



Simposio de Informática en el Estado

Categoría: Monografía

La Agencia de Recaudación de la Provincia de Buenos Aires hacia un Gobierno Cognitivo

Autor: Mg. Sandra D'Agostino¹

Introducción:

Tradicionalmente la computación analiza datos estructurados existentes en formularios, con datos formateados, como CUIT, nombre, dirección, teléfono, almacenados en base de datos relacionales y posteriormente utilizados mediante aplicativos de la organización para la toma de decisión.

En este sentido nos encontramos que la cantidad de información disponible en la Agencia de Recaudación de la Provincia de Buenos Aires (en adelante ARBA) se multiplica, existen bases de datos con información estructurada que data de la década del 70 y contiene millones de registros vinculados a los millones de contribuyentes, lo que nos lleva en la actualidad a 170 Tera bytes de información estructurada.

Pero también se observa que la información no estructurada tiene un crecimiento importante y las fuentes de esa información relevante para la Agencia se diversifican. Ya no reside solamente en las bases de datos de los sistemas propios o de bases de datos de otros organismos públicos. Esa información está diseminada en registros de transacción de decenas de sistemas (incluyendo aquellos que se conectan a sensores, maquinaria y electrodomésticos, en lo que se dio en llamar la Internet de las Cosas), redes sociales, papeles, videos, incluso en el pronóstico del tiempo o el monitoreo del tránsito.

Para la Administración Tributaria, la disponibilidad de información oportuna y confiable es indispensable para lograr sus objetivos. Los sistemas de información tributaria implementan estrategias de recolección, tratamiento y distribución de la información requerida, y son uno de los factores condicionantes para el éxito de una administración.

¹ Gerente General de Tecnología e Innovación – Sub Dirección Tecnológica y Administrativa ARBA-
Profesora Facultad de Informática - UNLP



“ ... La cantidad de información que estamos produciendo nosotros (las personas) es mucho mayor que nuestra capacidad de consumirla..” [1]

El 80% de la información que circula en los sistemas es no estructurada" afirmó el presidente de Cognitiva, Roberto Cruz. [2]

Esto significa que "de cada cinco datos que hay en los sistemas, cuatro no están estructurados y están en lenguaje natural" desde las conversaciones por chat, las imágenes en los muros de las redes sociales, hasta las compras -sean electrónicas o no- y las búsquedas, entre otras opciones.

El uso masivo de dispositivos digitales alrededor del mundo ha permitido la creación exponencial de este tipo de datos no estructurados. A la fecha existe un trillón de dispositivos digitales conectados en el planeta. A manera de ejemplo, el mundo genera 2,5 billones de gigabytes de datos cada día y el 80 por ciento de esos datos corresponde a información no estructurada. [3]

Se estima que para el año 2020, un 93% de la información no será estructurada. Como ejemplo, la cantidad de publicaciones científicas en medicina se duplica cada 1,5 años. En el 2020, se estima que esa información se va a duplicar cada 73 días.

También es importante destacar que en los últimos años, el uso de Internet ha experimentado un crecimiento muy elevado, ampliándose los mercados tradicionales al ámbito virtual. Esta circunstancia se ha traducido en nuevos comportamientos y se ha caracterizados por un gran volumen de datos que deben ser convertidos rápidamente en información útil para las organizaciones.

Las neuronas y el cerebro han constituido siempre un misterio para los investigadores. El geólogo y paleontólogo Henry Fairfield Osborn [4] decía que “el cerebro humano es el objeto más maravilloso y misterioso de todo el universo”. Quizás la intriga que suscita entre los científicos el funcionamiento de este órgano haya provocado que otras ramas de la tecnología busquen en la neurociencia inspiración para su desarrollo. Este es el caso de la informática de tercera generación, que como acuñó Wang en 2002, se sirve del estudio del cerebro, para crear la llamada “computación cognitiva”.



La computación cognitiva es un aspecto de la inteligencia artificial que involucra actividades de razonamiento, pensamiento, entendimiento, interpretación, aprendizaje, autoaprendizaje y adaptación, entre otras. La computación cognitiva está cambiando la forma como se interactúa con las aplicaciones [5]

La computación cognitiva analiza la información estructurada y no estructurada disponible en los sistemas y el paso evolutivo es que da respuestas y dialoga con quienes la utilizan.

Relevancia para el Interés Público: La importancia de este trabajo es para los organismos públicos que procuran mejorar la interacción con los ciudadanos a partir del uso de la información no estructurada, que ingresa a los organismos mediante redes sociales, call center y otros medios y la misma está expresada en lenguaje natural. Este trabajo pretende dar un enfoque sobre el cambio de gestión que deberán realizar las administraciones para poder medir efectivamente que es lo que ocurre con un sitio web de forma integral, analizando todos sus aspectos más relevantes: recursos y procedencia del tráfico, experiencia del usuario, patrones de navegación, puntos de entrada y fuga, modelos de atribución, medición real time de analytics, configuración de eventos, audiencias,

Análisis y Redacción:

Mediante la Ley N° 13.766 se creó la Agencia de Recaudación de la Provincia de Buenos Aires (Arba), un organismo técnico especializado, dotado de las herramientas necesarias para ejecutar la política tributaria a través de la aplicación, recaudación y fiscalización de los tributos y accesorios dispuestos por las normas legales y la administración del catastro territorial de la provincia de Buenos Aires, entre otras acciones. Los pilares sobre los cuales se crea la Agencia son los siguientes:

Misión : Ejecutar eficientemente la política tributaria mediante la determinación, fiscalización, y percepción de los tributos y accesorios dispuestos por las normas legales, administrar el catastro territorial y brindar servicios a otros organismos públicos, promoviendo el cumplimiento voluntario y contribuyendo al bienestar económico y social de la ciudadanía .

Visión: Ser un organismo impecable en la calidad de los servicios y en la administración de las relaciones con el ciudadano necesarias para cumplir con las metas de recaudación, conformado por un equipo profesional y comprometido que actúa en un marco de integridad y de permanente búsqueda de la innovación mediante el uso de tecnología de avanzada.

Estos pilares se sustentan en los siguientes valores:

- Servicio al ciudadano
- Transparencia y no discrecionalidad
- Eficacia y eficiencia en la operación diaria
- Participación y compromiso del personal y de la gestión
- Ética

Y alineados absolutamente con estos y con sus lineamientos estratégicos:

- Optimizar la recaudación
- Optimizar el sistema de servicios
- Incrementar la percepción de riesgo
- Fomentar el fortalecimiento institucional

Los impuestos y contribuciones que administra son los siguientes:

- Ingresos Brutos
- Inmobiliario
- Automotor
- Embarcaciones Deportivas
- Sellos
- Planes de Regularización de Deudas
- Contribución Provincial de Energía
- Contribución Fondo Provincial de la Vivienda
- Contribución Fondo Provincial Educación

Es de suma importancia incorporar tecnología que simplifique el pago de los impuestos y dinamice los trámites que realizan los contribuyentes. Acercar buenos servicios a la gente y ahorrarle tiempo termina favoreciendo un mejor cumplimiento tributario. Por este motivo desde la existencia de la Subsecretaría de Ingresos Públicos al comienzo y la creación de la Agencia después, el uso de las TICs se fue incorporando en distintas fases para cumplir con el objetivo.

Si bien el uso de las TICs fue cambiando la forma de interactuar entre la Agencia y el contribuyente para facilitar los trámites destinados al cumplimiento tributarios, nos encontramos en la actualidad



con una gran cantidad de información, no estructurada, proveniente de distintas fuentes que resulta imposible utilizarla para la toma de decisiones.

En la actualidad nos encontramos que la agencia:

- Recibe 110.000 llamadas mensuales al Call Center
- Tiene 15.000 usuarios en Facebook
- 15.600 seguidores en Twitter

La información recibida por estos medios es de suma importancia para poder solucionar en tiempo y forma los reclamos e inquietudes de los contribuyentes, y se observa un crecimiento constante día a día.

Nos encontramos entonces con la imposibilidad de procesar dado que la misma no está en una base de datos con información estructurada.

En la Fig.1 podemos observar la evolución del descubrimiento de datos a partir de la computación cognitiva.

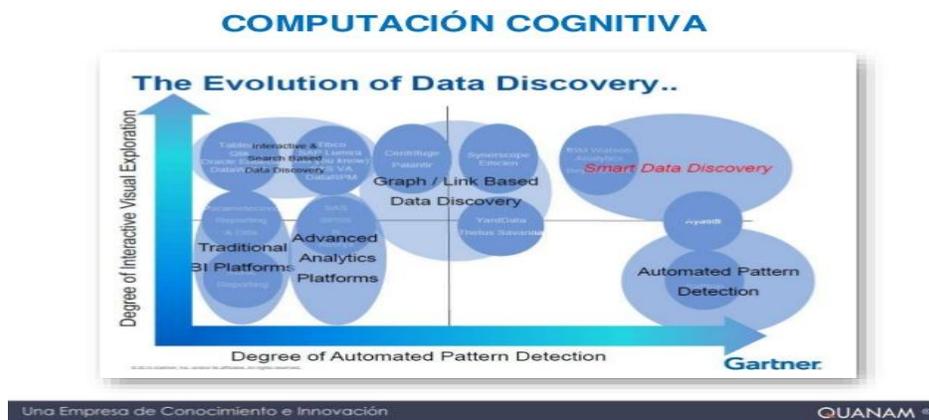


Fig. 1 -

Fuente :Quantam

- El objetivo de la computación cognitiva [7] es simular procesos de pensamiento humano en un modelo computarizado. Utilizando algoritmos de auto aprendizaje que utilizan la minería de datos, reconocimiento de patrones y procesamiento del lenguaje natural, la computadora puede imitar la forma en que funciona el cerebro humano.



Los sistemas cognitivos consideran conjuntos de datos complejos y desarrollan entendimiento, razonamiento y aprendizaje contextualizado. La meta de los sistemas cognitivos es iluminar aspectos del mundo que han sido invisibles previamente, reconociendo patrones en los datos no estructurados y posibilitar la toma de decisiones basada en la inferencia de nuevos conocimientos. El verdadero potencial de la era cognitiva se logrará mediante la combinación del análisis de datos y el razonamiento estadístico de máquinas con singulares cualidades humanas, tales como metas auto dirigidas, sentido común y valores éticos[6] . Los sistemas de esta era rompen la barrera de la interacción de eras anteriores. La interacción con los sistemas cognitivos se vuelve natural y su objetivo es aumentar las capacidades humanas.

Entonces, encontramos una solución posible dentro de la Agencia para poder dar una respuesta a este desafío, la computación cognitiva, lo que nos permitirá dar paso a una nueva generación de gobierno: **Gobierno Cognitivo**.



En esta nueva etapa ARBA deberá cambiar su interacción con el contribuyente donde deberá:

- **Percibir**, las demandas de un ciudadano/contribuyente cada vez más familiarizado con el uso de las TICs
- **Razonar**, sobre las demandas no satisfechas e internamente analizar las barreras que nos impiden avanzar
- **Actuar**, para producir las mejoras necesarias
- **Adaptar**, las herramientas y la comunicación utilizando computación cognitiva, llevando la atención hacia el contribuyente de una manera personalizada.

Esto nos permitirá generar:

- **Autoservicio** (Acceso en todos lados, todo el tiempo)
- **Agente Asistido** (Respuesta en todos lados, todo el tiempo)
- **Personalizado** (Guía en todos lados, todo el tiempo).

El primer paso dado por la Agencia de Recaudación en este aspecto fue la investigación de las herramientas de inteligencia artificial en el mercado en torno a la posibilidad de implementar un Agente Virtual o Chat Bot .

Un agente virtual es un programa de software que interpreta y responde a las declaraciones realizadas por los usuarios en lenguaje natural corriente. Integra técnicas de lingüística computacional y la comunicación se puede establecer a través de Internet, Mensajes instantáneos, e-mail, foros, etc.

El término Agente Virtual también puede ser definido como: “agente conversacional”, que puede ser analizado por las palabras que lo componen:

La palabra “agente” es un sistema software, el cual, viene definido por su flexibilidad, entendiendo por flexible que un agente sea:

- Reactivo, responda al entorno en que se encuentra.
- Proactivo, que sea capaz de intentar cumplir sus propios objetivos.
- Social, sea capaz de comunicarse con otros agentes mediante algún tipo de lenguaje

La palabra conversacional significa: Perteneciente o relativo a la conversación, entonces un agente conversacional es una entidad artificial diseñada para tener conversaciones con seres humanos u otros agentes conversacionales.

En la actualidad existen muchos chatbots con variadas características, utilidades, complejidad y capacidad de respuesta. Podemos encontrar desde sencillos chatbots pensados como juego y diversión, que pretenden responder de forma más o menos cómica, hasta otros muy complejos y técnicos en algún campo que tienen un uso práctico elevado o simplemente que han sido construidos con fines comerciales.

Las herramientas analizadas fueron:

- ✓ AgentBot (AIVO) [10]
- ✓ BotGenes EQUIFAX [11]
- ✓ Watson (IBM) [8]

Los parámetros que se tuvieron en cuenta para la comparación de las herramientas fueron las siguientes:

- 1) Canales de atención en los que se puede implementar
- 2) Características funcionales de la herramienta
- 3) Características Técnicas

Esta investigación dio como resultado los siguientes cuadros comparativos:

Cuadro 1 - Canales de atención en los que se puede implementar

| | AgentBot | BotGenes | Watson |
|--|---|---|---|
| Web ARBA | Si | Si | Si |
| Chat ARBA | Si | Si | Si |
| Aplicaciones específicas (SIC, SURyC, COT, etc). | ARBA debe disponer de una API particular. | Si, se integra fácilmente, a través de webservices, | Si, se puede integrar a través del desarrollo de web Services |
| Facebook Messenger | Si | Si | Si |
| WhatsApp | Se puede pero todavía no tiene un canal corporativo | Se puede pero todavía no tiene un canal corporativo | Se puede pero todavía no tiene un canal corporativo |
| Chat de Mesa de Ayuda | Si | Si | Si (habría que entrenarlo en esas cuestiones particulares) |
| Intranet | Si | Si | Si (habría que entrenarlo en esas cuestiones particulares) |
| Call Center | Si, pero responde | Si (con el chatbot de la página de | Se podría hacer con las API de |



| | | | |
|---------------|---|---|---|
| | mediante grabaciones pre armadas. | botgenes se puede entablar una conversación mediante el habla) | Texto a Voz y Voz a Texto |
| Otros canales | Telegram, Skype, android, ios,webmobile | Se puede integrar con cualquier canal mediante webservices. Además provee la posibilidad de utilizar validadores de identidad sobre la base del Veraz y utilizar el experto de scoring que toma información de la misma base para brindar información de la situación crediticia de cada contribuyente. | Se puede integrar a todas las herramientas de IBM y tiene un gran potencial para cualquier cuestión en la que se lo desee entrenar, pero requiere de mucho tiempo de desarrollo/ implementación y/o entrenamiento |

Cuadro 2 – Características funcionales de la herramienta

| | AgentBot | BotGenes | Watson |
|--|--|--|--|
| Texto a Voz: ¿Permite dar respuesta en formato de audio? | Si | Si, capacidad de verbalizar la respuesta via TTS, síntesis de voz | Si, implementando la API específica para ello |
| Voz a Texto: ¿Permite procesar conversaciones ingresadas en formato no escrito? | No (Se encuentra en desarrollo en una versión Beta) | Si, capacidad para reconocer la voz humana, ASR (authomatic speech recognition) | Si, implementando la API específica para ello |
| Responsivo: ¿Se adapta a los distintos dispositivos desde los que se pueda acceder? | Si | Si | Si |
| Accesible: ¿Puede ser utilizado por personas con discapacidades? | Se puede desarrollar, habría que definir el alcance. | Si, y tiene un Bot ya implementado para un centro de atención de gente con discapacidad. | Se puede mediante la integración de las distintas APIs |
| Customizable: ¿Se puede personalizar la imagen que se muestra como "Avatar"? | Se puede personalizar en base a una serie de avatar predefinidos | Si, en forma animada e incluso se puede personalizar a alguien real | A implementar con Banco Santander (Gala) |
| Realismo: ¿Permite seguir el hilo de una conversación, como si se estuviese hablando con un | Si, pero en todo momento se sabe que se | Si, excelente capacidad para poder recordar durante toda | Si, pero requiere de entrenamiento o consultoría. |



| | | | |
|--|--|--|--|
| humano? | está interactuando con un chatbot. | la conversación y poder contextualizar la respuesta. Establece un dialogo muy natural con el contribuyente. | |
| Adaptabilidad: ¿Reconoce los distintos modismos o errores de escritura que pudieran utilizarse al ingresar las preguntas? | Si | Si, tiene un diccionario común ya desarrollado que reconoce errores ortográficos, modismos, acentos, etc. | Si, pero requiere de entrenamiento o consultoría. |
| Análisis Estadístico: ¿Incluye herramientas que permitan realizar análisis de los diálogos y de tendencias en los cambios de comportamiento? | Si. | Si, permite armar distintos tipos de reportes exportables en formatos que pueden ser analizados con cualquier herramienta. | A través de la API Discovery permite identificar patrones, tendencias y conocimientos útiles que permitan una mejor toma de decisiones. También tiene la API Natural Language Understanding para analizar texto y para extraer metadatos de contenido. |
| Reconocimiento de imágenes y/o archivos: ¿Se pueden subir archivos para cargar la información? | Se pueden subir archivos en formato texto | Si, enlaces, archivos, imágenes, hay que configurarlo previamente. | Cuenta con APIs específicas que permiten reconocer imágenes y archivos en distintos formatos. |
| Capacitaciones: ¿Incluye capacitaciones para el personal que deba llevar adelante la carga de contenido y para aquellos que quieran hacer uso de la información generada? | Si, están incluidas en la Propuesta comercial. | Si, cuenta con capacitaciones para el equipo implementador y para el personal que vaya a encargarse de la supervisión y de la carga de contenidos. | Tiene soporte a través de consultoría. |
| Soporte: ¿Existen canales de asistencia para evacuar cualquier duda que pueda surgir en el uso de la herramienta una vez implementada? | Si | - Manuales de procedimiento y - mail dentro de 24 hs. Contacto directo con representantes las 24 hs ante una EMERGENCIA | Si. |
| Capacidad de poder llegar a integrarse con cualquier tipo de servicios cognitivos | No | Face detection, Face Recognition, Emotional Depth, | Si, se puede integrar a través de los distintos |

| | | | |
|-----------------------------------|---|---|--|
| | | Emotions present, SDK, Ethnicity”, etc) de las siguientes empresas: IBM, Amazon, Cognitec, Microsoft, Google, OpenCV, NEC, Kairos y Face | servicios ofrecidos por IBM en Bluemix y con cualquier otra herramienta de IBM. |
| Otras características funcionales | Permite ir cambiando la URL de la Web para entregar respuestas mas compeltas, asistiendo al usuario en la resolución de sus tramites. | Permite hablar de distintos temas sin perder el hilo de la conversación, manteniendo en todo momento la sensación de estar comunicándose con un humano. | Infinidad de funcionalidades a través de las distintas APIs previamente explicadas en el apartado que de análisis de la herramienta. |

Cuadro 2 – Características técnicas

| | AgentBot | BotGenes | Watson |
|---|---|---|---|
| Procesamiento de Lenguaje Natural | Si | Si | Si |
| Machine Learning | Si, pero limitado. | Si | Si |
| Lenguajes de programación. | JAVA | Conocimiento almacenado en lenguaje AIML estándar y open source | Java y otros |
| Capacidad de Integración con Redes Sociales. | Si | Si | Si, cuenta con APIs específicas que permiten reconocer el estado de ánimo de los usuarios en función de tendencias de lo que se publica en las redes sociales |
| Capacidad de integración con Aplicaciones. | Si | Si | Si |
| Propiedad de la información que se construye. | Propiedad del cliente, confidencial | Propiedad del cliente, confidencial | Propiedad del cliente, confidencial |
| Despliegue y operación (SaaS, PaaS, On Premise). | SaaS | SaaS, PaaS, On Premise | SaaS |
| Posibilidad de autenticación en base a preguntas. | No. Se podría hacer siempre y cuando le proveamos | Si. Permite la Integración con servicios de autenticación de | Si (siempre y cuando le proveamos la base de datos) |



| | | | |
|--|--------------------------------------|--|---|
| | las bases de datos. | identidad con Equifax/Veraz | Va a enviar ejemplo de Bco en Perú. |
| Cumplimiento de estándares de seguridad internacionales. | SI (Amazon PSI) | Si, con respaldo del Veraz que cumple los más altos estándares de seguridad por las propias normas que se le imponen. | Los servicios en la nube |
| Escalabilidad a otros proyectos. | | Se puede integrar con cualquier tipo de plataforma, con validadores, repositorios, analíticos de EQUIFAX. Puede extraer datos de cualquier base de datos, de cualquier servicio cognitivo. - Proximamente podrá dar avisos por ejemplo campañas en texto, imágenes videos. (integración con google adsense) | Si, permite escalarse en todos los ámbitos relacionados con servicios cognitivos, con cuestiones de ciberseguridad, big data y cualquier otro tema, debido a que la empresa que lo provee es líder en el mercado mundial de todas estas cuestiones. |
| Otras características técnicas | Adjunta documentación SLA y Security | Mantenimiento / aumento permanente de las reglas lingüísticas para maximizar el desempeño del agente. | Permite la escalabilidad a infinidad de temas a través de las APIs de Bluemix y de los distintos servicios que ofrece IBM. |

Conclusión y Sugerencias.

La cantidad de información que se produce de manera permanente es mucho mayor a la capacidad de consumirla. Un alto porcentaje de esta información es no estructurada y esencial para las organizaciones que deben utilizarla para mejorar los servicios, lograr sus objetivos de manera más eficiente y tomar decisiones en el momento oportuno.

Los organismos públicos no están ajenos a esta situación y su deber es mejorar la relación con los ciudadanos y en el caso particular de ARBA con sus contribuyentes.

Hoy la industria ofrece una respuesta a las potencialidades de Big Data y Analytics como el caso de IBM [8] con Watson, AgentBoT de AVIO [10] y Botgenes – EQUIFAX [11], otras ofrecen productos que se acercan un poco más a las capacidades cognitivas humanas- tanto en lo intelectual como en lo social como la empresa de tecnología IPSoft donde desarrollaron una plataforma de inteligencia artificial para la atención de call center. También encontramos en el mercado muchas herramientas que nos posibilitan análisis y medición del tráfico de un sitio web y la toma de decisiones en función de dichos datos.

En esta era cognitiva que comenzó, cada organización tanto pública como privada puede distinguirse a partir de generar capacidades y modelos de negocio cognitivos que mejoran sustancialmente todo proceso que involucre la toma de decisiones y comprender con más exactitud qué está pasando [9]. Entendemos que la única forma de encontrar valor y darle sentido a la inmensa cantidad de información será a través de la computación cognitiva.

Entendemos que transformar la gestión hacia un Gobierno Cognitivo debe ser el próximo objetivo a cumplir en ARBA, la tecnología ya está disponible, sólo hay que comenzar a usarla.



Referencias:

- [1] <http://www.sophimania.pe/tecnologia/redes-sociales/watson-y-la-nueva-era-de-computacion-cognitiva/>
- [2] Roberto Cruz, Presidente de Cognitiva - IBM
- [3] Rometty, M. Virginia. IBM Annual Report: "What will we make of this moment?" 2013.
- [4] https://es.wikipedia.org/wiki/Henry_Fairfield_Osborn
- [5] Computación Cognitiva - Luis Javier Chavarría Sánchez -coordinador del Área Académica de Administración de Tecnologías de Información del Instituto Tecnológico de Costa Rica
- [6] John E. Kelly III. Computing, cognition and the future of knowing. How humans and machines are forging a new age of understanding. 2015
- [7] COMPUTACIÓN Y CONOCIMIENTO. Hacia una fundamentación de la ciencia cognitiva. Zenon W PYLYSHYN
- [8] <http://blogthinkbig.com/beta-publica-de-watson/>
- [9] What Everyone Should Know About Cognitive Computing - Bernard Marr: Founder and CEO, Advanced Performance Institute
- [10] <http://aivo.co/agentbot/>
- [11] <http://www.botgenes.com/BotGenes/>

Bibliografía:

- Christense, C., Antony, S., and Roth, E. Seeing What's Next: Using the theories of innovation to predict industry change. HBS Press Book. 2004.
- John E. Kelly III. Computing, cognition and the future of knowing. How humans and machines are forging a new age of understanding. 2015
- Computación Cognitiva - Luis Javier Chavarría Sánchez - coordinador del Área Académica de Administración de Tecnologías de Información del Instituto Tecnológico de Costa Rica
- Roberto Cruz, Presidente de Cognitiva - IBM
- Rometty, M. Virginia. IBM Annual Report: What will we make of this moment? 2013.