

## Estudio y análisis de técnicas de modelado de grandes volúmenes de datos jurídicos

**Lilia Palomo, Norma Lesca y Juan Mulki**

Facultad de Ciencias para la Innovación y el desarrollo

Universidad Católica de Santiago del Estero

[lilia.palomo@ucse.edu.ar](mailto:lilia.palomo@ucse.edu.ar) – [norma.lesca@gmail.com](mailto:norma.lesca@gmail.com) – [jmulki@ucse.edu.ar](mailto:jmulki@ucse.edu.ar)

### RESUMEN

En los últimos años, a raíz de la aplicación de las tecnologías de la información, se ha producido la proliferación masiva de datos, lo que dificulta a los científicos, empresas y organizaciones mantenerse al día con los resultados y tecnologías desarrolladas, llevando a que se malgaste tiempo, esfuerzo y recursos

Actualmente se puede contar con grandes repositorios de datos disponibles a costos bajos e incluso gratuitos, listos para ser explotados por procesos informacionales.

A partir de este aumento de información, se da la necesidad de extraer de ella patrones, tendencias y/o conocimiento de forma rápida y eficiente, para lo cual los métodos tradicionales han tenido que evolucionar en busca de rendimiento y escalabilidad. El gran contenido de valor que genera este tipo de información está permitiendo a las organizaciones una mejora en la toma de sus decisiones, lo que conlleva a la obtención de ventajas competitivas en los diferentes campos de acción.

Este impacto social y la rápida evolución de las tecnologías determinan también una nueva visión en el Derecho, que ha procurado responder a esta realidad como responsable por garantizar el uso adecuado de aquéllas en beneficio de la sociedad en general, y por la necesidad de satisfacer adecuadamente a las demandas de la sociedad que exige a los operadores del derecho, tanto judiciales como legislativos, respuestas más rápidas, más acertadas y en mayor cantidad, lo que hace que el estudio de herramientas informáticas aplicadas al mundo jurídico suponga una

acción muy importante, dentro del sector informático.

Por esta razón, este proyecto propone una investigación exploratoria del modelado y visualización de Big Data para documentación y datos jurídicos, basado en la tarea de realizar la recolección, análisis y crítica de los diferentes métodos propuestos en la literatura sobre el manejo de grandes volúmenes de datos en el proceso de razonamiento y respuesta a consultas, y su consolidación en un trabajo que relacione estas perspectivas diferentes, analizando las ventajas y desventajas, así como los retos de implementación que presentan.

**Palabras clave:** BigData, datos jurídicos, grandes volúmenes de datos, modelado, visualización.

### CONTEXTO

La facultad de Ciencias para la Innovación y el Desarrollo de la Universidad Católica de Santiago del Estero impulsa la implementación de proyectos que apuntan a incentivar la investigación desde las cátedras promoviendo la interacción vertical y horizontal entre ellas, y posibilitando a los docentes obtener resultados que puedan ser aplicados en las aulas.

Por ello es que, a fines del año 2017, desde las asignaturas de Programación I, Estructuras de Datos y Análisis Numérico surge el proyecto “Big Data: Modelado y visualización de grandes volúmenes de datos jurídicos” con el objetivo de promover la investigación aplicada, la formación de recursos humanos, la innovación de los

contenidos de las cátedras y de las prácticas profesionales.

Este artículo presenta una línea de investigación que surge de ese proyecto, ante la necesidad de los operadores del derecho de disponer de instrumentos capaces de brindar el aprovechamiento de las oportunidades que genera el Big Data como herramienta predictiva y en la tarea de toma de decisiones.

## 1. INTRODUCCIÓN

La gran cantidad de información disponible y su heterogeneidad han sobrepasado la capacidad de las tecnologías actuales de gestión de datos. El tratamiento con grandes volúmenes de datos estructurados y no estructurados, a menudo referido como Big Data, es un tema de investigación de actualidad así como un importante desafío tecnológico.

Por tratarse de un fenómeno reciente, Big Data no se halla todavía suficientemente delimitado, aunque se puede encontrar varios puntos en común entre las definiciones planteadas por la comunidad académica, los técnicos y los profesionales especializados en el área. McKinsey [1] ofrece una definición “intencionalmente subjetiva” y dinámica a partir de las limitaciones tecnológicas en la actualidad: “Big Data refiere a un conjunto de datos cuyo tamaño está más allá de la capacidad que tienen los software de base de datos para capturar, almacenar, administrar y analizar”. De esta manera, a medida que la frontera tecnológica se va modificando, el tamaño del conjunto de datos que califica como Big Data también irá cambiando. Se habla de Big Data cuando el tamaño se vuelve el principal problema. Si bien el volumen es un punto en común entre los especialistas que definen Big Data, otros conceptos incorporan más dimensiones al análisis. Gartner [2] y UN Global Pulse [3] suelen caracterizar el tema a partir de los desafíos que Big Data representa no sólo a partir del volumen, sino también de la variedad de la información y la velocidad

de su tratamiento<sup>1</sup>, incluso dándole mayor importancia a estas últimas. Con *variedad* se suele hacer referencia a la heterogeneidad de la representación e interpretación semántica de los datos, es decir que no están directamente listos para ser integrados a una aplicación; mientras que con *velocidad* se hace referencia tanto a la frecuencia de los datos como al tiempo de respuesta: la importancia reside en la velocidad del feedback, utilizando los datos para la toma de decisiones.

En sus primeros años, la informática jurídica fue, sustancialmente, informática documental, y abarca la creación, gestión y recuperación de datos en bancos de datos que contienen informaciones específicamente jurídicas, tales como leyes, doctrina y jurisprudencia, que fue recibida de buen grado por abogados y juristas.

Pero ante la existencia de un gran volumen de documentos e informaciones jurídicas, es necesario que los operadores del derecho dispongan de instrumentos capaces de brindar las oportunidades que genera el Big Data, como una herramienta predictiva de auxilio a sus tareas decisorias.

Frente a este desafío surgen las siguientes preguntas: ¿se encuentra el país, y en particular los poderes judiciales provinciales, en condiciones de aprovechar los beneficios que promete esta nueva fase de grandes volúmenes de datos, basados en el conocimiento?; si el análisis de grandes datos requiere de una base tecnológico-comunicacional: ¿están los poderes judiciales en condiciones de dar respuesta a esta exigencia?; ¿cómo impacta el escaso desarrollo local de infraestructura de alta complejidad para el desarrollo de Big Data?

## 2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación del presente trabajo tiene como eje central el estudio y análisis de

<sup>1</sup> Volumen, Velocidad y Variedad, tres términos conocidos como las 3 Vs

las técnicas de visualización de grandes volúmenes de datos jurídicos.

Entre los supuestos que guían el proyecto se encuentran:

- En las últimas décadas, el sector de servicios de informática e información de la Argentina se posicionó entre los más dinámicos de la región [4], por lo que existen capacidades técnicas reales y potenciales que permiten aprovechar este avance.
- Asociados al fenómeno del Big Data existen tres tipos de problemas: *tecnológicos*, relacionados con el almacenamiento, la seguridad y el análisis de los volúmenes crecientes de datos; *comerciales*, relacionados con el valor añadido que se puede generar; y *sociales*, relacionados con la privacidad de la información personal [5].
- Los retos que genera Big Data están asociados a cambios de paradigma que implican una mayor importancia de la disponibilidad y el acceso a los datos, aceptación de ciertos niveles de imprecisión y desorden, y concentración en las correlaciones. [6]

Se trabajará en modelos de repositorios institucionales que se ajusten a los requerimientos de visualización de Big Data de carácter jurídico.

### 3. RESULTADOS ESPERADOS

Considerando que el mercado general de Big Data en Argentina muestra una demanda escasa y de baja complejidad, explicado en parte por la falta de conocimiento sobre el tema, problemas institucionales y limitaciones de infraestructura [7], el objetivo general del proyecto es realizar un estudio y análisis de técnicas de modelado y visualización de grandes volúmenes de datos jurídicos.

Este objetivo general se desarrollará mediante los siguientes objetivos específicos:

- Entender la problemática de la gestión del Big Data.
- Identificar las características más relevantes en la gestión de Big Data que deben guiar la elección de una solución para el ámbito jurídico.
- Analizar las principales herramientas de gestión de Big Data existentes en el mercado.
- Identificar los modelos más adecuados para el problema de los grandes volúmenes de datos jurídicos.
- Adquirir conocimientos específicos sobre el uso de Big Data para la toma de decisiones.
- Emplear herramientas de modelización de grandes volúmenes de datos y documentación del ámbito jurídico.

Para alcanzar el objetivo planteado, se llevarán a cabo las actividades siguientes:

- Buscar, recopilar y clasificar publicaciones, artículos y libros relacionados con la problemática planteada.
- Investigar los recursos técnicos existentes en el mercado, estudiando las metodologías de modelado, las plataformas y software para el tratamiento de Big Data y los diferentes métodos de visualización, estableciendo una comparación entre ellos.
- Identificar las características más relevantes de la gestión de Big Data que deben guiar la elección de una solución arquitectónica para datos jurídicos, seleccionando los modelos estadísticos o de aprendizaje automatizado más adecuados para el problema.
- Validar la solución escogida, evaluando el impacto de la utilización de Big Data en el ámbito jurídico.

Se espera que los resultados de esta investigación se incorporen a los contenidos de las cátedras relacionadas y al espacio curricular correspondiente.

Como resultados indirectos se espera la consolidación del grupo de investigación, la formación de nuevos investigadores y la motivación y entrenamiento en investigación de los estudiantes de grado.

#### 4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo está conformado por tres docentes de la carrera de Ingeniería en Informática, con dedicación simple.

El grupo hace difusión y formación de recursos humanos desde las cátedras: Programación I, Estructura de Datos y Análisis Numérico.

Asimismo, se considera de gran interés la incorporación de becarios, para motivar a los alumnos de la carrera de Ingeniería en Informática a realizar su trabajo final de grado en el área de este proyecto.

#### 5. BIBLIOGRAFIA

- 1] MCKINSEY, (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
- 2] GARTNER, (2011). Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data
- 3] UN GLOBAL PULSE (2012). Big Data for Development: Challenges & Opportunities. United Nations.
- 4] BARLETTA, PEREIRA, ROBERT & YOGUEL, (2013). "Argentina: Dinámica reciente del sector de software y servicios informáticos", Revista CEPAL N° 110
- 5] NUNAN, DI-DOMENICO, (2013). "Market research and the ethics of big data". International journal of market research, v. 55, n. 4, pp. 505-520. Disponible en: <http://dx.doi.org/10.2501/IJMR-2013-015>
- 6] MAYER-SCHÖNBERGER, CUKIER (2013). Big data. La revolución de los datos masivos. Madrid: Turner
- 7] MALVICINO & YOGUEL (2015). "Descubriendo Big Data en Argentina. Encuesta Digital 2014". AGRANDA 1era ed. 44va Jornadas de Informática (JAIIO)