
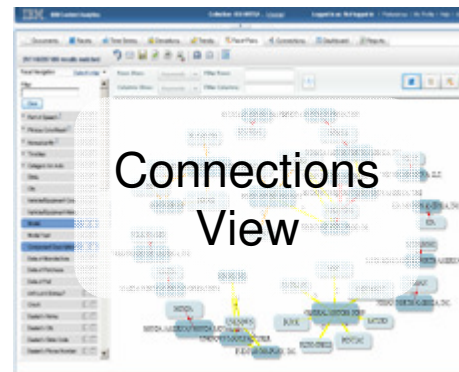
**Facet View**

Keywords	Frequency	Correlation
MILWAUKEE	10	0.95
INDEPENDENT	8	0.92
FILE A COMPLAINT	5	0.88
SCORE	3	0.85
WIPON	2	0.82
THREE TIMES	2	0.80
HAVE BEEN	2	0.78
ON BY A	2	0.75
...	...	...



**Facet Pairs View**

Facet 1	Facet 2	Count
Part of Question?	Phrase Occurrence?	100
Part of Question?	Name/Entity?	80
Part of Question?	Toolbox	60
Part of Question?	Category for Axis	40
Part of Question?	Title	30
Part of Question?	City	20
Part of Question?	Manufacturer	15
Part of Question?	Model	10
Part of Question?	Component Description	8
Part of Question?	Date of Purchase	5
Part of Question?	Date of Failure	3
Part of Question?	APN/Case Number?	2
Part of Question?	Owner	1
Part of Question?	Owner's Name	1
Part of Question?	Owner's City	1
Part of Question?	Owner's State Code	1
Part of Question?	Owner's Phone Number	1





## Demonstrations

- Data Extraction Demonstration
- Visualisation and Data mining Demonstration

