Smarter Analytics 2012

Analysis of Unstructured Content enhances data analytics and improves decision making

Lloyd Parata
ECM Solution Specialist
Singapore

Please Note:

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.
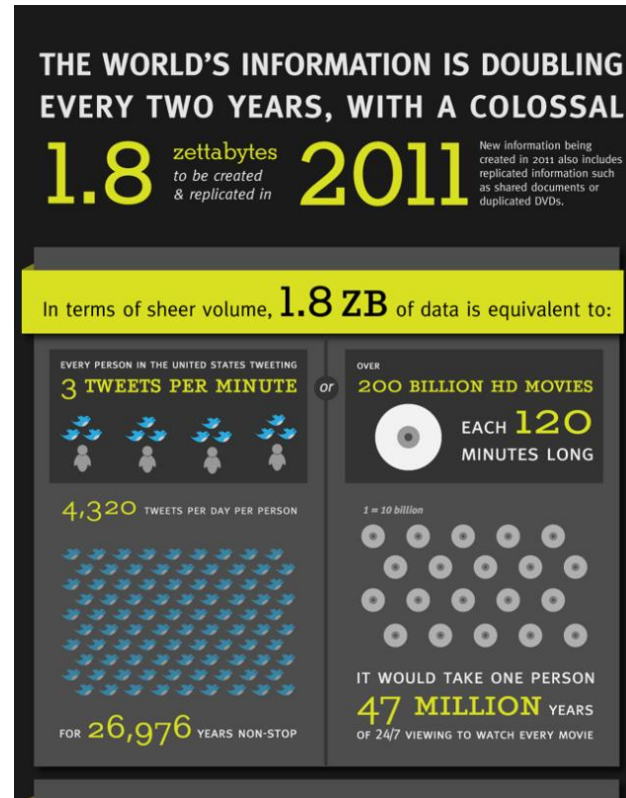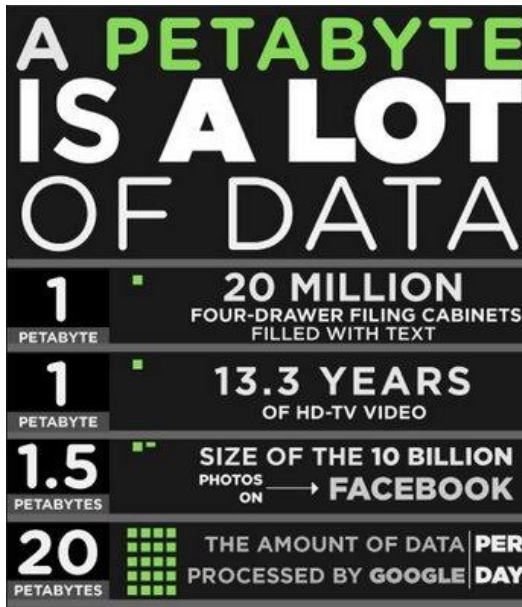
Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

All the data that existed up to 2003, we generate now in 2 days.

Every day, 5 Exabytes of data are created

2 MILLION BLOG POSTS are written.

Enough posts to fill Time Magazine for 770 years.

TIME

A PETABYTE IS A LOT OF DATA

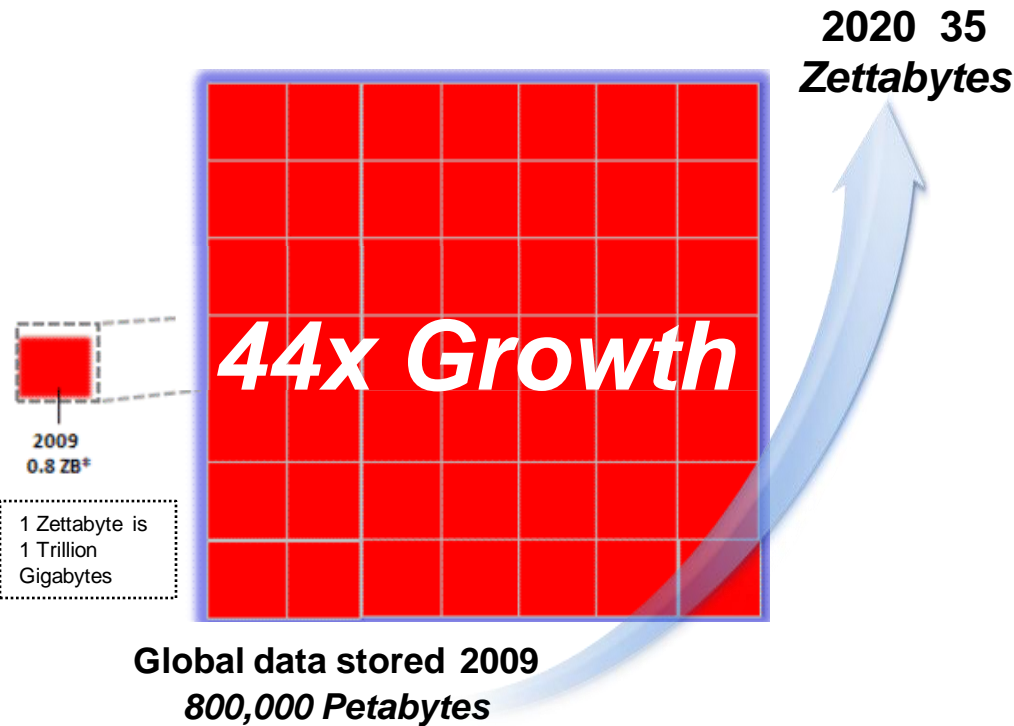|	
1 PETABYTE	**20 MILLION** FOUR-DRAWER FILING CABINETS FILLED WITH TEXT
1 PETABYTE	**13.3 YEARS** OF HD-TV VIDEO
1.5 PETABYTES	**SIZE OF THE 10 BILLION** PHOTOS ON → **FACEBOOK**
20 PETABYTES	THE AMOUNT OF DATA PER PROCESSED BY **GOOGLE** DAY

THE WORLD'S INFORMATION IS DOUBLING EVERY TWO YEARS, WITH A COLOSSAL

1.8 zettabytes to be created & replicated in **2011**

New information being created in 2011 also includes replicated information such as shared documents or duplicated DVDs.

In terms of sheer volume, **1.8 ZB** of data is equivalent to:

EVERY PERSON IN THE UNITED STATES TWEETING
3 TWEETS PER MINUTE

or

OVER
200 BILLION HD MOVIES
EACH **120** MINUTES LONG

4,320 TWEETS PER DAY PER PERSON

1 = 10 billion

FOR **26,976** YEARS NON-STOP

IT WOULD TAKE ONE PERSON
47 MILLION YEARS
OF 24/7 VIEWING TO WATCH EVERY MOVIE

294 BILLION emails are sent.

SENT!

It would take **2 years to process** that many pieces of mail in the US.

MAR 2014

In one day, enough information is consumed by internet traffic to fill **168 MILLION DVDS.**

= 1 MILLION

1 Gigabyte	= 1,000,000,000
1 Terrabyte	= 1,000,000,000,000
1 Petabyte	= 1,000,000,000,000,000
1 Exabyte	**= 1,000,000,000,000,000,000**
1 Zetabyte	= 1,000,000,000,000,000,000,000
1 Yottabyte	= 1,000,000,000,000,000,000,000,000
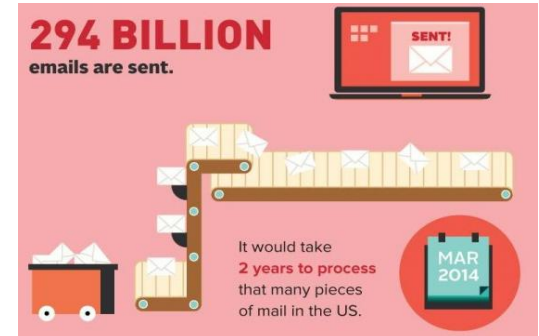
All this information is at the Center of a New Wave of Opportunity…

… and Organizations Need Deeper Insights

2020 35 Zettabytes

44x Growth

2009 0.8 ZB*

1 Zettabyte is 1 Trillion Gigabytes

Global data stored 2009
800,000 Petabytes

1 in 3 Business leaders frequently make decisions based on information they don't trust, or don't have

1 in 2 Business leaders say they don't have access to the information they need to do their jobs

60% of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions

80% Of the world's data is unstructured

Sources: IDC and Smarter Planet series: Smarter Analytics episode; BAO client presentation using: The Guardian, May 2010

Why should you care about the growth in data and information being created?

Because your competitors who are using Analytics today are improving their business decisions…

5.4x Top performers are 5.4X more likely to use an analytic approach over intuition

83% of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

3x Organizations that lead in analytics outperform those who are just beginning to adopt analytics
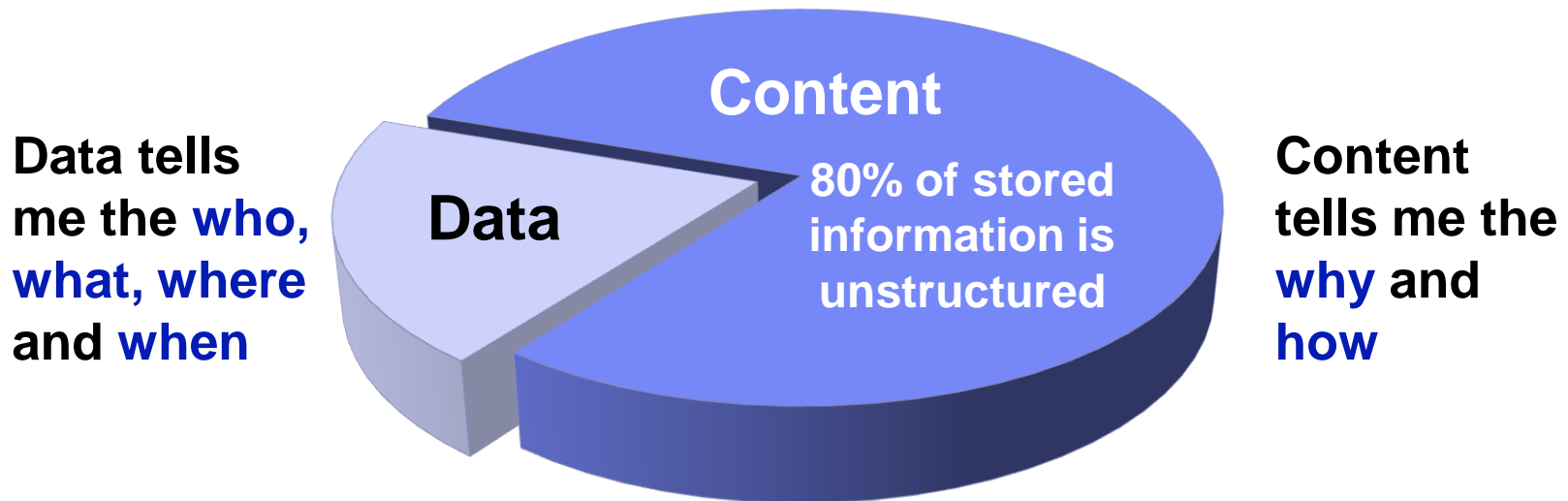
57% increase in respondents saying analytics creates a competitive advantage: increased from 37% in 2010 to 58% in 2011

2.2x Organizations achieving a competitive advantage with analytics are 2.2x more likely to substantially outpreform their industry peers

Together, data and content represent the complete business context, but Content is 4/5 of all information

Data tells me the who, what, where and when

Data

Content

80% of stored information is unstructured

Content tells me the why and how

Most business analytics initiatives are about data.

Analysing your Unstructured Content requires a new approach….

Traditional approaches to Analytics are converging
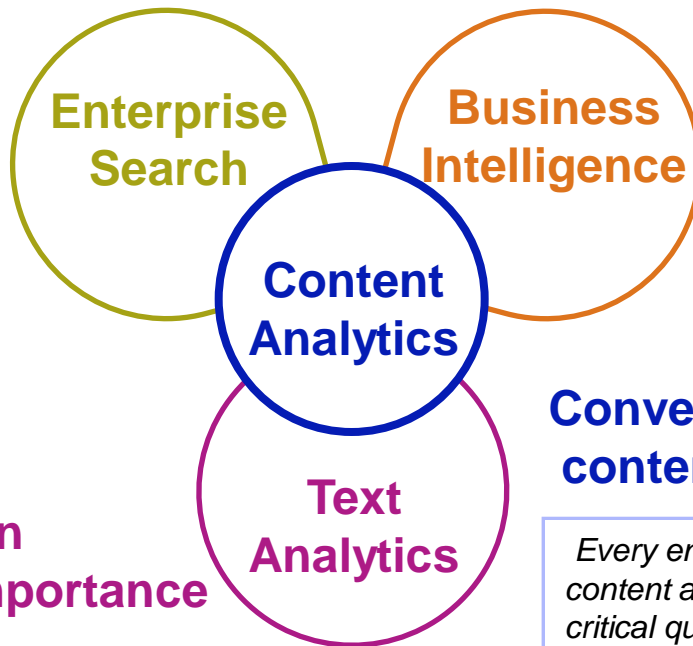
More than keyword search is needed

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

Enterprise Search

Business Intelligence

Content Analytics

Text Analytics

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

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

How do you 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.**

social media

documents

images

web forms

video

e-mail

mobile devices
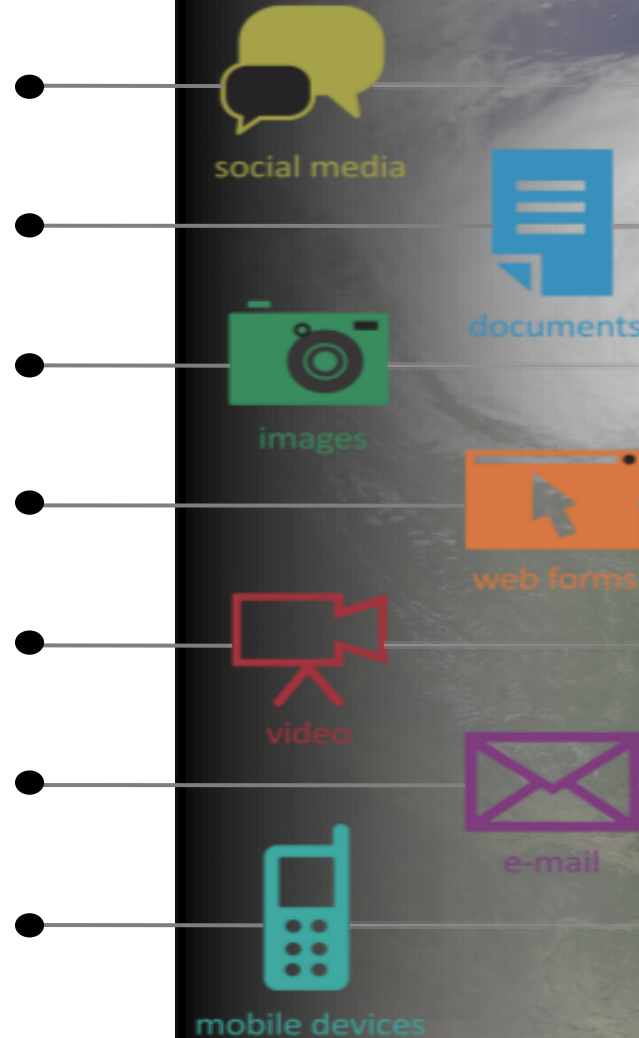
IBM Content Analytics combines the power of Analytics with the scale of Enterprise & Website Text Search

Mine unstructured information to improve service levels and build relationships

Transform raw information into business insight quickly without building models or deploying complex systems.

Visually identify and explore trends, patterns, and statistically relevant facts.

Gain deeper understanding of existing customers

Discover cross-sell or up-sell opportunities

Key Point – Content Analytics is about the discovery of information, not keyword search

Rapidly Derived Insight

Search and Explore

Analyze and Visualize

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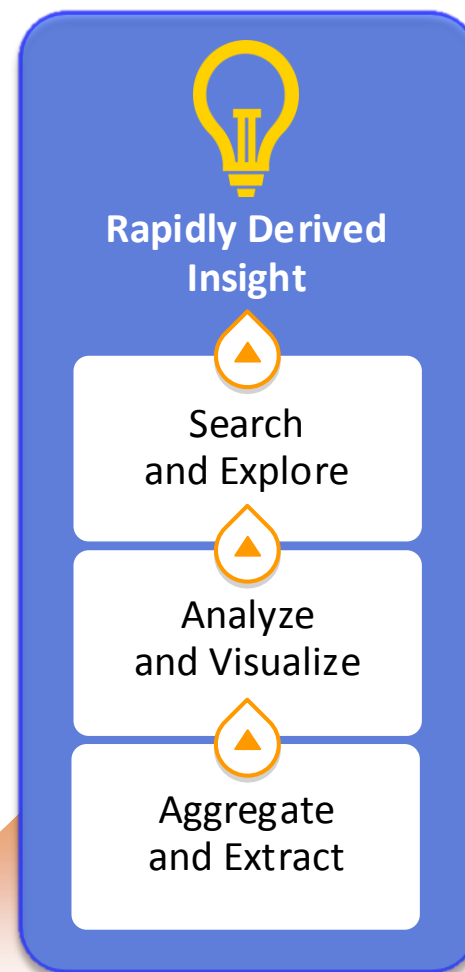
External and Internal Content (and Data) Sources including Social Media and More

IBM Leadership in Search, Text Analysis and Classification

IBM has a 50+ year history in text analysis and discovery

In 1957, IBM published pioneer research done on text classification (and related topics, such as text search, and automatic creation of text abstracts)

IBM invests ~$50M annually in research and development for search and text analytics

Over 200 people actively engaged in R&D

IBM holds over 200 patents in information access with more each year

H. P. Luhn

A Statistical Approach to Mechanized Encoding and Searching of Literary Information*

Abstract: Written communication of ideas is carried out on the basis of statistical probability in that a writer chooses that level of subject specificity and that combination of words which he feels will convey the most meaning. Since this process varies among individuals and since similar ideas are therefore relayed at different levels of specificity and by means of different words, the problem of literature searching by machines still presents major difficulties. A statistical approach to this pr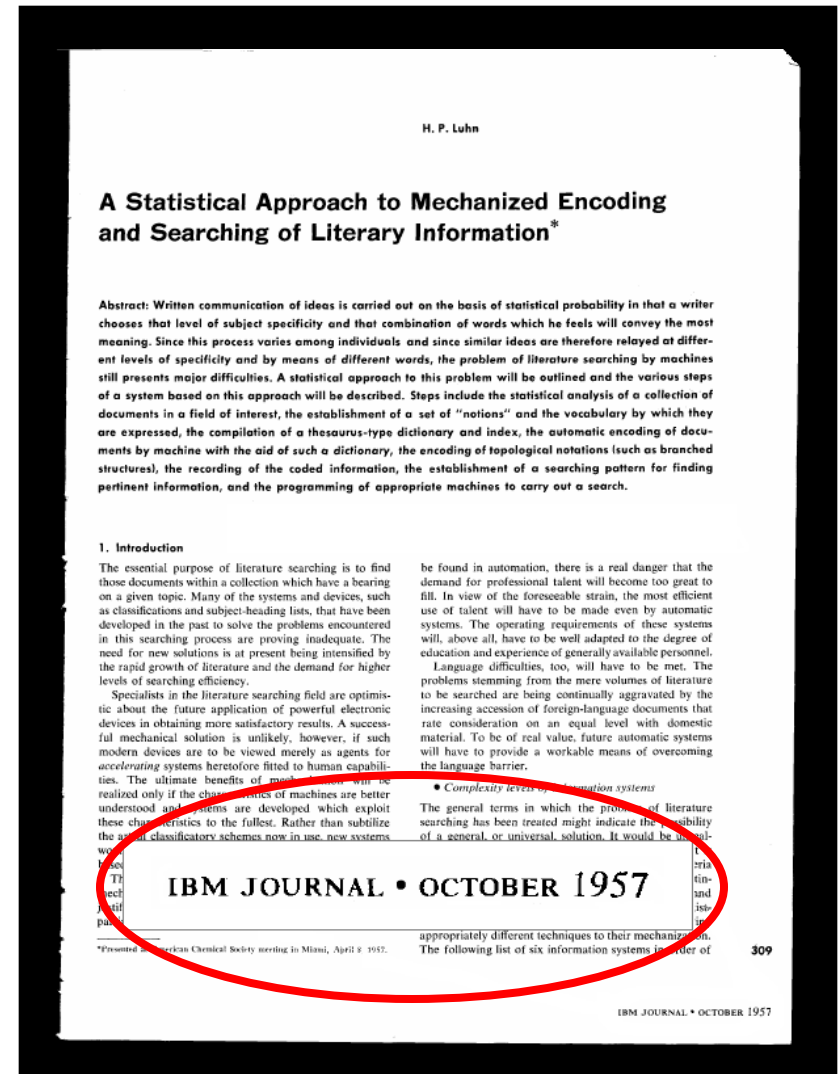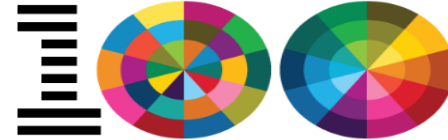oblem will be outlined and the various steps of a system based on this approach will be described. Steps include the statistical analysis of a collection of documents in a field of interest, the establishment of a set of "notions" and the vocabulary by which they are expressed, the compilation of a thesaurus-type dictionary and index, the automatic encoding of documents by machine with the aid of such a dictionary, the encoding of topological notations (such as branched structures), the recording of the coded information, the establishment of a searching pattern for finding pertinent information, and the programming of appropriate machines to carry out a search.

1. Introduction

The essential purpose of literature searching is to find those documents within a collection which have a bearing on a given topic. Many of the systems and devices, such as classifications and subject-heading lists, that have been developed in the past to solve the problems encountered in this searching process are proving inadequate. The need for new solutions is at present being intensified by the rapid growth of literature and the demand for higher levels of searching efficiency.

Specialists in the literature searching field are optimistic about the future application of powerful electronic devices in obtaining more satisfactory results. A successful mechanical solution is unlikely, however, if such modern devices are to be viewed merely as agents for accelerating systems heretofore fitted to human capabilities. The ultimate benefits of mechanization will be realized only if the characteristics of machines are better understood and systems are developed which exploit these characteristics to the fullest. Rather than subtilize the age-old classificatory schemes now in use, new systems...

be found in automation, there is a real danger that the demand for professional talent will become too great to fill. In view of the foreseeable strain, the most efficient use of talent will have to be made even by automatic systems. The operating requirements of these systems will, above all, have to be well adapted to the degree of education and experience of generally available personnel.

Language difficulties, too, will have to be met. The problems stemming from the mere volumes of literature to be searched are being continually aggravated by the increasing accession of foreign-language documents that rate consideration on an equal level with domestic material. To be of real value, future automatic systems will have to provide a workable means of overcoming the language barrier.

Complexity levels of information systems

The general terms in which the problem of literature searching has been treated might indicate the possibility of a general, or universal, solution. It would be appropriately different techniques to their mechanization. The following list of six information systems in order of...

309

IBM JOURNAL • OCTOBER 1957

*Presented at American Chemical Society meeting in Miami, April 8, 1957.

IBM JOURNAL • OCTOBER 1957

IBM: 100 Years of Innovation

巧み = skillful, clever

匠 = craftsman, artisan

"*I didn't invent TAKMI to do something humans could do, better. I wanted TAKMI to do something that humans could not do.*"

Nasukawa-san

TAKMI
Bringing Order to Unstructured Data

IBM EXPERT

Tetsuya Nasukawa,
Dr. Nasukawa joined IBM in 1989, and has been leading text mining projects since 1997, including sentiment analysis, conversation mining and cross-lingual text mining. Prior to text mining, he was involved in machine translation and digital library projects.

TAKMI: BRINGING ORDER TO UNSTRUCTURED DATA

IBM
Solutions Today

IBM ContentAnalytics is the commercially available version of TAKMI.

IBM Content Analytics
CRM Analytics Software and the Contact Center of the Future

In 1997, IBM researchers at the company's Tokyo Research Laboratory pioneered a prototype for a powerful new tool capable of analyzing text. The system, known as TAKMI—for Text Analysis and Knowledge Mining—was a watershed development: for the first time, researchers could efficiently capture and utilize the wealth of buried knowledge residing in enormous databases of text.

Home > Podcasts > Business & Management
Mining the Talk: Unlocking the Business Value in Unstructured Information – Part 1 (audio)
Scott Spangler describes how to unlock the business value hidden in unstructured data (word processing docs, websites, emails, instant messages). Learn about breakthrough opportunities to become responsive, agile, & competitive. Part 1 of 2.

BizTech
Topics: Business & Management, Home & Office Computing

UIMA: Unstructured Information Management Architecture

..an open, industrial-strength platform for unstructured information analysis and search. Developed by IBM. Now Apache Open Source project and OASIS standard

OASIS
Advancing open standards for the information society

IBM Content Analytics adds value to…

Healthcare Analytics
- *Analyzing:* Patient care, hospital reports
- *For:* Clinical analysis; treatment protocol optimization
- *Benefits:* Better management of chronic diseases; optimized drug formularies; improved patient

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

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
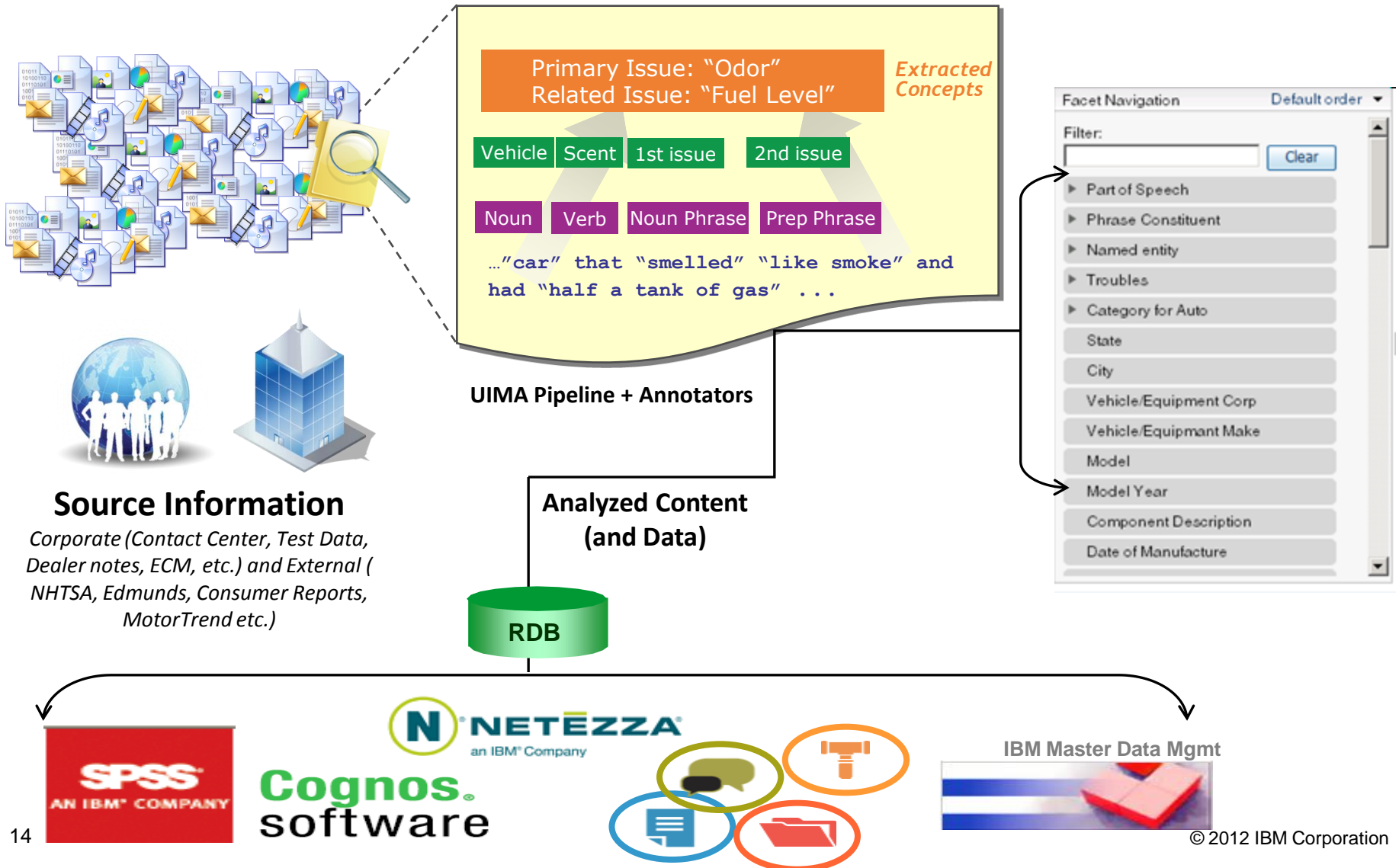
New York Police Department
Smart is: reducing crime

- Search and analyze complaints, police reports, 911 records, arrest records, and data marts

- All of these types of Text have 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 unsolved (years old) murder cases were solved

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

The PoliceChief
The Professional Voice of Law Enforcement

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

Content Analytics: How does it work? (Top USA Car Model problems)

Extracted Concepts

Primary Issue: "Odor"
Related Issue: "Fuel Level"

| Vehicle | Scent | 1st issue | 2nd issue |

| Noun | Verb | Noun Phrase | Prep Phrase |

…"car" that "smelled" "like smoke" and had "half a tank of gas" ...

UIMA Pipeline + Annotators

Facet Navigation — Default order

Filter: _____ Clear

▶ 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

Source Information

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

Analyzed Content (and Data)

RDB

SPSS AN IBM COMPANY

Cognos software

NETEZZA an IBM Company

IBM Master Data Mgmt

Content Analytics explore relationships between Car problems: Free Text Values

By exploring the relationship between documents from different sources you can quickly identity unusual correlations or anomalies that might require further investigation.
Example Oil Leaks, Windshield Cracks, Flat Tyre or Vibration

SmarterAnalytics

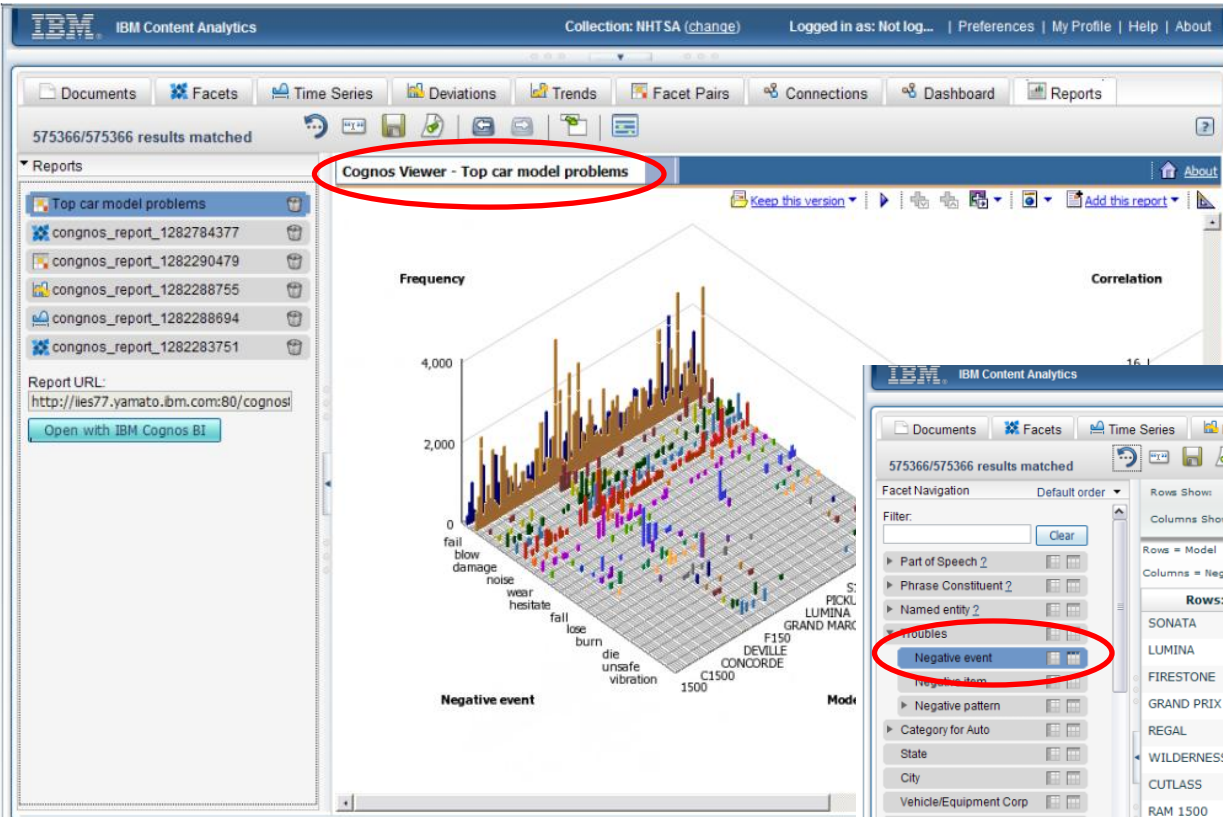Content Analytics identifies highly correlated terms – Ford, blow & Firestone
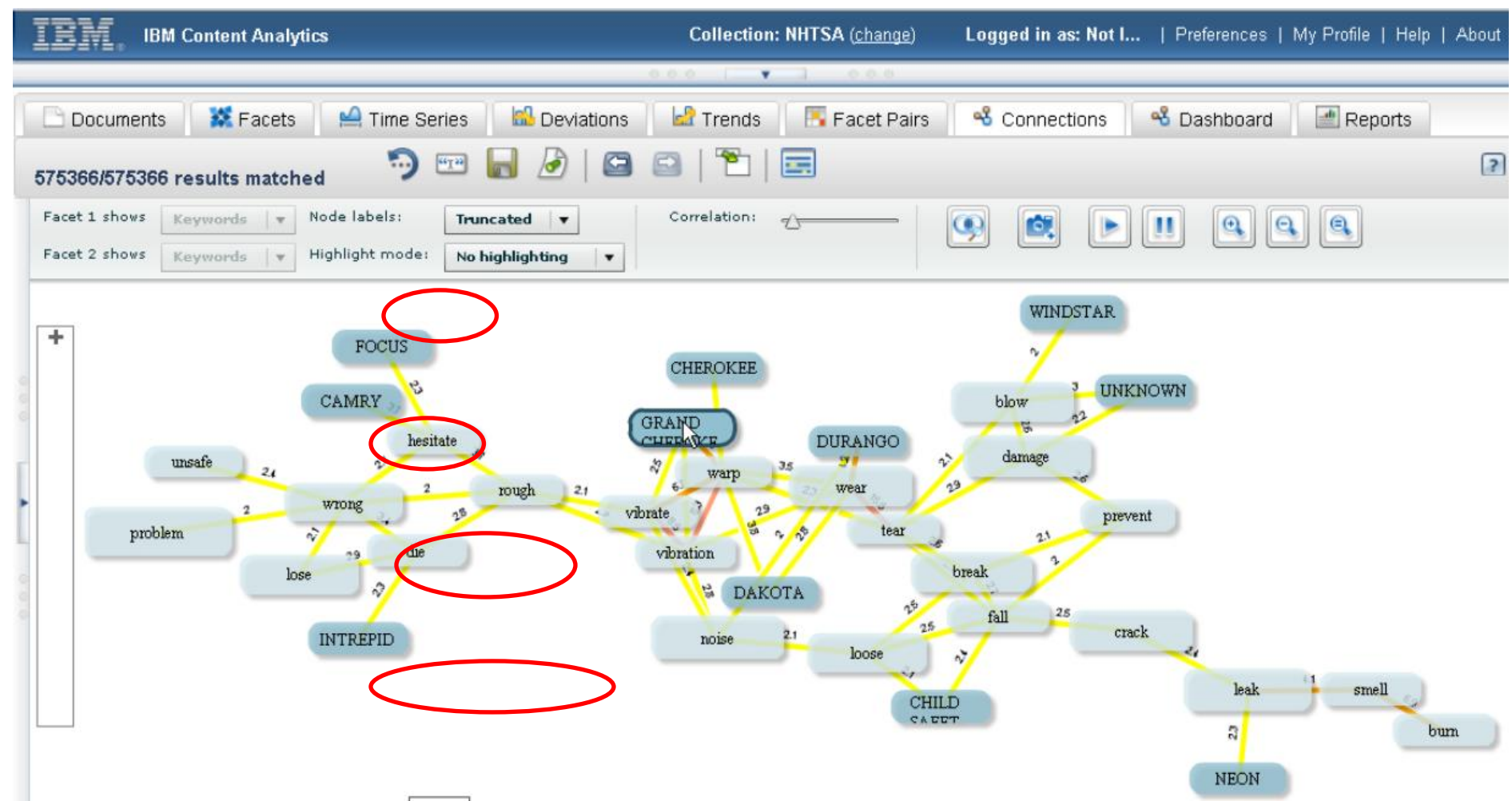
Combine IBM Content Analytics with Cognos BI

Creating reports in Content Analytics, Business Analysts can gain unique insights by examining the unstructured data which is usually in a Free Text format in using Cognos.

Content Analytics: Top Car Model Problems Dashboard

Search Query Exploration

Views, Filters and Thresholds

Automatically Extracted and Analyzed Concepts, Entities, Relationships, Meta Data and Classifications
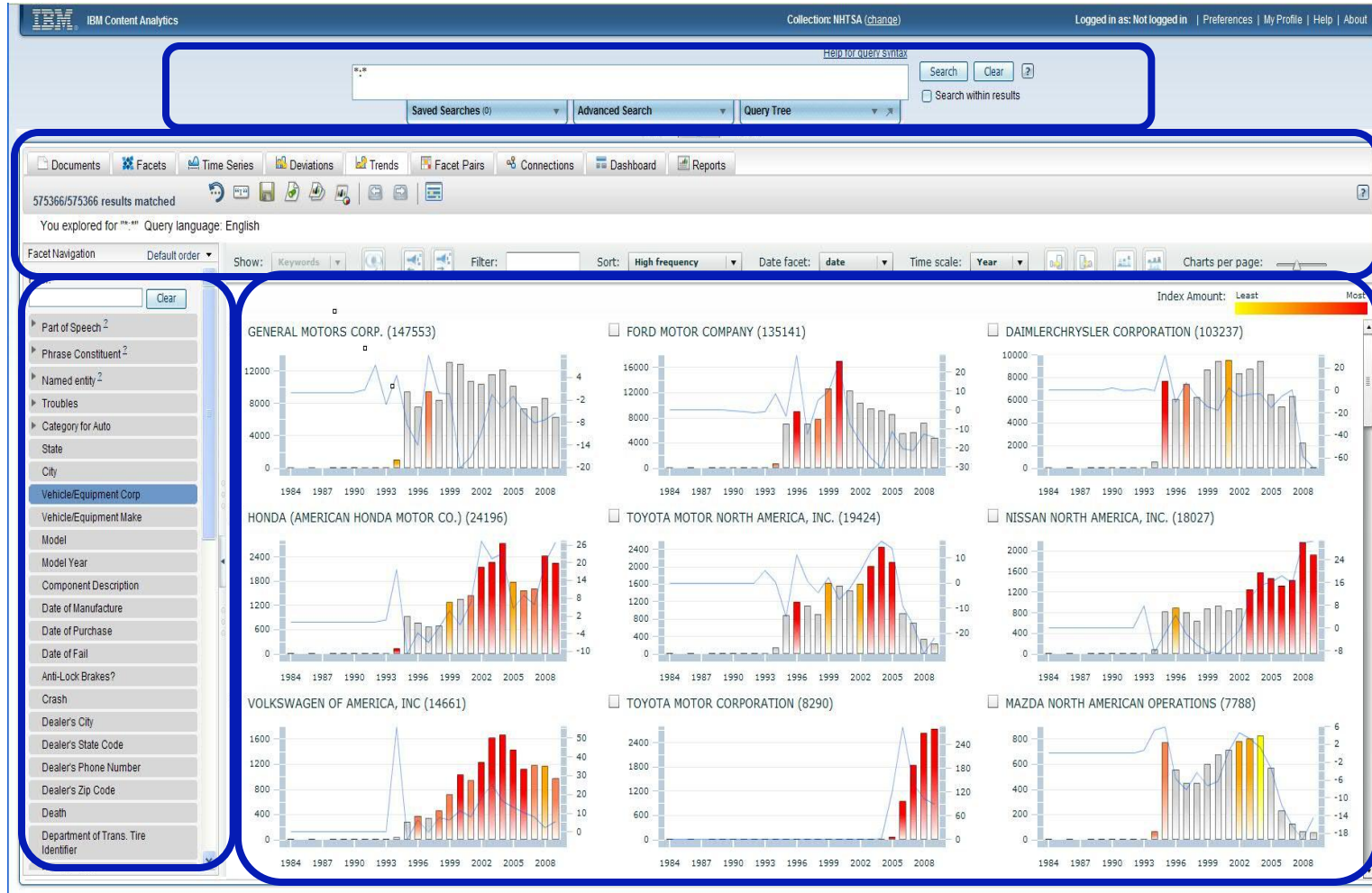
Visualization with Drill Down for Exploration and Assessment

Government Agency: Responsible for global threat identification

This organization is a government intelligence agency responsible for identifying and mitigating global terrorist threats to enhance the security of their citizens at home and abroad.

Uncovering new insights

The details

Business Challenge

How to build a social network analysis of all terrorists and related persons of interest from multiple information sources and documents

What's Smart?

Implemented IBM Content Analytics to extract key facts and entities from unstructured information to create a social network hub that combines structured and unstructured information of latest terrorist activity.

Smarter Business Outcomes

Information reaches agents in hours, not days or weeks. Previously unknown relationships between suspects and new persons of interests automatically uncovered and viewable in a single version of the truth.

Structured Data ETL

Information Server provided high performance and quality extraction and cleansing of the structured data, workflow definition and management.

Unstructured Content ETL

IBM Content Analytics provided high precision and recall extraction of facts and entities from all text creating indexes for quick search and analysis. It then passed this new structured content to Information Server for conventional ETL.

Single data warehouse to layer analytics

All cleansed information populated an InfoSphere data warehouse. The organization used key analytical applications like Identity Insights and i2 to associate relationships between terrorists and analyze their activity.

Content Analytics – Voice Of the Customer (VoC)

A set of linguistic, statistical, and machine learning techniques to analyze and extract key facts, entities and concepts from text used to…

Examination:

</Product>

</Sentiment>

I charged my **Smart 3G** through out the night and when I took it out of the charge it was at full. But, it runs out of **battery** in 3 hours. **It is annoying**. How do I preserve the **battery** life?

</Component>

</Question>

Model	**Smart 3G**
Component **Sub-Component**	**Battery** **Battery Life**
Sentiment Concept: Detail:	Negative Poor Battery life **Annoying, 3 hours**
Question	How do I preserve **battery life?**
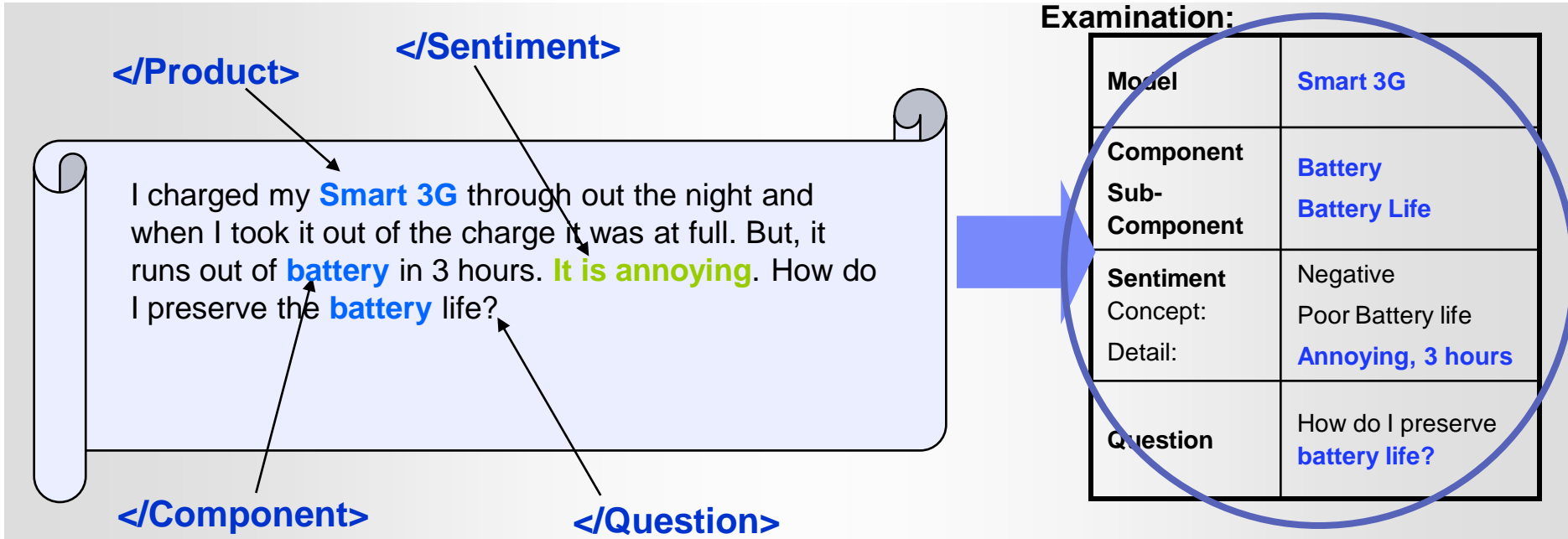
…visually search, identify and explore trends, patterns, and statistically relevant facts found in various types of content spread across internal and external content sources.

Analysing the Voice Of the Customer (VoC)

1 Customer → **Shop** | **Sales** | **Web** | **Call Center**

Customized for Business Scenario

IBM Content Analytics

2

Lexical Analysis → Dictionary Lookup → Pattern Matching → Rule-based Categorization | ICA Function

| Word | Phrase | Product | Service | Voice of Customer | Question Category | Extracted Information |

3

Category = *Service > Data communications > xxx discount*

I heard about that you are going to change the estimation method of "**xxx discount**" service. Please **give me more details** on the changes made.

Product = *Smart 4G*

Category = *Product > Smart 4G > Battery*

I charged my **Smart 4G** through out the night and when I took it out of the charge it was at full. But, it runs out of **battery** in 3 hours. **It is annoying**. How do I preserve the **battery** life?

Service = *Data Communication*   VOC = *Question*

Component = *battery*   VOC = *Unfavorable*

4

2D Map Analysis
Product / Voice of Customer

Trends Analysis
Category A
Category B
Category C
Category D
Time

Time Series Analysis
Frequency / Time

And More…

A mobile telecommunication provider utilizes ICA to analyze large volumes of information from multiple customer service centers for better customer support.

Analyzer → **Customer Service Sales / R&D Dep.** **5**

A Telecommunications Company
Smart is: reducing customer churn using VoC

"Insight into customer interaction logs is an information gold mine for us."

— General Manager
Japanese Telecomm Business

Industry context: telecommunications
Value driver: improve customer service
Solution: Content Analytics

Business Challenge
Adopt a customer-oriented business strategy to offer highly satisfying products and services based on real voice of customers (VoC).

What's Smart?
They process call center notes and customer emails to detect likely candidates for customer churn. A rules-based text analysis engine in IBM Content Analytics detects the customer churn candidates. An alerting engine then automatically sends reports to a department that deals specifically with customer churn situations.

Smarter Business Outcomes
Improved rates for model and service upgrades to loyal customers. Started new Premium Club points program based on VoC. Set initial parameters of mobile phones based on VoC.

SmarterAnalytics

IBM

Smart is: finding content that matters

Drive users to the knowledge they seek and enhance the visibility of content and context of unstructured information

Industry context: Scientific research
Value driver: Connecting researchers to content
Solution : Content Analytics

Business Challenge
How to enable 13,000 staff scientists and engineers from 56 Institutes to securely search for and discover relevant scientific research papers to enable new business innovations

What's Smart?
In 4 months, this research organization fully implemented Content Analytics to enable secure semantic search and classification of research assets for their internal portal that services all Institutes, as well as their public facing website

Smarter Business Outcomes
Scientists and engineers from this institute are able to find the most relevant research papers and knowledge experts, leading to continued innovations

Content Analytics analyses free text within emails

Make detailed distinctions between long-form texts with very different intent
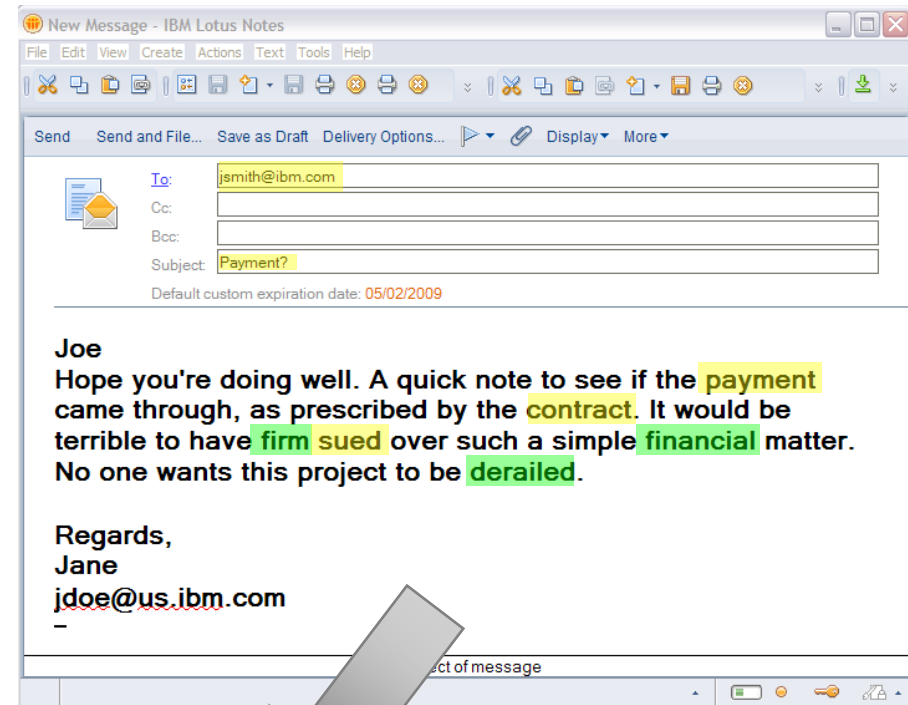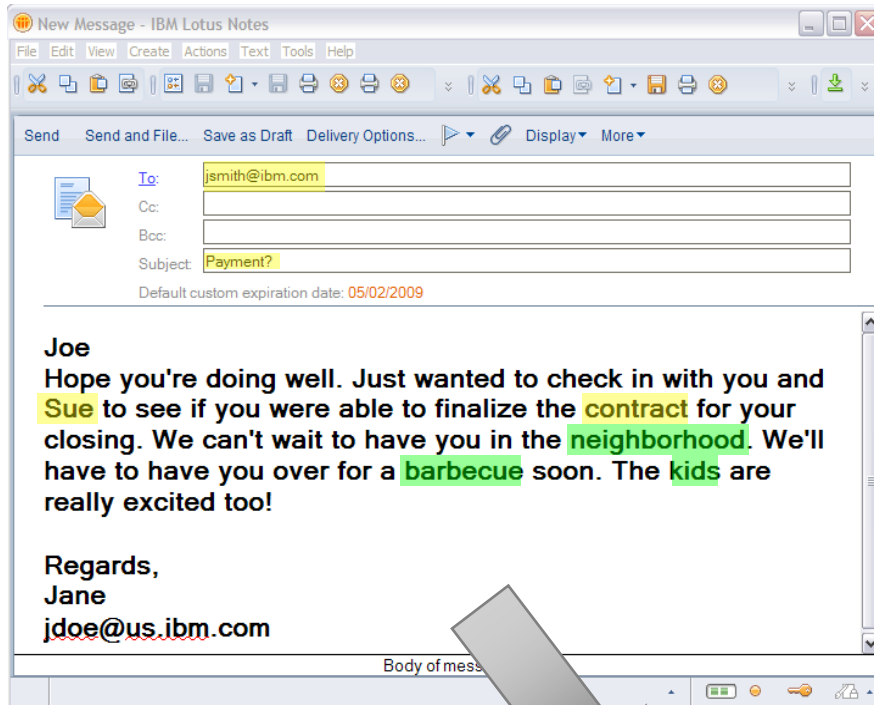
Seton Healthcare Family Hospital
Reducing Congestive Heart Failure readmissions to improve care

"IBM Content and Predictive Analytics for Healthcare uses the same type of natural language processing as IBM Watson, enabling us to leverage information in new ways not possible before. We can access an integrated view of relevant clinical and operational information to drive more informed decision making and optimize patient and operational outcomes."

Charles J. Barnett, FACHE, President/Chief Executive Officer, Seton Healthcare Family

Business Challenge

Seton Healthcare strives to reduce the occurrence of high cost Congestive Heart Failure (CHF) readmissions by proactively identifying patients likely to be readmitted on an emergent basis.

What's Smart?

IBM Content and Predictive Analytics for Healthcare solution will help to better target and understand high-risk CHF patients for care management programs by:

- Utilizing natural language processing to extract key elements from unstructured History and Physical, Discharge Summaries, Echocardiogram Reports, and Consult Notes

- Leveraging predictive models that have demonstrated high positive predictive value against extracted elements of structured and unstructured data

- Providing an interface through which providers can intuitively navigate, interpret and take action

Smarter Business Outcomes

Seton will be able to proactively target care management and reduce re-admission of CHF patients.

Teaming unstructured content with predictive analytics, Seton will be able to identify patients likely for re-admission and introduce early interventions to reduce cost, mortality rates, and improved patient quality of life.

IBM solution

- IBM Content and Predictive Analytics for Healthcare

- IBM Cognos Business Intelligence

- IBM BAO solution services

Medical Transcription Discharge Summary Sample # 2:

DATE OF ADMISSION: MM/DD/YYYY
DATE OF DISCHARGE: MM/DD/YYYY
ADMITTING DIAGNOSIS: Syncope.

CHIEF COMPLAINT: Vertigo or dizziness.

HISTORY OF PRESENT ILLNESS: This is an (XX)-year-old male with a past medical history of coronary artery disease, EKG done a few years ago, atrial fibrillation, peripheral neuropathy, atherosclerosis, hemorrhoids... [obscured] ...The patient came to the ER for an episode of vertigo while reaching for some books. The patient was able to reach the books, to support self, but did not have any syncope. No nausea or vomiting. No chest pain. No shortness of breath. Came to ER and had a CT head, which was within normal limits. The impression was atrophy with no ischemic new findings. No acute intracranial findings. No focal weakness... [obscured] ...since one year and not relieved with multiple medications. The patient also complains of weight loss of 25 pounds in the last 6 months. No colonoscopy done. Recent history of hematoch... [obscured] ...appetite. No nausea, vomiting, no abdominal pain.

PROCEDURES PERFORMED: The patient had a chest x-ray, ... thickening and small pleural effusion, a left costophrenic an... pattern. The patient also had a head CT which showed atro...

CONSULTS OBTAINED: A rehab consult was done.

PAST MEDICAL/SURGICAL HISTORY: Positive for atrial fibri... hypertension, peripheral neuropathy, atherosclerosis, hemo...

FAMILY HISTORY: Positive for atherosclerosis, hypertension...

SOCIAL HISTORY: Never smoked. Alcohol socially. No drug...

ALLERGIES: NO KNOWN DRUG ALLERGIES

REVIEW O... [obscured] hemorrho... and feet.

PHYSICA... Appearan... masses. ... rhythm. ... smooth ... within no...

LABORAT... neutroph... 1.6, PTT 3...

The patie... effusion, ... head CT, ...

HOSPITA...

1. Syncop... and the p... ordered a... Doppler ... 59%. The... 1.38 and... informed... rate cont... suggested... cardiac re... valve rep... showed p... arterioscl... whatsoev...

Echocardiogram Sample Report:

DATE OF STUDY: MM/DD/YYYY

DATE OF INTERPRETATION OF STUDY:

Echocardiogram was obtained for assessm... function. The patient has been admitted ... syncope. Overall, the study was suboptim...

FINDINGS:

1. Aortic root appears normal.
2. Left atrium is mildly dilated. No gross i... recognized, although subtle abnormalities... atrium is of normal dimension.
3. There is echo dropout of the interatrial ... could not be excluded.
4. Right and left ventricles are normal in i... ventricular systolic function appears to be... fraction is around 55%. Again, due to poo... abnormalities in the distribution of lateral ... excluded.
5. Aortic valve is sclerotic with normal exc... Doppler study demonstrates trace aortic r...
6. Mitral valve leaflets are also sclerotic w... imaging and Doppler study demonstrates ... regurgitation.
7. Tricuspid valve is delicate and opens n... clearly seen. No evidence of pericardial e...

CONCLUSIONS:

Cardiology Consultation Transcribed Medical Transcription Sample Reports

FERRING PHYSICIAN: John Doe, MD

CONSULTING PHYSICIAN: Jane Doe, MD

HISTORY OF PRESENT ILLNESS: This (XX)-year-old lady is seen in consultation for Dr. John Doe. She has been under consideration for ventral hernia repair and has a background of aortic valve replacement and known coronary artery disease. The patient was admitted with complaints of abdom... [obscured] ...vomiting. She underwent a CT scan of the abdomen and pelvis ... this sho... hernia involving the transverse colon, but without strangulation. There was an atrophic right kidney. She had bilateral renal cysts. The hepatic flexure wall was thickened. There was sigmoid diverticulosis without diverticulitis. It has been recommended to her that she undergo repair of the ventral hernia. For this reason, cardiology consult is obtained to assess... that she can be cared from the cardiac standpoint.

PAST CARDIAC HISTORY: Bypass surgery. She underwent echocardiography and cardiac ... ation. Echocardiography showed an ejection fraction of 50%. There ... ertrophy with septal wall 1.60 cm and posterior wall 1.55 cm. ... 90% stenosis in the anterior descending artery, situated distally ... ventricle. Only mild to moderate narrowing was seen elsewhere in...

Her father had an irregular heartbeat and her brother had a fatal ... d high blood pressure for 20 years. She has elevated cholesterol ... diabetes for 20 years. She is not a cigarette smoker. She does little...

CARDIOVASCULAR AND RESPIRATORY: She has no chest pain. She ... eath if she walks too far. No cough. She has occasional swelling of ... mildly lightheaded. Has not lost consciousness. She tends to be ... e is tired. She has no history of heart murmur or rheumatic fever. ... I symptoms as noted above, but she does not usually have such ... atemesis. She has no history of ulcer or jaundice. She sometimes ... n and no blood in the stool. GENITOURINARY: She tends to have ... once at night to pass urine. No dysuria, incontinence. She has had ... stones noted. NEUROLOGIC: She has occasional headaches. No ... hearing, or speech. No limb weakness. MUSCULOSKELETAL: She ... pains and has a history of gout. HEMATOLOGIC: No anemia, ... blood transfusion. GYNECOLOGIC: No gynecologic or breast...

e has had shoulder and hand injuries and has had carpal tunnel ... and has been on insulin. She has chronic renal insufficiency with ... had hypothyroidism. She has had morbid obesity. She has chronic ... BiPAP. She has had hysterectomy and oophorectomy in the past.

tal, she was taking glipizide XL 2.5 mg daily, metoprolol 50 mg ... roid 75 mcg daily, aspirin 81 mg daily, ... ntly, she is taking Lipitor 40 mg daily, Lantus 10 units at bedtime, ... rolol 50 mg b.i.d., and Zosyn 2.25 grams q.6h.

... no distress. She is alert, oriented, and ... react normally. No icterus. Mucous membranes well colored. ... opathy. Jugular venous pressure not elevated. Carotids equal. ... r minute and regular and the blood pressure 132/78. The cardiac ... here is a grade 3/6 ejection systolic murmur heard medial to the ... n well heard radiation to the neck vessels.

... ssion and auscultation. Normal respiratory effort. ... . The presence of a large ventral hernia is noted. ... ma. Posterior tibial pulses were felt bilaterally, but I did not feel the ... ions are noted.

LABORATORY AND DIAGNOSTIC DATA: Electrolytes are normal. BUN and creatinine 18/2.2. Blood sugar 150. White count is 7.6, hemoglobin 11.7 with hematocrit 34.9, platelets 187,000.

Cardiology Consultation Transcribed Medical Transcription Sample Reports

DATE OF CONSULTATION: MM/DD/YYYY

REFERRING PHYSICIAN: John Doe, MD

CONSULTING PHYSICIAN: Jane Doe, MD

REASON FOR CONSULTATION: Surgical evaluation for coronary artery disease.

HISTORY OF PRESENT ILLNESS: The patient is a (XX)-year-old female who has a known history of coronary artery disease. She underwent previous PTCA and stenting procedures in December and most recently in August. Since that time, she has been relatively stable with medical management. However, in the past several weeks, she started to notice some exertional dyspnea with chest pain. For the most part, the pain subsides with rest. For this reason, she was re-evaluated with a cardiac catheterization. This demonstrated 3-vessel coronary artery disease with a 70% lesion to the right coronary artery; this was a proximal lesion. The left main had a 70% stenosis. The circumflex also had a 99% stenosis. Overall left ventricular function was mildly reduced with an ejection fraction of about 45%. The left ventriculogram did note some apical hypokinesis. In view of these findings, surgical consultation was requested and the patient was seen and evaluated by Dr. Doe.

PAST MEDICAL HISTORY:

1. Coronary artery disease as described above with previous PTCA and stenting procedures.
2. Dyslipidemia.
3. Hypertension.
4. Status post breast lumpectomy for cancer with followup radiation therapy to the chest.

ALLERGIES: None.

MEDICATIONS: Aspirin 81 mg daily, Plavix 75 mg daily, Altace 2.5 mg daily, metoprolol 50 mg b.i.d. and Lipitor 10 mg q.h.s.

SOCIAL HISTORY: She quit smoking approximately 8 months ago. Prior to that time, she had about a 35- to 40-pack-year history. She does not abuse alcohol.

FAMILY MEDICAL HISTORY: Mother died prematurely of breast cancer. Her father died prematurely of gastric carcinoma.

REVIEW OF SYMPTOMS: There is no history ... [obscured] sures ... e to asthma. There is no hemoptysis or productive cough. No ... on ... or ... palpitations. She notes no nausea, vomiting, constipation, diarrhea, but immediately prior to admission, she did develop some diffuse abdominal discomfort. She says that since then, this has resolved. No diabetes or thyroid problem. There is no depression or psychiatric prob... [obscured] ...al disorder ... asis ... problems or blood dyscrasias. No ... lumpectomy procedures for this with followup radiation therapy. She has been followed in the past 10 years and mammography shows no evidence of any recurrent problems. There is no recent fevers, malaise, changes in appetite or changes in weight.

PHYSICAL EXAMINATION: Her blood pressure is 126/70, pulse is 80. She is in a sinus rhythm on the EKG monitor. Respirations are 18 and unlabored. Temperature is 98.2 degrees Fahrenheit. She weighs 160 pounds, she is 5 feet 4 inches. In general, this was an elderly-appearing, pleasant female who currently is not in acute distress. Skin color and turgor are good. Pupils were equal and reactive to light. Conjunctivae clear. Throat is benign. Mucosa was moist and noncyanotic. Neck veins not distended at 90 degrees. Carotids had 2+ upstrokes bilaterally without bruits. No lymphadenopathy was appreciated. Chest had a normal AP diameter. The lungs were clear in the apices and bases, no wheezing or egophony appreciated. The heart had a normal S1, S2. No murmurs, clicks or gallops. The abdomen was soft, nontender, nondistended. Good bowel sounds present. No hepatosplenomegaly was appreciated. No pulsatile masses were felt. No abdominal bruits were heard. Her pulses are 2+ and equal bilaterally in the upper and lower extremities. No clubbing is appreciated. She is oriented x3. Demonstrated a good amount of strength in the upper and lower extremities. Face was symmetrical. She had a normal gait.

IMPRESSION: This is a (XX)-year-old female with significant multivessel coronary artery disease. The patient also has a left main lesion. She has undergone several PTCA and stenting procedures within the last year to year and a half. At this point, in order to reduce the risk of any possible ischemia in the future, surgical myocardial revascularization is recommended.

PLAN: We will plan to proceed with surgical myocardial revascularization. The risks and benefits of this procedure were explained to the patient. All questions pertaining to this procedure were answered.

Unstructured Hospital Data is Messy but Filled with Key Medical Facts

Medications, Diseases, Symptoms, Non Symptoms, Lab Measurements, Social History, Family History and Much More

But - What Really Causes Readmissions at Seton Hospital

The Data they thought would be useful … wasn't

- 113 candidate predictors from structured and unstructured data sources

- Structured data was less reliable then **unstructured data** – increased the reliance on unstructured data

New Unexpected Indicators Emerged … Highly Predictive Model

- 18 accurate indicators or predictors

97% at 80th percentile

49% at 20th percentile
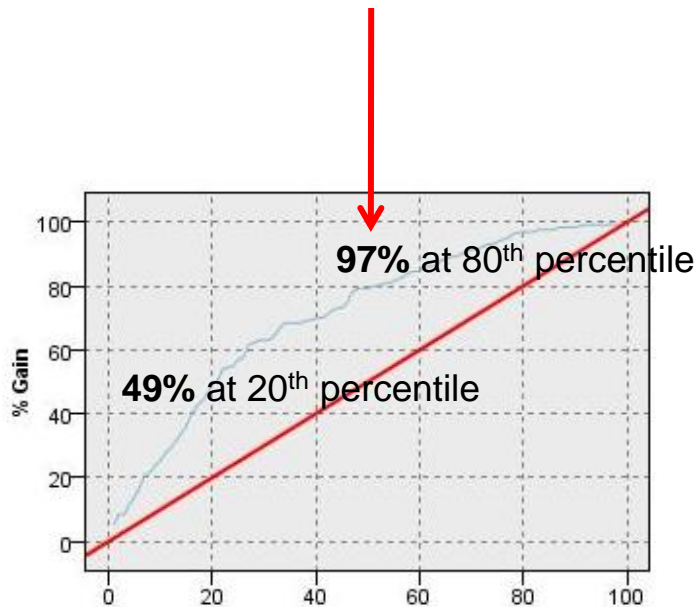
Predictor Analysis	% Encounters Structured Data	% Encounters Unstructured Data
Ejection Fraction (LVEF)	2%	74%
Smoking Indicator	35% (65% Accurate)	81% (95% Accurate)
Living Arrangements	<1%	73% (100% Accurate)
Drug and Alcohol Abuse	16%	81%
Assisted Living	0%	13%

BJC Healthcare and Washington University Partnership
Smart is: unlocking biomedical informatics answers

"We anticipate this solution to be a game changer in biomedical research and patient care. I believe that IBM Content Analytics will ultimately accelerate the pace of clinical and translational research through more rapid and accurate extraction of research relevant information from clinical documents"

Dr. Rakesh Nagarajan, M.D., Ph.D., Associate Professor, Department of Pathology and Immunology, Washington University.

Industry context: Healthcare
Value driver: Access to biomedical trends, insight
Solution : Content Analytics

Business Challenge
Existing Biomedical Informatics (BMI) resources were disjointed and non-interoperable, available only to a small fraction of researchers, and frequently redundant. No capability to tap into the wealth of research information trapped in unstructured clinical notes, diagnostic reports, etc.

What's Smart?
Capitalizing on **the untapped, unstructured information of clinical notes and reports by using IBM Content Analytics with IBM InfoSphere Warehouse.**

Smarter Business Outcomes
Researchers now able to answer key questions previously unavailable. Examples include *Does the patient smoke?, How often and for how long?, If smoke free, how long? What home medications is the patient taking? What is the patient sent home with? What was the diagnosis and what procedures performed on patient?*

An Insurance Company
Smart is: creating rapid insights from content

"The demo impressed the customer so much that the customer was ready to buy ICA in a few days."

— *ECM Sales Rep*

Industry context: insurance claims
Value driver: payment optimization
Solution: Content Analytics

Business Challenge

A Japanese Insurance Agency wanted to automatically assess and verify accurate payment of claims. The industry in general was facing challenges of over or under payments due to inaccurate coding of medical diagnosis. Additional regulation required insurance companies to verify paid payment by implementing a payment assessment and inspection solution.

What's Smart?

Using IBM Content Analytics, the customer is able to normalize medical diagnosis and procedures to a single representation by building a consistent dictionary against which all claims are checked in an automated workflow. Complex claims that required further evaluation can now be identified more quickly and accurately through this automated validation

Client Benefits
• Reduction of human cost and error of manually normalizing and categorize medical certificates and claims.
• More accurately and quickly identify over/under payments and take action.

Sample list of IBM Content Analytics Customers
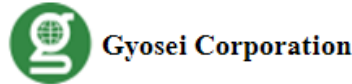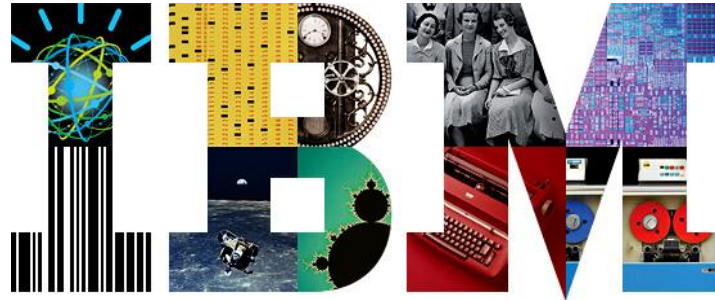
Lloyd Parata
ECM Solution Specialist
Singapore

Question?

What percentage of your organizations information is unstructured?

A) 60%

B) 70%

C) 80%

Backup

Where to find more information about ICA

Start unlocking the insight trapped in your content today

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

IBM Content Analytics

- Find relevant enterprise content quickly and securely
- Assess enterprise content to decommission the unnecessary and govern the content that matters
- Customize rapid insight to industry and customer specific needs
- Enable deeper insights through integration to other systems and solutions
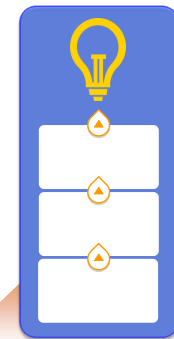
A Car Rental Company and Mindshare Technologies

Smart is: identifying customer satisfaction trends

"We wanted to leverage this insight at both the strategic level and the local level to drive operational improvements"

Industry context: travel services, car rental
Value driver: access to customer survey
Solution onramp: content analytics

Business Challenge
A car rental company needed to better understand customer feedback to adapt its business accordingly. Most of its valuable information was trapped inside free-form customer feedback surveys. This company's location managers read each customer comment submitted via email or phone and then manually categorized it, proving to be very labor-intensive and inconsistent.

What's Smart?
Transforming customer information into actionable intelligence. Using IBM Content Analytics together with Mindshare Technologies sentiment-based tagging solution, the company created a "Voice of the Customer" analytics system to automatically capture customer experiences in real-time.

Smarter Business Outcomes
The company realized improved accuracy and speed of the customer feedback analysis process, almost doubling what had been achieved manually.

Social Content Management… Why should I care?

Content needs proactive management, governance, and discovery.

25%
1 out of every 4 page views are FACEBOOK!

22%
Social networking accounts for 22% of all online time

55%
Make better decisions based on insights from like-minded professionals.[1]

30B
New pieces of content shared EVERY month on Facebook

550M+
Socialize, share, and collaborate on social tools everyday… (Facebook, LinkedIn etc)

800%
Increase in the number of virtual workers over the past 5 years

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

Insight for Multiple Lines of Business

customer service

- Corporate Reputation
- Voice of the Customer
- CSR Training & Monitoring

- Loyalty Alerts
- Monitor FAQs
- Market Sentiment

product management

service management

Rapidly Derived Insight

Search and Explore

Analyze and Visualize

Aggregate and Extract

sales

partners

marketing
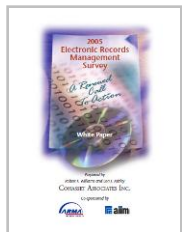
industry reports

market research transcripts
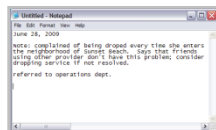
call logs

internal docs and reports

web

blogs

email
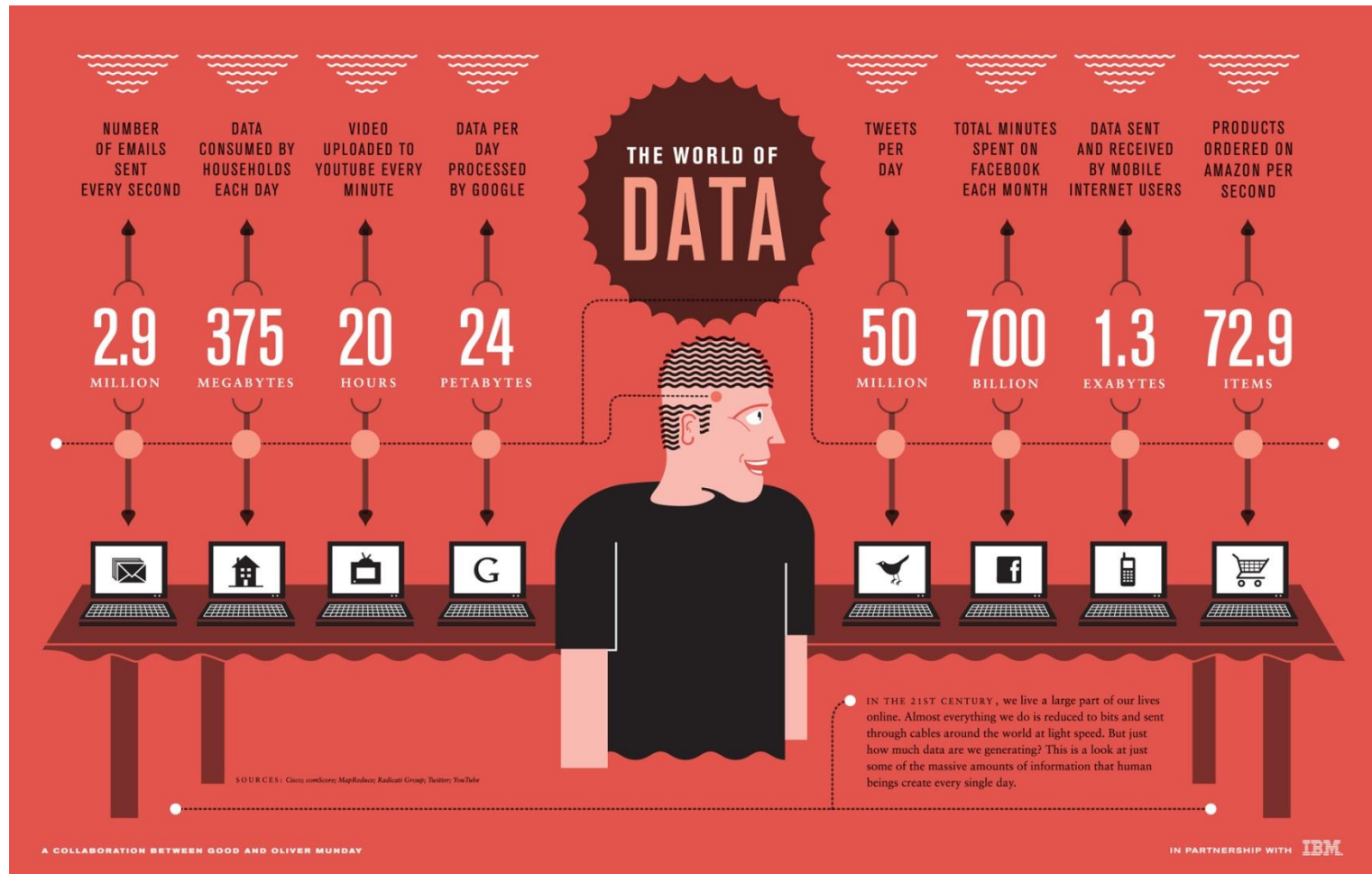
Content Created in 60 Seconds

"The most sought-after CIOs will have a keen understanding of how companies can put to use the oceans of information they now collect…"

HBR, March 2011 – The New Path to the C-suite

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

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.

42

Content Analytics includes Text 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.
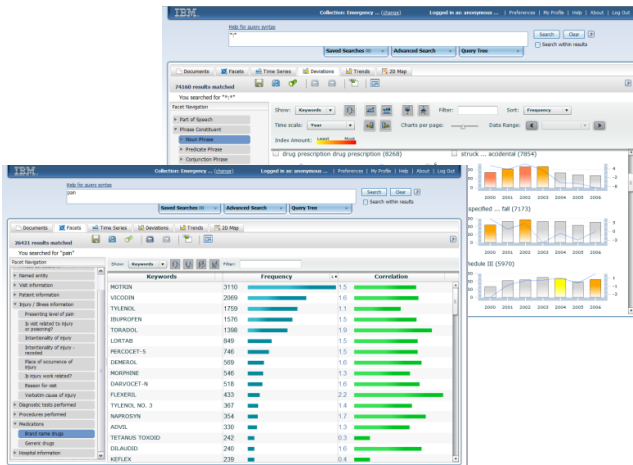
* Natural Language Processing

Unlock valuable insight and make the right decision

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.

Reduce re-admission rates of patients into hospital.

Are you unlocking the value of your unstructured content?