Cognitive Analytics: A Step Towards Tacit Knowledge?

Fred A. MAYMIR-DUCHARME, Ph.D.,
Executive Architect, IBM US Federal CTO Office, FredMD@us.ibm.com
Adjunct Professor, University of Maryland University College (UMUC), USA

and

Lee A. ANGELELLI
Analytics Solution Center Architect, IBM US Federal CTO Office, LAngelel@us.ibm.com

ABSTRACT

Tacit Knowledge (TK) generally refers to information that is difficult to convey, store, or transfer explicitly. KT is a key challenge for corporations interested in capturing information in Knowledge Management (KM) systems that is generally lost with attrition or other human factors (e.g., dimensia). In particular, the challenge is in the capture of implicit information (e.g., additional related data, perspectives, and other frames of reference) – in a manner in which it can later be utilized. This paper suggests the use of Cognitive Computing (Analytics) as an advanced approach to capture and extract tacit knowledge. KM involves the process of identifying, capturing, extending, sharing, and ultimately exploiting individual or organizational knowledge. Today's KM requires a multi-disciplinary approach, capable of extending itself to deal with large volumes of disparate data types and emerging technologies that provide a broad set of search and analytics capabilities to meet an organization's need to innovate and thrive. Many organizations have extended their KM to include a variety of unstructured text (e.g., documents and web pages) and multimedia (e.g., pictures, audio and video). The last decade has shown a strong focus on analytics. Analytics provide large organizations the ability to deal with the exponential growth in data volumes and the complexities associated with effectively and efficiently exploiting corporate or organizational data – thus allowing them to dynamically meet internal goals, as well as survive in very competitive environments. This paper provides an overview of various analytic approaches that have been applied to KM over the years, and the state of the art in analytics (Cognitive Computing); and it identifies additional capabilities and technologies in the horizon.

Keywords: Tacit Knowledge, Knowledge Management, Advanced Analytics, Cognitive Computing, Analytics Taxonomy, Watson

1.0 INTRODUCTION

The focus of this paper is on the role of advanced search and analytics, and the application of these technologies to address the Tacit Knowledge (TK) challenge. Polanyi described TK as implicit information difficult to capture linguistically. [1] This challenge is further exacerbated when attempting to automate the capture, process and exploit TK with computers. Nonaka and Takeuchi created a variety of models as a means of capturing and communicating TK. [2]

The size and complexity of a KM System (KMS) continuously grows over time – at an alarming rate. The increasing volumes of data that individuals and organizations store in their KMS tend to grow exponentially, adding complexity size challenges into the mix. Very large volumes of data, and dealing with various types of data (structured, semi-structured, and unstructured text – as well as multimedia) adds complexity to the fundamental need to search and extract information from the KMS. Multimedia has become increasingly important to KMS, recognizing the value of capturing TK in non-linguistic forms. Gourlay's KM framework stresses the value of non-verbal modes of information (e.g., behaviours and processes) to convey a variety of perspectives. [3] Similarly, Gal's model of "Tacit Knowledge and Action" include other sensory modalities used to represent various frames of reference an individual uses to assess information and decide which action is best suited for the situation at hand. [4] Gal's model includes interaction graphs that guide actions based on the tacit knowledge base.

Search is a fundamental element of a KMS. Structured data stored in a database management system (DBMS) on block storage is natively searched and extracted through a DBMS query; whereas unstructured data (e.g., document or a web page) is generally stored on a file system (or made available on the internet through a variety of internet protocols (e.g., TCP/IP, HTTP) and use a search engine to find and ingest/extract the data. And with the recent growth of multimedia (e.g., images, audio and video), new search and analytics systems have evolved over the last decade that add the ability to either search for multimedia through pre-processed metadata (e.g., tagged or represented in the file name), or automatically analyze the file for the search criteria (e.g., image recognition, speech-to-text). Given the rapid pace of technology, there is no single KMS that integrates all of the evolving KM capabilities into a single store, and provides a single search and analytics user interface; hence, the most effective KMS today act as a federation of KMS systems – supporting yesterday's KMS, exploiting today's KM capabilities, and are designed to be flexible in supporting the next generation of search and analytics technology.

Search and analytic capabilities have evolved considerably over the last couple of decades. Search was the initial focus, since one must logically "find" the data before one can analyze the data. Hence, the early work focused on a variety of "relevancy ranking techniques" (e.g., key word proximity, cardinality of key words in document, thesaurus, stemming, etc.) and federated search supporting the ability to search across a variety of DBMS, XML-stores, file stores, and web pages with a single search interface. In parallel, but at a slower pace, the analytics capabilities have followed suit in their evolution – extending traditional search engine reverse-indices to store and exploit search criteria (and other metadata captured during previous searches or data crawls) for analyses.

A variety of analytics have been developed, extending traditional business intelligence and data warehousing techniques to provide predictive analysis, and even stochastic analysis. Predictive models depend on a large corpus of data and can be calibrated by modifying the model attributes and the parameter ranges. Stochastic modeling applies a variety of techniques to discover new aggregations of data (and other data affinities) that can later be used to augment the organization's predictive models. These techniques rely on capabilities such as object-based computing, contextual computing, and cognitive computing.

The remainder of this paper will elaborate on the concepts (e.g., Predictive Analytics, Stochastic Analytics) and technology (e.g., Cognitive Computing), and their use in the capture and exploitation of TK in KMS. The authors recognize that many of the concepts and approaches discussed in this paper also apply to the field of Artificial Intelligence (AI); but adding that topic would broaden the scope of this paper well beyond page limits. The technologies discussed reflect some of the state of the art in cognitive computing capabilities available today, and provide a view into some of the capabilities on the horizon.

2.0 ADVANCED ANALYTICS TAXONOMY

The creation and description of advanced analytics taxonomy is well beyond the scope of this paper – and the subject of numerous doctoral dissertations. The authors are part of a team in the IBM Federal CTO Office that is working on taxonomy of advanced analytics technology. [5] This section briefly describes some of the relevant concepts – and their applicability to TK and KMS.

2.1 ANALYTIC TECHNIQUES

As discussed in the Introduction, there is quite a variety of analytic concepts and technologies one can apply to KMS. Traditional Data Warehousing and Business Intelligence (BI) technologies (e.g., Master Data Management) provide the ability to integrate multiple data sources into a single repository and into a single lexicon (e.g., using traditional Extract, Translate and Load (ETL) or “Data Model to Data Model translations”) in order to better understand the data an organization has in their KM repositories – to help better understand and manage the associated assets, products, or services.

The next level of maturity is embodied by Predictive Analytics, which use models of the data (e.g., entity and relationship maps) to predict current or future events, meeting objectives, or even identifying risks such as fraud, market changes, etc. Predictive analytics models are typically represented as aggregations of data (i.e., models) that generally represent “observable insights.” When sufficient data is available and mapped to a model (reaching a level of assurance) an organization is then able to make predictions. Organizations in mature industries with a large corpus of empirical data can effectively calibrate their predictive models by changing the models’ attributes and ranges – and then assessing the accuracy and fidelity of the model (e.g., “false-positives” & “false negatives”) against the corpus of data.

Stochastic modeling goes beyond the deterministic aspects of predictive modeling, introducing non-determinism in the creation (or rather, discovery) of new models. Stochastic approaches generally include affinity link analysis (e.g., petri-net or semantic modeling) that extends predictive models with new data clusters (e.g., new associations that co-occur, are temporally or geodetically associated, or are statistically related by an additional data element or parametric such as data attributes and ranges).

Section 3 describes Object-based, Contextual and Cognitive computing – all of which support the above analytical concepts and can be applied to capture, manage and exploit TK.

2.2 DATA ATTRIBUTES

There are numerous data attributes that must be considered for analytics. Structured data is typically physically stored on block storage, and preferably Direct Attached Storage (DAS) if latency is an issue; whereas unstructured is physically stored on file storage, quite often on Storage Attached Network (SAN) or Network Attached Storage (NAS) – which have a variety of impacts on the analytics capabilities. For example, the implicit clustering of Hadoop’s Map-Reduce heuristics require a tight coupling of processors and storage (i.e., DAS) – which is limiting if your KM is widely distributed and policy or governance prohibits duplicating the data outside of where it resides (e.g., Personally Identifiable Information (PII)).

Many other data attributes come into play as well. This paper won’t belabor the differences between structured and unstructured text, and the additional complexities of searching and analyzing multimedia. IBM generally refers to the “4 V’s” when referring to other “big data” factors that should be taken into consideration; these include:

Volume: Today’s systems (particularly KM) go well beyond terabytes and petabytes – into exabytes… Twitter estimates that 12 terabytes of tweets are created daily; and the utility industry is struggling with the challenge of converting 350 billion annual meter readings in order to design and assess “smarter utilities” strategies.

Velocity: Data is not only being created at a much faster pace than ever before, it is becoming available at a much faster pace. The US Securities Exchange Committee (SEC) needs to scrutinize more than 5 million trade events created daily, to identify fraud and apply other analytics.

Variety: The variety of data extends beyond the set of “structured & unstructured text, and multimedia” that most organizations are dealing with today. Marine biologists ingest and analyze petabytes of sonar, creating beam forms and other electronic products that can each be in the terabyte range. And referring back to the smarter utilities data volume example, that industry also has to deal with a variety of different data types as utility metering is transformed from analog data to digital data for transmission, measurement, analysis, and ultimately – optimized management.

Veracity: The accuracy, as well as the genealogy of data also plays a major role in analytics. The more important the decision (e.g., mission critical systems or those with lives at stake) require much higher fidelity and accuracy, as well as assurance (or prioritization) based on the pedigree of the data (e.g., where it came from, who created it, and who has modified it).

There are other data attributes to consider, and one more in particular is worth noting. A new paradigm has evolved over the last decade. Stream computing differs from traditional systems in that it provides the ability to process (analyze) **data in motion (on the network)**, rather than **data at rest (in storage)**. Figure 1 below illustrates the parallel nature of stream processing, as well as the “pipe and filter” architecture that allows a variety of analytics to be applied to data in motion. [6]

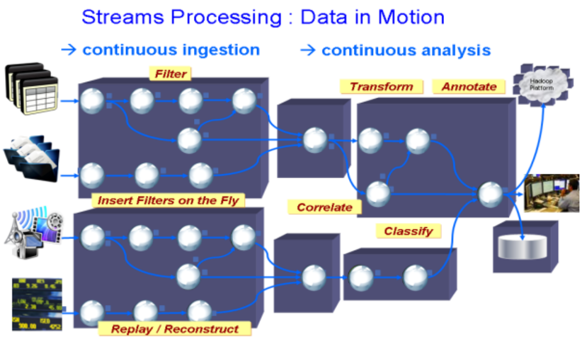


Figure 1: Streams Processing

Stream processing dynamically supports analytics on data in motion. In traditional computing, you access relatively static information to answer evolving and dynamic analytic questions. With stream processing, you can deploy an application that continuously applies analysis to an ever-changing stream of data before it ever lands on disk – providing real-time analytics capabilities not possible before.

Stream computing is meant to augment, not replace current data at rest analytic systems. The best stream processing systems have been built with a data centric model that works with traditional structured data as well as unstructured data - including video, image, and digital signal processing. Stream processing is especially suitable for applications that exhibit three application characteristics: compute intensity (high ratio of operations to I/O), data parallelism allowing for parallel processing, and ability to apply data pipelining where data is continuously fed from producers to downstream consumers. As the number of intelligent devices gathering and generating data has grown rapidly, alongside numerous social platforms in the last decade, the volume and velocity of data organizations can exploit have mushroomed (e.g., consumer insights & trends) . Organizations need to make more timely decisions faster than ever before. Organizations that want to analyze data as it arrives from monitors and equipment (measurements and events) as well as text, voice transmissions, and video feeds – are good candidates for stream processing.

2.3 OTHER ANALYTIC CONCEPTS

As previously mentioned, this Advanced Analytics Taxonomy is currently being developed. In the interest of time (and space), this paper briefly describes two additional areas for consideration when assessing analytics for a KMS.

System Element Topologies: The distributed or centralized nature of both, the data and the processing can play a key role in the analytic capabilities and algorithms. Traditional MDM and Data Warehousing are built on the premise that data must be centralized in order to effectively manage and exploit it. But there are many cases in which some of the data cannot be ingested into the central repository (e.g., due to policies, or technical limitations such as proprietary legacy systems requiring that require the data be provided by the native system – instead of directly from the data store). One of the challenges of not being able to centralize and control (e.g., ETL) the data, is in anomalies introduced by different data owners with different goals, objectives, and governance. An approach that's been successfully applied and fielded is to introduce (leverage) the use of standards based data modeling (e.g., in XML). If each of the peripheral data owners / providers agrees to maintain and publish their data model, the centralized system (e.g., KMS) can then create and maintain data model translations to the KM's data model. This allows for dynamic updates to both the centralized data, as well as the decentralized data – as long as the data model updates are shared / published, and the translations are kept up to date.

Similarly, processing can be done centrally or peripherally. The above MDM example requires coordination across multiple organizations, as well as an implicit structuring of the data (e.g., DBMS or XML-labeled data stores). There are many cases in which this is not possible, nor feasible. In these cases, the challenge is to apply analytics peripherally, as well as internally. Section 3 will get into much more details regarding the processing of data centrally and peripherally.

Visualization & Navigation: There are two key concepts in this area. The first is "visualization." As the amount and complexity of the data increases (as well as the richness of the associated analytics), it becomes more and more difficult to fit, organize, and illustrate the results of a large search, or analytics on hundreds or thousands of elements onto a screen. Likewise, recognizing one can represent search results and analytics at various levels of abstractions, or from multiple perspectives – the ability to effectively and intuitively navigate multi-level results (e.g., visualizations) becomes a critical capability. Visualization and navigation are particularly important to multimedia analytics. Consider the many Geospatial Information System (GIS) visualizations that are combined on a single screen at times (e.g., mapping, charting, and geodesy). These two concepts are key areas IBM Research is focused on.

> *Multimedia analytics and visual analytics address two emerging needs in analyzing data. Multimedia analytics is about computers making sense of images and videos, and being able to extract information and insights from those sources, whereas visual analytics is about humans using visual interfaces to consume and make sense of complex data and analytics.*
>
> *Multimedia analytics will require systems to learn which image features are important in these different settings and industries, and recognize variations of those features so they can be properly labeled. Visual analytics will require systems to automatically determine what to visualize, pick the right visual metaphor based on user context and show changes over time and uncertainty.*
>
> *Innovation in four key areas is needed to address visual analytics requirements: visual comprehension, visualizing aspects of time, visual analytics at scale, and visualizing uncertainty and predictions. Industries should explore different applications of visual analytics to their data and use cases, with a view to transform their decision-making and analytics. [7]*

3.0 COGNITIVE ANALYTICS

Advanced analytics has been the primary focus of the authors for the last decade. Having previously worked on traditional DBMS On-Line Analytical Processing (OLAP) and a variety of Data Warehousing projects prior to 2000, the new millennium challenge was "unstructured data" and multimedia.

This section describes cognitive analytics, which the authors define to include contextual analytics. Contextual analytics is used to assess information within a confined set of data sources (e.g., a set of 200 documents resulting from a federated search query and ingest). Contextual analytics applies techniques such as relevancy ranking, entity extraction & entity-relationship modeling, parts of speech tagging, etc. to analyze the data within the context of those 200 documents. Cognitive analytics extends the scope of the analyses to include "implicit knowledge" and perspectives that may be represented by lexicons, taxonomies, models, or rules-based computations tailored to the areas of interest represented by the data in the KMS. These frames of reference are then used to build understanding and insights about the explicit knowledge in the KMS. Cognitive Analytics requires four major capabilities:

Collection: the ability to identify and ingest relevant information from a variety of sources, and a variety of forms

Context: the ability to implicitly or explicitly extract features and relationships from diverse data and data sources, and creating metadata to continuously build and update context around the data elements

Cognition: the analysis of data within a variety of contextual perspectives, delivering understanding and insights about the information and metadata (e.g., relationships.)

Exploitation: the application of Collection, Context and Cognition to guide actions – e.g., decision support or automation.

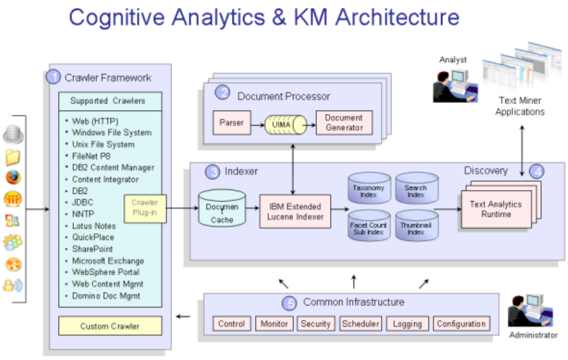


Figure 2: Cognitive Analytics Architecture

Figure 2 illustrates the functional (software) architecture of IBM's Content Analytics (ICA), a Cognitive Analytics solution. The architecture above does not reflect networks, physical or logical views, or the storage (traditionally a combination of DBMS for the structured data, and a content management system for the unstructured text and multimedia.) The Crawler Framework provides the ability to acquire (identify and ingest) data from a variety of sources, and in various formats. ICA uses an IBM search engine that extends the open source Lucene indexer for efficient storage of the information content, metadata, and other analytical results.

The first phase of cognitive computing involves ingest and indexing of the KM's data set. General indexing may include metadata such as: URL, URI, document name, type of document, date stamp or date indexed, etc. The original data documents may or may not become persistent in the KMS – depending on organizational (e.g., policy), economic (e.g., storage costs), and legal (e.g., data rights) constraints.

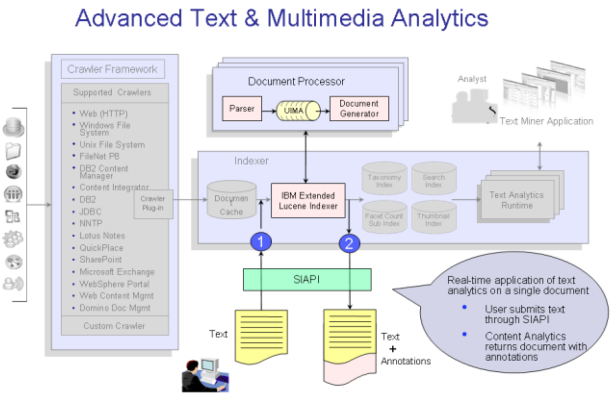


Figure 3: Contextual Analytics Components

Figure 3 highlights the contextual analytics components. The second phase of cognitive computing involves applying various analytics to individual documents, and ultimately across all documents (KM data). This includes the ability to implicitly or explicitly extract features and relationships from diverse data and data sources, creating metadata to continuously build and update context around the data elements. The figure above illustrates the "stand-alone" ability to apply a unique set of analytics (annotators are described below) on a select document (or group of documents) – which is at times is a valuable capability outside of the broader analytical framework. The Search and Index API (SIAPI) is used to analyze one or more documents, resulting in the "annotations" illustrated as output from the SIAPI.

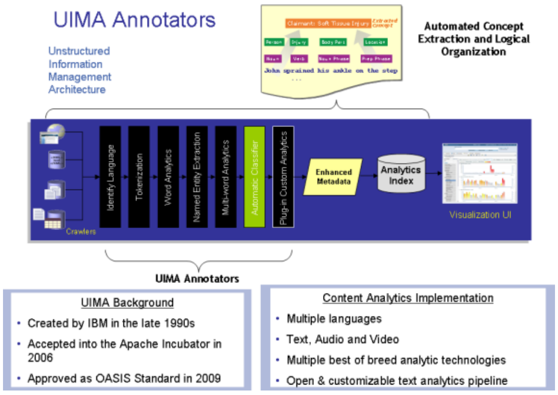

UIMA Annotators
Unstructured Information Management Architecture
Claimant: Soft Tissue Injury
Extracted Concept
Person
Injury
Body Part
Location
Noun
Verb
Noun Phrase
Prep Phrase
John sprained his ankle on the step
Automated Concept Extraction and Logical Organization
Crawlers
Identify Language
Tokenization
Word Analytics
Named Entity Extraction
Multi-word Analytics
Automatic Classifier
Plug-in Custom Analytics
Enhanced Metadata
Analytics Index
Visualization UI
UIMA Annotators
UIMA Background
Created by IBM in the late 1990s
Accepted into the Apache Incubator in 2006
Approved as OASIS Standard in 2009
Content Analytics Implementation
Multiple languages
Text, Audio and Video
Multiple best of breed analytic technologies
Open & customizable text analytics pipeline

Figure 4: UIMA Framework

The UIMA illustrated in figure 4 provides the core contextual analytics through a variety of annotators. Annotators support an array of analytical processing capabilities, such as: language identification, entity extraction, entity type extraction, parts of speech tagging, tokenization, machine translation, speaker identification and tagging, etc. The annotations (e.g., metadata) are used for both, contextual analytics, as well as cognitive analytics.

As the figures below reflect, the UIMA framework can be used for structured and unstructured data – as well as for multimedia (e.g., images, audio and video). And in addition, IBM has created numerous translingual annotators that can be used to enhance search (e.g., supporting transliterations, foreign character sets in UTF-16, multi-lingual search, and relevancy ranking algorithms such as stemming and polymorphic analysis).

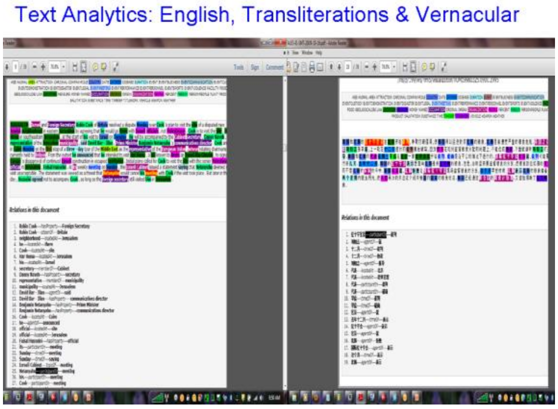
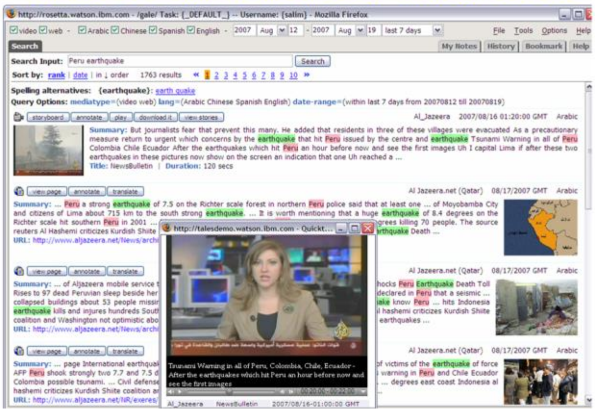


Figure 5: Text Analytics

Figure 5 illustrates some of the results of text analytics annotators. Note that the keywords have been labeled (e.g., entity type labels below the text) and color coded to facilitate user's finding the terms of interest within a document. The "Relations in this text" also identified relationships between the entities extracted. Behind the scenes are tokenizers that recognize multiple spellings of the same name (refining associative analytics), as well as parts of speech tagging that recognizes pronouns and is therefore able to include name-relationships associated with an individual that would not have been possible without the analytics needed to map "she" and "her" to Robin Cook.

Note IBM has extended these text analytics to support multiple foreign languages – e.g., English, Russian, Chinese, Arabic. Our translingual technology includes two types of machine translation (MT): Rules Based MT, and Statistical MT. And the search engine supports searching in English, in a foreign language, or in multiple languages – and it can search foreign language transliterations, as well as in the native language (vernacular) and foreign character sets.

Multimedia Analytics: Audio, Video, Imagery, Sensors



http://rosetta.watson.ibm.com - /gale/ Task: {_DEFAULT_} -- Username: {salim} - Mozilla Firefox

☑ video ☑ web - ☑ Arabic ☑ Chinese ☑ Spanish ☑ English - 2007 Aug 12 - 2007 Aug 19 last 7 days

Search Input: Peru earthquake

Sort by: rank | date | in ↓ order 1763 results « 1 2 3 4 5 6 7 8 9 10 »

Spelling alternatives: {earthquake}: earth quake

Query Options: mediatype=(video web) lang=(Arabic Chinese Spanish English) date-range=(within last 7 days from 20070812 till 20070819)

Al_Jazeera 2007/08/16 01:20:00 GMT Arabic

Summary: But journalists fear that prevent this many. He added that residents in three of these villages were evacuated As a precautionary measure return to urgent which concerns by the earthquake that hit Peru issued by the centre and earthquake Tsunami Warning in all of Peru Colombia Chile Ecuador After the earthquakes which hit Peru an hour before now and see the first images Uh I capital Lima if after these two earthquakes in these pictures now show on the screen an indication that one Uh reached a ...

Title: NewsBulletin | Duration: 120 secs

Al Jazeera.net (Qatar) 08/17/2007 GMT Arabic

Summary: ... Peru a strong earthquake of 7.5 on the Richter scale forest in northern Peru police said that at least one ... of Moyobamba City and citizens of Lima about 715 km to the south strong earthquake. ... It is worth mentioning that a huge earthquake of 8.4 degrees on the Richter scale hit southern Peru in 2001 ... grees killing 70 people. The source reuters Al Hashemi criticizes Kurdish Shiite ... rthquake Death ...

URL: http://www.aljazeera.net/News/archi

Al Jazeera.net (Qatar) 08/17/2007 GMT Arabic

Summary: ... of Aljazeera mobile service t ... hocks Peru Earthquake Death Toll Rises to 97 dead Peruvian sleep beside her ... declared in Peru that a seismic ... collapsed buildings about 53 people missin ... ake know Peru ... hits Indonesia earthquake kills and injures hundreds Sout ... l hashemi criticizes Kurdish Shiite coalition and Washington not optimistic abo ... earthquakes ...

URL: http://www.aljazeera.net/News/archi

Al Jazeera.net (Qatar) 08/17/2007 GMT Arabic

Summary: ... page International earthqua ... of victims of the earthquake of force AFP Peru shook strongly two 7.7 and 7.5 d ... warning in Peru and Chile Ecuador Colombia possible tsunami. ... Civil defens ... degrees east coast Indonesia al hashemi criticizes Kurdish Shiite coalition a ...

URL: http://www.aljazeera.net/NR/exeres/

http://talesdemo.watson.ibm.com - Quickt...

Tsunami Warning in all of Peru, Colombia, Chile, Ecuador - After the earthquakes which hit Peru an hour before now and see the first images

Al_Jazeera NewsBulletin 2007/08/16-01:00:00 GMT

Figure 6: Multimedia Analytics

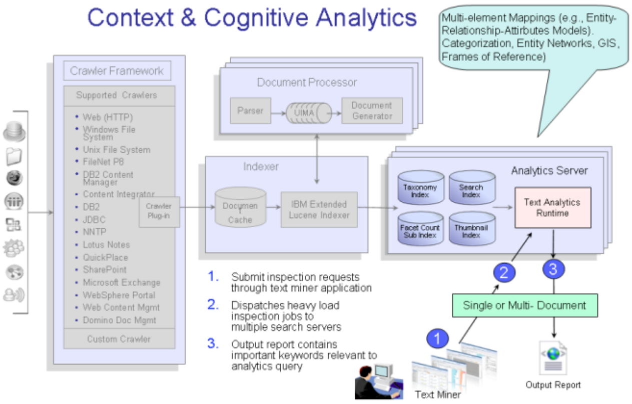
Figure 6 above illustrates the ability to apply the text analytics previously described to multimedia. One can then apply the same set of analytics on video, audio and imagery as one is able to do with text. To accomplish this, one uses multimedia annotators such as:

- Language Identifier: able to identify one of 128 languages within the first three phonemes
- Speech to text translations : Converting speech into text – and then applying the appropriate text analytics
- Speaker Identifier: able to identify and distinguish multiple speakers in an audio or video clip, and tag conversations appropriately
- Speaker Authentication: able to authenticate a speaker if their voice has been enrolled into the system as an identity's voice

The figure shows a newscaster speaking Arabic, with a dynamic speech to text conversion below her video, and then dynamic translation from Arabic to English in the window below the Arabic text. The screen shot above also shows annotated text from a resulting search across data sources including both text and multimedia.

An important observation is that this system now enables users to search multimedia natively (as opposed to the traditional limitations of static tagging techniques) and apply the same analytics (e.g., implicit knowledge and discovery) to any combination of structured text, unstructured text, images, audio, and video – concomitantly.

Context & Cognitive Analytics



Multi-element Mappings (e.g., Entity-Relationship-Attirbutes Models). Categorization, Entity Networks, GIS, Frames of Reference)

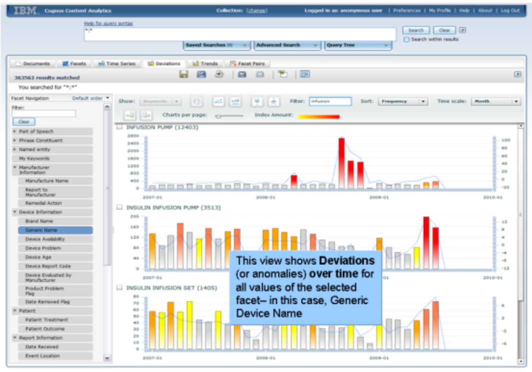
1. Submit inspection requests through text miner application
2. Dispatches heavy load inspection jobs to multiple search servers
3. Output report contains important keywords relevant to analytics query

Figure 7: Cognitive Analytics

The Analytics Server in Figure 7 above illustrates the ability to extend the text and multimedia analytics previously discussed, to support lexicons, taxonomies, and other analytical models. The "text miner" in the analytics server uses a multi-element mapping structure to create analytic models. An element can be an entity, an entity type, a relationship, or an attribute (e.g., tagged or derived metadata). For example, various analytic tools provide an entity-relationship (E-R) model that one can then visualize a number of individuals and their direct and transitive relationships graphically. One can then create a variety of models (e.g., entity social network models, hierarchical, organizational, risk/threat, etc.) representing the organizations' predictive models or categorization schemes. Valuable attributes such as time, geo-location, demographics, etc. can be used to provide unique analytic models and visualizations.

Experience shows that the quality of the lexicon and taxonomy created for the KM directly impacts the soundness of the resulting models and analyses. These multi-element mappings are processed by the text miner and result in an XML'ed / tabularized list which can itself be used for analysis, or as input to a variety of visualizations.

Trends & Anomalies (Predictive Analytics)



This view shows **Deviations** (or anomalies) **over time** for all values of the selected facet– in this case, Generic Device Name

363562 results matched

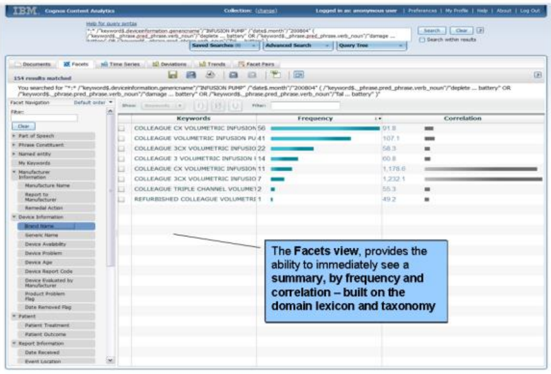
You searched for "*:*"

Filter: infusion Sort: Frequency Time scale: Month

Figure 8: Trending and Anomaly Visualizations

There are numerous types of visualizations one can then derive from the multi-element mapping results (models). ICA includes Entity - Relationship Model Views, Entity-type – Relationship Model Views, Hybrid Entity-type – Relationship Model Views, Categorization Model Views, and traditional Measurement & Metrics Views (e.g., Pie charts). The figure above illustrates a Trending and Anomaly view, which extends the aggregation of entities and relationships with attributes such as time, to visually analyze trends or anomalies over time. The height of the bars in the figure above represent the cardinality of indices for individual entities, whereas colors are used to easily differentiate entities or entity types.

Deterministic Discovery (Predictive Analytics)



The **Facets view**, provides the ability to immediately see a **summary, by frequency and correlation – built on the domain lexicon and taxonomy**

154 results matched

You searched for "*:* /"keyword$.deviceinformation.genericname"/"INFUSION PUMP" /"date$.month"/"200804" (/"keyword$._phrase.pred_phrase.verb_noun"/"deplete ... battery" OR /"keyword$._phrase.pred_phrase.verb_noun"/"damage ... battery" OR /"keyword$._phrase.pred_phrase.verb_noun"/"fail ... battery")"

Keywords	Frequency	Correlation
COLLEAGUE CX VOLUMETRIC INFUSION	56	91.8
COLLEAGUE VOLUMETRIC INFUSION PU	41	107.1
COLLEAGUE 3CX VOLUMETRIC INFUSIO	22	58.3
COLLEAGUE 3 VOLUMETRIC INFUSION I	14	60.8
COLLEAGUE CX VOLUMETRIC INFUSION	11	1,178.6
COLLEAGUE 3CX VOLUMETRIC INFUSIO	7	1,232.1
COLLEAGUE TRIPLE CHANNEL VOLUME	2	55.3
REFURBISHED COLLEAGUE VOLUMETRI	1	49.2

Figure 9: Facets Visualization

Facets represent different views a user can define to visualize either search results (e.g., category views based on a taxonomy), or analytics results (e.g., multi-element mapping results mapped to a taxonomy categorization scheme.) The example in Figure 9 above reflects the results of a search query, mapped to a device's categorization scheme. It provides a quick glance at where there is a majority of device-problem issues (reflecting trouble areas), as well as different clustering that may reflect the specific device-problem issue an analyst is interested in pursuing. Facets provide the ability to model various clusters of data elements, to help intuitively guide the prioritization, evaluation, and focus areas of search results or analytics results.

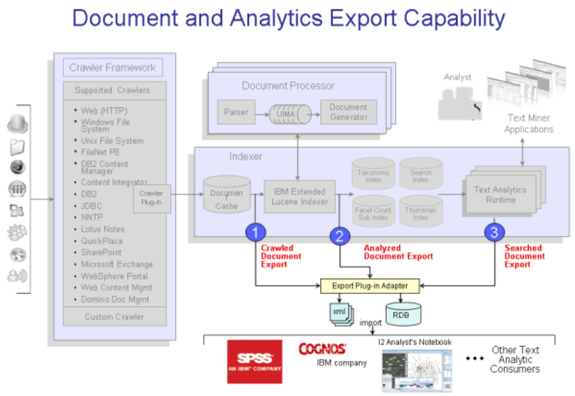


Figure 10: Cognitive Analytics Export

Recognizing the speed with which analytics technology is emerging and maturing, a KMS Analytics solution needs to be extensible and provide interfaces for new inputs – as well as output to other systems. The figure above reflects the ability to input and crawl a variety of data sources (the list is not an exhaustive list of ICA connectors and crawlers.) The system also provides the ability to export data in either an XML document, or directly into a relational DBMS. There are three different export opportunities in the architecture above:

1. Crawled data (e.g., URI, document name, keywords matched) can be exported before indexing or performing any other text analytics
2. Indexed data (e.g., reverse search index, and metadata/annotations resulting from UIMA processing) can be exported before applying multi-element mappings and other analytics
3. Analyzed data (e.g., results including documents, metadata and annotations, as well as multi-element mappings) can be exported to be used by different visualization technologies, or additional analytic engines.

Section 1 discusses stochastic analysis, which is not supported by this Cognitive Analytics solution. One approach to stochastic analytics could be to export the predictive models from ICA into another modeling engine such as SPSS Modeler, which includes petri-net and semantic models that could be used to discover new aggregations – extending the exported multi-element mappings and relevant metadata.

4.0 BEYOND COGNITIVE ANALYTICS

Some organizations are using KMS based on Cognitive Analytics to capture, structure, manage, and disseminate knowledge throughout an organization enabling employees to work faster, reuse best practices, and reduce costly rework from project to project. Human society has now entered an era where the complexity of our world and the risks thereof demand a capacity for reasoning and learning far beyond individual human capability. Today's world is creating an explosion of Big Data – structured data and new unstructured data e.g. social, email, multimedia, sensor, etc. that organizations struggle with using traditional technology to capture and exploit this knowledge in an easily accessible natural language processing (NLP) form to help employees overcome human limitations to make successful decisions in today's complex world.

To help organizations apply KM to solve complex problems in today's world - the IBM Watson team took the Cognitive Analytics technology described in Section 3 to a new level, adding advanced natural language processing, automated reasoning, and machine learning to the Cognitive Analytics components (e.g., information retrieval, knowledge representation, and analytics). Watson used databases, taxonomies, and ontologies to structure its knowledge, enabling the processing of 200 million pages of unstructured and structured text (stored on four terabytes of disk storage.) Watson used the UIMA framework as well as the Apache Hadoop framework to support the required parallelism of the distributed system – which consisted of ninety IBM Power 750 processors (each an eight core 3.5GHz processor), and sixteen terabytes of RAM. More than 100 different techniques and technology were used to provide natural language analytics, source identification, hypothesis generation and discovery, evidence discovery and scoring, and hypotheses merging and ranking in Watson.

> *In 2007, IBM Research took on the grand challenge of building a computer system that could compete with champions at the game of Jeopardy!. In 2011, the open-domain question-answering (QA) system, dubbed Watson, beat the two highest ranked players in a nationally televised two-game Jeopardy! Match.. This paper provides a brief history of the events and ideas that positioned our team to take on the Jeopardy! challenge, build Watson, IBM Watson, and ultimately triumph. It describes both the nature of the QA challenge represented by Jeopardy! and our overarching technical approach.. The main body of this paper provides a narrative of the DeepQA processing pipeline to introduce the articles in this special issue and put them in context of the overall system. Finally, this paper summarizes our main results, describing how the system, as a holistic combination of many diverse algorithmic techniques, performed at champion levels, and it briefly discusses the team's future research plans.*[8]

IBM is leveraging its Watson technology to create the concept of Cogs computing - designed to follow and interact with people (and other cogs & services) inside and across cognitive environments. A "cog" represents a specific frame of reference and the associated data. IBM is using Cogs computing to create **"Industry of Knowledge"** expert advisors available to every worker – dedicated to their success of the job.

For example, Memorial Sloan-Kettering Cancer Center (MSKCC) is using Watson to help oncology physicians battle cancer. Traditionally, oncology physicians diagnose cancer using a patient's chart, x-rays, laboratory data, a few medical books; and they might then recommend either the general radiation therapy or three types of chemotherapy. Today, oncology physicians face a perpetually growing sea of data in their efforts to effectively deal with in every aspect of their patients' care. The associated medical information doubles every five years -- e.g., thousands of books and articles, electronic patient and family medical records, over 800 different cancer therapies ,sequencing 340 cancer tumors (each with multiple mutations), analyzing 20,000 genes, correspondence with over 1,000 physicians, and the exponential rise in medical publications. Traditional processes for cancer prognosis and the recommendation of therapies are no longer able to effectively harness all of the available data. Keeping up with medical literature could take up to 160 hours per week – an unrealistic option. Hence physicians are turning to Watson to develop precision based medicine in cancer.

MSKCC and IBM are training Watson to compare a patient's medical information against a vast array of treatment guidelines, published research and other insights to provide individualized, condensed, scored recommendations to physicians. Watson's NLP capabilities enable the Watson system to leverage this sea of unstructured data, including journal articles, multiple physicians' notes, as well as the guidelines and best practices from the National Comprehensive Cancer Network (NCCN). The evolving IBM Watson Oncology Diagnosis and Treatment Advisor includes supporting evidence with every suggestion, in order to provide transparency and to assist in the doctor's decision-making process and patient discussions. Watson will interactively point out areas in which more information is needed and update its suggestions as new data is added. [11]

For example, MSKCC had a cancer case involving a 37 year-old Japanese non-smoking patient, diagnosed with lung Adenocarcinoma cancer. The physician asked Watson for a recommend therapy. Watson's initial case analysis recommended Chemo-Erlotinib treatment at a 28% confidence interval. Watson needed more information and recommended the physician perform a molecular pathology test to detect if there are any EGFR mutations (57% of all EGFR mutation in women with Adenocarcinoma cancer would be missed). Lab results came back, identifying the presence of an EGFR exon 20 mutation. Watson referenced a medical paper citing an exception, where the EGFR exon 20 mutation doesn't respond to Erlotinib treatment. Analyzing the new lab information along with the medical article, Watson then recommended Cisplatin / Pemetrexed treatment with 90% confidence. "There are only about 2 or 3 physicians in the world who would know this information" said Dr. Jose Baselga (Physician-in-Chief at MSKCC) at the IBM Watson Group Launch in New York Event (9 January, 2014).

The DARPA SyNAPSE Project further extends Watson capabilities in other ways. The project is called "The Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE);" and it represents another innovative application of Cognitive Analytics. [9] As the name infers, this project's goal is to simulate the brain, automating the neural network of the brain. One can envision automating human senses to augment "frames of reference," with additional data and TK with audio analytics for hearing, video analytics for seeing, and a variety of other sensor technologies. This will further extend the ability to capture and exploit the TK humans are able to identify and process naturally. Through these "senses," SyNAPSE can apply additional contextual analytics, augmented by cognitive analytics to exploit the TK in the KMS. SyNAPSE requires the invention of a new "non Von Neumann" architecture. This new architecture will be composed of trillions of "neurosynaptic chips" and connectors -- i.e., simulating the 10 billion neurons and hundred trillion synapses in the human brain. [10].

Both Watson and SyNAPSE are applying advanced Cognitive Analytics furthering the fascinating and evolving field of Artificial Intelligence and Robotics.

5.0 CONCLUSIONS

This paper provided a variety of perspectives on analytics and the impact of analytics on KM systems. The advanced analytics overview provided a view into traditional, commodity analytics -- as well as advance analytics. And the advanced analytics taxonomy provides a framework upon which to evaluate specific KM analytic capabilities, based on a variety of system element perspectives.

The Cognitive Analytics exemplar system provided a detailed overview of the basic components of an advanced analytics solution, and the unique capabilities such a solution offers to a KM system. Page and time limitations limited a variety of additional views (e.g., the power of analytics on social software platforms, and the evolving "Internet of Things"). Yet, this view into cognitive computing can help envision the future KM systems in our horizon – e.g., the reigning Jeopardy! champion (a computer named Watson) for NLP-based decision support, and the embryonic beginnings of an intelligent robot capable of cognitive processing and learning.

ACKNOWLEDGEMENTS

The authors wish to thank Thomas McCreary (Manager, Software Architects Group, IBM Federal CTO Office) and Frank Stein (Director, Analytics Solution Center, IBM Federal CTO Office) for their technical review and discussions – as well as Todd Leyba (Senior Technical Staff Member: IBM Software Group, Watson) for technical consulting and references provided. The authors also wish to thank Susan M. Carlson for insights provided regarding Small and Medium Business Knowledge Management practices and environments.

REFERENCES

[1] Polanyi, M. (1966) *The Tacit Dimension*, London: Routledge & Kegan Paul

[2] Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company*, Oxford: Oxford University Press

[3] Gourlay, Stephen (2002) "Tacit Knowledge, Tacit Knowing or Behaving," *3rd European Organizational Knowledge, Learning and Capabilities Conference;* 5-6 April 2002, Athens, Greece.

[4] Gal, Y. et al. "A Model of Tacit Knowledge and Action," IEEE Computational Science and Engineering, 2009 International Conference p. 463-468.

[5] Maymir-Ducharme, F. and Porpora, G. "IBM Advanced Analytics Portfolio," IBM draft paper and presentations, and conversations 2012 – 2014

[6] Maymir-Ducharme, F. and Ernst, R. "Optimizing Distributed and Parallel TCPED Systems," US Geospatial Intelligence Foundation (USGIF) Technical Workshop, Denver CO, July 17-19, 2013

[7] IBM Global Technology Outlook, IBM Research. (unpublished)

[8] Ferrucci, D. "Introduction to 'This is Watson'," IBM Journal of Research and Development, Vol 56 May/July 2012

[9] DARPA SyNAPSE Program, http://www.artificialbrains.com/darpa-synapse-program

[10] IBM Research, "Neurosynaptic Chips, Building Blocks for Cognitive Computing," http://www.research.ibm.com/cognitive-computing/neurosynaptic-chips.shtml

[11] Memorial Sloan-Kettering Cancer Center IBM Watson Helps Fight Cancer With Evidence-Based Diagnosis and Treatment Suggestions, February 8, 2013, http://www-935.ibm.com/services/multimedia/MSK_Case_Study_IMC14794.pdf