

# Beyond Watson: The Business Uses of Natural Language Processing

Randall Wilcox  
jStart – IBM Emerging Technologies  
[randywil@us.ibm.com](mailto:randywil@us.ibm.com)  
[www.ibm.com/jstart](http://www.ibm.com/jstart)

August 11, 2011 8:00am-9:00pm

# Truly **understanding natural language** is the next great computing challenge



- Over **80%** of information today is unstructured and based on **natural language**
- The impact of **Systems of Engagement** both inside and outside the firewall is dramatic ... such masses of information **not easily understandable** by humans
- Legacy approaches have all failed; “**searching**” **not the right approach**
- A new approach is needed, leveraging **content analysis** and natural language processing



# The Next Grand Challenge



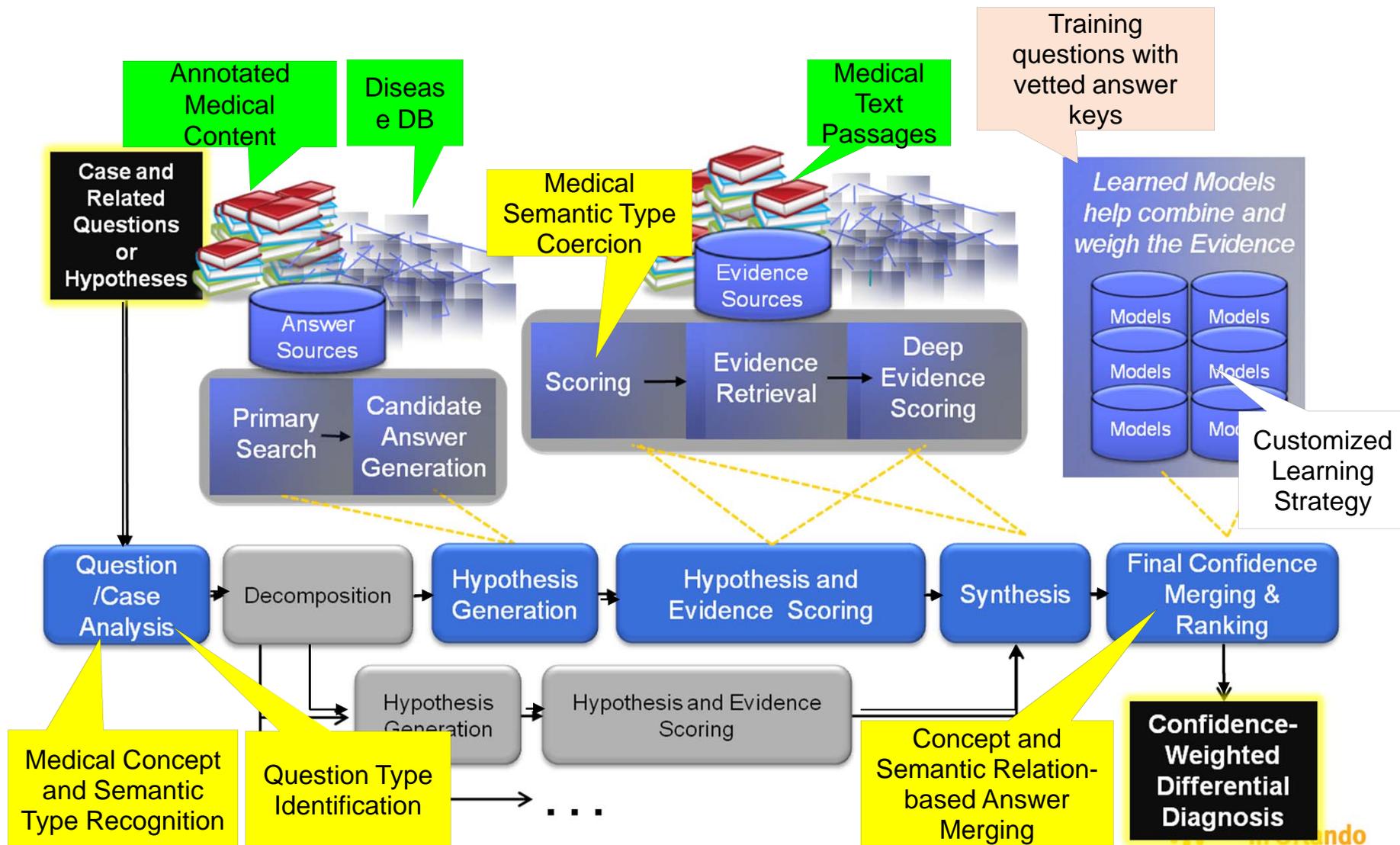
**JEOPARDY!**  **The IBM Challenge**



## Watson in Healthcare



# IBM Watson for Healthcare Pipeline



# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis



| Relations  | Entity Types / Roles     |
|------------|--------------------------|
| causeOf    | FAMILY-DISEASE           |
| modifierOf | FAMILY-SUBSTANCE-ABUSE   |
| negationOf | FINDING-BLOODPRESSURE    |
| partOf     | FINDING-GENERIC          |
| remedyOf   | FINDING-HEARTRATE        |
| resultOf   | FINDING-HEIGHT           |
|            | FINDING-OXYGEN-SATURATIO |
|            | FINDING-RESPIRATORYRATE  |
|            | FINDING-TEMPERATURE      |
|            | FINDING-WEIGHT           |
|            | MODIFIER-ANATOMY         |
|            | MODIFIER-GENERIC         |
|            | MODIFIER-NEGATION        |
|            | MODIFIER-TIME            |
|            | PATIENT-ACTIVITY-EVENT   |
|            | PATIENT-AGE              |
|            | PATIENT-ALLERGY          |
|            | PATIENT-FEMALE           |
|            | PATIENT-HAZARD-EXPOSURE  |
|            | PATIENT-HEALTHSTATE      |
|            | PATIENT-LOCATION         |
|            | PATIENT-MALE             |
|            | PATIENT-NAME             |
|            | PATIENT-OCCUPATION       |



1 Chamarthi, Bindu; Morris, Charles A.; Kaiser, Ursula B.; Katz, Joel T.; Loscalzo, Joseph

2 Stalking the Diagnosis

3 362/9/834

4 [http://content.nejm.org/cgi/content/full/362/9/834</citation\\_fulltext\\_html\\_url>](http://content.nejm.org/cgi/content/full/362/9/834</citation_fulltext_html_url>)

5 A 58-year-old woman presented to her primary care physician after several days of dizziness, anorexia, dry mouth, increased thirst, and frequent urination. She had also had a fever and reported that food would "get stuck" when she was swallowing. She reported no pain in her abdomen, back, or flank and no cough, shortness of breath, diarrhea, or dysuria. Her history was notable for cutaneous lupus, hyperlipidemia, osteoporosis, frequent urinary tract infections, three uncomplicated cesarean sections, a left oophorectomy for a benign cyst, and primary hypothyroidism, which had been diagnosed a year earlier. Her medications were levothyroxine, hydroxychloroquine, pravastatin, and alendronate. She lived with her husband and had three healthy adult children. She had a 20-pack-year history of smoking but had quit 3 weeks before presentation. She reported no alcohol or drug abuse and no exposure to tuberculosis. Her family history included oral and bladder cancer in her mother, Graves' disease in two sisters, hemochromatosis in one sister, and idiopathic thrombocytopenic purpura in one sister.

# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis



### Applying Watson to the Real World

### Continuous Evidence-Based Diagnostic Analysis

A 58-year-old woman presented to her primary care physician after several days of **dizziness**, **anorexia**, **dry mouth**, **increased thirst**, and **frequent urination**. She had also had a **fever** and reported that food would “**get stuck**” when she was swallowing. She reported no pain in her abdomen, back, or flank and no cough, shortness of breath, diarrhea, or dysuria

**difficulty swallowing**

**fever**

**dry mouth**

**thirst**

**anorexia**

**frequent urination**

**dizziness**

Symptoms

#### Pervasive Probabilistic Framework

| Diagnosis Models | Symptoms | Confidence |
|------------------|----------|------------|
| Renal failure    |          |            |
| UTI              |          |            |
| Diabetes         |          |            |
| Influenza        |          |            |
| hypokalemia      |          |            |

Most Confident Diagnosis: **Influenza**

# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis




### Applying Watson to the Real World Continuous Evidence-Based Diagnostic Analysis

A 58-year-old woman presented to her primary care physician after several days of **dizziness, anorexia, dry mouth, increased thirst, and frequent urination**. She had also had a **fever** and reported **that food would “get stuck” when she was swallowing**. She reported no pain in her abdomen, back, or flank and no **cough, shortness of breath, diarrhea, or dysuria**

**Symptoms**

- difficulty swallowing
- fever
- dry mouth
- thirst
- anorexia
- frequent urination
- dizziness
- no abdominal pain
- no back pain
- no cough
- no diarrhea

#### Pervasive Probabilistic Framework

| Diagnosis Models | Symptoms       | Confidence     |
|------------------|----------------|----------------|
| Renal failure    | [Progress bar] | [Progress bar] |
| UTI              | [Progress bar] | [Progress bar] |
| Diabetes         | [Progress bar] | [Progress bar] |
| Influenza        | [Progress bar] | [Progress bar] |
| hypokalemia      | [Progress bar] | [Progress bar] |

Most Confident Diagnosis: **Diabetes**

•Identify negative Symptoms

# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis



### Applying Watson to the Real World

### Continuous Evidence-Based Diagnostic Analysis

A 58-year-old woman presented to her primary care physician after several days of **dizziness**, **anorexia**, **dry mouth**, **increased thirst**, and **frequent urination**. She had also had a **fever** and reported **that food would "get stuck" when she was swallowing**. She reported no pain in her abdomen, back, or flank and no **cough**, **shortness of breath**, **diarrhea**, or **dysuria**

**difficulty swallowing**

**fever**

**dry mouth**

**thirst**

**anorexia**

**frequent urination**

**dizziness**

**no abdominal pain**

**no back pain**

**no cough**

**no diarrhea**

### Pervasive Probabilistic Framework

| Diagnosis Models | Symptoms                               | Confidence                          |
|------------------|--|-------------------------------------|
| Renal failure    | [Bar chart showing low symptom match]  | [Bar chart showing low confidence]  |
| UTI              | [Bar chart showing high symptom match] | [Bar chart showing high confidence] |
| Diabetes         | [Bar chart showing low symptom match]  | [Bar chart showing low confidence]  |
| Influenza        | [Bar chart showing low symptom match]  | [Bar chart showing low confidence]  |
| hypokalemia      | [Bar chart showing low symptom match]  | [Bar chart showing low confidence]  |

Most Confident Diagnosis: **UTI**

- Identify **negative Symptoms**
- Reason with mined relations to explain away symptoms (thirst is consistent w/ UTI)

# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis




### Applying Watson to the Real World Continuous Evidence-Based Diagnostic Analysis

A 58-year-old woman presented to her primary care physician after several days of **dizziness**, **anorexia**, **dry mouth**, **increased thirst**, and **frequent urination**. She had also had a **fever** and reported **that food would "get stuck" when she was swallowing**. She reported no **pain** in her abdomen, back, or flank and no **cough**, **shortness of breath**, **diarrhea**, or **dysuria**. Her family history included **oral** and **bladder cancer** in her mother, **Graves' disease** in two sisters, **hemochromatosis** in one sister, and idiopathic thrombocytopenic **purpura** in one sister

**Symptoms**

- difficulty swallowing
- fever
- dry mouth
- thirst
- anorexia
- frequent urination
- dizziness
- no abdominal pain
- no back pain
- no cough
- no diarrhea

**Family History**

- Oral cancer
- Bladder cancer
- Hemochromatosis
- Purpura
- Graves' Disease (Thyroid Autoimmune)

#### Pervasive Probabilistic Framework

| Diagnosis Models | Stump | Confidence |
|------------------|-------|------------|
| Renal failure    |       |            |
| UTI              |       |            |
| Diabetes         |       |            |
| Influenza        |       |            |
| hypokalemia      |       |            |

Most Confident Diagnosis: Diabetes

- Extract **Family History**
- Use Medical Taxonomies to generalize medical conditions to the granularity used by the models

# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis




### Applying Watson to the Real World Continuous Evidence-Based Diagnostic Analysis

**Symptoms**

difficulty swallowing  
fever  
dry mouth  
thirst  
anorexia  
frequent urination  
dizziness

**Family History**

no abdominal pain  
no back pain  
no cough  
no diarrhea

**Patient History**

cutaneous lupus  
osteoporosis  
hyperlipidemia  
frequent UTI  
hypothyroidism

**Pervasive Probabilistic Framework**

| Diagnosis Models | Symp | Hist | Confidence |
|------------------|------|------|------------|
| Renal failure    | Low  | Low  | Low        |
| UTI              | High | High | High       |
| Diabetes         | Low  | Low  | High       |
| Influenza        | Low  | Low  | Low        |
| hypokalemia      | Low  | Low  | Low        |

Most Confident Diagnosis: **UTI**

•Extract Patient History

# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis




### Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis

A 58-year-old woman presented to her primary care physician after several days of **dizziness, anorexia, dry mouth, increased thirst, and frequent urination**. She had also had a **fever** and reported **that food would "get stuck" when she was swallowing**. She reported no **pain** in her abdomen, back, or flank and no **cough, shortness of breath, diarrhea, or dysuria**. Her family history included **oral and bladder cancer** in her mother, **Graves' disease** in two sisters, **hemochromatosis** in one sister, and idiopathic thrombocytopenic **purpura** in one sister. Her history was notable for **cutaneous lupus, hyperlipidemia, osteoporosis, frequent urinary tract infections, three uncomplicated cesarean sections, a left oophorectomy for a benign cyst, and primary hypothyroidism**, which had been diagnosed a year earlier. Her medications were **levothyroxine, hydroxychloroquine, pravastatin, and alendronate**.

**Symptoms**

- fever
- dry mouth
- thirst
- anorexia
- frequent urination
- dizziness
- difficulty swallowing

**no abdominal pain**  
**no back pain**  
**no cough**  
**no diarrhea**

**Family History**

- Oral cancer
- Bladder cancer
- Hemochromatosis
- Purpura
- Graves' Disease (Thyroid Autoimmune)

**Patient History**

- cutaneous lupus
- osteoporosis
- hyperlipidemia
- frequent UTI
- hypothyroidism

**Medications**

- alendronate
- pravastatin
- levothyroxine
- hydroxychloroquine

### Pervasive Probabilistic Framework

| Diagnosis Models | Symptoms    | Family History | Patient History | Medications | Confidence       |
|------------------|-------------|----------------|-----------------|-------------|------------------|
| Renal failure    | [Bar Chart] | [Bar Chart]    | [Bar Chart]     | [Bar Chart] | [Confidence Bar] |
| UTI              | [Bar Chart] | [Bar Chart]    | [Bar Chart]     | [Bar Chart] | [Confidence Bar] |
| Diabetes         | [Bar Chart] | [Bar Chart]    | [Bar Chart]     | [Bar Chart] | [Confidence Bar] |
| Influenza        | [Bar Chart] | [Bar Chart]    | [Bar Chart]     | [Bar Chart] | [Confidence Bar] |
| hypokalemia      | [Bar Chart] | [Bar Chart]    | [Bar Chart]     | [Bar Chart] | [Confidence Bar] |
| Esophagitis      | [Bar Chart] | [Bar Chart]    | [Bar Chart]     | [Bar Chart] | [Confidence Bar] |

Most Confident Diagnosis: Diabetes and Esophagitis

- Extract Patient History
- Extract Medications
- Use database of drug side-effects
- Together, multiple diagnoses may best explain symptoms

# Applying Watson to the Real World

## Continuous Evidence-Based Diagnostic Analysis



### Applying Watson to the Real World

### Continuous Evidence-Based Diagnostic Analysis

A 58-year-old woman presented to her primary care physician after several days of **dizziness, anorexia, dry mouth, increased thirst, and frequent urination**. She had also had a **fever** and reported that **food would "get stuck" when she was swallowing**. She reported no **pain** in her abdomen, back, or flank and no **cough, shortness of breath, diarrhea, or dysuria**. Her family history included **oral and bladder cancer** in her mother, **Graves' disease** in two sisters, **hemochromatosis** in one sister, and idiopathic thrombocytopenic **purpura** in one sister. Her history was notable for **cutaneous lupus, hyperlipidemia, osteoporosis, frequent urinary tract infections, three uncomplicated cesarean sections, a left oophorectomy for a benign cyst, and primary hypothyroidism, which had been diagnosed a year earlier**. Her medications were **levothyroxine, hydroxychloroquine, pravastatin, and alendronate**. A **urine dipstick was positive for leukocyte esterase and nitrites**. The patient was given a prescription for **ciprofloxacin** for a urinary tract infection and was advised to drink plenty of fluids. On a follow-up visit with her physician 3 days later, her fever had resolved, but she reported continued weakness and dizziness despite drinking a lot of fluids. She felt better when lying down. Her supine blood pressure was 120/80 mm Hg and her pulse was 88 beats per minute; on standing, her systolic blood pressure was 84 mm Hg, and her pulse was 92 beats per minute. A urine specimen obtained at her initial presentation had been cultured and grew more than 100,000 colonies of *Escherichia coli*, which is sensitive to ciprofloxacin.

**Symptoms**  
 difficulty swallowing  
 fever  
 dry mouth  
 thirst  
 anorexia  
 frequent urination  
 dizziness  
 no abdominal pain  
 no back pain  
 no cough  
 no diarrhea

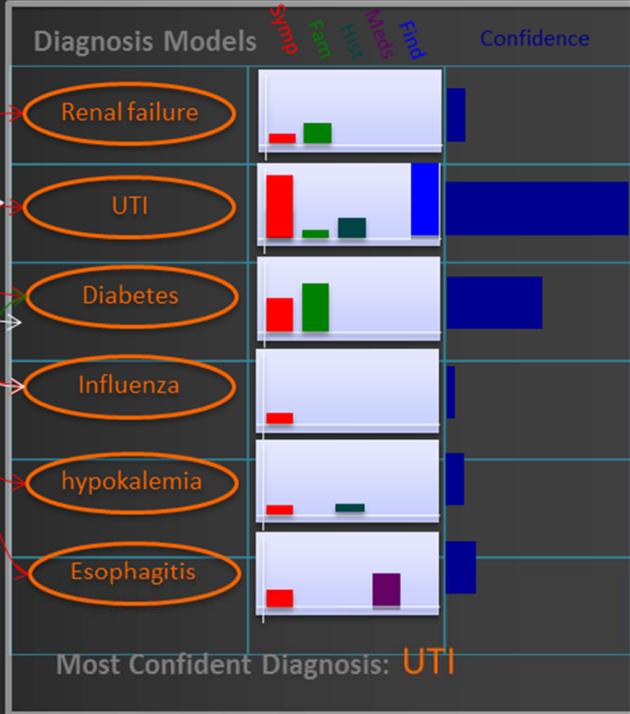
**Family History**  
 Oral cancer  
 Bladder cancer  
 Hemochromatosis  
 Purpura  
 Graves' Disease (Thyroid Autoimmune)

**Patient History**  
 cutaneous lupus  
 osteoporosis  
 hyperlipidemia  
 frequent UTI  
 hypothyroidism

**Medications**  
 alendronate  
 pravastatin  
 levothyroxine  
 hydroxychloroquine

**Findings**  
 urine dipstick:  
 leukocyte esterase  
 supine 120/80 mm HG  
 heart rate: 88 bpm  
 urine culture: E. Coli

### Pervasive Probabilistic Framework



- Extract Patient History
- Extract Medications
- Use database of drug side-effects
- Together, multiple diagnoses may best explain symptoms
- Extract Findings: Confirms that UTI was present

# Watson and IBM Today



- **Natural Language Processing (NLP)** is the cornerstone to translate interactions between computers and human (natural) languages
  - Watson uses **IBM Content Analytics** to perform critical NLP functions
- **Unstructured Information Management Architecture (UIMA)** is an open framework for processing text and building analytic solutions
  - Several IBM ECM products leverage UIMA text analytics processing:
    - **IBM Content Analytics**
    - **OmniFind Enterprise Edition**
    - **IBM Classification Module**
    - **IBM eDiscovery Analyzer**



# IBM at 100: Innovation for Over 50 Years



SHARE  
Technology • Connections • Results

## Beginning in 1957 ...

Searching and Classifying

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

H. P. Luhn

**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 problem 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 artful 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 unreal-

IBM JOURNAL • OCTOBER 1957

\*Presented at American Chemical Society meeting in Miami, April 9, 1957.

appropriately different techniques to their mechanization. The following list of six information systems in order of

# Definitions

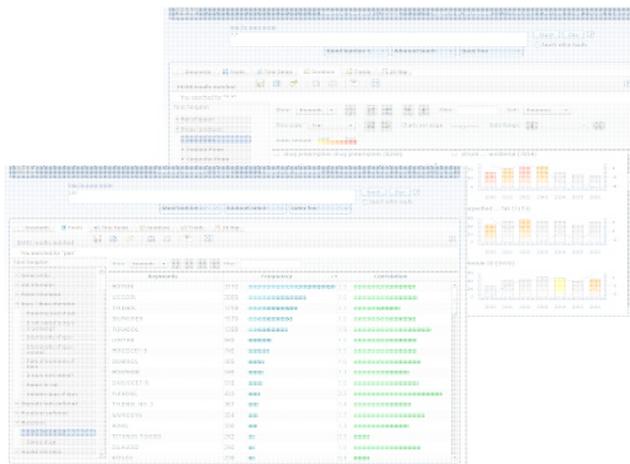
## What is Text Analytics?

*Text Analytics* (NLP\*) describes a set of linguistic, statistical, and machine learning techniques that allow text to be analyzed and key information extracted 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     |

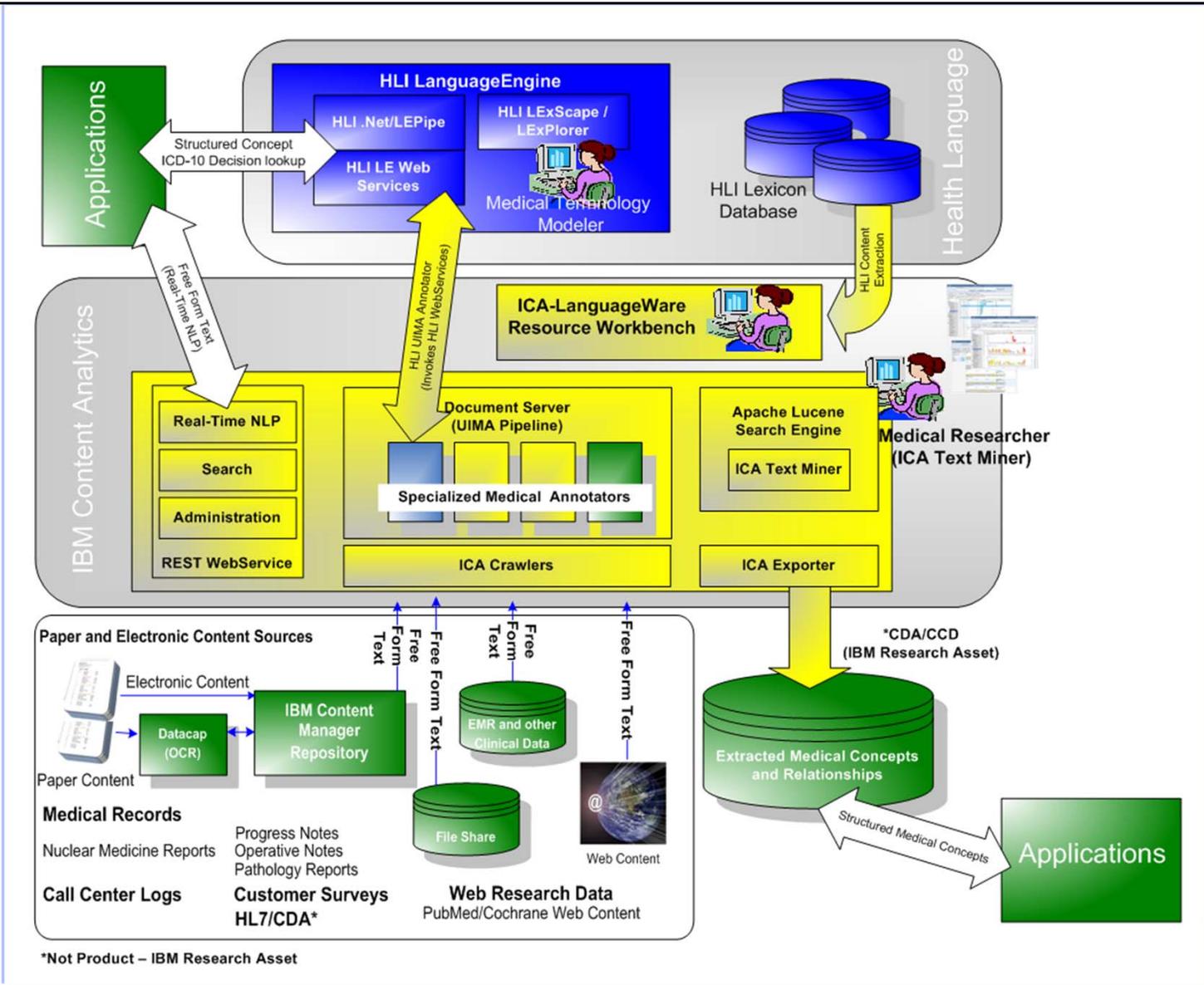


\*NLP = Natural Language Processing

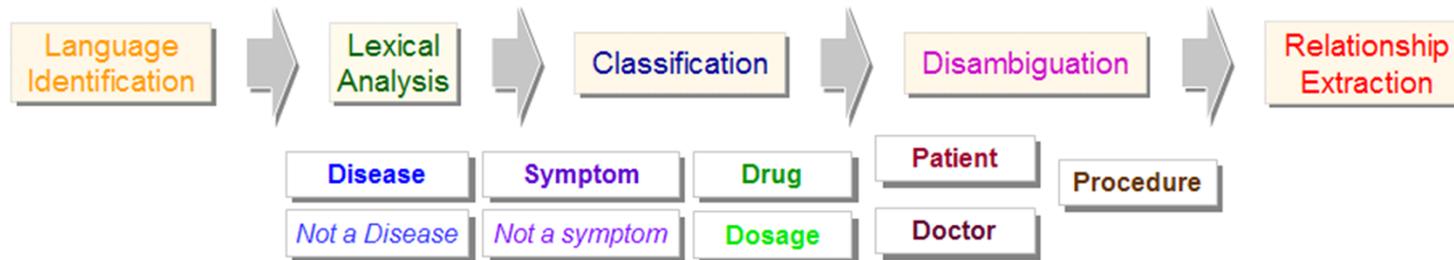
## What is Content Analytics?

*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.

# Medical Text Analytics IBM Content Analytics and Health Language Terminology Server



# Medical Text Analytics IBM Content Analytics LanguageWare Resource Workbench

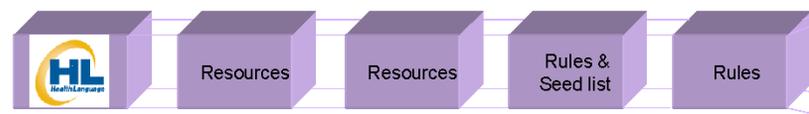


I reviewed **Elizabeth** at our clinic today. **She** was referred to us by **Professor O'Mahony** for further evaluation. **She** complains of **breathlessness with exertion** and **decreased exercise tolerance** with **intermittent palpitations**. **She** did not have any **cough** or **sputum production**, also no **orthopnea** or **paroxysmal nocturnal dyspnea**, or **angina**. **She** had a suspected **myocardial infarction** in 2001, **echocardiography** showed good left ventricular systolic function and normal valves. There is no history of **hypertension**, **diabetes** or **hyperlipidaemia**. Her current medications include **Eltroxin 50 ugs daily**, **Nu-Seals Aspirin 75 mgs daily**, **Lanoxin PG 0.25 mgs daily**, **Salbutamol inhaler 2 puffs prn** and **Beclazone 250 mgs two puffs twice daily**.



English

Customizable Domain Resources



LanguageWare Workbench  
(Medical Records Specialist)



<http://alphaworks.ibm.com/tech/lrw/download>

# Medical Text Analytics IBM Content Analytics LanguageWare Resource Workbench



Annotations Index - Mozilla Firefox: IBM Edition

file:///C:/discovery/healthcare/Output/index.html

Annotations Index

CLIA # 3X13177744  
Patient: Sarah Doe  
PT. # 9833399  
Acc # B11223554  
Physician: Professor John Franks  
Acct # 3332-02

HISTORY AND PHYSICAL

DATE OF ADMISSION: 10/04/2008

This patient is being admitted for observation.

CHIEF COMPLAINT: Vertigo and possible syncope.

HISTORY OF PRESENT ILLNESS: The patient is a 70 year old woman visiting from Quebec who has been here for 3 weeks helping to take care of her son's 18-month-old child. She began having some pain in her right ear and some ringing for the last several days and then today had this episode where she said she kind of collapsed, maybe passed out, and found herself on the floor. She also has had some abdominal cramps for the last several days with some diarrhea earlier this week that has since resolved. No nausea. No vomiting. No chest pain. No palpitations. No swelling in the legs. No fever, but she has had episodes of shaking kind of chills.

PAST MEDICAL HISTORY: 1. Hypertension. 2. Hypothyroidism.

ALLERGIES: NO KNOWN DRUG ALLERGIES.

MEDICATIONS: 1. Synthroid 0.05 mcg. She actually takes 3/4 of a tab of this every day. 2. She takes a French medicine called Coversyl, which is some type of antihypertensive.

SOCIAL HISTORY: Occasional wine. No tobacco.

FAMILY HISTORY: No heart disease. History of cancer.

PHYSICAL EXAMINATION: VITAL SIGNS: Blood pressure 165/71. Temperature 36.8. Respiratory rate 20.

GENERAL: She is an elderly woman in no acute distress.

HEENT: Pupils are equal, round, and reactive to light. Extraocular movements are intact. Cranial nerves II-XII are intact. Tympanic membranes were both clear. The left side showed some more scarring than the right side, but no effusions.

NECK: No jugular venous distention (JVD). No adenopathy or bruits.

LUNGS: Clear to auscultation.

CARDIAC: Regular rate and rhythm. No murmurs, gallops, or rubs.

ABDOMEN: Soft and nontender. No hepatosplenomegaly.

H&P1

- com.HLI.DictFinding (43)
- com.HLI.DictGender (3)
- com.HLI.DictProcedure (9)
- com.IBM.Address (1)
- com.IBM.CessationEducation (0)
- com.IBM.Date (1)
- com.IBM.DictDrug (6)
- com.IBM.DictDrugForm (1)
- com.IBM.DictDrugRoute (0)
- com.IBM.Doctor (1)
- com.IBM.Drug (3)
- com.IBM.Finding (24)
- com.IBM.FindingAbsent (19)
- com.IBM.FootExamFinding (0)
- com.IBM.FootExam\_FilamentTest (0)
- com.IBM.FootExam\_Structural (0)
- com.IBM.FootExam\_Touch (0)
- com.IBM.FootExam\_Vascular (0)
- com.IBM.LOINC\_Date (1)
- com.IBM.LOINC\_Section (9)
- com.IBM.PatientName (1)
- com.IBM.PatientNumber (1)
- com.IBM.Person (0)
- com.IBM.RELname2drug (1)
- com.IBM.RELname2finding (1)
- com.IBM.RELname2findingabsent (0)
- com.IBM.RELname2number (1)
- com.IBM.SmokingNegative (0)
- com.IBM.SmokingPositive (0)
- com.IBM.SmokingQuitIntent (0)
- com.IBM.SmokingQuitIntentNegativ (0)
- com.IBM.TestResult (13)
- com.IBM.VISIT\_DATE (0)

# Medical Records Text Analytics IBM Content Analytics LanguageWare Resource Workbench



Annotations Index - Mozilla Firefox: IBM Edition

file:///C:/discovery/healthcare/Output/index.html

Annotations Index

CHIEF COMPLAINT: Abdominal pain.

HISTORY OF PRESENT ILLNESS: The patient is a 46-year-old woman who noted the onset of abdominal pain this morning of sharp and felt gassy-like. She tried to alleviate the pain with Gas-X and with an enema but these did not alleviate the pain and it became increasingly worse. She came into the emergency room. CT scan was performed and she was diagnosed with a ureteral stone. She has had some nausea and some vomiting but no diarrhea, no constipation. No fever, chills, no dysuria, no chest pain, no shortness of breath, no palpitations.

PAST MEDICAL HISTORY: For gastroesophageal reflux, bleeding ulcer several years ago, jerking legs, migraines.

PAST SURGICAL HISTORY: Hysterectomy, right ovary due to adhesions.

MEDICATIONS: 1. Nexium. 2. Prozac 60 mg daily. 3. Lamictal one oral daily. 4. Multivitamin.

SOCIAL HISTORY: No tobacco, occasional alcohol. She is from Texas.

FAMILY HISTORY: Unremarkable.

PHYSICAL EXAMINATION: VITAL SIGNS: Pulse 88, 96% on 2 liters. She is afebrile. Blood pressure 127/74.

GENERAL: This is a 43-year-old woman appearing in no acute distress. Pupils equal and reactive to light. Extraocular movements intact.

NEUROLOGICAL: Cranial nerves grossly intact.

LUNGS: Clear to auscultation.

CARDIAC: Regular rate and rhythm, no murmurs, gallops or rubs.

ABDOMEN: Soft, nontender, no hepatosplenomegaly, no rebound, guarding, no peritoneal signs. No CVA tenderness. She was examined by me after several doses of pain medications.

EXTREMITIES: No clubbing, cyanosis or edema.

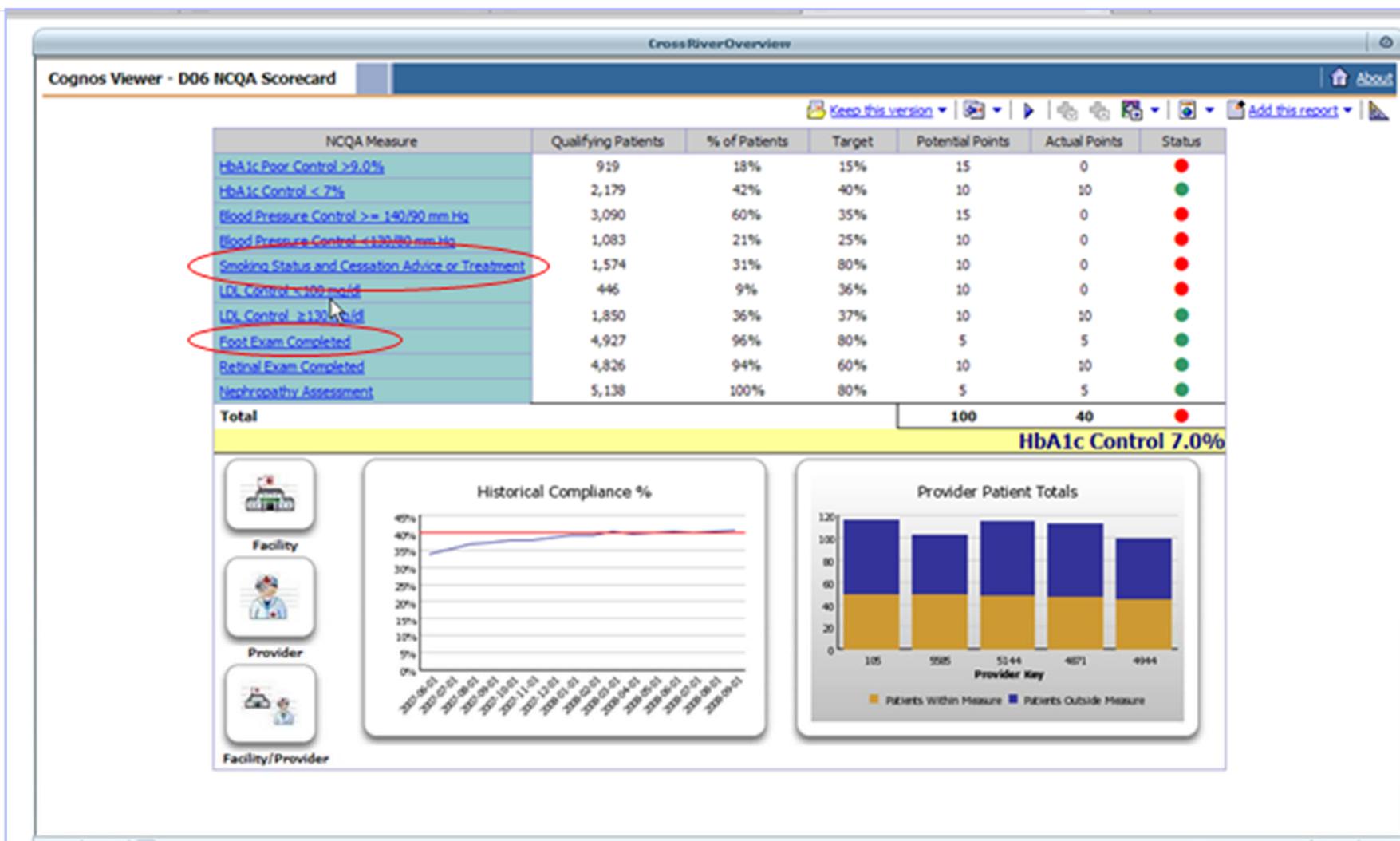
LABORATORY STUDIES: CT scan: A 2 mm nonobstructing stone in the right kidney and 3 mm obstructing stone in the distal left ureter with considerable peripelvic and perinephric edema on the left suggesting some urine extravasation from the obstructing collecting system. White count 10.0, hematocrit 37, note MCV 100, sodium 138, potassium 3.5, chloride 103, bicarb 19, anion gap 20, creatinine 1.1, liver function tests are unremarkable. Calcium 9.9. A urinalysis 15 of ketones but otherwise, unremarkable.

ASSESSMENT AND PLAN: A 46-year-old woman with ureteral stone.

1. Ureteral stone. The emergency room physician spoke with the on call urologist who felt that there is a reasonable chance that the stone could pass on her own and that at this point will treat with IV hydration and pain control. Follow her exam clinically. At this point, there is no sign or symptoms of infection. We will treat her with some intravenous fluid and pain control through the night as an observation status.
2. Increased MCV. She has a family history of hypothyroidism. She has been checked before some time last year. I will recheck the TSH now but I have told her to have B12 levels checked when she returns to Texas. The case was discussed with the patient and the patient's family.
3. History of depression. We will continue her Prozac and her Lamictal.
4. Gastroesophageal reflux disease. Will continue her Nexium.

Done

# Medical Text Analytics IBM Content Analytics LanguageWare Resource Workbench



# Medical Text Analytics IBM Content Analytics LanguageWare Resource Workbench



Annotations Index - Mozilla Firefox: IBM Edition

file:///C:/discovery/healthcare/Output/index.html

Annotations Index

Account#: 7X4553C  
MRN: 179-79-723  
OFFICE VISIT  
Exam Date: 07/20/09

Sex: F DOB: 7/13/61

HX: This 38 year diabetic is here for a follow-up evaluation following adjustment of her hypoglycemic medication. Glyburide was had been increased from 5mg to 10mg daily. She reports that her glucometer readings at home have improved, with first morning glucose around 100. Her hypertension and hyperlipidemia have been under reasonable control with medication. Ms. Doe has no specific complaints.

SH: Married, two children ages 12 and 14, A&W.  
Alcohol: occasional social drinker  
Tobacco: 1 ppd, wishes to stop

PHYS EX:  
HEENT: normocephalic, without masses, non-tender  
CN: II-XII intact and symm  
Fundus: minimal arteriolar nicking, without hemorrhages  
Neck: supple, full ROM, nontender, full ROM  
Chest: clear to P&A, nontender  
Cardiac: S1, S2, gr II/VI systolic ejection murmur PMI L4th ICS MCL  
Abd: soft, nontender, without masses, without organomegaly  
Pelvic: deferred  
Rectal: nl rectal tone, no masses, brown heme neg stool.  
Ext: full ROM, DTRs AJ, KJ, B, T 2+ bilat

Radial pulses 2+ bilat, nl hand neuro exam  
Femoral and popliteal artery pulses 2+ bilat  
DP pulse, 2+ R, 1+ L  
Without ulcers, wounds, erythema  
Nl proprioception both halluces  
Nl light touch on feet  
10g monofilament sensed 3/4 locations left plantar surface, 4/4 locations right plantar surface

Neuro: F to N, RAMs intact  
Gait, tandem gait normal

Rombergs normal  
LABS: reviewed  
IMPRESSION: Improved diabetic management  
No acute abnormalities  
PLAN: Continue current medication regimen, glucometer readings at home four times a day.  
Referred to Smoke-enders  
Follow-up two months

PhysExam

- com.HLI.DictFinding (11)
- com.HLI.DictGender (0)
- com.HLI.DictProcedure (2)
- com.IBM.Address (0)
- com.IBM.CessationEducation (1)
- com.IBM.Date (2)
- com.IBM.DictDrug (4)
- com.IBM.DictDrugForm (0)
- com.IBM.DictDrugRoute (2)
- com.IBM.Doctor (0)
- com.IBM.Drug (3)
  - Glyburide
  - glucose
    - drug
      - com.IBM.DictDrug (1)
      - glucose
- Alcohol
- com.IBM.Finding (9)
- com.IBM.FindingAbsent (1)
- com.IBM.FootExamFinding (1)
- com.IBM.FootExam\_FilamentTest (1)
- com.IBM.FootExam\_Structural (1)
- com.IBM.FootExam\_Touch (1)
- com.IBM.FootExam\_Vascular (1)
- com.IBM.LOINC\_Date (1)
- com.IBM.LOINC\_Section (4)
- com.IBM.PatientName (0)
- com.IBM.PatientNumber (0)
- com.IBM.Person (2)
- com.IBM.RELName2drug (0)
- com.IBM.RELName2finding (0)
- com.IBM.RELName2findingabsent (0)
- com.IBM.RELName2number (0)
- com.IBM.SmokingNegative (0)

## 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 onramp:** 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?*



# Hertz

## Mining customer experiences to improve service

### The need:

Hertz gathers an amazing amount of customer insight daily, including thousands of text comments. However, categorizing comments was laborious and subjective. Each customer comment had to be viewed separately and then categorized for reporting by Hertz location managers. Categorization consistency was only 43 percent and the activity distracted managers from other important duties.

### The solution:

Working with Mindshare Technologies, Hertz implemented enterprise feedback management technology to capture customer experiences in real time and transform them into actionable intelligence. Unstructured feedback is analyzed using linguistic rules that annotate and “tag” comments with descriptive terms (e.g., vehicle cleanliness, staff courtesy, etc.). Managers can filter comments by category and easily identify customers who request a call.

### What makes it smarter:

- Enables managers to spot service trends across customer comments to dramatically improve the effectiveness of Hertz’s Voice-of-the-Customer program
- Provides reliable analysis of customer comments to increase categorization consistency from 43 to 85 percent
- Enables managers to spend less time reading comments and more time making operational improvements based on data previously locked inside surveys

*“We can better focus on improvements that our customers care about, while removing a time-consuming burden from location managers.”*

— Brian Dickerson, Vice President, Customer Care, Hertz

### Solution components:

- IBM® Content Analytics
- MindShare Reveal™



# Mindshare Reveal using IBM ICA



Not only was the **pick-up line** at the **counter** **very long**, but I **waited 30 minutes** just to talk to a **rude representative** who gave me a **car** that **smelled like smoke**, had **stained floor mats**, a **dented fender**, and only **half** a **tank of gas**.

Pickup/Delivery

Attitude/Helpful/Friendly

Body Damage

Speed of Service

Odor

Fuel Level

Counter

Interior Cleanliness

# Mindshare Reveal using IBM ICA

Not only was the **pick-up line** at the **counter** **very long**, but I **waited 30 minutes** just to talk to a **rude representative** who gave me a **car** that **smelled like smoke**, had **stained floor mats**, a **dented fender**, and only **half** a **tank of gas**.

Pickup/Delivery

Attitude/Helpful/Friendly

Body Damage

Speed of Service

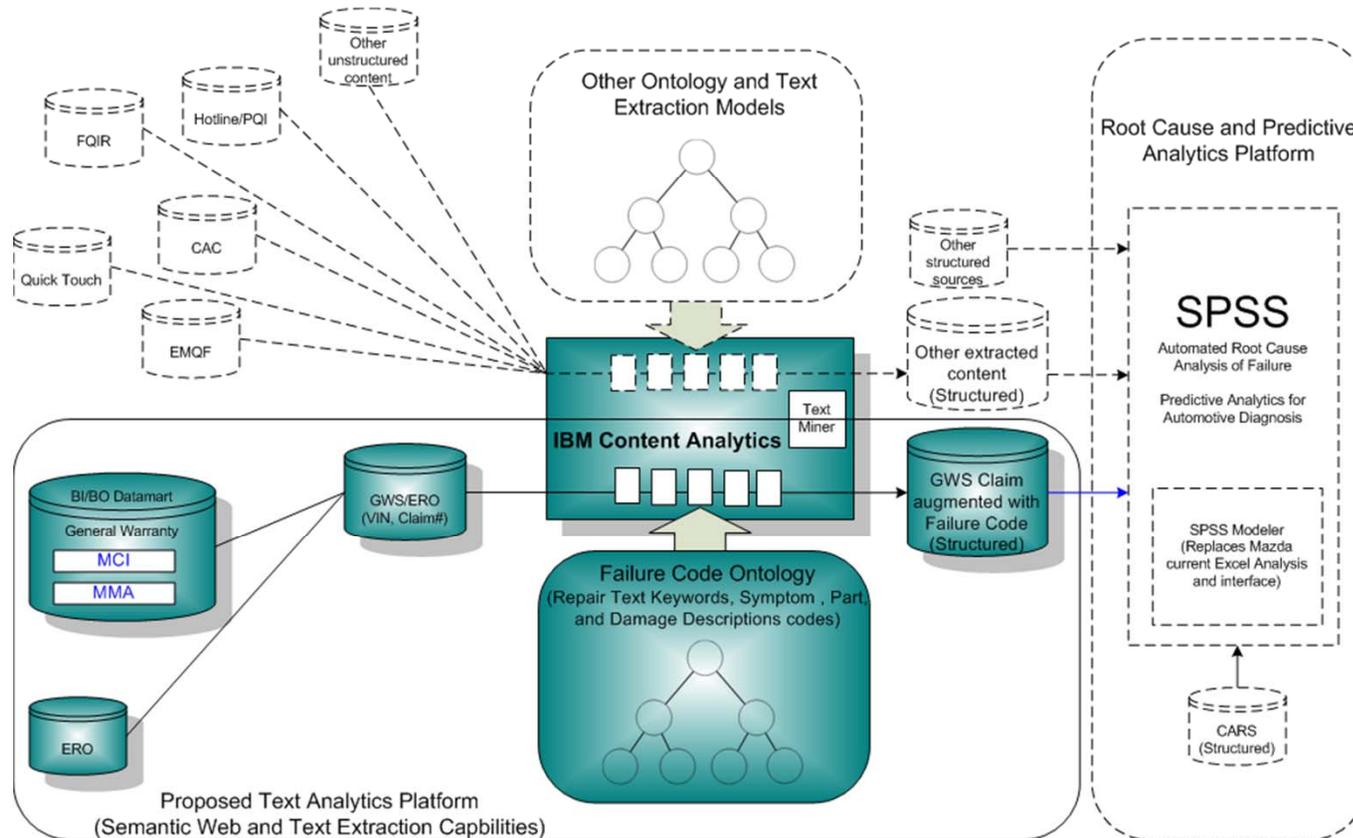
Odor

Fuel Level

Counter

Interior Cleanliness

This automotive manufacturer's Quality Management team is looking to implement IBM Content Analytics and SPSS software in order to analyze unstructured content from a variety of internal and external data sources. The goal is to derive greater insight into issues related to their products and services, improve operational productivity, and provide predictive analytics capabilities.



# Definitions

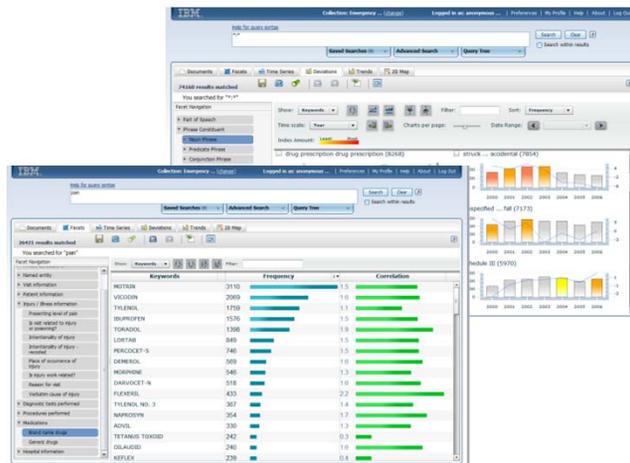
## What is Text Analytics?

*Text Analytics* (NLP\*) describes a set of linguistic, statistical, and machine learning techniques that allow text to be analyzed and key information extracted 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     |



\*NLP = Natural Language Processing

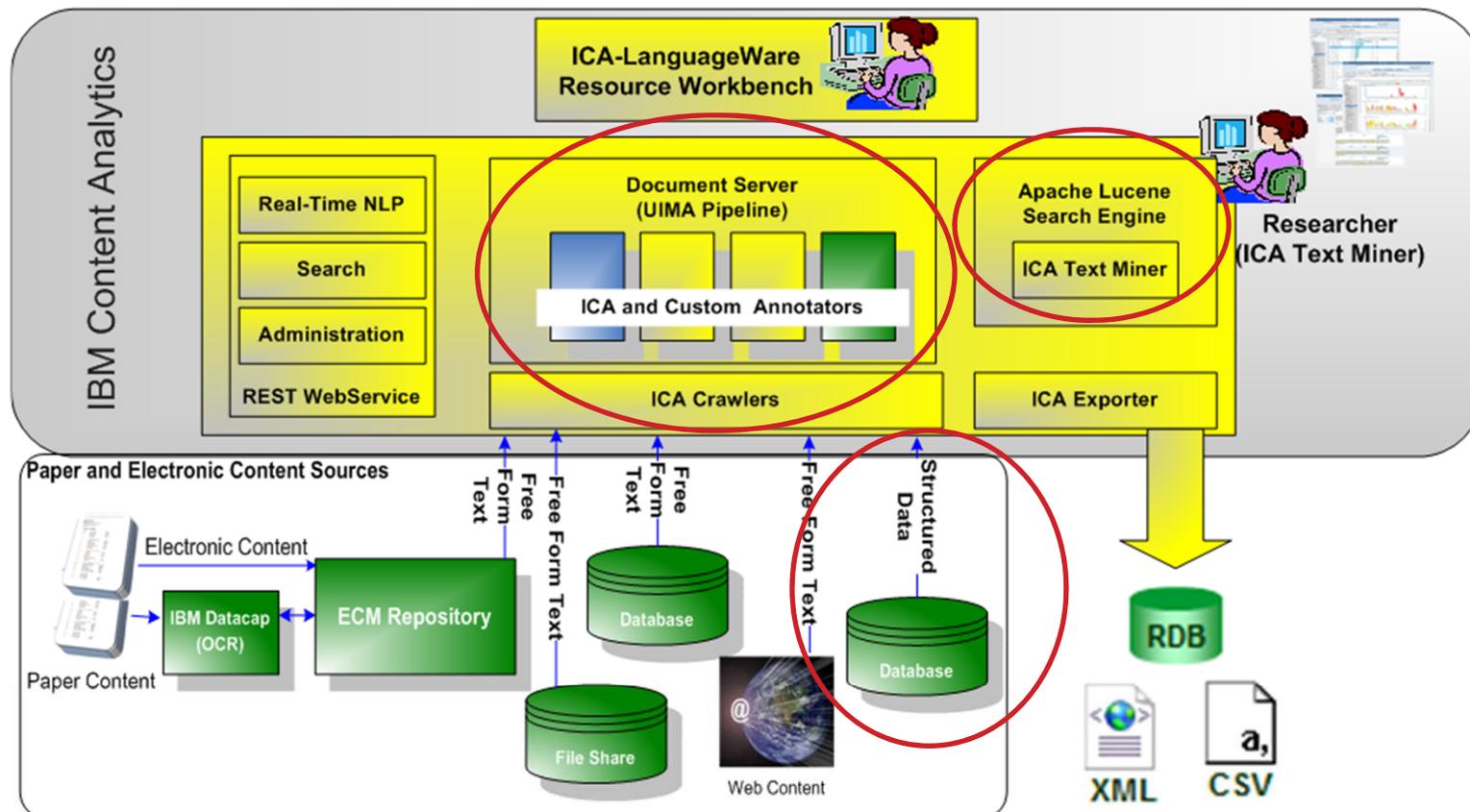
## What is Content Analytics?

*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.

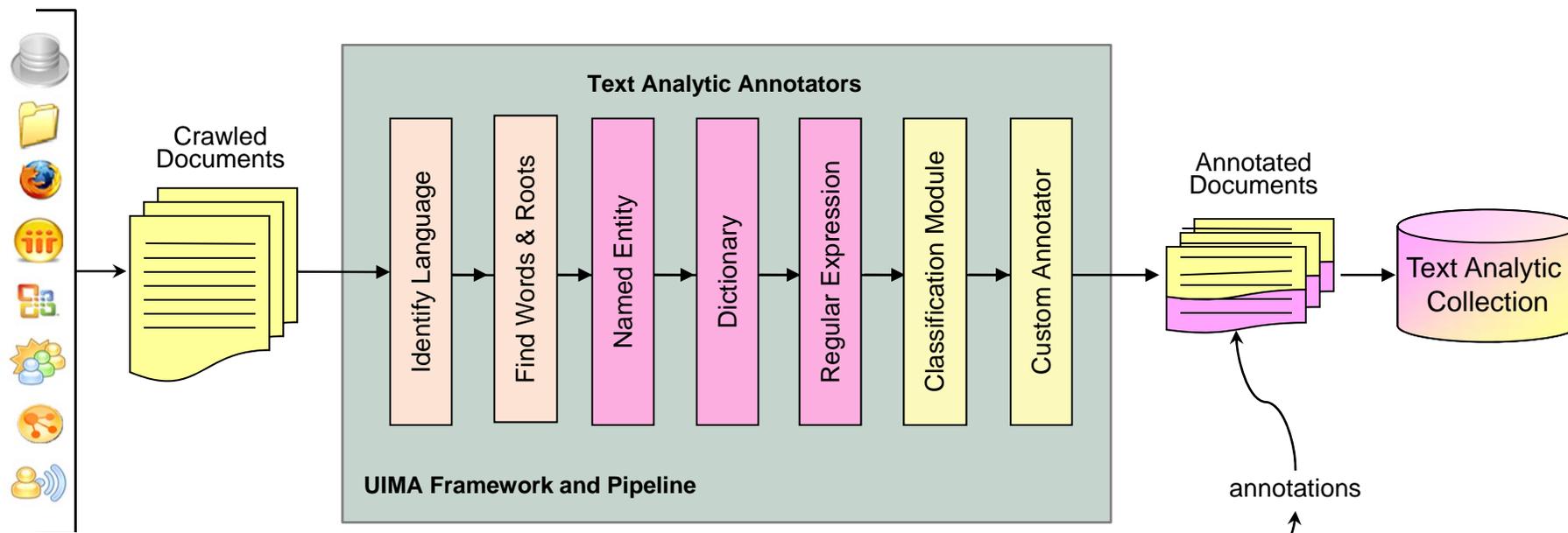
# IBM Content Analytics



**SHARE**  
Technology · Connections · Results

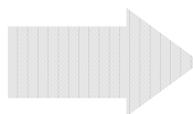


# IBM Content Analytics: How does it work?



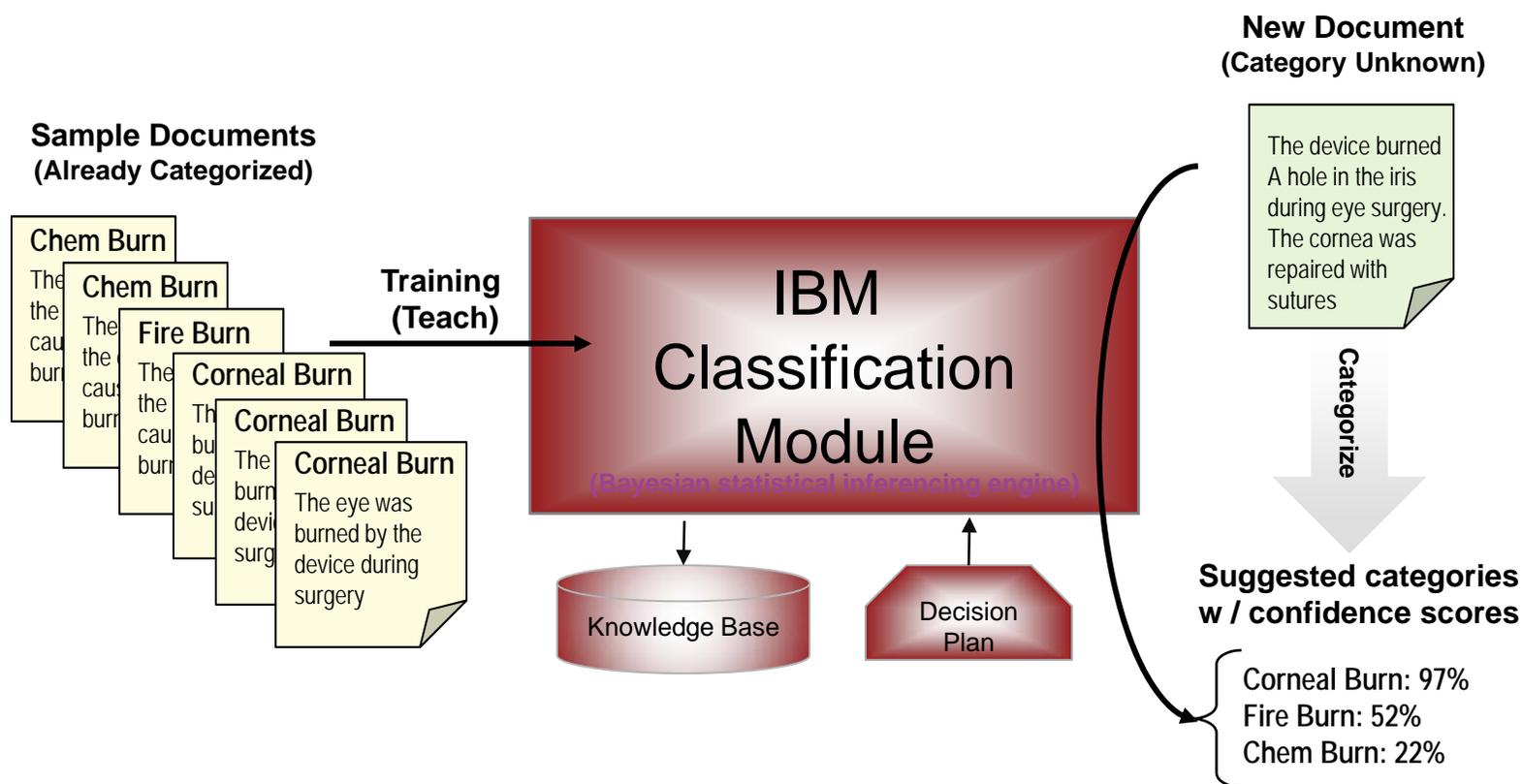
FDA MedWatch Form

**Device Malfunction Description:**  
 It was reported that during a gastric bypass roux-en-y procedure, on the 3<sup>rd</sup> firing with a blue load on the stomach there was bleeding. Not sure if staples were formed properly. They over sewed the staple line. There was no PT consequence.

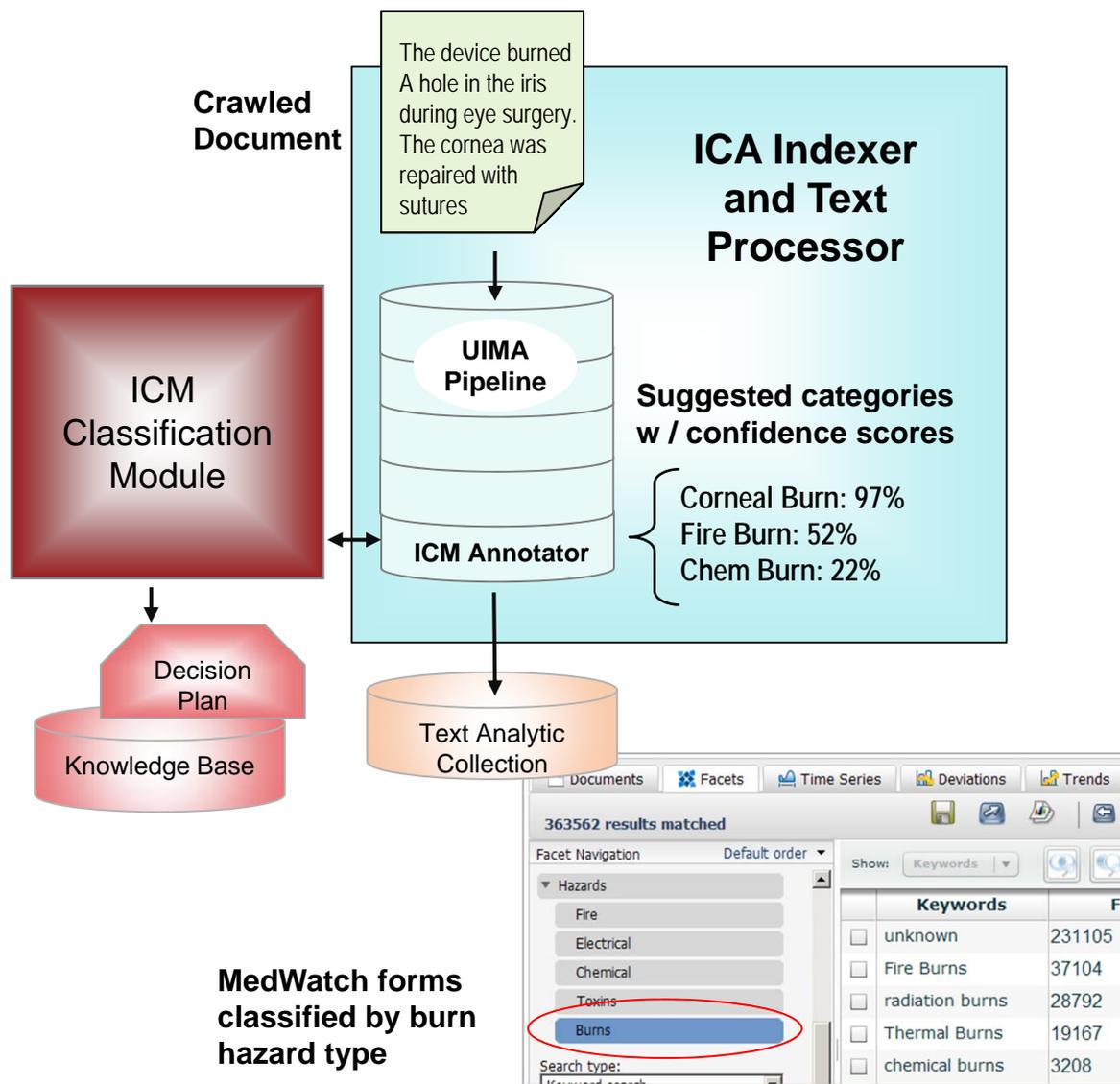


|                     |                        |
|---------------------|------------------------|
| Involved Body Part  | stomach                |
| Type of Injury      | bleeding               |
| Procedure Performed | gastric bypass surgery |

# IBM Content Analytics: How does it work?



# IBM Content Analytics: How does it work?





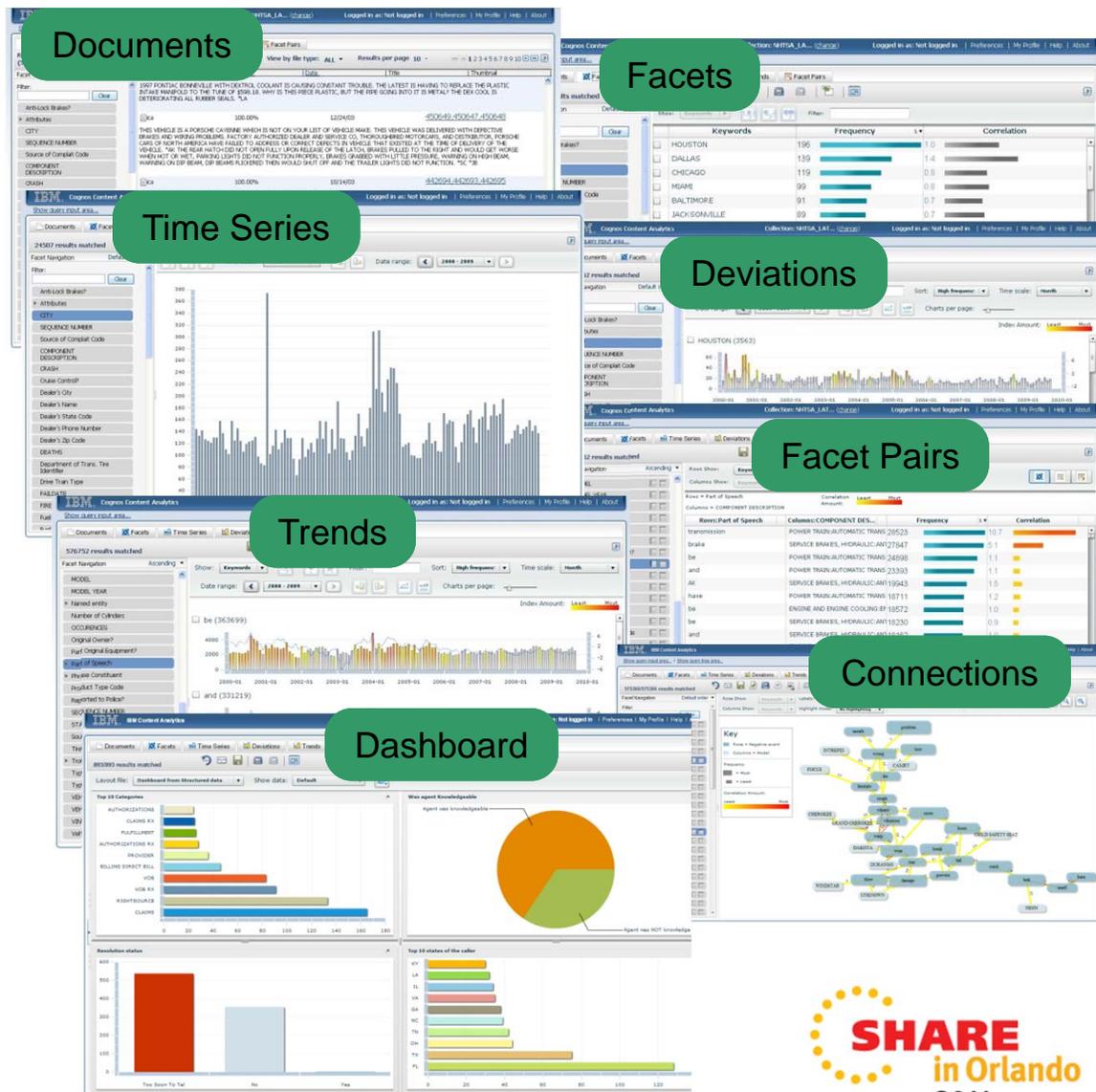
# IBM Content Analytics Text Miner User Interface

## • Interactive text-mining

- Easy to find documents that have distinct index values against certain facets
- Multiple views depending on facets of analytics
- Colorful and interactively rendered charts

## • 8 views

- **Documents** lists documents limited by a query
- **Facets** lists keywords in a facet
- **Time Series** shows frequency changes over time
- **Deviations** shows deviation of keywords on cyclic timeline
- **Trends** detects sharp increase over time
- **Facet Pairs** shows 2 dimensional facet correlation
- **Connections** shows multiple facet values intuitively
- **Dashboard** Show multiple analysis results in various types of visualization at one view



*Transform raw information  
into business insight quickly and easily*



*Industry context: retail customer surveys*

*Value driver: faster, deeper insights*

*Solution onramp: content analytics*

### **Business Challenge**

Provide immediate feedback from surveys, which is then translated into actionable intelligence through powerful and incisive reporting.

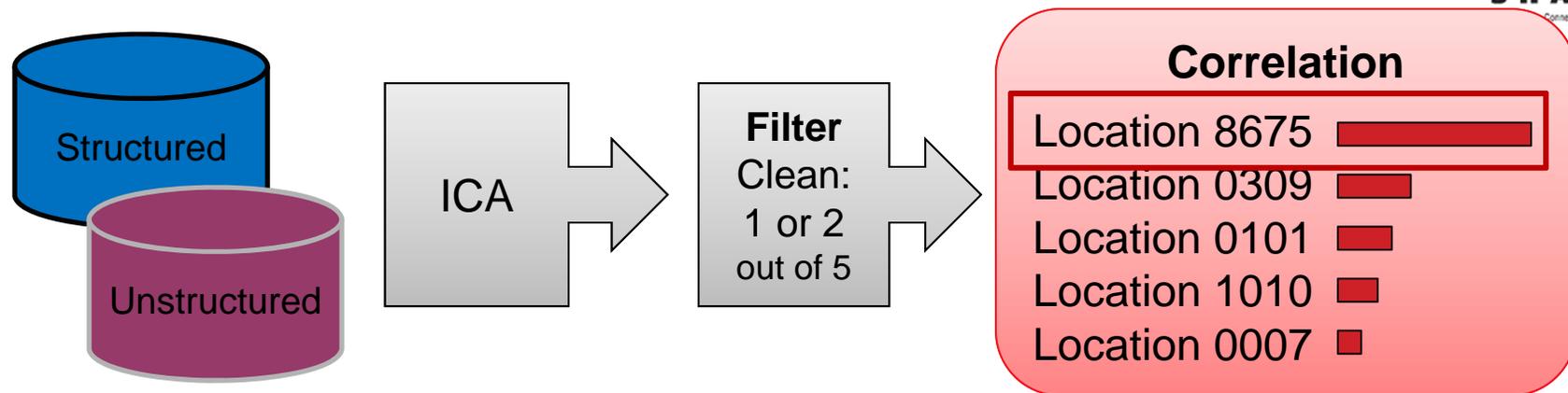
### **What's Smart?**

Exposing additional insight currently not available through their manual survey review process

### **Smarter Business Outcomes**

Mindshare Technologies, using IBM Content Analytics, received a prestigious partnership award from the company for discovery of new emerging trends/patterns and delivering key insights into new products, customer demographics and visit frequency.

# Combining Structured and Unstructured Data



“The restaurant was **appalling**. The entire restaurant has a very **strong odor of urine**. The **bathroom was disgusting** and the **tables and seats were dirty**. I would appreciate you keeping a closer eye on your stores. This restaurant needs a major overhaul for me to even consider eating there again.”

“You have too many **managers** but they are **not focusing on customers**, instead they are **talking** about their personal things. Your **staff** are not ready to work I guess because most of them are **standing** and your **washroom is smelly and dirty**.”

“**Staff are not polite** at all and **too much talking** for nonsense. Place is **very dirty**, especially the **washroom**.”

“Upon entering the restaurant, a **fowl smell of vomit** filled the air. This restaurant is in a central part of downtown and desperately needs to be **cleaned up**. Overall **very unsanitary**.”



Utilize IBM's text analytics technology to help NC State's Office of Technology Transfer identify companies that would be good candidates to license NCSU innovations

Discovery of individuals in those companies  
Discovery of contact information of those individuals

## Use Cases

1. **Smart Inhaler technology**  
Crawl a series of pharma and pharma oriented web sites to find potential licensees of a smart inhaler NCSU has developed where targeted drug deposition is achieved by injecting the drug aerosols from an optimal release position in the mouth inlet cross section by means of a controllable nozzle  
Look for evidence of failed inhaler oriented clinical trials, mine company names and contact info
2. **New Husbandry Vaccine**  
Crawl a series of pharma and pharma oriented web sites to find potential licensees for a new strain of Salmonella enterica serovar Typhimurium NCSU has developed  
Look for evidence of vaccine related R&D, involving animals (not humans), mine company names and contact info

# Beyond Watson: The Business Uses of Natural Language Processing

Randall Wilcox  
jStart – IBM Emerging Technologies  
[randywil@us.ibm.com](mailto:randywil@us.ibm.com)  
[www.ibm.com/jstart](http://www.ibm.com/jstart)



The screenshot shows the BBC website's World Cup 2010 section. It features a navigation bar with 'SPORT' and 'WORLD CUP 2010' prominently displayed. Below this, there are tabs for 'SPORT', 'FOOTBALL', 'WORLD CUP 2010', 'GROUPS & TEAMS', and 'FIXTURES & RESULTS'. The main content area is dedicated to David Villa, a Spanish striker. It includes a 'Latest matches' sidebar with four match reports (URU 2-3 NED, GER 0-1 ESP, URU 2-3 GER, NED 0-1 ESP). The central section for David Villa provides his position (Striker), squad number (7), date of birth (3 December, 1981), and height (5'9"). It also displays 'Tournament totals' with various statistics: Games played (7), Shots on target (19) and off target (10), As substitute (0), Total Spain (7), Goals (5), and Assists (1). A 'Features' section includes an article titled 'Who will win the Golden Boot?'. An advertisement for ANZ Go Transaction Account is also visible. At the bottom, there is a 'Top 5 World Cup stories' section with a list of news items.

| Category         | David Villa | Total Spain | Other Player  |
|------------------|-------------|-------------|---|
| Games played     | 7           |             |   |
| Shots on target  | 19          | 52          | Diego Forlan (20)                                     |
| Shots off target | 10          | 57          | Asamoah Gyan, Cristiano Ronaldo, Wesley Sneijder (12) |
| As substitute    | 0           |             |   |
| Goals            | 5           | 8           | Best Diego Forlan (6)                                 |
| Assists          | 1           | 5           | Best Leo Messi, Bastian Schweinsteiger (4)            |

The BBC World Cup site featured over 700 webpages and was powered by a semantic publishing framework. The site boasted a comprehensive ontology (a map of concepts), that output "automated metadata-driven web pages" created on-the-fly.