

Obtenga beneficios incorporando el conocimiento de la información no estructurada

IBM Content Analytics Workshop



Agenda

- 11:30** **Bienvenida e introducción**
Gregorio Gómez Amor - Director de Soluciones Smarter Content, IBM

- 11:35** **Convierta su información no estructurada en resultados de negocio**
Joaquín Lacambra - Director de Ventas de Software para Sectores, IBM

- 12:00** **IBM Content Analytics. La solución a su problemática**
Francisco Izquierdo - Client Solution Professional Smarter Content, IBM

- 12:30** **Demostración**
Francisco Izquierdo - Client Solution Professional Smarter Content, IBM

- 13:00** **Aplicabilidad para su negocio**
Jaime Berrocal - Experto en Soluciones Industriales, IBM

- 13:30** **Cóctel**



Convierta su información no estructurada en resultados de negocio



Los cambios en la Información

Más del 72% de los trabajadores tienen dificultades a la hora de encontrar información

Más del 60% de las organizaciones tienen problemas para capturar el 75% de la información de sus clientes

Los costes por incumplimientos en las normas superan en EEUU más de 115\$M



El 60% de las compañías cree que la organización de su contenido es un Caos

BIGDATA el nuevo paradigma

Volumen

12 terabytes

De tweets diarios

Análisis de Sentimientos

Velocidad

5 million

de acciones comerciales

Identificación posible Fraude

Variedad

4 terabytes/site/dia

Videos y seguridad

Monitorización eventos

15 petabytes

De nueva información diaria

Determinar que es importante

500 million

Registros de llamadas por día

Prevenir pérdida del cliente

80% crecimiento

Contenido no estructurado

Mejorar satisfacción cliente

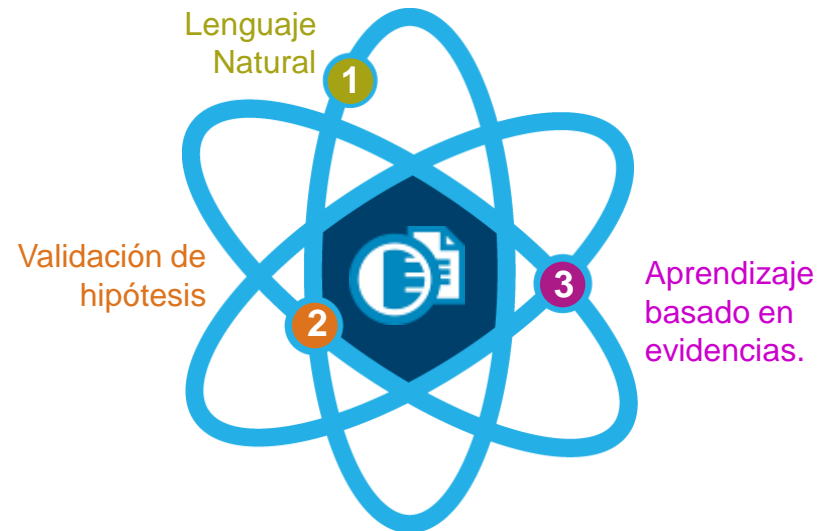
Durante esta presentación, **458.81 terabytes** de información se estarán creando

Analizar para extraer información de negocio.

Extraer nueva información de negocio accediendo, interpretando y analizando la información estructurada y no estructurada.

- **Buscador** corporativo.
- **Clasificación** automática.
- **Datos maestros.**
- Capacidades de **autoaprendizaje.**
- Analizar **grandes volúmenes** de información.
- Interpretación del **lenguaje natural.**
- **Fácilmente adaptable/** evolucionable a las necesidades específicas.

IBM ECM proporciona las más avanzadas tecnologías que definen la siguiente generación de soluciones Smarter Analytics que pueden razonar y autoaprender.



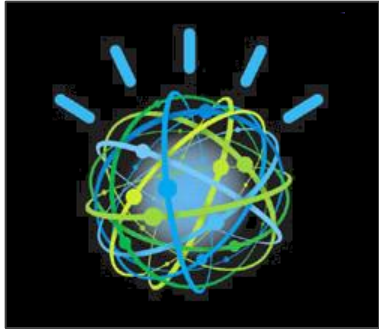
Evolucionar desde buscar a descubrir, de posibilidades a probabilidades y de resultados simples a opciones inteligentes.

¿Que es Watson?

Es un ordenador con un potente software analítico y de clasificación, capaz de entender y responder preguntas

Watson accedió a 200 millones de páginas de contenido estructurado y no estructurado consumiendo cuatro terabytes de almacenamiento en disco, algo similar a lo que incluye el texto completo de la Wikipedia. El ordenador no estaba conectado a Internet durante el juego. Para cada pista, tres posibles respuestas de Watson se mostraban en la pantalla del televisor. Watson siempre superó a sus oponentes humanos en los resultados de cada una de las respuestas,

El software que se utilizó está basado en tecnologías ECM de análisis y minería de la información.



IBM Content Analytics es la plataforma para obtener una visión rápida de la información



- Transformar la información en bruto en información empresarial de forma **rápida** sin necesidad de construir modelos o el despliegue de sistemas complejos
- Obtener **conocimiento** en horas o incluso días ... semanas o meses.
- **Fácil de usar** para todos los trabajadores para buscar y explorar el contenido
- **Flexible y extensible**



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

Thank
YOU

The image features the words "Thank YOU" in a large, 3D, light blue font. Each letter of the word "Thank" contains a different person's face. The "T" shows a man in a suit and tie. The "h" shows a woman in a green top. The "a" shows a man with a green face. The "n" shows a woman in a blue patterned top. The "k" shows a man with glasses. The word "YOU" is positioned below "Thank". The "Y" shows a man in a white shirt. The "O" shows a man in an orange shirt. The "U" shows a woman in a dark top.



IBM Content Analytics. La solución a su problemática



¿Qué es Analytics?



- Estudiar la información de forma **agregada** para **entender mejor** lo que está pasando en **mi empresa**
 - “Mining for Insights”

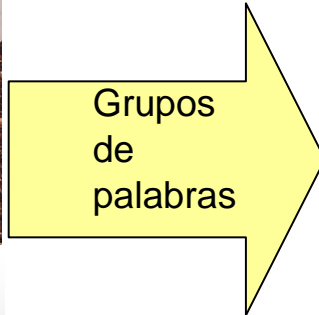
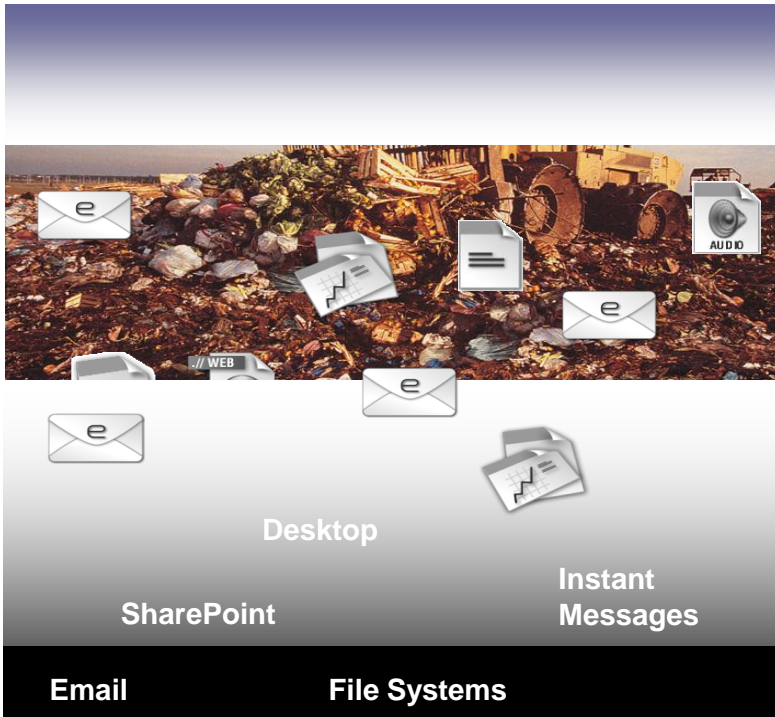
¿Y qué es Content Analytics?



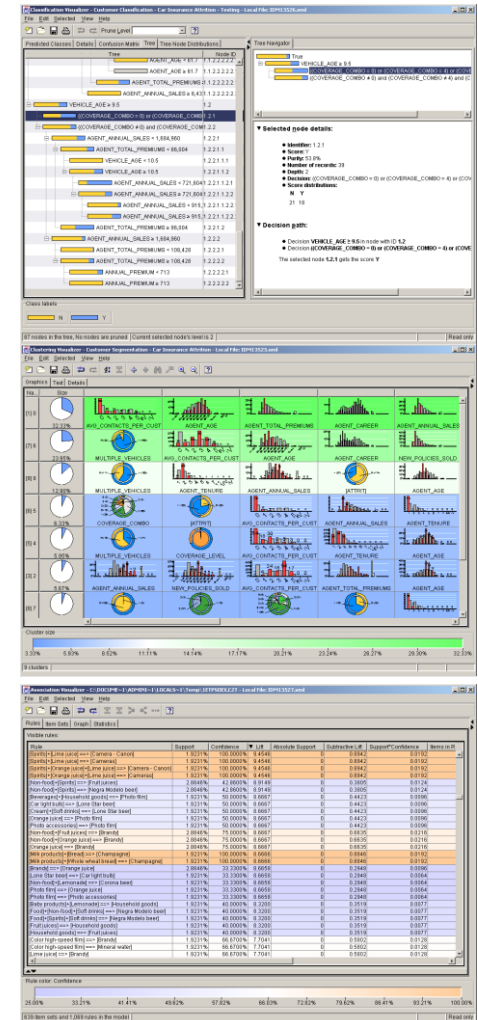
- Minería de contenidos no estructurado, es decir información en lenguaje natural, para entender aún mejor lo que está pasando en mi empresa.
 - El 80% de la información empresarial es no estructurada

¿Qué es la Minería de Textos?

- La Minería de Textos es el proceso de extracción de **información utilizable a partir de datos textuales no estructurados**, mediante la identificación de conceptos, opiniones y tendencias.



01100101
 11101011
 01101110
 00010100



Content Analytics

- 80% de la información que gestionan las empresas y organismos es no estructurada
- Y la información estructurada también puede tener errores
 - ¿Sabe realmente el conductor qué pieza es la que ha generado el problema?
 - Si un paciente dice que no es fumador, pero en una consulta dice que fuma un par de cigarros con los amigos cuando ven un partido de futbol ¿qué información es la verdadera?
 - Si la puntuación de un producto sólo es de 1 sobre 5, pero los comentarios se refieren a problemas en la entrega, ¿Cuál es la opinión real sobre el producto?

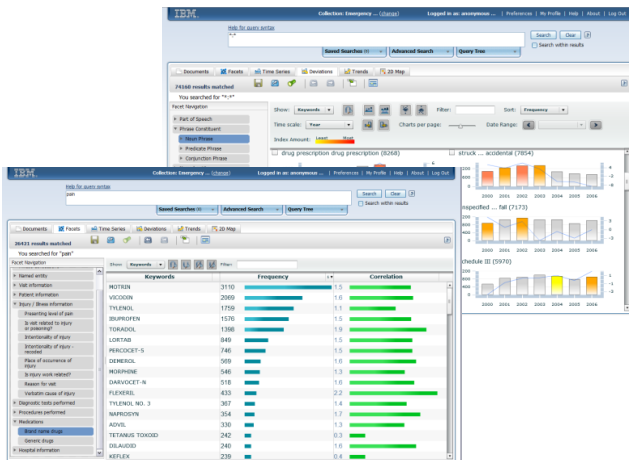
Content Analytics

- Cuando disponemos de información textual, lo más habitual es que sea más completa y más precisa que la información estructurada
- Hay ocasiones en las que SOLO se dispone de información textual
- La necesidad real consiste en transformar la información textual en información estructurada

Definiciones

¿Qué es el Análisis de Texto?

Text Analytics (PNL) describe un conjunto de técnicas estadísticas lingüísticas, y de aprendizaje automático que permiten que el texto a analizar y la extracción de información sea clave para la integración de negocios.



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.

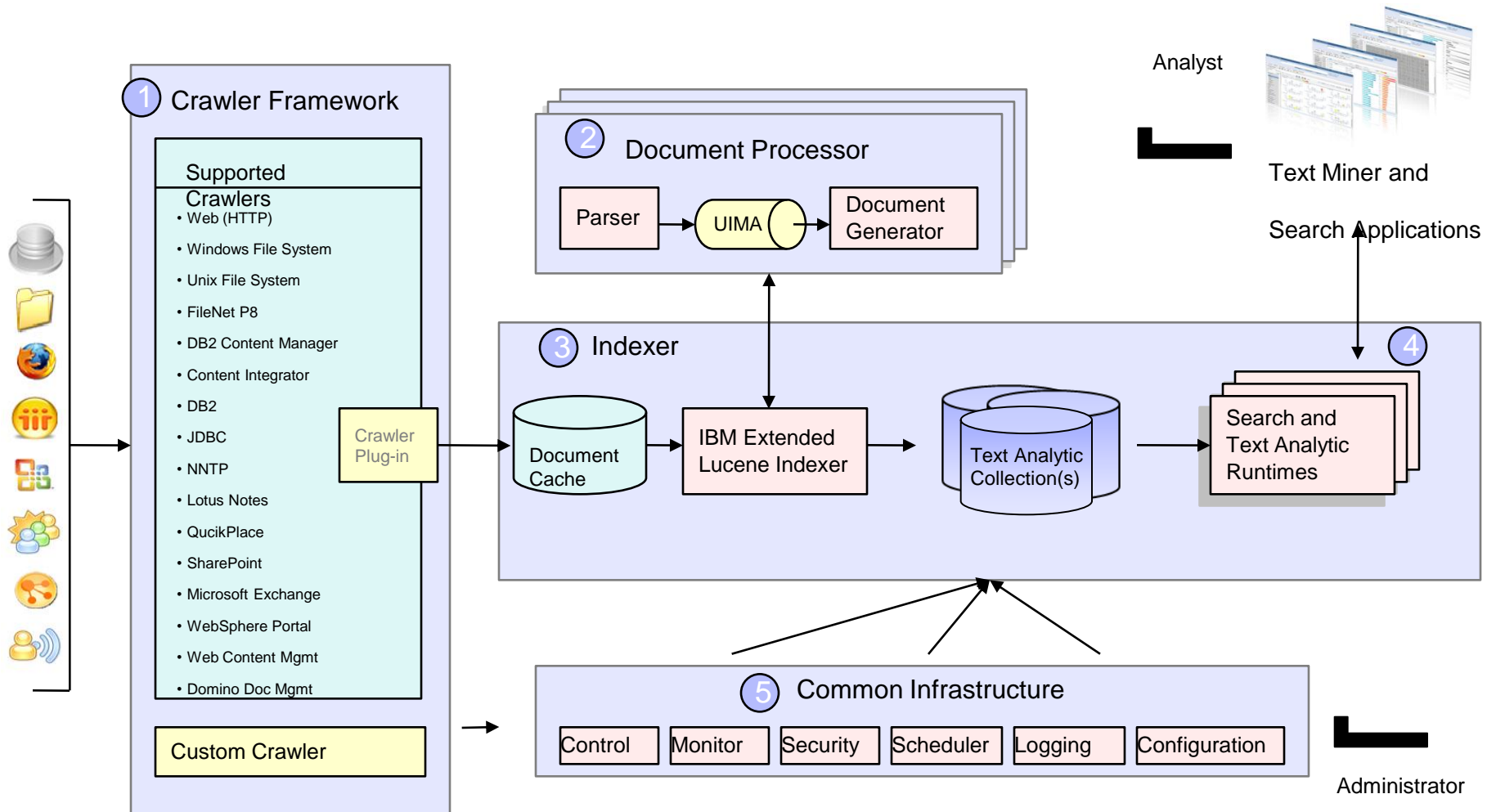


Arresting_Officer	PC 143
Arrest_Date_Time	15/06/2006 : 23:47
Suspect_Forename	John
Suspect_Surname	Setsuko
Suspect_VRN	W563WDL
Suspect_Vehicle_Color	White
Suspect_Vehicle_Make	Ford Mondeo
Suspect_Addr_Street	22 East Dene Ridge
Suspect_Addr_Town	Ipswich
Evidence_1_Description	1 oz Cannabis Resin
Classification	Drug possession

Qué es Content Analytics?

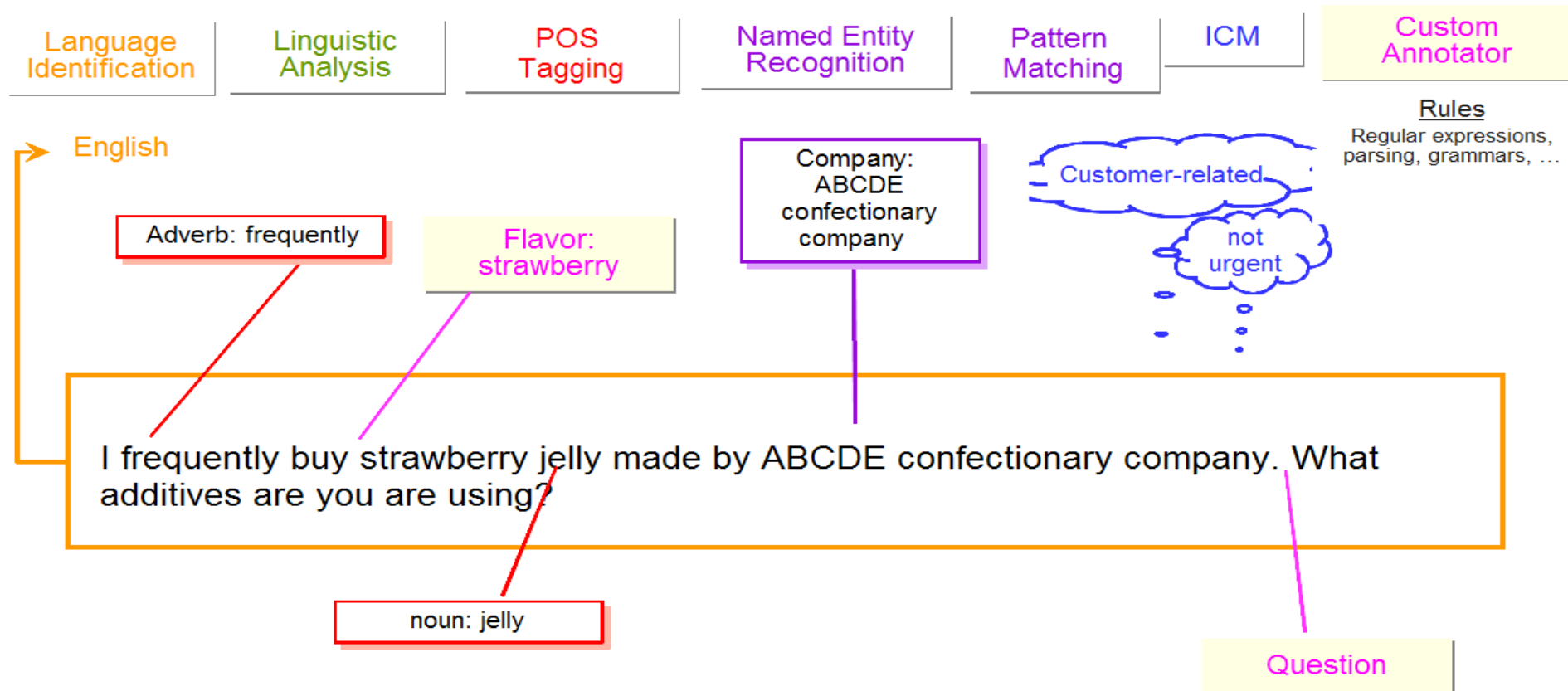
Content Analytics (Text Analytics + Mining) se refiere al proceso de análisis de texto además de la capacidad para identificar visualmente y explorar las tendencias, patrones, y los hechos estadísticamente relevantes encontrados en distintos tipos de difusión de contenidos a través de fuentes de contenido internas y externas

Arquitectura de ICA



¿Qué son los Anotadores ICA?

Anotador: Se trata de un componente software que realiza tareas de análisis lingüístico, y como resultado obtiene y registra **anotaciones**



¿Y qué es lo que hacen?

Language Identification

Segmentation

Normalization

Classification

Disambiguation*

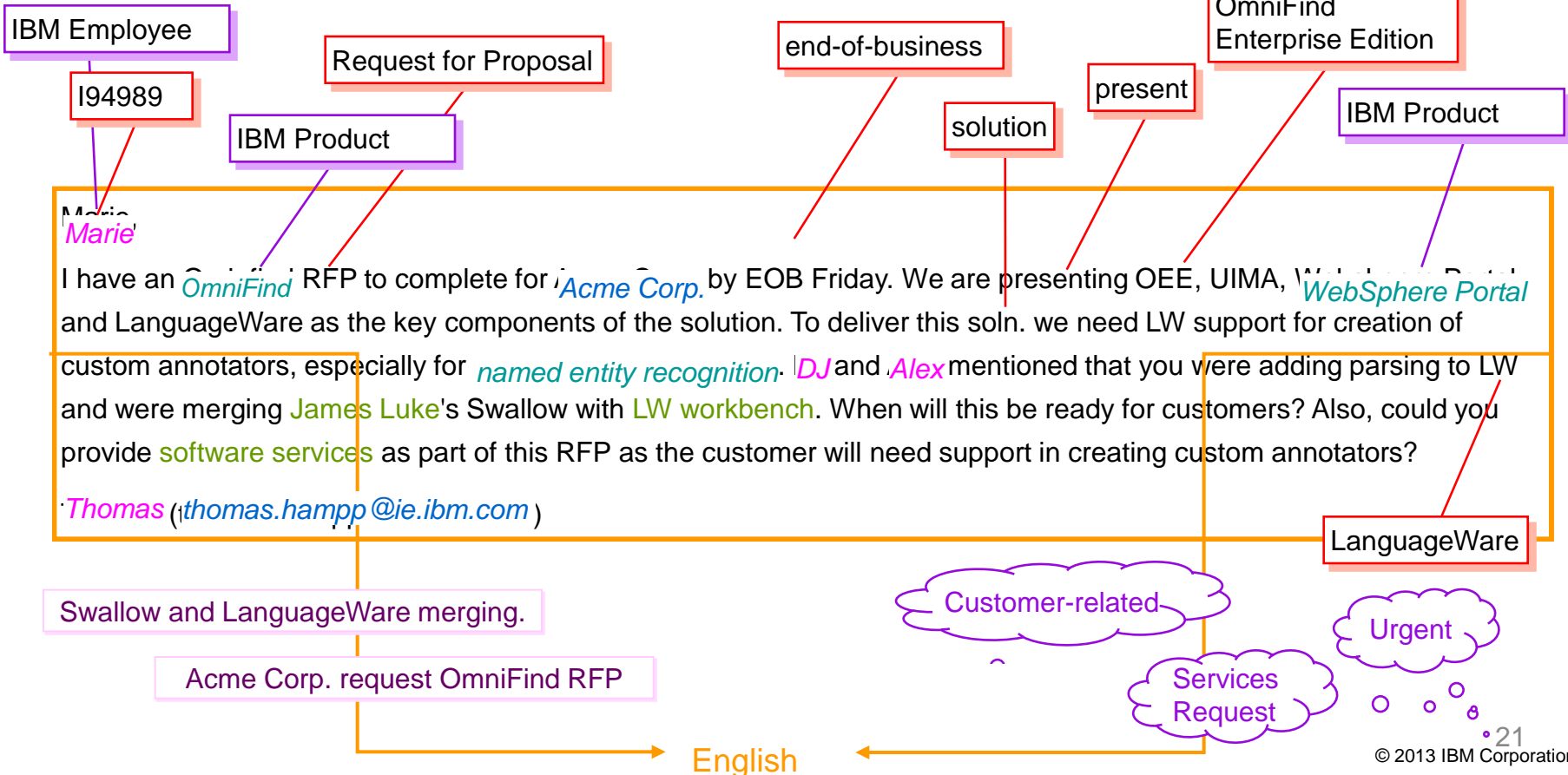
Relationship Extraction*

Rules

Regular expressions, parsing, grammars, ...

Fuzzy matching

Spelling correction, approx. lookup, hyphenation, ...

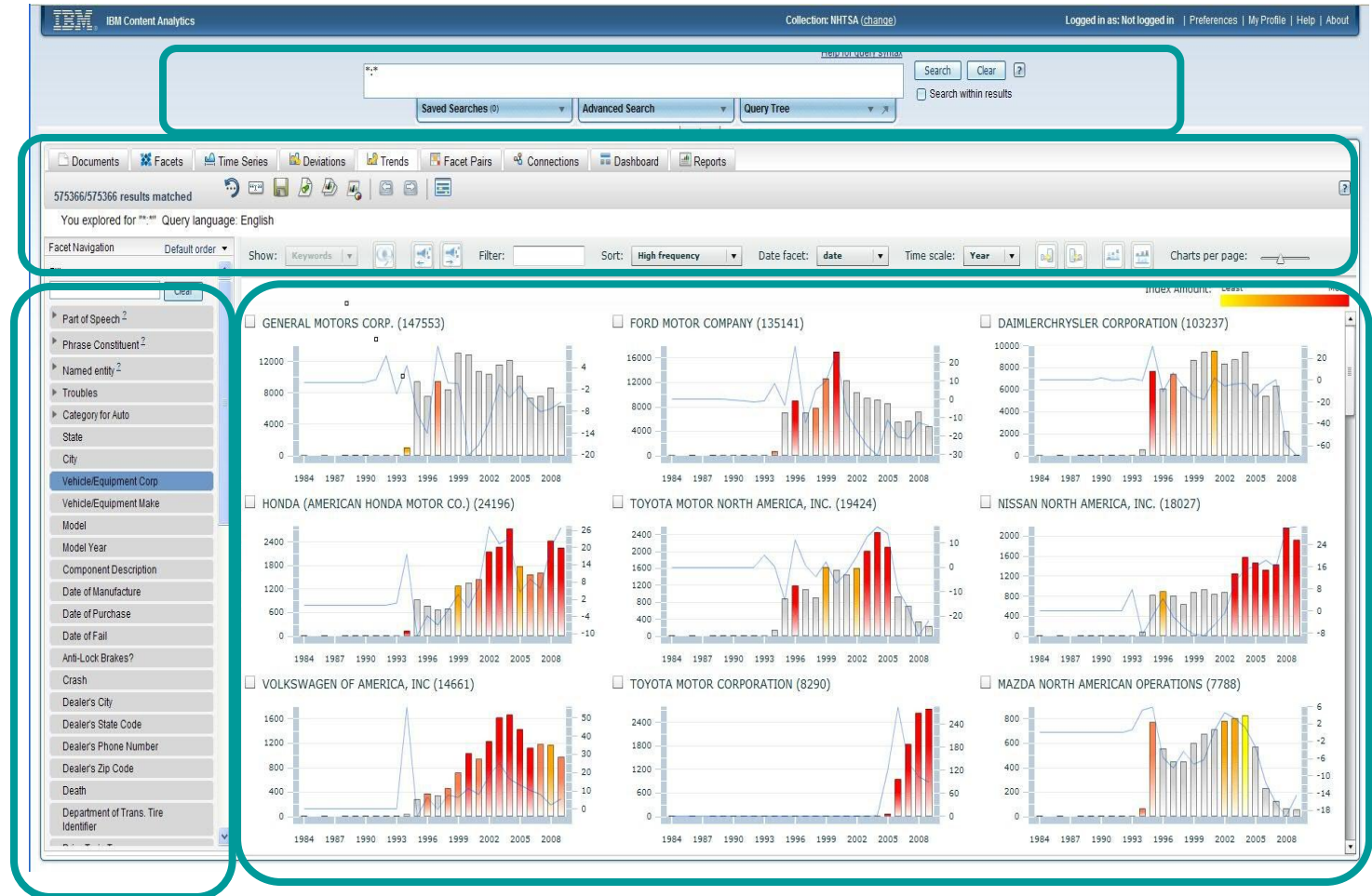


Vista del Análisis de los Contenidos

Query, búsqueda

Puntos de vista, filtros y humbrales

Automáticamente se extraen y analizan los conceptos, las entidades, relaciones, metadatos y Clasificaciones



Visualización con capacidades de Drill Down

Content Analytics Miner

- 8 vistas para búsqueda, análisis e investigación
- Funcionalidades de búsqueda habituales: type ahead, número de resultados, etc

Facetas

Dashboard

Time Series

Desviaciones / Tendencias

Document Analysis

Enterprise Search

Conexiones

Pares de Facetas

The image displays a collage of screenshots from the IBM Cognos Content Analytics interface, illustrating various features and data visualizations. The screenshots are arranged in a grid-like fashion, with green callout boxes highlighting specific capabilities.

- Facetas:** A screenshot showing a list of facets (Attributes, Sequence Number, Source of Complaint, etc.) and a table of results with columns for Keywords, Frequency, and Correlation.
- Dashboard:** A screenshot showing a dashboard with multiple charts and graphs, including a pie chart for City (Top 5) and a bar chart for State (Top 5).
- Time Series:** A screenshot showing a time series chart with a date range from 2000-01 to 2010-01.
- Desviaciones / Tendencias:** A screenshot showing a time series chart with a date range from 2000-01 to 2010-01, highlighting deviations and trends.
- Document Analysis:** A screenshot showing a document analysis view with a search bar and a list of results.
- Enterprise Search:** A screenshot showing an enterprise search interface with a search bar and a list of results.
- Conexiones:** A screenshot showing a network diagram with nodes and connections, representing relationships between data points.
- Pares de Facetas:** A screenshot showing a facet pair view with a table of results and a bar chart.

Conexiones: Vista con términos correlacionados

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

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

575366/575366 results matched

Facet Navigation: Documents | Facets | Time Series | Deviations | Trends | Facet Pairs | Connections | Reports

Rows Show: Keywords Labels: All Correlation: [Slider] Columns Show: Keywords Highlight mode: No Highlighting

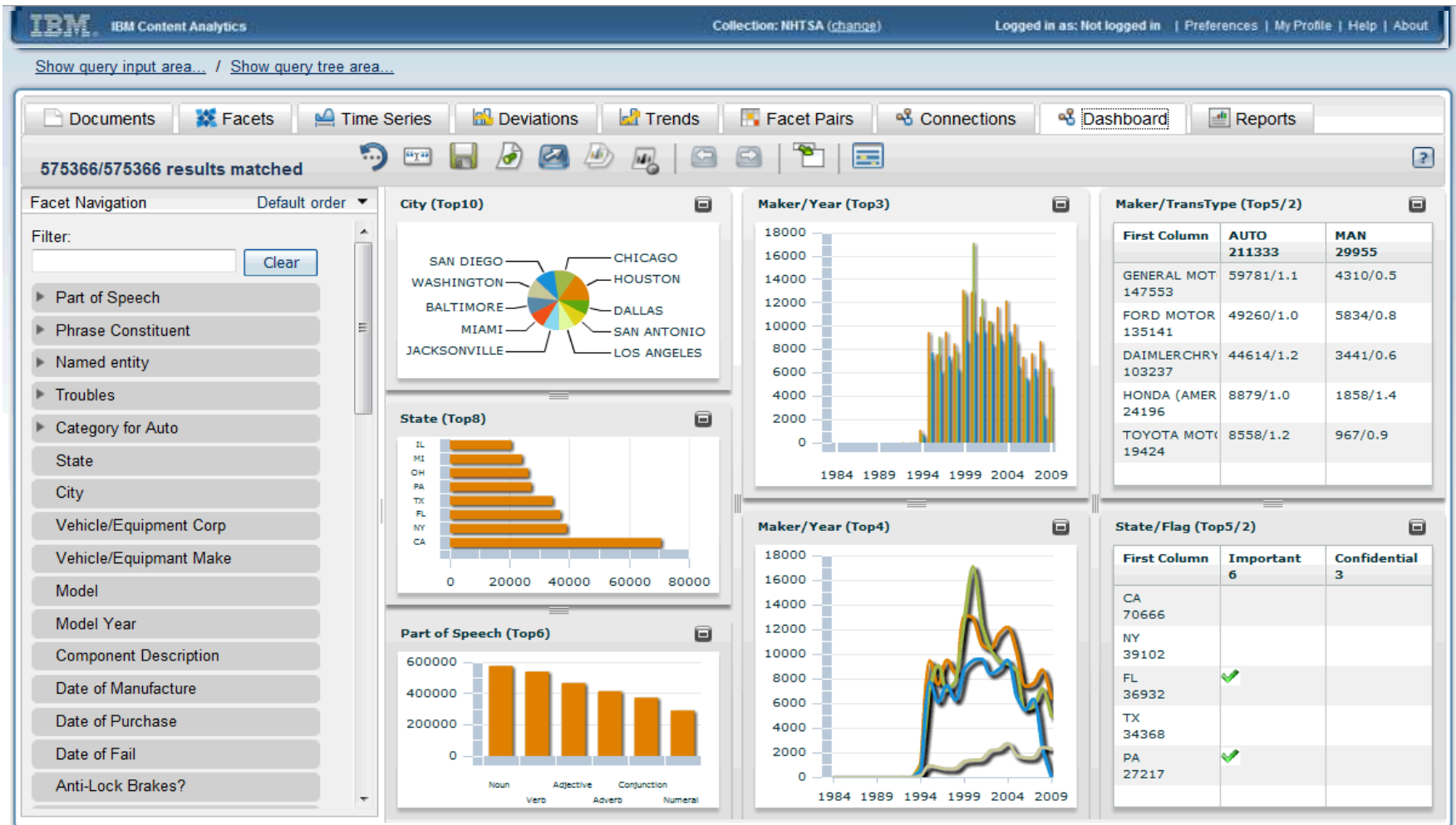
Filter: [Input] Clear

Facet Navigation: Part of Speech, Phrase Constituent, Named entity, Troubles (Negative event selected), Category for Auto, State, City, Vehicle/Equipment Corp, Vehicle/Equipmant Make, Model (selected), Model Year, Component Description, Date of Manufacture, Date of Purchase, Date of Fail, Anti-Lock Brakes?, Crash, Dealer's City, Dealer's State Code

Key: Rows = Negative event, Columns = Model. Frequency: Most, Least. Correlation Amount: Least to Most (color scale).

The network graph displays correlations between various terms. Nodes are labeled with terms such as 'unsafe', 'problem', 'wrong', 'lose', 'die', 'rough', 'noise', 'loose', 'break', 'fall', 'crack', 'leak', 'burn', 'smell', 'NEON', 'UNKNOWN', 'damage', 'prevent', 'fall', 'tear', 'wear', 'warp', 'vibration', 'noise', 'loose', 'CHILD SAFETY SEAT', 'CRACK', 'leak', 'smell', 'burn', 'NEON', 'UNKNOWN', 'damage', 'prevent', 'fall', 'tear', 'wear', 'warp', 'vibration', 'noise', 'loose', 'CHILD SAFETY SEAT', 'CRACK', 'leak', 'smell', 'burn', 'NEON'. Edges represent correlations, with thickness and color indicating the amount. A key explains the color coding for frequency and correlation amount.

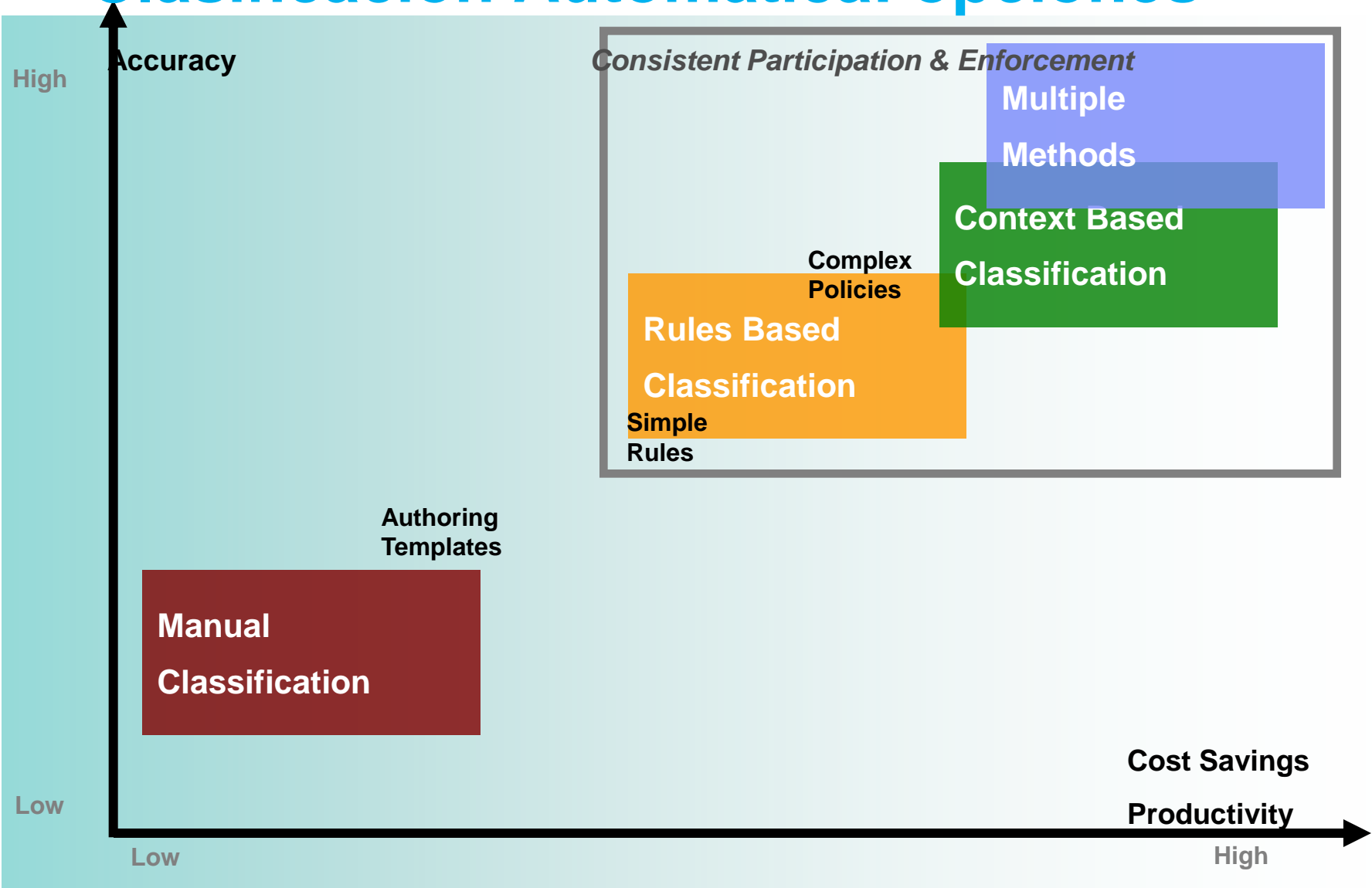
Dashboard ejecutiva



Características de la Clasificación

	Manual	Automática
Precisión	92% 46%	50 – 80%
Coste (por doc)	\$ 0.17	< \$ 0.01
Consistencia	<50%	100%
Mayor Volumen		

Clasificación Automática: opciones



Classification Workbench

Una “**Knowledge Base**” (KB) es un índice estadístico que permite clasificar documentos según una taxonomía

Se entrena mediante una **lista de categorías con ejemplos**, y a partir de ese momento se puede utilizar para proponer categorías

Un “**Decision Plan**” es una secuencia de reglas, que pueden utilizar una **KB** para clasificar los nuevos documentos que lleguen

The screenshot displays the Classification Workbench interface for a project named 'online retail'. The main window shows a table with columns for 'message_body', '_Categories', 'OriginalID', and five 'Match' columns (Match1 to Match5). The table contains 26 rows of data, each representing a document and its classification results. A right-hand pane shows a tree view of the project's fields, including 'message_body', '_Categories', 'OriginalID', and the five 'Match' fields. The bottom pane shows a log of system events, including messages about loading project categories, keywords, and field information, as well as the initialization of the Relationship Modeling Engine.

	message_body	_Categories	OriginalID	Match1	Match2	Match3	Match4	Match5
1	Do you have my ite...	Inventory	1	Out of Stock (68.29)	Invent...	Locatio...	NewInf...	Bridal ...
2	Does my purchase i...	Batteries	2					
3	How much does it c...	Gift Orders	3	Gift Wrap (26.96)	Shippi...	Bridal ...	Catalo...	Produc...
4	Do you ship to a Fle...	APO_FPO	4					
5	Are batteries include...	Batteries	5	Batteries (98.73)	Selecti...	Order ...	Web Br...	Shippi...
6	If I buy a mother's d...	Gift Orders	6					
7	What are your next d...	Shipping Charges	7	Shipping Charges (4...	Order ...	Prop65...	NewInf...	Locatio...
8	When was your com...	Company	8					
9	I don't think it's fair t...	Shipping Refund	9	Order not received (...	Order S...	Shippi...	Selecti...	Return ...
10	Is my item out of sto...	Inventory	10					
11	Please help. I've lost...	Password	11	Gift Certificates (48.1...	Passwo...	Order ...	Secure ...	Return ...
12	Is sales tax included?	Sales Tax	12					
13	How much sales tax ...	Sales Tax	13	Sales Tax (88.71)	NewInf...	Shippi...	Order ...	Accou...
14	Can I buy from over...	International Order	14					
15	Can I create my own...	Wish List	15	Wish List (97.21)	Prop65...	Accou...	Return ...	Order ...
16	What kind of wrappi...	Gift Orders	16					
17	I would like to pay b...	Online Payment	17	Online Payment (67...	Secure ...	Return ...	Order ...	Credit...
18	I placed an order on ...	Order not received	18					
19	Can I place an order ...	Order Pickup	19	Order Pickup (97.90)	Phone...	Partial ...	Locatio...	Return ...
20	Can I make an excha...	Return Policy	20					
21	What number can I ...	PhoneOrder	21	PhoneOrder (98.53)	Order ...	Web Br...	Order ...	Interna...
22	Do you sell tv's?	Selection	22					
23	Do you offer gift wra...	Gift Wrap	23	Gift Orders (98.64)	Gift Wr...	Gift Ce...	Selecti...	Bridal ...
24	Can I return or excha...	Return Policy	24					
25	What is your sales ta...	Sales Tax	25	Sales Tax (98.42)	NewInf...	Order ...	Return ...	Privacy...
26	I would like to excha...	Return Policy	26					

Viewed 802

Log:

- 08/08/2008 18:04:... Adding entry to fields: _MatchesScore
- 08/08/2008 18:04:... Setting current matches field to: _Matches
- 08/08/2008 18:04:... Setting current scores field to: _MatchesScore
- 08/08/2008 18:04:... Loading project categories from C:\IBM\ClassificationModule\Classification Workbench\Projects_Unicode\online retail\
- 08/08/2008 18:04:... Loading keywords data...
- 08/08/2008 18:04:... Loading field information from C:\IBM\ClassificationModule\Classification Workbench\Projects_Unicode\online retail\NVP_BootStrap_Catalog.xml
- 08/08/2008 18:04:... Loading views...
- 08/08/2008 18:04:... Project opened
- 08/08/2008 18:04:... Initializing Relationship Modeling Engine...

Active view: View all (2) #Content set items: 802 #Categories: 46/46

¿Quién se puede beneficiar de IBM Content Analytics?

1. ICA es una solución ideal para los casos en que hay información textual revisada y analizada por personas
2. ICA se puede utilizar en los casos en los que sabemos que en el texto hay hechos importantes, pero extraerlos se convierte en un proceso largo, lento y caro
3. La necesidad de análisis textual está presente en todas las disciplinas. No es necesario disponer de un “documento formal”, podemos trabajar con conversaciones grabadas, medios audiovisuales, etcétera

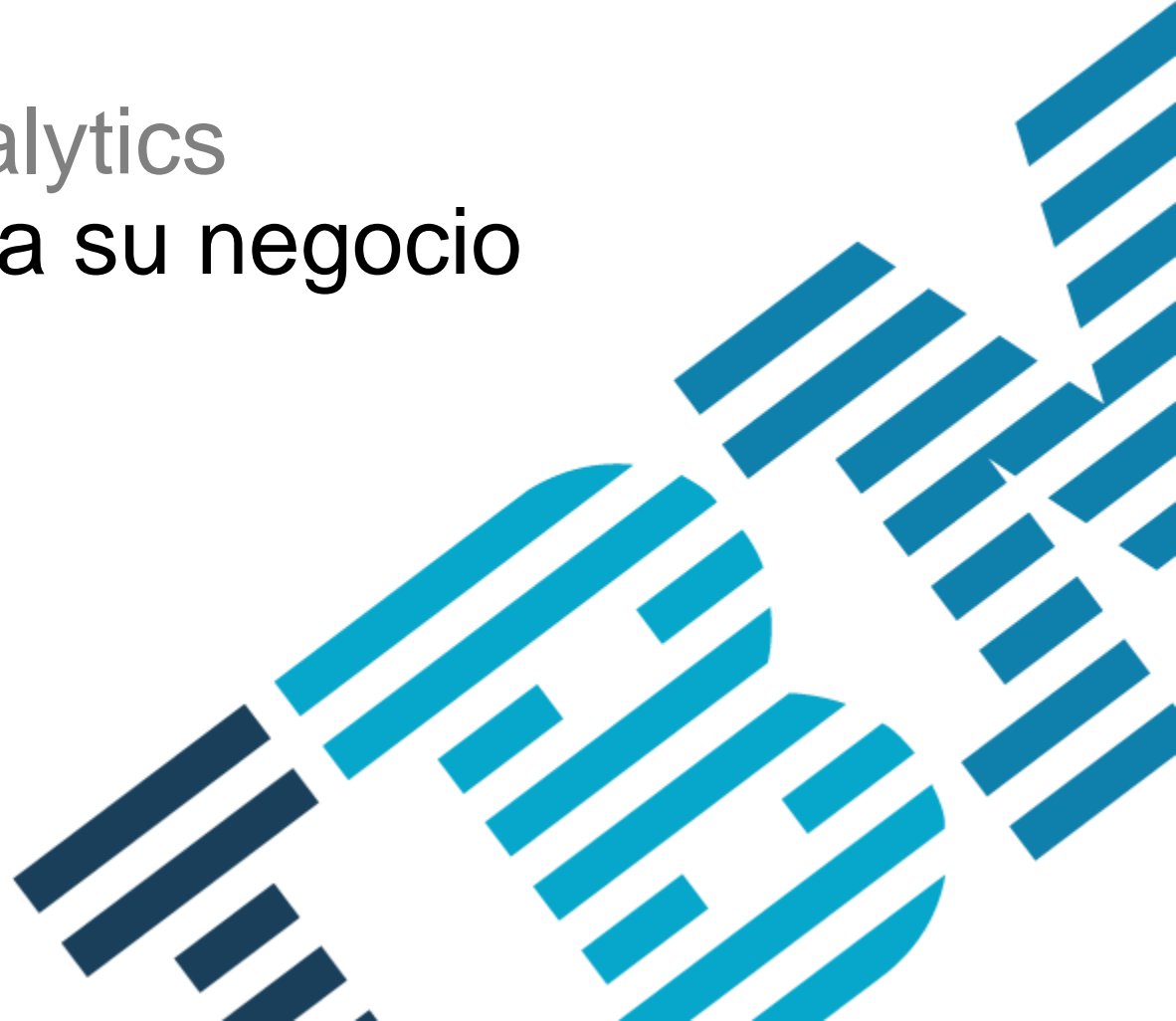
Thank
YOU

The image features the words "Thank YOU" in a large, 3D, light blue font. Each letter of the word "Thank" and "YOU" is filled with a different photograph of a person. The "T" shows a man in a white shirt and orange tie. The "h" shows a woman in a green top. The "a" shows a man with a green face. The "n" shows a woman in a blue top. The "k" shows a man with glasses in a blue top. The "Y" shows a man in a blue shirt looking at a document. The "O" shows a man in a white shirt and orange tie. The "U" shows a woman in a green top.



IBM Content Analytics

Aplicabilidad para su negocio



IBM Content Analytics:

Ámbitos generales de aplicación



Inteligencia comercial



Servicio al cliente

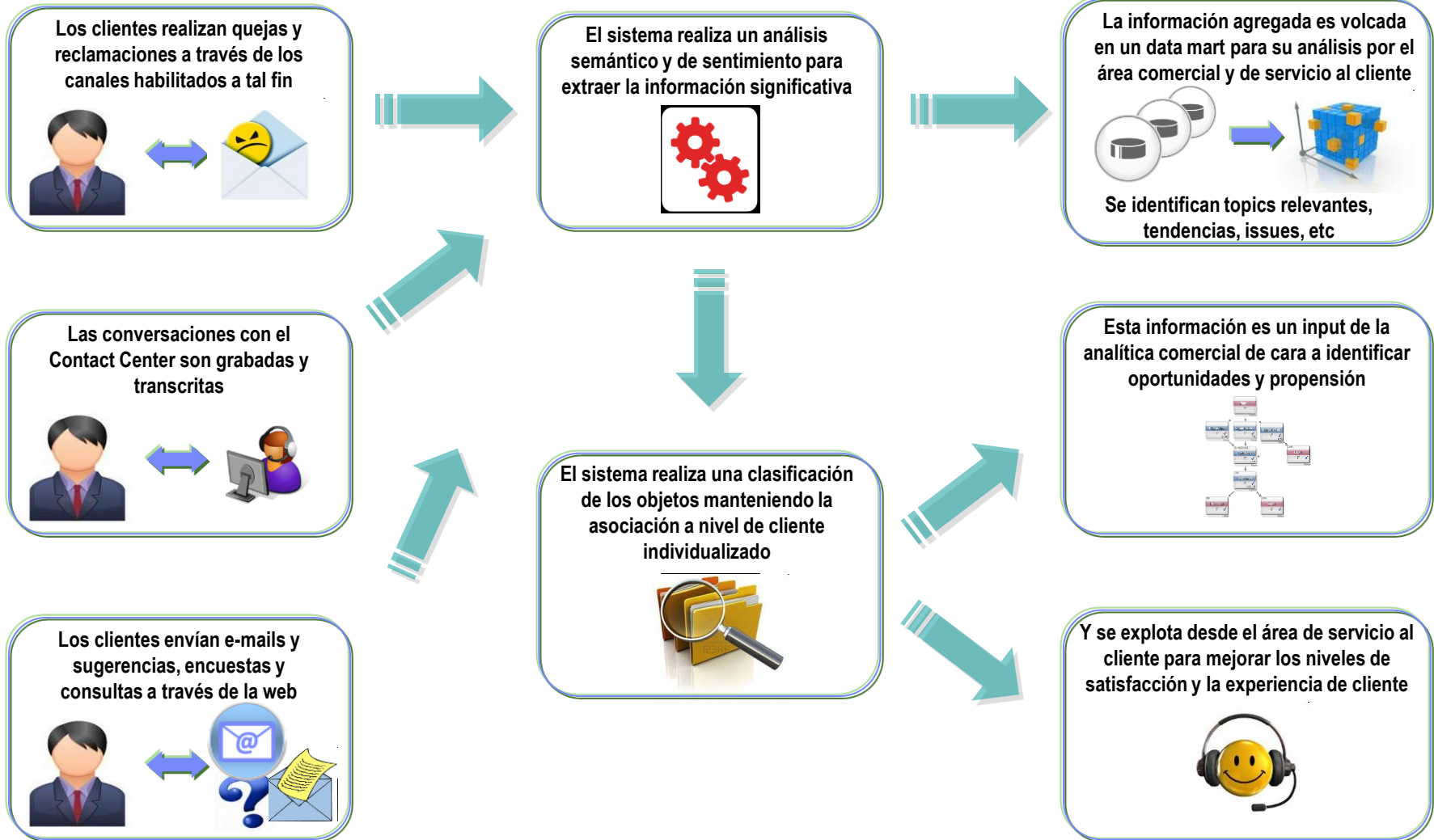


Fraude y seguridad

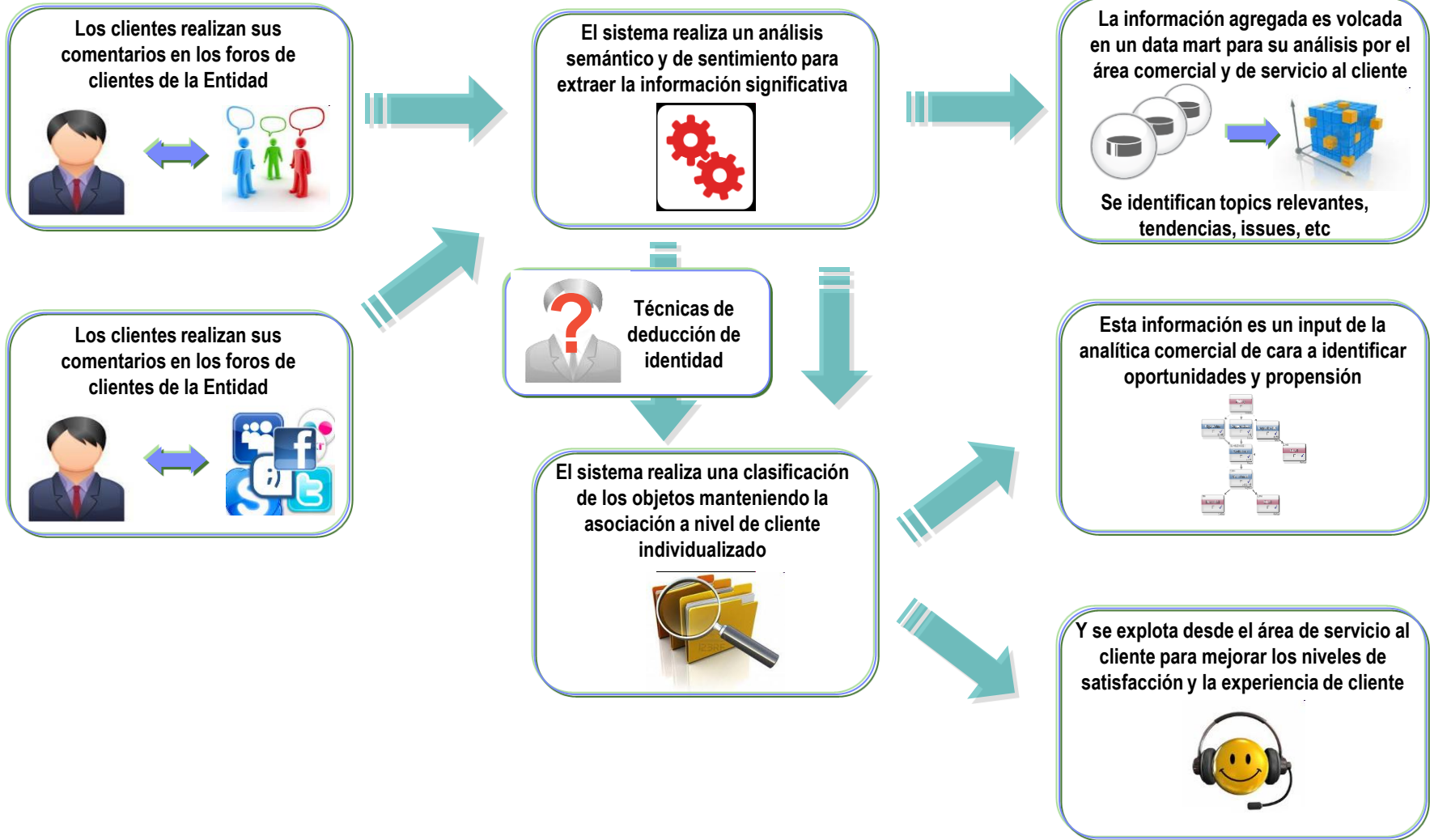


Clasificación y búsqueda

💡 Inteligencia comercial y 😊 Servicio al cliente



💡 Inteligencia comercial y 😊 Servicio al cliente



Gran cadena de distribución de bienes de consumo



Extrayendo tendencias de mercado de las redes sociales



Objetivos

- Compañía japonesa con más de 10.000 centros de distribución.
- Optimizar el stock en base a las dinámicas tendencias de compra de su público objetivo.

Solución

IBM Content Analytics aplicado a enormes volúmenes de información de foros y noticias:

- Identificación temprana de nuevos intereses y preferencias.
- Análisis ágil de necesidades y quejas de clientes.
- Planificación y desarrollo de nuevos productos y servicios.

Beneficios

- Incremento del beneficio neto en medio punto al año.
- Proporcionar soporte a las decisiones de marketing y cadena de suministro.

Compañía de Telecomunicaciones Asiática

Reduciendo la fuga de clientes



Inteligencia comercial

Servicio al cliente



Objetivos

- Entender la “voz del cliente y prevenir cancelación de contratos.
- Identificar nuevas oportunidades.
- Responder ágilmente a las incidencias.

Solución

Analizar transcripciones del call center, e-mails y quejas/reclamaciones del cliente para:

- Detectar clientes en riesgo de fuga.
- Identificar problemas de los clientes y sugerir FAQ candidatas a incorporar al canal web.

Beneficios

- Incremento de un 50% en la precisión de la detección de clientes en riesgo de fuga.
- Mejora en los ratios de lealtad de los clientes y de upselling.
- Incremento del nivel de resolución de problemas a través del sistema de FAQ.

Entidad financiera

Monetizando la información de las redes sociales



Inteligencia comercial

Servicio al cliente



Objetivos

- Enriquecer el conocimiento de los clientes con información de hábitos de consumo, preferencias y eventos vitales.
- Ofrecer servicios de valor añadido a los clientes pymes con foco en el sector de comercio.

Solución

Extracción y análisis de información de foros propios, redes sociales y TPVs para:

- Enriquecimiento del datamart de clientes con indicadores de propensión comercial.
- Operacionalización en los procesos de marketing y servicio al cliente.
- Servicio de públicos objetivo para comercios.

Beneficios

- Incremento de la tasa de conversión de las ofertas comerciales a clientes.
- Incremento de la lealtad de clientes Pymes y revenue adicional.

Proveedor Japonés de tele-servicios



Mejorando la preparación de los agentes



Objetivos

- Mejorar el entrenamiento de los agentes y proporcionar una atención al cliente mejorada.
- Procesar grandes volúmenes de información de las interacciones con los clientes.

Solución

- Utilización de IBM Content Analytics para analizar logs de conversaciones telefónicas, e-mails e interacciones web.
- Identificación de palabras clave en las consultas y evaluación de las respuestas de los agentes.

Beneficios

- 20% de reducción en la transferencia de llamadas a agentes de 2º nivel, gracias a la identificación de deficiencias en la formación de los operadores.
- Visión agregada de problemas relacionados con productos y servicios para planificar entrenamiento y mejora de los servicios.

IBM Content Analytics:
Ámbitos generales de aplicación



Inteligencia comercial



Servicio al cliente



Fraude y seguridad

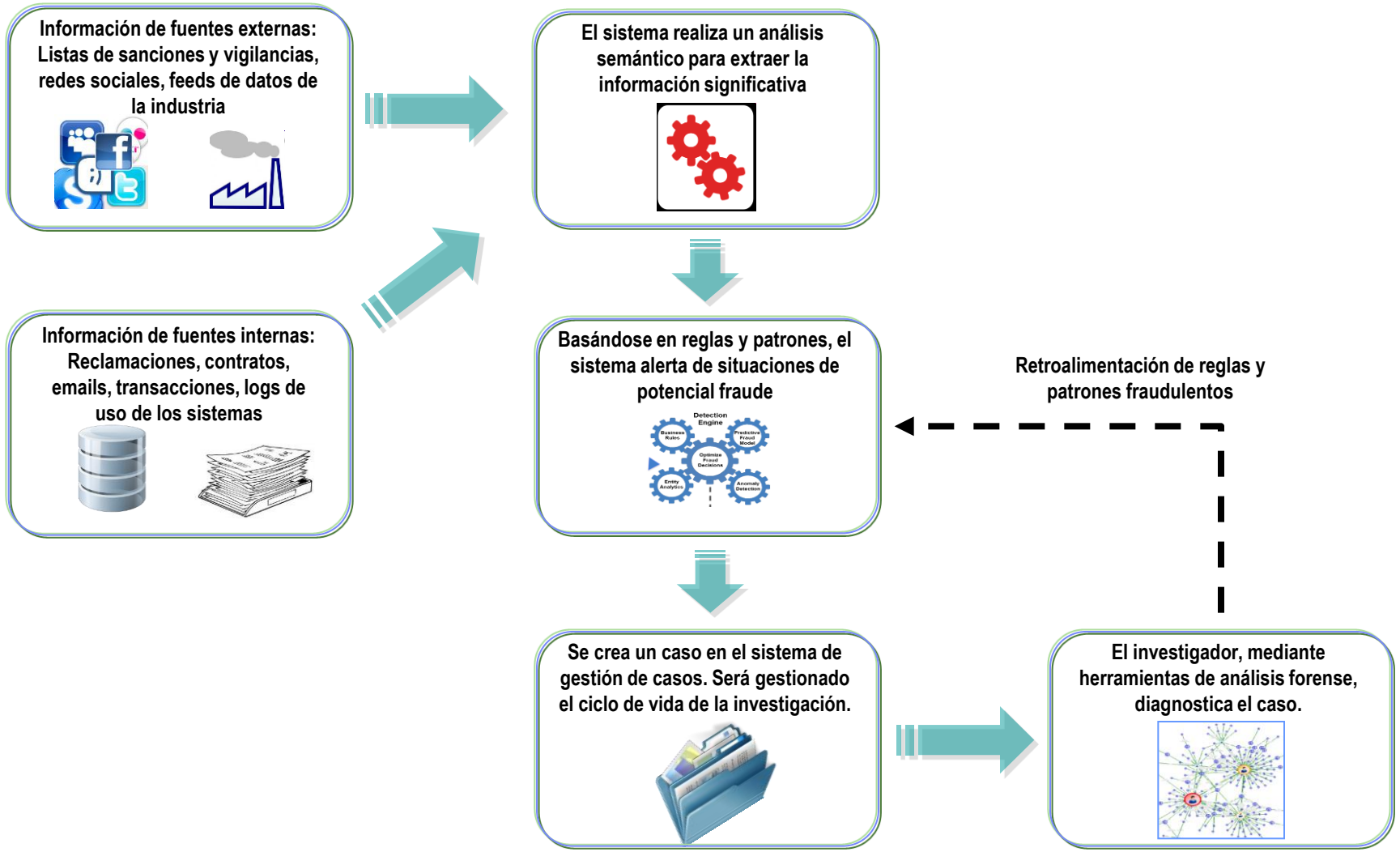


Clasificación y búsqueda



Fraude y seguridad

IBM Intelligent Investigation Manager



Institución Financiera de ámbito Europeo



Fraude y seguridad

Identificando potenciales amenazas de fraude



Objetivos

- Identificar posibles comportamientos y técnicas fraudulentas mediante la investigación de determinados sitios web.

Solución

- Utilización de IBM Content Analytics para analizar un conjunto seleccionado de sitios Web.
- Investigación de la información potencialmente relacionada con comportamientos fraudulentos.
- Reporte de las potenciales amenazas al departamento de seguridad.

Beneficios

- Identificación temprana de posibles tipos de intrusión.
- Proporcionar información al dpto de seguridad para la elaboración de escenarios de prevención.

Grupo internacional de seguros y servicios financieros

Reduciendo la exposición al riesgo



Fraude y seguridad



Objetivos

- Reducción del ratio de pérdidas por siniestros.
- Detección y prevención del fraude.

Solución

- Análisis de múltiples fuentes con información de siniestros, clientes y contratos con un histórico de 15 años.
- Búsquedas semánticas sobre información estructurada y no estructurada.
- Capacidad de incorporar de forma ágil nuevas fuentes de datos internas o externas.

Beneficios

- Mejora en los modelos de scoring de riesgo.
- Indicación de posibles situaciones de fraude mediante la identificación de patrones sobre siniestros.
- Mayor nivel de automatización en los procesos de scoring de riesgos.

Policía Federal en Latino América

Mejorando el servicio de inteligencia



Fraude y seguridad



Objetivos

- Fusionar la información de diversas fuentes para obtener una visión única de los sospechosos.
- Proporcionar acceso simplificado a gran cantidad de información dispersa.

Solución

- Extracción de hechos, entidades y conceptos de fuentes no estructuradas como informes de investigación.
- Búsqueda y correlación de información para resolución de identidades y visión única del sospechoso.
- Gestión de +500 millones de registros no estructurados.
- Implementado junto con I2.

Beneficios

- Ayuda en la identificación de sospechosos y sus relaciones.
- Ayuda en la identificación de ubicaciones de potencial actividad criminal.

IBM Content Analytics:
Ámbitos generales de aplicación



Inteligencia comercial



Servicio al cliente



Fraude y seguridad



Clasificación y búsqueda

Clasificación y búsqueda

Información dispersa en múltiples fuentes:
Registros, sistemas de ficheros, contenidos web, comunidades, foros, feeds, etc.



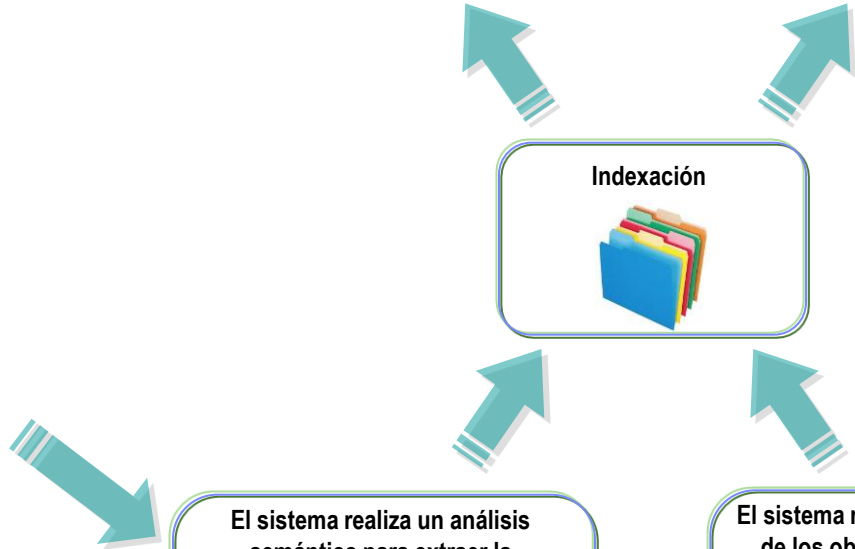
El usuario puede realizar búsquedas mediante lenguaje natural y palabras clave

El usuario puede realizar búsquedas mediante la exploración de taxonomías

Indexación

El sistema realiza un análisis semántico para extraer la información significativa

El sistema realiza una clasificación de los objetos por taxonomías



Gran compañía de telecomunicaciones

Facilitando la búsqueda de activos



Clasificación y búsqueda



Objetivos

Proporcionar un instrumento de búsqueda sencillo y preciso sobre una base de información:

- Incoherente y anticuada tras varios procesos de fusión.
- Dispersa en múltiples fuentes (bases de datos, ficheros, sitios web, gestores de contenido, etc).

Solución

- Buscador corporativo implementado con ICA.
- Devuelve resultados combinando múltiples fuentes.
- Procesos asistidos de registro y verificación de nuevas fuentes de información.

Beneficios

- Mayor agilidad en la localización de información a lo largo de la corporación.
- Los procesos de administración aseguran que la información disponible cumple las políticas de seguridad y estándares regulatorios.

Compañía de asesoría legal

Agilizando la actualización regulatoria



Clasificación y búsqueda



Objetivos

Proporciona servicios de información legal a más de 1.200 administraciones locales.

- Agilizar la actualización de la información regulatoria.
- Reducir los tiempos necesarios para hacer llegar las actualizaciones a sus clientes.
- Consolidar toda la información en una única plataforma.

Solución

- Transformación de la gestión de información regulatoria incorporando tecnología de análisis de lenguaje natural.
- Reconocimiento automático y actualización sistemática de modificaciones.

Beneficios

- Minimización de errores manuales.
- Reducción del tiempo de actualización en un 50%
- Incremento de la agilidad en la búsqueda de información legal.



Jaime Berrocal Prat
Industry Solutions
IBM Software Group

Santa Hortensia 26-28
28002 Madrid (Spain)

Phone: +34-913976611
jberrocal@es.ibm.com

Thank
YOU

The image features the words "Thank YOU" in a large, 3D, light blue font. Each letter of the word "Thank" contains a different person's face. The "T" shows a man in a white shirt and orange tie. The "h" shows a woman in a green top. The "a" shows a man with a green face. The "n" shows a woman in a blue top. The "k" shows a man with glasses in a blue top. The word "YOU" is positioned below "Thank". The "Y" shows a man in a blue shirt looking at a document. The "O" shows a man in an orange shirt. The "U" shows a woman in a green top.