

# INCRUSTACIÓN DE DATOS: LO MÁS MODERNO DE LAS EMPRESAS DE TECNOLOGÍA SEMÁNTICA

*Maddalena della Volpe\**

*Università degli Studi Suor Orsola Benincasa*

*Francesca Esposito\*\**

*Università degli Studi di Salerno*

---

## Introducción

La innovación y las nuevas tecnologías de la información están cambiando la manera como las compañías producen, se gestionan y se comunican. Hoy en día, las compañías están buscando un proceso comercial más creativo y eficiente. Carbone, Contreras, Hernández y Gómez-Pérez (2012) vieron la necesidad empresarial de posicionar mejores soluciones en el mercado en menor tiempo y costo. Las empresas empiezan a reconocer que las tecnologías informáticas basadas en la analítica pueden proveer una ventaja estratégica, ayudándoles a ser más competitivas y a adaptarse a los cambios rápidos (Stephens, 2007). Las compañías necesitan integrar información, la cual contribuye al mejoramiento del proceso, los productos y los servicios, y esto no es fácil. Los datos considerados no solo están estructurados y contenidos en sistemas de almacenamiento tradicionales, sino que las compañías los desestructuran y generan específicamente (por ejemplo, correo, documentos y reportes) o la web los recolecta. Las organizaciones están interesándose cada vez más en convertir los datos en ventajas competitivas; han cambiado sus modelos comerciales tradicionales, adoptando unos que son más colaborativos e involucran una cantidad creciente de socios estratégicos, los cuales pueden contribuir

---

\* Doctora en Sociología en la Universidad de Nápoles. Profesora Asociada de Economía y Gestión de las Empresas en la Università degli Studi Suor Orsola Benincasa. Contacto: maddalena.dellavolpe@unisob.na.it.

\*\* Doctoranda en Ciencia de la Comunicación en la Università degli Studi di Salerno. Contacto: fraesposito@unisa.it.

El presente artículo es resultado de un proyecto de investigación desarrollado entre la Università degli Studi Suor Orsola Benincasa y la Università degli Studi di Salerno.

Fecha de recepción: 10 de agosto de 2016; fecha de aceptación: 1 de septiembre de 2016.



al proceso de innovación. Dirigir empresas a un enfoque cognitivo mundial es la herramienta principal para enfrentar los retos de innovación y competir en los mercados internacionales. Además, este enfoque debe involucrar a toda la empresa, que trata de entender cómo, al compartir la tecnología y el conocimiento y los datos de recursos humanos a través de todas las unidades organizacionales, logra las metas analíticas generales. Para las empresas emergentes innovadoras y las empresas pioneras, es necesario coordinar la recolección de datos en todos los niveles de la compañía (Davenport, 2014).

Los costos de la infraestructura requerida son altos y el tiempo dedicado por el experto está justificado. En casos donde se requiere una rápida respuesta y distintos departamentos proveen datos en formatos semiestructurados, se necesita una solución más ágil (Casas-Bayona & Ceballos, 2014). La competencia por la recuperación y el análisis de datos estratégicos está aumentando la tendencia de subcontratar las actividades de análisis de datos, permitiéndole al mercado de tecnología semántica definirse mediante la explosión de STE.

Este documento provee un esquema de las STE alrededor del mundo que ofrecen minería de textos (MT) o descubrimiento de conocimiento en textos (KDT, por sus siglas en inglés). Por este motivo, se consultaron páginas web de compañías para entender mejor las características que este tipo de empresas comparten, operando en distintos sectores. Después, se creó un conjunto de datos para observar el enfoque del mercado de las STE, identificando los sectores y las aplicaciones, desarrollos y estrategias.

## **Tecnologías semánticas para entender la empresa**

Díaz, Aguilar y García (2009) notan cómo las nuevas generaciones han entendido que para crear conocimiento se debe ir paso a paso junto al crecimiento tecnológico y la evolución de la web. Las estructuras organizacionales están soportando el entorno colaborativo, cambiando desde sus sistemas informáticos hasta la forma en la que operan todas las acciones relacionadas con los mismos. La Web 3.0 y la Web 4.0 son las propuestas para el futuro, la red les ha dado más significado y más semántica (p. 244). Dominigue, Fensel y Hendler (2011) definen la *semántica* como aquellas tecnologías que proveen descripciones de datos, programas e infraestructura comprensibles por computadores (o mejor procesables por computadores), permitiéndoles a los computadores entender dichos artefactos (p. 11). La web semántica, como



Tim Berners-Lee (2001) lo imaginó, sigue parcialmente sin cumplirse. Sin embargo, en los últimos años, muchos avances han surgido en el desarrollo de herramientas que soportan la infraestructura semántica. Las tecnologías centrales han sido identificadas y diseñadas, proveyendo algunos entornos de *software* para ser adoptados por compañías; un usuario que use la web semántica sigue siendo una utopía, las herramientas siguen siendo muy sofisticadas para servirle al público y requieren un soporte por parte de profesionales especializados. La web semántica sigue teniendo un largo camino por recorrer para lograr un buen nivel de adopción generalizada por parte de los usuarios. Sin embargo, la investigación en este campo ha producido resultados notables hasta la fecha, a menudo creando herramientas valiosas y adoptables, en especial dentro de las compañías. La meta de los desarrolladores de tecnologías semánticas es crear un lenguaje, un código universal, que pueda expresar y combinar los datos presentes en la web para crear nuevos conocimientos.

Las tecnologías semánticas no son una idea reciente, aunque la atención que las compañías, las instituciones y la sociedad están prestando a las oportunidades que podrían surgir del análisis de estos datos no es un lado conocido de la historia. IBM ([www.ibm.co](http://www.ibm.co)) establece que cada día creamos 2,5 trillones de *bytes* de datos, estos datos no estructurados no pueden ser ignorados. Las compañías sienten cada vez más una fuerte necesidad de adoptar sistemas de gestión de conocimiento que integren los datos internos de sistemas de almacenamiento tradicionales y datos externos que pueden ser recuperados mediante el análisis de los recursos que la web ofrece.

Los negocios desempeñan un papel clave en este escenario, ya que usualmente están buscando nueva información para utilizarla como un recurso estratégico. Chaudhuri, Dayal y Narasayya (2011) sugieren que es difícil encontrar una empresa exitosa que no haya aprovechado la tecnología de inteligencia comercial para conducir sus negocios (p. 88). Sin embargo, una razón que está siendo crucial para los negocios es la capacidad de las compañías para monetizar los datos vendiéndoselos a otras compañías (Gartner, 2015).

Una tendencia bastante interesante se está desarrollando hoy en día: por una parte, las compañías que tienen productos de *software* innovan y transforman sus sistemas, ofreciendo soluciones de análisis de datos fuera de línea y en línea; por otra parte, se está presenciando la creación de varias empresas emergentes innovadoras, las cuales están incursionando en los mercados de la analítica. Las tecnologías semánticas pueden ser usadas para integrar conjuntos de datos



heterogéneos y formalizar la estructura subyacente de la información, para permitirle a un computador entender las semánticas de la misma (Shadbolt, Berners-Lee & Hall, 2006). La integración de herramientas de tecnologías semánticas consiste en una verdadera revolución para todos los niveles de compañías; explica por qué la mayoría de ellas no están preparadas para este cambio disruptivo y hace referencia a otras empresas, las cuales les ofrecen productos limitados. Específicamente, implica:

1. Conversión de sistemas de almacenamiento y análisis de datos tradicionales;
2. Contratación de profesionales especializados: científicos de datos y gerentes de proyectos que diseñen la adopción y la implementación de tecnologías semánticas;
3. Tecnologías de incrustación en cada unidad empresarial.

A menudo se les dificulta a las compañías instalar estas tecnologías dentro de sus negocios. Probablemente sucede porque la alta gerencia no reconoce el valor derivado de su uso; o porque su adopción requiere esfuerzos considerables en cuanto a recursos económicos; o simplemente porque la estrategia comercial no está muy acostumbrada al cambio. La capacidad de anticipar, abordar y gestionar el cambio es la única manera de hacer que las empresas sean proactivas ante las necesidades cambiantes del mercado. El factor de la no aceptación de cambios implica el rechazo al entendimiento del mercado e ignora el equilibrio de triunfo que lleva a las empresas al fracaso seguro (Foglio, 2011).

Una alternativa para la integración de herramientas de tecnologías semánticas para el análisis y la administración de datos dentro de la compañía consiste en hacer que las compañías gestionen y adquieran un paquete acabado de información orientado a las necesidades de la compañía. Esto permite adquirir una ventaja competitiva en un menor tiempo, evitando los costos de transacción ocasionados por la incertidumbre y a menudo no recuperados mediante inversiones. De hecho, se está difundiendo una tendencia de subcontratación de actividades de análisis y recuperación de datos, lo que se traduce en una expansión del mercado de analítica en campos de aplicación específicos para ofrecer soluciones adecuadas y sofisticadas.

Un creciente número de STE está respondiendo a la necesidad crítica de gestionar e integrar una gran cantidad de fuentes de datos y la aplicación en las empresas. Por lo general, las STE son organizaciones



que han desarrollado herramientas de tecnología semántica para extraer el significado oculto de los datos. En específico, por una parte el mercado está atendiendo la conversión de sistemas de compañías antiguas que siempre compiten en el mercado mundial de las TIC; por otra parte, el surgimiento de empresas emergentes en el mercado ocurre con servicios innovadores y un alto perfil profesional. De acuerdo con el cliente de la compañía, las STE normalmente proponen dos tipos de soluciones semánticas: un *software* con licencia para integrarlo en el proceso comercial o un reporte con datos finales con respecto al análisis que fue llevado a cabo. Es probable que la segunda solución sea más amplia.

Aunque las aplicaciones son muy numerosas y específicas para cada sector, el enfoque estará en las STE que se proponen a ellas mismas como soporte para los usuarios o las compañías en cuanto al análisis de grandes cantidades de datos textuales. La creciente disponibilidad de datos de texto, como reportes, correos y documentos, trae nuevos retos para los negocios (Chaudhuri *et. al.*, 2011, p. 90). Esta cantidad de textos se está convirtiendo en un recurso valioso de información y conocimiento. La heterogeneidad de los datos y las fuentes complica el proceso de descubrimiento de conocimiento, es evidente que mucha información disponible es muy difícil de manejar. Por este motivo, se usa MT y KDT para identificar cualquier tipo de análisis textual (Feldman & Dagan, 1995; Feldman & Hirsh, 1997; Kodratoff, 1999, 2000; Loh, Wives & de Oliveira, 2000), lo cual puede ayudar a las compañías a extraer el significado oculto de los textos y también puede ayudarles a tomar mejores decisiones de manera más rápida.

### **Emprendimiento de tecnología semántica en América Latina**

En este estudio se pretende mejorar la perspectiva mundial del desarrollo y la comercialización de tecnologías semánticas en el mercado de la innovación. En específico, se aborda la descripción y el estudio profundo de sectores y campos de aplicación relacionados con las STE. Se ha omitido el uso específico del lenguaje y su relación con el mercado de referencia de cada compañía analizada, ya que se ha pospuesto para la próxima investigación.

Es costumbre, en especial en los negocios, considerar el idioma inglés como el idioma más ampliamente hablado, debido a que es la lengua franca entre los 6.900 idiomas existentes en el mundo. La utilidad de poder comunicarse en inglés en todos los mercados ha



superado con rapidez cualquier legado de orgullo nacional o tradición. Pero aunque es el segundo idioma más hablado en el mundo, el uso de cualquier otro idioma no es un problema insignificante.

Por ejemplo, el español en las transacciones comerciales y, en este caso, aplicado a las tecnologías semánticas no puede ser ignorado. De acuerdo con el reporte *Ethnologue* (2009), el español es el segundo idioma más usado en el mundo después del chino mandarín; se habla en 31 países, la mayoría en América Latina, donde muchas compañías de alta tecnología están ubicadas. La lengua española tiene aproximadamente 400 millones de hablantes nativos y casi la misma cantidad de hablantes no nativos, un dato que no puede ser subestimado por los empresarios comerciales.

Comúnmente, la actividad empresarial en sus fases tempranas es mayor en economías informales y menos competitivas, pero los empresarios raramente son ambiciosos e innovadores. Así que más negocios comienzan, pero los empresarios fallan en la creación de empleo. En el lado opuesto, los empresarios son más innovadores y ambiciosos en economías más formalizadas y competitivas, pero raramente comienzan nuevos negocios. Así que la mayoría de los 44 países analizados por el Foro Mundial Económico en colaboración con el *Global Entrepreneurship Monitor* (Monitor Mundial de Emprendimiento) (2015), si bien tienen buenas características en un campo para lograr su potencial empresarial, no poseen otras. Solo dos economías, Colombia y Chile, están en el grupo «integral», porque combinan todas las dimensiones descritas: una alta proporción de empresarios innovadores y ambiciosos con alta actividad empresarial en la etapa temprana, gracias a los fuertes planes de formulación de políticas empresariales.

Varias compañías reconocieron el papel primordial que el español ocupa en la cultura latina general en los mercados de la innovación: para adaptarse y seguir a sus clientes han expandido su entorno de análisis lingüístico con el estudio del español (por ejemplo, la compañía italiana *Expert System*) o han tenido que implementar directamente oficinas secundarias en países como Colombia o México. El emprendimiento ha demostrado ser un motor del crecimiento económico, la innovación y la creación de empleo estable y robusto durante las últimas décadas. Varias empresas demuestran el alto potencial y la alta calidad de las empresas emergentes de Colombia y de América Latina, así como también han facilitado su desarrollo y su proyección internacional. Además, el mercado latino es tan amplio y está tan lleno de oportunidades que varias compañías eligen en particular esos territorios para expandir sus negocios. Por ejemplo, Séntisis ([www.sentisis.org](http://www.sentisis.org))



es una compañía emergente fundada en el año 2012: en un comienzo exclusivamente ubicada en España, decidió expandirse en Colombia y en México. Séntisis ofrece servicios de tecnología semántica a las compañías: mide conversaciones, interpreta el contexto y los aspectos lingüísticos para comprender la complejidad del idioma español.

Además, incluso el mundo académico se ha interesado en la grandeza del mercado latinoamericano para las nuevas tecnologías. En la última década, la necesidad de invertir creció, presentando enormes oportunidades en cuanto a los recursos humanos y el avance de tecnologías semánticas. En Colombia, las universidades están financiando proyectos basados en las aplicaciones de la tecnología semántica, alineados a la perfección con los sectores del Gobierno colombiano con respecto a la interoperabilidad entre las instituciones y las empresas para promover el desarrollo de las tecnologías de la información y la comunicación, de acuerdo con un enfoque más colaborativo y eficiente (Ministerio de las Tecnologías de la Información y las Comunicaciones, 2010).

De hecho, las compañías en estos países comenzaron a desarrollar elementos claves para la gestión del conocimiento; aún más, en la formación de capital intelectual, como también en el uso de herramientas de *software*, pero les falta integración para identificar, generar, retener, compartir y aplicar el conocimiento (Echeverry, Trujillo y Giraldo, 2013).

## Explorando el fenómeno: sectores y aplicaciones

Como se mencionó anteriormente, varias aplicaciones se desarrollan en este campo, con herramientas específicas para cada uso. Las STE han expandido los sectores tradicionales con aplicaciones innovadoras. Las STE pertenecen a los sectores enumerados a continuación, los cuales por lo general ofrecen servicios de descubrimiento de conocimiento en textos.

*Ciencias de la Vida y Asistencia Médica.* Las organizaciones de asistencia médica están usando cada vez más la analítica de datos y están produciendo una gran cantidad de datos (Raghupathi, 2010). Las aplicaciones promueven un mayor cuidado del paciente y apoyan la investigación. Algunos emprendimientos involucrados en este tipo de análisis incluyen registros médicos, exámenes y demás tipos de registros médicos para entender, por ejemplo, cuál es la evolución de una enfermedad específica, el cuidado apropiado y para verificar las





correlaciones con otras enfermedades. El análisis de los datos confidenciales de cada paciente puede llevar al descubrimiento de una predisposición a una enfermedad y prevenir su evolución. Este tipo de servicio se ofrece, en lo usual, a centros de investigación, de diagnóstico y centros clínicos, o a los pacientes mismos.

*Mercadeo y Comunicaciones.* Esta actividad está normalmente dirigida a las compañías, el punto fundamental es conocer el mercado y entender las necesidades de sus clientes actuales y potenciales. Las herramientas que analizan los datos de las redes sociales enriquecen el incremento analítico, estudiando las conversaciones y las opiniones de los usuarios para personalizar cada vez más la oferta de productos y servicios. Por ejemplo, el *análisis de sentimientos* es una de las herramientas más utilizadas para obtener retroalimentación de evaluación sobre los productos y servicios. Esta aplicación define las opiniones de los usuarios con respecto a un tema en particular. Se extrae el significado de la expresión usada por el usuario con la ayuda de MT y el procesamiento de lenguajes naturales (PLN), con el fin de aprender acerca de las opiniones, las evaluaciones, los estados emocionales y las intenciones. La clasificación de una revisión se predice por la orientación semántica promedio de las frases en la revisión que contienen adjetivos o adverbios. Una frase tiene una orientación semántica positiva cuando tiene buena asociación, y una orientación semántica negativa, cuando tiene malas asociaciones (Turney, 2002).

*Integración de la Tecnología.* Esta categoría incluye todas las aplicaciones de *software* basadas en tecnología semántica, las cuales pretenden mejorar y automatizar aún más el proceso de extracción de significado. Grandes compañías, como IBM, Oracle y Microsoft, que trabajan en la tecnología de la información en general han desarrollado servicios semánticos como una forma de mejorar los resultados de investigación de la web.

*Información, Medios de Comunicación y Entretenimiento.* Muchas compañías como editoriales, agencias de prensa y análisis de los medios extraen dicha información de varias fuentes internas y externas en tiempo real y después proveen a los usuarios información más detallada. Burger (2008) observó que también en la industria mediática existe una gran demanda de la reutilización de contenido. Sin embargo, varios objetos multimedia son creados de la nada, debido a las capacidades insuficientes de reutilización de las herramientas existentes.

*Servicios Jurídicos.* Incluye el análisis de grandes cantidades de documentos relacionados con procesos legales, en efecto, la creación de un sistema automatizado de extracción de significado reduciría el





tiempo requerido para revisar los expedientes de casos. El *análisis de patentes* ha probado ser una herramienta poderosa para el monitoreo de negocios y tecnologías. En este aspecto, algunos negocios analizaron directamente textos de patentes, llenos no solo de terminología jurídica, para verificar la disponibilidad de mandatos judiciales tecnológicos y para identificar otras compañías que operan en el mismo sector. La traducción de datos de patente en inteligencia competitiva les permite a las compañías calibrar su competitividad técnica actual, pronosticar tendencias tecnológicas y hacer planes en cuanto a la competencia potencial, basándose en las nuevas tecnologías (Fleisher y Bensoussan, 2002).

*Seguridad y Seguros.* Las tecnologías semánticas se aplican a este sector para obtener una mejor gestión de riesgos debido a la transparencia de los instrumentos. El *análisis predictivo* se ha convertido con rapidez en una de las mejores prácticas de la industria de seguros. Las aseguradoras usan técnicas analíticas predictivas para dirigirse a clientes potenciales, determinar precios más precisos y para identificar reclamaciones de potencial fraudulento (Nyce, 2007). Además, los Gobiernos se valen de herramientas como la *analítica predictiva* para estudiar los hechos actuales e históricos y así predecir eventos futuros y garantizar la seguridad nacional: fenómenos políticos y sociales, rastrear actividad terrorista y riesgos de bienestar.

*Fabricación, Logística y Servicios.* El análisis de varios recursos, como documentos comerciales, reportes, planes de negocios, le ayuda a las compañías a agilizar el proceso de toma de decisiones y planeación estratégica. Aunque una gran cantidad de datos constituye un recurso invaluable de conocimiento, no todo el conocimiento es útil, sobre todo el conocimiento recuperado de la web. Cualquier compañía tiene que construir su propia gestión de conocimiento con base en su sistema de necesidades comerciales; las herramientas MT reducen el tiempo de búsqueda, emparejándola a la eficiencia de recursos de interés. Todo puede ser compartido, pero no todo el conocimiento es útil para los procesos comerciales.

*Gestión de las Relaciones con los Clientes.* Las soluciones de *gestión de clientes* le permiten monitorear y analizar el comportamiento de los usuarios y a veces identificar las acciones óptimas para interactuar de mejor manera con los clientes actuales y potenciales. La experiencia de los clientes puede mejorarse mediante una mayor personalización; las necesidades, en la mayoría de los casos, no son explícitas pero se debe investigar con profundidad para entender lo que desean, cómo y cuándo.



*Educación.* La MT aplicada en el campo de la educación se refiere a aquellos sistemas utilizados para mejorar y optimizar el proceso de aprendizaje. Extraer el significado del texto, el *software*, provee una simplificación y una perspectiva general de los conceptos abordados en una publicación específica.

*Financiero y Bancario.* Los análisis MT están integrados con otros análisis estadísticos para apoyar al sector financiero, por ejemplo, las evaluaciones necesarias para otorgar una hipoteca en el caso de un banco, o para asumir una inversión en el mercado bursátil, o para prevenir una crisis financiera potencial. Se puede encontrar una aplicación reciente de todo esto en el colapso del mercado de hipotecas de alto riesgo. «Las organizaciones financieras no pudieron identificar rápidamente y cuantificar la exposición a las hipotecas de alto riesgo que pudieron haber existido en sus propias carteras» (Cataldo, 2009, p. 2).

## Metodología

La encuesta aplicada contribuye al entendimiento del nivel de difusión, las características compartidas y las mayores tendencias del mercado de las STE. El objetivo de la investigación es proveer los avances de las STE, que ofrecen MT y KDT específicamente para apoyar y ayudar a los negocios. Se descubrieron 210 compañías internacionales por su grado de visibilidad en la web: 154 se encontraron en los motores de búsqueda más grandes (Google, Yahoo) y 56 por fuentes secundarias, como artículos, foros en la web y proveedores de *software* en línea. Se realizó un censo en un período de cuatro meses: desde marzo de 2015 hasta julio de 2015. Para cada STE, se han rastreado las siguientes características: nombre de la compañía, página web de la compañía, año de fundación y fase de desarrollo, reconocimiento geográfico, aplicaciones principales, tecnología utilizada, productos o servicios ofrecidos, fuentes de datos, mercado objetivo y sector.

Posteriormente, se creó un conjunto de datos conforme a la tabla 1. El primer registro del conjunto de datos contiene el nombre de las STE identificadas. El segundo y el tercero están interconectados: identifican el año de fundación y, en lo posterior, la fase de desarrollo de las compañías. Los registros cuatro y cinco contienen respectivamente el país y la ciudad en la cual la compañía tiene su oficina central, en particular solo la primera oficina. El sexto registro demuestra las aplicaciones y servicios principales ofrecidos por la compañía.



En el séptimo registro, se identificaron los dominios en el sector antes clasificados. El octavo campo demuestra el *software* utilizado. El noveno campo contiene la información relacionada con el mercado objetivo: B2B o B2C. El décimo registro expone las fuentes de datos: datos externos o internos suministrados por los clientes de la compañía o una combinación de ambos, la cual se denominó *datos combinados*. Finalmente, se indica la página web de la compañía.

**Tabla 1**  
**Extracto del conjunto de datos de las STE**

STE	Dev. phase	Foundation	Country	City
Apache Foundation	Company	1999	USA	Los Angeles
360pi	Startup	2008	Canada	Ottawa
Abbyy developers	Company	1989	Germania	Monaco
Abzooba	Startup	2010	USA	Sunnyvale
Acetic	Company	1994	Francia	Parigi
Adaptive Blue	Company	2006	USA	NY

Application	Sector	Software	Target	Data Source	Website
public sectors, open c	Healthcare	cTakes	B2B	Combined data	<a href="http://www.apa">http://www.apa</a>
assortment intelligen	Marketing	360pi online	B2B	Combined data	<a href="http://www.360">http://www.360</a>
Imaging&texting Anal	Tech Integration	Abbyy Comprenc	B2C	ED	<a href="http://www.abb">http://www.abb</a>
automated distillatio	Marketing	Xpresso	B2B	ED	<a href="http://www.abz">http://www.abz</a>
technological watch,	Tech Integration	Tropes	B2B	ED	<a href="http://www.serr">http://www.serr</a>
social tagging, sentim	Media	AdaptiveBlue	B2B	ED	<a href="http://tvttag.com">http://tvttag.com</a>

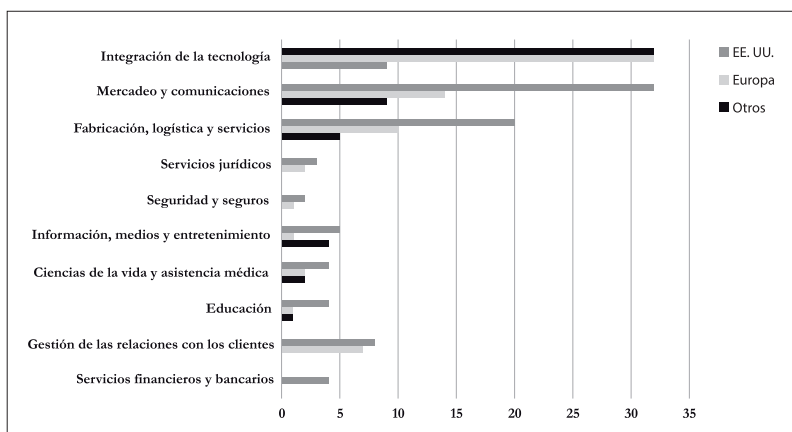
## Encuesta y hallazgos

Las STE están en 32 países, con un mayor enfoque en el Occidente que en el Oriente, en el Norte más que en el Sur del mundo. Las STE están ubicadas en los Estados Unidos (115); Gran Bretaña (17); Italia (10); Canadá (9); España (7); Alemania e Israel (5); Austria, Francia y la India (4); Suecia (3); Chile, Irlanda, Países Bajos, Portugal y Suiza (2); Australia, Bélgica, Bulgaria, China, Colombia, República Checa, Dinamarca, Egipto, Estonia, Hungría, Japón, Malta, Nueva Zelanda, Polonia, Rusia, Turquía, Uruguay (1). De la observación de datos, existe una evidente concentración más alta de STE en los Estados Unidos (54,3 %) y en Europa (31,4 %), seguida de los otros países (14,3 %).

La figura 1 reporta la cantidad de STE para cada sector por las tres áreas: Los Estados Unidos, Europa y los demás países identificados atrás. Con respecto al valor Global de cada sector: La Integración de la Tecnología (32,9 %); Mercadeo y Comunicaciones (26,2 %); Fabricación, Logística y Servicios (16,7 %); Gestión de las Relaciones con



los Clientes (7,2 %); Información, Medios de Comunicación y Entretenimiento (4,8 %); Ciencias de la Vida y Asistencia Médica (3,8 %); Educación (2,9 %); Servicios Jurídicos (2,4 %); Servicios Financieros y Bancarios (1,9 %); Seguridad y Seguros (1,4 %). Conforme se demuestra en la figura 1, los sectores más concurridos son específicamente el *Business Insider*: demuestra la importancia de las tecnologías semánticas en el proceso de innovación.



**Figura 1. Sectores de STE por distribución geográfica**

Fuente: Elaboración propia.

Enfocándose en las STE identificadas, usan aplicaciones MT y KDT para cumplir con los requisitos del sector de referencia. Estos tipos de tecnologías se pueden aplicar a las diferentes herramientas, partiendo del objetivo específico de análisis: gestión de conocimiento (KM); recuperación de información (RI); gestión de contenido (CM); inteligencia de negocios (BI); análisis predictivo (AP); análisis de sentimientos (AS); análisis de redes sociales (ARS); aprendizaje automático (AA); computación en la nube (CC); procesamiento de lenguaje natural (PLN). Globalmente, las aplicaciones más utilizadas por las STE son PLN (25,2 %), RI (14,3 %) y BI (13,8 %), mientras que los valores más bajos son los de CM (3,5 %), CC (3,8 %) y KM (4,5 %). PLN es la única aplicación presente en todos los sectores con el mayor valor; esto es claro si se considera que las herramientas MT y KDT requieren un estudio profundo del lenguaje natural como punto de arranque.

Considerando que algunos sectores no funcionan con todas las aplicaciones, en la figura 2 se pueden ver las más usadas. En los Servicios Financieros y la Banca es el AA (33,5 %), seguida por PLN (33,3 %), BI (16,6 %) y AP (16,6 %); la Gestión de las Relaciones con los Clientes



tes tiene en el primer puesto a PLN (26,7 %), después BI (20 %), el mismo porcentaje de AS y AA (13,4 %), CM (6,7 %), RI (6,6 %) y ARS (6,6 %). En cambio, el campo de la Educación está dividido entre PLN (50 %), KM (20 %), RI (10 %), AP (10 %) y ARS (10 %). Con respecto a los sectores de las Ciencias de la Vida y la Asistencia Médica, la aplicación más difundida es AP (30 %), seguida de KM (24 %), RI (12,5 %), PLN (12,5 %) y finalmente por AA y CC, las cuales tienen los menores porcentajes (10 %). La Información, los Medios de Comunicación y el Entretenimiento tienen a ARS en el primer puesto (26 %); después PLN (25 %), RI (18 %) AS (17 %) y CM (14 %). El sector de Seguridad y Seguros le apuesta a AP (45 %), PLN (30 %) y BI (25 %), mientras que los Servicios Jurídicos trabajan con RI (40 %), PLN (21 %), CM (20 %), CC (13 %) y AA (6 %). Los últimos sectores son más concurridos con varias aplicaciones. Esto denota una complejidad en el proceso de análisis que requiere una cooperación entre varias herramientas. Por lo tanto, el campo de la Fabricación, la Logística y los Servicios se aventajan en PLN (40 %), BI (34%), RI (10,3 %) AP (10,3 %), KM (6,9 %), y tanto AS como ARS (3,5 %). El Mercadeo y la Comunicación se enfocan primordialmente en ARS (21 %), después trabajan en conjunto con PLN (19 %), AA y AS (15,4 %), BI (10,7 %), AP y RI (6,2 %), CM (4,6 %) y finalmente CC (1,5 %). Por otra parte, la Integración de Tecnología usa RI (32,3 %), PLN (28 %), BI (11,3 %), AA (9,4 %), KM (7,6 %), AS y CC (3,8 %), CM y ARS (1,9 %). Para terminar, se omite el Análisis de Patente entre las aplicaciones, debido a que solo una de estas STE, la cual opera en los Servicios Jurídicos, la utiliza.

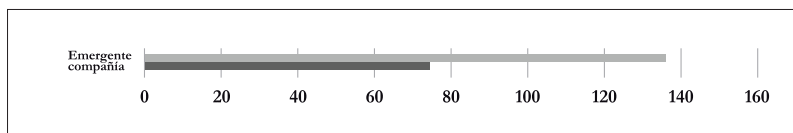
	BUSINESS INTELLIGENCE	CLOUD COMPUTING	CONTENT MANAGEMENT	RECUPERACION DE LA INFORMACION	KNOWLEDGE MANAGEMENT	APRENDIZAJE AUTOMATICO	PROCESAMIENTO DEL LENGUAJE NATURAL	ANALISIS PREDICTIVO	ANALISIS DE LOS SENTIMIENTOS	ANALISIS DE REDES SOCIALES
Servicios financieros y bancarios	■	.	.	.	.	■	■	■	.	.
Gestión de las relaciones con los clientes	■	■	■	■	.	■	■	.	■	■
Educación	.	.	.	■	■	.	■	■	.	■
Ciencias de la vida y asistencia médica	.	■	.	■	■	■	■	■	.	.
Información, Medios y entretenimiento	.	.	■	■	.	.	.	.	■	■
Seguridad y seguros	■	.	.	.	.	.	■	■	.	.
Servicios jurídicos	.	■	■	■	.	■	■	.	.	.
Fabricación, logística y servicios	■	.	.	■	■	.	■	■	.	■
Mercadeo y comunicaciones	■	.	■	■	.	■	■	■	■	■
Integración de la tecnología	■	■	.	■	■	■	■	.	.	.

**Figura 2. Área de aplicaciones en los sectores de las STE.**

Fuente: Elaboración propia.

Cada STE identificada opera en el mercado B2B y, como se puede ver en la figura 3, 74 STE (35 %) son empresas emergentes, empresas

en sus etapas tempranas de existencia, mientras que 136 STE (65 %) son empresas consolidadas. Como ya se ha observado, varias compañías han reconvertido sus propios sistemas, mientras que el gran interés en la tecnología semántica ha generado varios negocios nuevos, los cuales se definen como *empresas emergentes innovadoras*. Si se considera que las empresas emergentes están definidas como impulsores de innovación, no se debería estar sorprendido por su impresionante presencia dentro del mercado de tecnología semántica. Estos hallazgos sugieren que en la última década el interés en las oportunidades de tecnología semántica creció de manera considerable y varios negocios han enriquecido el mercado con varias aplicaciones. Sin embargo, varias compañías no se han embarcado hacia la evolución de la web. Únicamente 17 % de las STE recolectadas existía antes del 2001, cuando el artículo sobre *web semántica* se publicó. Estas compañías, que pertenecen a la antigua generación de compañías TIC, han evolucionado con el tiempo y aprovechado los retos de la innovación. En particular, entre el año 2001 y el año 2014, la cantidad de STE incrementó de manera exponencial: en promedio, 12 STE fueron creadas cada año.



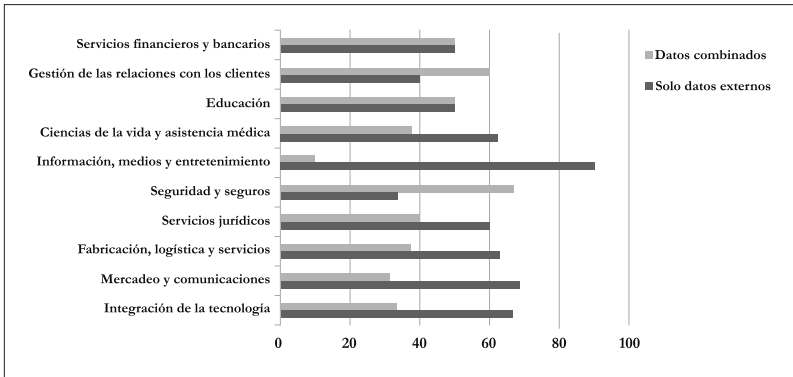
**Figura 3. Empresas emergentes y compañías en el conjunto de datos de las STE.**

Fuente: Elaboración propia.

Se hace referencia a las fuentes de información como una simple fuente de datos. Los datos pueden estar ubicados en una base de datos de los computadores de la compañía, o en otro computador, o tal vez pueden ser encontrados en una red como flujos de datos. En este estudio, en particular, se distinguen los datos internos (dentro de la compañía) de los datos externos (no contenidos en una base de datos de la compañía) y se consideran los datos combinados como una mezcla de ambos. Para entender el comportamiento de cada sector, se ha creado la figura 4, en la cual se examina qué tipo de datos usan las STE en un porcentaje. Entre las compañías analizadas, no se encontraron empresas que operen únicamente con datos internos. Por lo tanto, en la siguiente figura solo se observará la relación entre datos externos y datos combinados. El uso exclusivo de datos externos es más común en casi todos los sectores, excepto en los de Gestión de las



Relaciones con los Clientes y de Seguridad y Seguros, mientras que los Servicios Financieros y Bancarios y la Educación tienen una relación de igualdad (50 %). Otro punto de discusión está relacionado con la Información, los Medios de Comunicación y el Entretenimiento, las STE en este campo son intensas en cuanto a la información y están buscando en su mayoría datos nuevos continuos; probablemente analizan flujos de datos de la web con una relación crítica entre quien usa únicamente datos externos (90 %) y datos combinados (10 %).



**Figura 4. Fuentes de datos de las STE en distintos sectores.**

Fuente: Elaboración propia.

### Conclusión

El uso de tecnología semántica para analizar grandes cantidades de datos textuales producidos en la web se está incrementando con rapidez. Las organizaciones están reconociendo el papel de la información nueva para integrarla como procesos dentro de las unidades individuales de la empresa. Aunque la integración de la tecnología semántica en empresas requiere grandes esfuerzos en cuanto a recursos económicos, existe una tendencia creciente con respecto a la subcontratación de actividades de análisis de datos, permitiendo así que el mercado de tecnología semántica se defina a sí mismo con el incremento de las STE. Para entender de mejor forma las características y las tendencias del desarrollo de una compañía que trabaje con TS, se provee lo más moderno en cuanto a STE, analizando datos textuales mediante herramientas MT y KDT, clasificándolas de acuerdo con su grado de visibilidad en la web. Comenzando por las páginas web de las compañías, se identificaron las características principales de es-





tas compañías. En primer lugar, la distribución geográfica presenta una mayor concentración de STE en los Estados Unidos, seguido por Europa y compañías esporádicas alrededor del mundo. América Latina está sujeta a un fenómeno particular: incluso si se registran algunas actividades empresariales basadas en el territorio y enfocadas en el desarrollo de tecnología semántica, parece ser un mercado muy fértil para emprendedores alrededor del mundo. Más allá del uso del lenguaje, como ya se mencionó, estaría sujeto a un estudio posterior, el mercado latino es tan grande que es imposible ignorarlo cuando se tiene que expandir el negocio. Un caso identificado interesante es *Nethub*: una plataforma en la nube especialmente diseñada para almacenar y analizar datos de clientes de las interacciones en mercadeo y publicidad con marcas de productos y servicios. Su rápido crecimiento y el enfoque continuo en la innovación mejoran la calidad de la emergencia latina.

Independientemente del crecimiento del mercado, con varias empresas emergentes (65 % de las compañías reconocidas), la tecnología semántica poco a poco establece una relación con la gerencia. Pero las STE enfocan sus actividades en las fases de recuperación de datos, lo que demuestra que, con relación a la incrustación, las aplicaciones de minería y análisis son más avanzadas que la gestión de datos; las organizaciones están más interesadas en investigar y obtener datos, utilizados para mejorar el desempeño de las unidades comerciales, pero no se esfuerzan en gestionarlos. Los datos demuestran que la mayoría de STE suelen analizar flujos de información en vez de eventos individuales y que estos datos provienen de afuera de las empresas clientes. De acuerdo con la definición planteada, los datos que no están categorizados en bases de datos corporativas se encuentran principalmente por medio de redes sociales, como lo es en el caso de la Gestión de las Relaciones con los Clientes o el Mercadeo, de registros médicos y estudios científicos del sector de Asistencia Médica, en la industria jurídica de actos legales y patentes, y así en lo sucesivo.

Otro problema lo representa el hecho por el cual los productos y las aplicaciones actuales de MT siguen siendo herramientas diseñadas para especialistas de conocimiento capacitados y, como se mencionó anteriormente, requieren mucho esfuerzo en cuanto a recursos humanos y económicos. Como parte de los sistemas de gestión del conocimiento, herramientas MT futuras deben ser fáciles de usar por usuarios técnicos, así como también por ejecutivos gerenciales (Tan, 1999). La necesidad no solo de adquirir datos, sino también de pro-



cesarlos y almacenarlos, de desarrollar tareas operativas modestas y analizar e interpretar principalmente dichos datos de la manera adecuada se está convirtiendo cada día más en una necesidad compartida, lo cual establece la intervención de un profesional específico, a quien varios llaman científico de datos, y quien tiene distintas destrezas y distintas experiencias.

En un contexto de evolución tecnológica continua, en el cual las limitaciones han cambiado de un nivel nacional a uno mundial, la gestión de datos se ha vuelto más estratégica para identificar nuevos clientes, para la gestión de los clientes actuales y del área de compañías de adquisición. Dentro de este contexto, es esencial saber cómo identificar e interpretar de manera correcta los datos en los mercados y la gente que trabaja en los mismos. Por lo tanto, la administración empresarial requiere en todo momento conocimiento, y entrar en contacto con datos útiles reduce el tiempo requerido para aprender acerca de este escenario de mercado, a veces para predecirlo.

Para finalizar, el presente análisis demuestra que las STE se están enfocando en y están trabajando con aplicaciones de recuperación de datos, ya sean dinámicas o estáticas, pero las compañías aún no están listas para integrar dentro de las mismas proyectos de gestión de datos que operen a través de la fase de asimilación del conocimiento de la compañía. Probablemente, este aspecto sigue estancado en la fase de información comercial y es imposible tener una visión a largo plazo al integrar los datos internos con los externos. Ya se ha visto que las compañías prefieren recuperar datos fuera de ellas, pero se considera que el resultado más importante en este documento es que únicamente el 10 % de las compañías usan datos combinados, esto significa que hay un gran espacio en el mercado apuntando en esta dirección. Las compañías que aprendan a utilizar datos combinados tendrán una fuerte ventaja competitiva si logran tomar estas oportunidades lo más pronto posible.

## Referencias

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, 284 (5), pp. 28-37.
- Bürger T. (2008). The Need for Formalizing Media Semantics in the Games and Entertainment Industry. *Revista Journal of Universal Computer Science*, 14 (10), pp. 1775-1791.



- Bussler, C. (2003). The role of Semantic Web technology in enterprise application integration. *IEEE Data Eng. Bull*, 26(4), pp. 62-68.
- Carbone, F., Contreras, J., Hernández, J. Z., & Gómez-Pérez, J. M. (2012). Open Innovation in an Enterprise 3.0 framework: Three case studies. *Expert Systems with Applications*, 39(10), pp. 8929-8939.
- Cardoso, J., Hepp, M., & Lytras, M. D. (Eds.). (2007). *The semantic web: real-world applications from industry*, vol. 6. Springer Science & Business Media.
- Casas-Bayona, A. & Ceballos, H. G. (2014, agosto). Integrating semi-structured information using Semantic Technologies, An Evaluation of Tools and a Case Study on University Rankings. En *DATA2014: Tercera Conferencia Internacional sobre Tecnologías y Aplicaciones de Gestión*. Universidad Tecnológica de Viena, Viena, Austria.
- Cataldo M. (2009, noviembre). The Semantic Web's the Next Frontier. *American Banker*. Recuperado de [http://www.americanbanker.com/btn/22\\_11/the-semantic-webs-the-next-frontier-1003476-1.html?pg=1](http://www.americanbanker.com/btn/22_11/the-semantic-webs-the-next-frontier-1003476-1.html?pg=1).
- Chaudhuri, S., Dayal, U. & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), pp. 88-98.
- Davenport, T. H. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Estados Unidos: Corporación Editorial de la Escuela de Negocios de Harvard.
- Domingue, J., Fensel, D. & Hendler, J. A. (Eds.) (2011). *Handbook of semantic web technologies*. Springer Science & Business Media.
- Díaz Piraquive, F. N., Aguilar, L. J., & García, V. H. M. (2009). Taxonomía, ontología y folksonomía, ¿qué son y qué beneficios u oportunidades presentan para los usuarios de la web? *Universidad & Empresa*, 11(16), pp. 242-261.
- Feldman, R. & Dagan, I. (1995, agosto). *KDT knowledge discovery in texts [Descubrimiento de conocimiento en textos KDT]*. Documento presentado en la Primera Conferencia Internacional sobre el Descubrimiento de Conocimiento (KDD), 95, pp. 112-117.
- Feldman, R. & Hirsh, H. (1996, agosto). *Mining Associations in Text in the Presence of Background Knowledge*. En *Segunda Conferencia Internacional sobre Descubrimiento de Conocimiento KDD* (pp. 343-346).
- Feldman, R. & Hirsh, H. (1997). Exploiting background information in knowledge discovery from text. *Journal of Intelligent Information Systems*, 9(1), pp. 83-97.



- Fleisher, C. & Bensoussan, B. (2002). *Strategic and Competitive Analysis: Methods and Techniques for Analyzing Business Competition*. 1a ed. Londres, Reino Unido: Pearson.
- Gartner (2015, diciembre). *How to monetize your customer data?* Recuperado de <http://www.gartner.com/smarterwithgartner/how-to-monetize-your-customer-data/>.
- Global Entrepreneurship Monitor & WEF (2015, enero). Leveraging Entrepreneurial Ambition and Innovation: a Global perspective on Entrepreneurship, Competitiveness and Development. *World Economic Forum 2015*. Recuperado de [http://www3.weforum.org/docs/WEFUSA\\_EntrepreneurialInnovation\\_Report.pdf](http://www3.weforum.org/docs/WEFUSA_EntrepreneurialInnovation_Report.pdf).
- IBM Corporation (s.f.) *What is Big Data? Bringing Big Data to the Enterprise*. Recuperado de <https://www-01.ibm.com/software/au/data/bigdata>.
- Kodratoff, Y. (1999, junio). *Knowledge discovery in texts: a definition, and applications*. En *Simposio Internacional de Metodologías para Sistemas Inteligentes* (pp. 16-29). Springer Berlin Heidelberg.
- Lewis, M. P. (Ed.) (2009). *Ethnologue: Languages of the World*. Sixteenth edition. Dallas, Texas: SIL International.
- Loh, S., Wives, L. K. & de Oliveira, J. P. M. (2000). Concept-based knowledge discovery in texts extracted from the web. *ACM SIGKDD Explorations Newsletter*, 2(1), pp. 29-39.
- Marulanda C. E., Trujillo, M. L. & Giraldo, M. M. (2013). Análisis de la gestión del conocimiento en pymes de Colombia. *Revista Gerencia, Tecnológica informática*, 12(33), pp. 33-43.
- Ministerio de las Tecnologías de la Información y las Comunicaciones (2010). *Manual para la Interoperabilidad del Gobierno en línea*. Recuperado de [http://programa.gobiernoenlinea.gov.co/apc-aa-files/da4567033d075590cd3050598756222c/Marco\\_de\\_Interoperabilidad\\_GEL.pdf](http://programa.gobiernoenlinea.gov.co/apc-aa-files/da4567033d075590cd3050598756222c/Marco_de_Interoperabilidad_GEL.pdf).
- Nyce, C. (2007). *Predictive Analytics White Paper*. Recuperado de <http://www.hedgechatter.com/wp-content/uploads/2014/09/predictivemodelingwhitepaper.pdf>.
- Polikoff, I. & Allemang, D. (2003, septiembre). Semantic technology. *TopQuadrant Technology Briefing*. Recuperado de <https://lists.oasis-open.org/archives/regrep-semantic/200402/pdf00000.pdf>.
- Raghupathi, W. (2010). Data Mining in Health Care. En S. Kudyba (ed) *Healthcare Informatics: Improving Efficiency and Productivity* (pp. 211-223). Reino Unido: Taylor & Francis Group LLC.
- Shadbolt, N., Berners-Lee, T. & Hall, W. (2006). The Semantic Web Revisited. *IEEE Intelligent Systems*, 21 (3), pp. 96-101.



- Stephens, S. (2007). The Enterprise Semantic Web. En *The Semantic Web* (pp. 17-37). Estados Unidos: Springer.
- Tan, A. H. (1999, abril). Text mining: The state of the art and the challenges. *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, (8) pp. 65-70.
- Turney, P. D. (2002, julio). *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews* (pp. 417-424). En *Reunión Anual 40 de la Asociación de Lingüística Computacional (ACL)*, Filadelfia, Estados Unidos.