

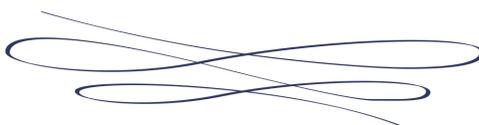
UNIVERSIDAD PABLO DE OLAVIDE



**EL HORIZONTE TECNOLÓGICO:
UNA VISIÓN PROSPECTIVA DEL
ANÁLISIS INTELIGENTE DE DATOS
MASIVOS**

**LECCIÓN INAUGURAL
CURSO 2016 / 2017**

Dr. D. Raúl Giráldez Rojo



SEVILLA 2016

EDITA:
Universidad Pablo de Olavide

UNIVERSIDAD PABLO DE OLAVIDE



**EL HORIZONTE TECNOLÓGICO:
UNA VISIÓN PROSPECTIVA DEL ANÁLISIS
INTELIGENTE DE DATOS MASIVOS**

LECCIÓN INAUGURAL
CURSO 2016 / 2017

Dr. D. Raúl Giráldez Rojo
Profesor Titular de
Lenguajes y Sistemas Informáticos

SEVILLA
2 0 1 6

a Alicia, Gonzalo y Guillermo

ÍNDICE

PREÁMBULO	5
¿ÉPOCA DE CAMBIOS O CAMBIO DE ÉPOCA?	9
LA NECESIDAD DE ANALIZAR LA INFORMACIÓN	13
CONCEPTOS ASOCIADOS AL ANÁLISIS INTELIGENTE DE DATOS	17
FUENTE DE DATOS: EL MUNDO VIRTUAL	20
EL NEGOCIO DEL CONOCIMIENTO	23
EL RETO DE LA UNIVERSIDAD	29
EPÍLOGO.....	33

PREÁMBULO

Es difícil expresar con palabras el orgullo que me invade al encontrarme hoy ante tan distinguida audiencia con el honroso encargo académico de impartir la solemne Lección Inaugural. Hacerlo representando a la Escuela Politécnica Superior acrecienta aún más si cabe la satisfacción personal que siento. Gracias a las autoridades académicas por concederme tan excelso honor.

Esta disertación que hoy efectúo ante ustedes es una mirada al futuro tecnológico desde el conocimiento y el análisis de los avances que hoy ya son una realidad patente. El riesgo de la prospectiva radica en el componente predictivo de la misma, acrecentado en este caso por la estrechez de la ventana temporal en la que la realizo. Este bosquejo, que hoy puede revelarse innovador, pasará a ser una reseña obsoleta en el mismo instante en el que alcancemos el horizonte tecnológico del que versa. No exagero cuando afirmo que tal momento acontecerá antes de que los estudiantes que hoy precisamente empiezan su vida universitaria se reúnan en este mismo Paraninfo para recibir sus becas como egresados. Ruego sean entonces magnánimos, pues son varias las contingencias de las que hoy les hablaré. Afortunadamente, el contexto en el que las expongo marca inequívocamente el momento.

SR. RECTOR MAGNÍFICO DE LA UNIVERSIDAD PABLO DE OLAVIDE, DE SEVILLA.

EXCMO. SR. CONSEJERO DE ECONOMIA Y CONOCIMIENTO DE LA JUNTA DE ANDALUCÍA.

ILMA. SRA. PRESIDENTA DEL CONSEJO SOCIAL.

EXCELENTÍSIMAS E ILUSTRÍSIMAS AUTORIDADES.

MIEMBROS DE LA COMUNIDAD UNIVERSITARIA.

DISTINGUIDOS AMIGOS.

SEÑORAS Y SEÑORES:

El desarrollo del ser humano, de la sociedad que hoy conocemos, y el desarrollo de la ciencia y la tecnología han ido inseparablemente de la mano a lo largo de la historia, tanto que, francamente, me cuesta discernir entre ambos. Sin embargo, el avance de la ciencia y el desarrollo tecnológico han conformado una espiral de retroalimentación que ha provocado el progreso exponencial de ambas. Una aceleración que las ha llevado a avanzar más en las últimas dos décadas que en toda la historia de la Humanidad. Sin ir más lejos, yo mismo he sido víctima de la vertiginosa velocidad a la que se suceden estos avances durante la preparación esta lección. Parafraseando a un afamado periodista deportivo cuando narraba las carreras de Fórmula 1: “Si parpadeas, te lo pierdes”. Me planteo entonces si quizá hoy ha llegado el momento en el que el progreso científico y tecnológico ha adelantado al desarrollo social y cultural, o lo que es lo mismo, si la sociedad está preparada para un futuro tecnológico que, sin darnos cuenta, ya es el presente.

Cada nuevo avance, lejos de aliviar la sed de tecnología de la sociedad, la acrecienta. Incluso me atrevería a decir que alimenta nuestro deseo de vivir en un mundo cibernético, hiperconectado, y robotizado. Como decía Oscar Wilde, “ten cuidado con lo que deseas, se puede convertir en realidad.”

¿ÉPOCA DE CAMBIOS O CAMBIO DE ÉPOCA?

Son tres las revoluciones industriales de las que tenemos constancia, o tres etapas de la misma, según se interprete. La primera estuvo marcada por la invención de la máquina de vapor, el ferrocarril y las máquinas de textiles, en el siglo XVIII. En la segunda, desarrollada desde finales el siglo XIX hasta la Primera Guerra Mundial, además del acero, la electricidad, los derivados del petróleo y la producción en cadena, fueron las tecnologías de las comunicaciones las que adquirieron especial protagonismo. El telégrafo, el teléfono, el cine y la radio abrieron un nuevo horizonte al intercambio de información y sentaron las bases para los mecanismos de interacción social que aún hoy seguimos utilizando. Pocos años más tarde, a mediados del siglo XX, la denominada revolución digital o de las nuevas tecnologías fue testigo del nacimiento de la Informática y de Internet, que junto a la robótica y las energías renovables conforman el corazón de esta tercera revolución industrial en la que aún nos encontramos.

Y cuando ya parece que estamos comenzando a asimilar este nuevo escenario, cuando la sociedad comienza a estabilizarse en esta nueva era, el ocaso de la misma parece descubrirse ante a nosotros. La robotización a gran escala, los teléfonos inteligentes, el Internet de las cosas, las redes sociales, la impresión 3D, las *smart cities*, la realidad aumentada... En apenas una década, el mundo ha asistido a un incesante bombardeo de novedades cuyo resultado es la transformación radical de muchas industrias. Parece que hay consenso internacional en aceptar como un hecho la llegada inminente de la cuarta revolución industrial. Si se dan realmente o no las circunstancias necesarias y suficientes para elevar el innegable cambio a la categoría de revolución es una disyuntiva que personalmente confieso no haber podido resolver. Permítanme, no obstante, usar el término aprovechando la aparente conformidad generalizada.

Esta cuarta revolución industrial, que también se ha convenido en llamar Industria 4.0¹, estará sustentada en las llamadas *smart factories* o fábricas inteligentes, caracterizadas fundamentalmente por dos aspectos: por un lado la hibridación entre el mundo físico y el virtual mediante la interconexión entre las máquinas y los sistemas a través del Internet Industrial de las Cosas (I2oT); y por otro lado, por un fluido y constante intercambio de información con el resto de actores: con los mercados, con sus competidores, con otras fábricas inteligentes... y sobre todo, con el cliente. En definitiva, pasamos de una producción automática a una *producción inteligente*, impulsada por sistemas ciber-físicos hiperconectados. Los primeros avances en este ámbito están propiciando una mayor flexibilidad de los procesos de fabricación, posibilitando llevar hasta el extremo la personalización de los productos y servicios. En otras palabras, el cliente consume lo que quiere, como lo quiere y en el momento que quiere, tomando así el mando en las relaciones comerciales. La principal consecuencia económica de esta innovación disruptiva es el desplazamiento del eje del sistema desde la oferta hacia la demanda.

En esta nueva industria inteligente, el avance en el desarrollo e incorporación de sensores y sistemas de comunicación, así como las nuevas capacidades de hiperconectividad e interacción entre los diferentes actores, está elevando el potencial de generación de información a niveles asombrosos. Sin embargo, la mejora real en la eficiencia y la flexibilidad de producción dependerá de la capacidad de las empresas para gestionar y analizar estas ingentes cantidades de datos, para lo cual, el mayor desafío estará en el lado del software. De ahí que el desarrollo de soluciones potentes en el ámbito de la ciencia de los datos en general y del *big data* en particular, se está volviendo un factor determinante en las organizaciones. Es este concepto, *big data*, en torno al cual girará gran parte de mi exposición, por lo que

¹ El término Industria 4.0 fue acuñado por el gobierno alemán para describir la fábrica inteligente, una visión de la fabricación informatizada con todos los procesos interconectados por Internet de las Cosas (IOT).

veo conveniente adelantar una breve definición. *Big data* es una parte de la ciencia de datos que reúne al conjunto de tecnologías que se diseñan para el procesamiento de datos masivos y la extracción de conocimiento útil a partir de los mismos.

El pasado mes de enero, líderes del ámbito político, empresarial y científico de países de todo el mundo se dieron cita el Foro Económico Mundial de Davos¹, cuya temática principal giró en torno a esta cuarta revolución industrial con el objetivo de determinar las causas y el alcance de la misma así como su impacto en todos los aspectos de la sociedad. Como era predecible, el consenso apunta a que la introducción de aplicaciones tecnológicas disruptivas en el campo de la inteligencia artificial, la robótica o la nanotecnología, entre otros muchas, están detrás del cambio hacia la Industria 4.0. Estas aplicaciones, denominadas *habilitadores digitales*, hacen posible que esta nueva industria explote todo su potencial y posibilitan la mencionada hibridación entre el mundo físico y el digital.

Sin duda, estamos ante una revolución que cambiará por completo la forma en que vivimos, trabajamos, y nos relacionamos. El impulso de la productividad, la racionalización o el ahorro energético, son ventajas que contrastan con los desarreglos económicos, y sobre todo laborales, que ocasionará este auténtico tsunami tecnológico. En sólo un lustro, la disrupción tecnológica interactuará con otras variables socioeconómicas, geopolíticas y demográficas para generar un entorno que cambiará el mercado laboral para siempre. Las previsiones² apuntan a la destrucción de más de 7 millones de puestos de trabajo antes del año 2020. Y esto sólo en los 15 países más industrializados del mundo, entre los que, por cierto, no se encuentra el nuestro. La mayoría de estos trabajos estarán relacionados con

¹ El Foro Económico Mundial (*World Economic Forum*, WEF), o también llamado Foro de Davos, es una fundación sin ánimo de lucro con sede en Ginebra que desde 1991 reúne anualmente en el Monte de Davos (Suiza) a los principales líderes empresariales, políticos e intelectuales internacionales para analizar los problemas más relevantes a los que se enfrenta el mundo.

² Datos extraído del informe del Foro Económico Mundial de Davos 2016:
http://www3.weforum.org/docs/WEF_AMNC16_Report.pdf

tareas administrativas, aunque también aquellos empleos implicados en procesos de fabricación y producción se verán fuertemente afectados. Sin embargo, el estrecho pero existente margen de mejora en su cualificación deja abierta la posibilidad de reconversión de estas profesiones si la reacción es oportuna. Por el contrario, se crearán también en torno a 2 millones de nuevos puestos de trabajo en áreas relacionadas con la informática, la ingeniería, las matemáticas y la arquitectura, la mayoría de ellos, evidentemente, de alta cualificación. En concreto, las profesiones más demandadas serán aquellas que, incidiendo en lo ya adelantado anteriormente, rodean al análisis de datos masivos. Parece claro que, a corto plazo, el escenario no es muy alentador, existiendo un temor generalizado ante este desafío. Sin embargo, quiero pensar que, a largo plazo y tras un duro proceso de ajuste y aprendizaje, este nuevo periodo puede ser una gran oportunidad, sobre todo para las sociedades que mejor preparadas lleguen a ese momento.

Como se podrán imaginar, el papel que juegan en esta revolución las Tecnologías de la Información y las Comunicaciones – las llamadas TICs– es más que protagonista, ya que son precursoras del cambio a la vez que motor de progreso del mismo. La informática, en comunión con otras disciplinas afines como la robótica y las telecomunicaciones, han posibilitado el desarrollo de los anteriormente mencionados habilitadores digitales. De todos ellos, hoy les hablaré de uno en concreto: el *big data*; no sólo por mi evidente afinidad profesional, sino por su peso específico y su carácter transversal a la Industria 4.0. Intentaré no sólo mostrarles qué es el *big data* y por qué es necesario, sino demostrarles, con un ejemplo con el que todos nos identificaremos, cómo ya forma parte de nuestras vidas.

LA NECESIDAD DE ANALIZAR LA INFORMACIÓN

Mi predecesor representando a la Escuela Politécnica Superior en esta misma tribuna, el profesor Aguilar, hizo un recorrido de la evolución tecnológica en su magistral Lección Inaugural del curso 2009-2010, dejando dicha evolución prácticamente en el mismo punto donde hoy yo la retomo. Él nos habla de la leve pero sustancial diferencia entre conocimiento – entendimiento, conjunto de saberes sobre una temática– e información –comunicación o adquisición de conocimientos. A esta dupla, yo añado dos conceptos más, datos e inteligencia, pues la diferencia y relación entre los cuatro marca el sentido de esta parte de la lección. Los datos son simplemente valores recolectados sin un significado por sí mismos. Cuando los datos son procesados para establecer relaciones entre ellos dan lugar a información, la cual tiene un significado y un propósito. La información utilizada en un contexto determinado y unida a la percepción y la experiencia, se convierte en conocimiento. Cuando el conocimiento es validado, orientado hacia un objetivo y reutilizado para la toma de decisiones, obtenemos inteligencia.

Medir o dimensionar el conocimiento o la inteligencia es una cuestión más filosófica que científica. Por el contrario, la información, en el sentido anteriormente mencionado, es un concepto medible. Así, Claude Shannon, padre de la Teoría Matemática de la Comunicación, proporcionaba a mediados del siglo XX una magnitud basada en la entropía para cuantificar la cantidad de información, denominada *binary digit* y que popularmente es conocida como *bit*. El byte, unidad de información básica utilizada en computación, no es más que un octeto de bits. Megabyte, gigabyte, incluso terabyte son términos que ya forman parte de nuestro vocabulario común cuando queremos referirnos a la capacidad de almacenamiento de un dispositivo. Sin embargo, es difícil imaginar cuánta información es, por ejemplo, un megabyte. Algunos dirán, sin riesgo de errar demasiado, que es un millón de bytes. Pero para hacernos una idea más tangible, podemos decir que un megabyte es lo que ocupa aproximadamente El Quijote.

Por tanto, un gigabyte son mil Quijotes y un terabyte un millón de Quijotes. Hasta aquí creo que prácticamente todos han podido identificar las unidades de medida y asociarlas a una idea. Pero como veremos, la cantidad de información generada y almacenada por el ser humano crece exponencialmente, por lo que en poco tiempo deberemos familiarizarnos con el resto de unidades. Petabyte y exabyte ya comienzan a sonar, mientras que pronto llegarán los zettabyte y yottabyte.

Esta breve incursión en el Sistema Internacional de Unidades tiene como único propósito poder interpretar la respuesta siguiente pregunta: ¿cuánta información existe actualmente almacenada en todo el mundo y cuánta se genera diariamente? Son muchos los que han dado una respuesta a esta pregunta y, en ocasiones, con gran repercusión en los medios y foros especializados. Sin embargo, en mi modesta opinión, la mayoría están contaminadas por intereses comerciales o han sido utilizadas como meros instrumentos de marketing. De hecho, creo que es imposible dar una respuesta fiable, ya que existen multitud de fuentes de datos que actualmente son herméticas, o cuanto menos de difícil acceso, y que sin embargo contribuyen significativamente a engordar el grueso de datos generados globalmente. Sirva como ejemplo, las redes sociales chinas, las cuales a menudo nos empeñamos en soslayar cuando presentamos estudios en este contexto.

En el año 2010, el por entonces CEO de Google, Eric Schmidt, afirmó que la Humanidad había creado hasta el año 2003 una cantidad de información equivalente a cinco exabytes. Según lo expuesto anteriormente, eso equivale a un cinco billones de Quijotes. Esta afirmación cuenta con muchos detractores, principalmente por la falta de justificación por parte de Schmidt. Sin embargo, un estudio menos susceptible de contaminación comercial publicado en la revista Science en el año 2011 estimaba que la cantidad de información generada y almacenada por la Humanidad entre los años 1986 y 2007 ascendía a 295 exabytes. A pesar de la diferencia notable entre ambas estimaciones, la magnitud de ambas es la misma, por lo que parece

razonable establecer la estimación del orden exabytes, es decir, trillones de bytes.

Hoy existen más de 1.000 millones de páginas web¹, un 44% más que hace un año. Ni siquiera Google puede abarcar esa inmensidad, pues solo tienen indexados el 0,004% de toda la información disponible. Cada minuto en todo el mundo se realizan 2,4 millones de búsquedas en Google, más de 20 millones de mensajes son enviados por WhatsApp y más de 150 millones por email, 2,8 millones de vídeos son vistos en Youtube, 700.000 usuarios inician sesión en Facebook, 350.000 nuevos twits son emitidos, 570 nuevos sitios web son creados. Y todo ello, cada sesenta segundos, el mismo tiempo que he tardado en darles estos datos. Un estudio publicado por la empresa Cisco Systems el pasado mes de junio revela que el tráfico en internet sobrepasará el zettabyte en 2016, alcanzando los 2,3 zettabyte en 2020, año en el que el tráfico generado por los dispositivos móviles superará al de los ordenadores. No es extraño que, a este ritmo, la información generada por la Humanidad se duplique cada 2 años. Esta escalada exponencial me recuerda a la afamada leyenda matemática sobre el inventor del juego del ajedrez y los granos de trigo. Comenzando por un grano de trigo en la primera casilla de un tablero ajedrez, si por cada casilla duplicamos el número de granos de la casilla anterior, la cantidad resultante tras aplicar la progresión a las 64 casillas del tablero es más de 18 trillones de granos de trigo, que equivalen a más de 700.000 toneladas. El mundo necesitaría más de 1000 años para producir tal cantidad de trigo. Traslada la leyenda al contexto del que hoy les hablo y dígame el resultado.

Mucho más sencillo que imaginar ese guarismo de magnitud superlativa es advertir los efectos que tendrá en la sociedad la posibilidad de acceder a tal cantidad de información de manera casi instantánea. Una consecuencia palpable de este fenómeno es la denominada *infoxicación* o intoxicación por exceso de información, que nos obliga a estar siempre alerta y procesar diariamente cientos de

¹ Dato extraído de *internetlivestats.com*.

mensajes e informaciones en las que no podemos profundizar. En 1970, Alvin Toffler define el término *shock del futuro* como un estado psicológico que afecta tanto a individuos como a sociedades enteras y que no es más que la percepción personal de estar ante “demasiados cambios en un período de tiempo demasiado corto”. Todo ello es el resultado de un mundo donde se prima la exhaustividad –cantidad– frente a la relevancia –calidad–.

En este mundo saturado de información, en esta sociedad infoxicada, la importancia de poder extraer el conocimiento útil inherente y escondido en esa ingente cantidad de información cobra una relevancia preeminente. Es en este caldo de cultivo donde aflora el *big data*, un término relativamente moderno que esconde un concepto que no lo es tanto, como veremos a continuación.

CONCEPTOS ASOCIADOS AL ANÁLISIS INTELIGENTE DE DATOS

D*ata science* (o ciencia de los datos), *big data*, *data mining* (o minería de datos), *open data*, inteligencia artificial, *business intelligence*, e incluso *data warehouse*, son vocablos que en los últimos años han trascendido más allá de la ingeniería informática y que han pasado a formar parte cualquier conversación en la que se hable de tecnologías de la información. Sin embargo, en muchos casos, estos términos son confundidos entre sí y erróneamente utilizados.

John McCarthy¹ acuñó en 1956 el término *inteligencia artificial*, definiéndola como “la ciencia y la ingeniería de crear máquinas inteligentes, especialmente programas de computación inteligentes”. Sesenta años después de esa definición, podemos hablar de dos tipos de inteligencia artificial. Por un lado máquinas que piensan, también llamada *inteligencia artificial fuerte*, respecto a la cual estamos aún muy lejos de aprobar el controvertido Test de Turing². Y por otro lado, las máquinas con conducta inteligente, y que hoy día son una realidad, como coches que conducen y aparcan de manera autónoma, ordenadores que juegan a la ajedrez o robots que reconocen objetos, entre otros muchos ejemplos. En este grupo se engloban los algoritmos de aprendizaje automático, capaces de aprender a partir de la experiencia mediante la generalización y búsqueda de patrones en los datos, y que forman, junto a bases de datos y las técnicas estadísticas, los pilares de la minería de datos.

¹ John McCarthy fue un prestigioso informático que recibió el Premio Turing en 1971 por sus importantes contribuciones en el campo de la Inteligencia Artificial. También se le atribuye el *cloud computing*.

² Test de Turing es una prueba de la habilidad formulada por Alan Turing en 1950 en la que se somete a una máquina para que ésta exhiba un comportamiento inteligente similar e indistinguible del de un humano. El test se basaba en la idea de que una computadora puede ser llamada “inteligente” si logra engañar a una persona haciéndole creer que es un humano.

En 1996 se define formalmente la extracción de conocimiento en base de datos, o KDD¹ por sus siglas en inglés, como “[...] el proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y fundamentalmente comprensibles en los datos”. Por tanto, el KDD se refiere al proceso completo que nos permite escalar desde los datos hasta la inteligencia, englobando: la definición de objetivos, es decir, qué queremos saber; la preparación de los datos para transformarlos en información; el análisis de la misma para extraer el conocimiento; la representación y visualización del mismo; y la aplicación del modelo obtenido a la toma de decisiones. Minería de datos, en términos estrictos, hace referencia únicamente a la fase de análisis en la que se aplican los anteriormente mencionados algoritmos de aprendizaje automático, aunque en muchas ocasiones el término se generaliza semánticamente y se utiliza para referirse también al proceso completo. *Business intelligence* o inteligencia de negocio no es más que una adaptación y aplicación concreta del KDD en contextos empresariales.

En 1998 John Mashey² define por primera vez el concepto de *big data* o datos masivos, asociándolo al imparable tsunami de datos que ya se vislumbraba en el horizonte y para cuya gestión no estaban preparadas las infraestructuras informáticas del momento. La definición más común deriva directamente de esta idea, considerándose *big data* a aquellos conjuntos de datos que superan la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable. Sin embargo, cada vez es más habitual incluir también en el concepto a las tecnologías que se diseñan para el procesamiento de dichos datos masivos. Por tanto, bajo mi punto de vista, el *big data* simplemente el resultado de enriquecer el proceso KDD con arquitecturas y técnicas de análisis

¹ Definición original: “*Knowledge Discovery in Databases is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*” (Fayyad, Piatetsky-Shapiro, and Smyth 1996).

² John Mashey, informático estadounidense, publicó en 1998 el artículo *Big Data and the Next Wave of Infrastrass*, en el que hace referencia al concepto de big data con el significado aquí empleado. Pocos meses antes dos científicos de la NASA, Michael Cox y David Ellsworth, había usado el mismo término, aunque sólo para referirse a datos de gran tamaño.

capaces de gestionar datos masivos. De ahí que, como decía, es un término nuevo pero basado en una idea ya existente.

Hay tres magnitudes originalmente asociadas al *big data* y conocidas como las tres “uves”: volumen, variedad y velocidad. A éstas “uves”, se han añadido dos más, como son la veracidad y el valor de los datos. Estas dos “nuevas” características atribuidas al *big data* han sido interpretadas en algunos entornos empresariales como una evolución del concepto original, otorgándole incluso una nueva denominación: *smart data*, o datos inteligentes. En mi opinión, este no es más que un ejemplo más de la proliferante afición de las empresas por acuñar nuevos términos con impacto comercial añadiendo matices a conceptos ya existentes. En base a estas características, yo defino el *big data* como el conjunto de tecnologías que proporcionan la capacidad de procesar a alta velocidad grandes volúmenes de datos obtenidos de fuentes de información muy variadas y heterogéneas para generar conocimiento veraz y valioso.

El concepto *data science* o ciencia de los datos es más amplio, ya que es el campo interdisciplinario que se ocupa de la extracción de conocimiento a partir de los datos en su significado más extenso, es decir, que engloba todo lo anterior, incluido el *big data*. El científico de datos es una de esas nuevas profesiones antes mencionadas. Es tan nueva que la comunidad tecnológica, incluyendo universidades y empresas, aún no se ha puesto de acuerdo en cuáles son las fronteras que separan a un científico de datos de un ingeniero de datos o de un experto en *big data*. En cualquier caso, estén o no definidas formalmente sus competencias en la memoria Verifica de un hipotético título, la innegable apuesta por la ciencia de los datos desde el mundo empresarial augura un futuro más que prometedor a estos profesionales dentro de la cuarta revolución industrial.

FUENTE DE DATOS: EL MUNDO VIRTUAL

Hoy en día, prácticamente cualquier dispositivo es capaz de generar, almacenar y compartir datos de manera autónoma y prácticamente de forma instantánea. Es difícil imaginar una actividad, sea personal, empresarial o industrial, que no genere algún tipo de información digital. Internet permite que toda esa información, heterogénea y habitualmente desestructurada, sea accesible y de alguna manera esté interconectada, integrándola en un macro-repositorio distribuido que a menudo llamamos *Ciberespacio* o mundo virtual. Así, huyendo de los tópicos de la ciencia ficción, podemos definir el *Ciberespacio*¹ como el espacio relacional de interacción virtual implementado dentro de todos los ordenadores del mundo cuya realidad se construye a través del intercambio de información.

Podríamos entender este mundo virtual simplemente como un gigantesco almacén de información al que podemos acudir cuando queramos. Sin embargo, la irrupción de la denominada Web 2.0 en la primera década de este siglo marca un hito trascendental que lo ha convertido en mucho más que eso. El hecho de que usuarios, que hasta entonces eran meros espectadores y consumidores de información, se hayan transformado en participantes capaces de interactuar con otros usuarios y de alimentar el sistema con información propia, ha permitido establecer relaciones reales entre las personas, aunque los escenarios sean virtuales. En otras palabras, ha permitido la convivencia de sus usuarios, por lo que realmente podemos considerarlo como un verdadero universo paralelo. Es más, cada vez, las fronteras entre el mundo virtual y el mundo real son más difíciles de discernir, no sólo por el desarrollo y la emergente popularización de las tecnologías de realidad aumentada², que llenan últimamente las

¹ El ministro de Defensa de los Estados Unidos considera que el ciberespacio es el quinto dominio de las operaciones militares, junto a las terrestres, marítimas, aéreas y espaciales.

² Realidad aumentada se refiere a la visión a través de un dispositivo de un entorno físico real cuyos elementos se combinan con elementos virtuales para la creación de una realidad mixta en tiempo real.

calles de gente cazando *Pokemons*¹, sino por los cada vez más frecuentes casos de personas incapaces de diferenciar su entidad real de su proyección digital, dándole en muchos casos más importancia a ésta última. No nos corresponde a los ingenieros informáticos el análisis de este fenómeno, que descubre un fértil campo de investigación a las ciencias sociales y, en algunos casos a la psiquiatría. Pero me atrevo a decir que estamos ante una reacción patológica originada, entre otras causas, por el aburrimiento de la Humanidad.

Si consiguiéramos agrupar y analizar toda la información que hay sobre una persona en el ciberespacio, podríamos crear un modelo más o menos veraz de su personalidad, cuya abstracción podríamos denominar individuo virtual. Esa agrupación y análisis es el que el *big data* automatiza para ofrecer a las empresas el conocimiento necesario que permite captar y fidelizar clientes los cuales, aunque virtuales, consumen productos y servicios sufragados por el bolsillo real. Es evidente que, ni la aplicación del *big data* se limita a la información personal, ni su único objetivo es a la captación y fidelización de clientes. Sin embargo, lo cotidiano del ejemplo me facilitará mi propósito, que no es otro que hacer tangible el alcance real de estas tecnologías. Después, será sencillo extrapolar el ejemplo a cualquier otro ámbito.

Me atrevo a afirmar sin titubeo que todos los aquí presentes tenemos nuestro 'yo digital' más o menos definido y generado con mayor o menor consciencia. Sin darnos cuenta, constantemente se vierte información acerca de nosotros en el ciberespacio, sea por nosotros mismos o por terceros, creando ese individuo digital mediante el reflejo en la red de nuestras acciones, vivencias, gustos u opiniones. Sólo con reflexionar un poco encontramos multitud de artefactos que realizan este cometido. El más elemental es el ordenador personal, desde el que accedemos a los diferentes servicios de internet, principalmente al correo electrónico, los contenidos de la

¹ *Pokémon GO* es un videojuego de realidad aumentada desarrollado por Niantic Inc. para dispositivos móviles.

Web y las redes sociales. Pero el más importante, en términos de cantidad, calidad y variedad de información registrada, es sin duda el teléfono móvil. Con él, la información se transmite en el instante y desde el lugar en los que se produce la acción, lo que agrega, a menudo sin ser conscientes de ello, una componente espacial y temporal a la información que multiplica su valor. Por otro lado, la reciente llegada de la *wearable technology* o dispositivos ‘ponibles’ o ‘vestibles’ ha hecho que relojes, pulseras, gafas o anillos se conviertan en objetos inteligentes con sensores biométricos que recogen constantemente información sobre nosotros. Igualmente debemos contemplar todos los objetos que tenemos en casa conectados a internet, como las *smartTV* o los dispositivos domóticos, así como nuestro propio coche. Todo ello forma parte de lo que se ha acordado en llamar el Internet de las cosas (IoT). Tampoco podemos olvidar otros sistemas externos que recogen información cuando interactuamos con ellos, como por ejemplo cuando pagamos con una tarjeta de crédito en el supermercado, cuando un radar de tráfico captura una infracción o cuando el médico nos carga en nuestra tarjeta sanitaria una receta digital. En definitiva, un sinnúmero de pequeñas informaciones que van añadiéndose como líneas a un diario cibernético casi exhaustivo.

Afortunadamente aún estamos lejos de que toda esa información acerca de nosotros pueda ser relacionada desde un único punto, ya que, por un lado no existe la interconexión total entre los sistemas, y por otro, muchos de esos datos están protegidos legalmente. Sin embargo, cada vez estamos más cerca de ese estado, ya que las compañías cuyos servicios digitales consumimos a diario están consiguiendo que seamos nosotros mismos quienes les proporcionemos esas conexiones que faltan y les otorguemos esos permisos que necesitan, aprovechando la despreocupación generalizada y las lagunas existentes en las diferentes legislaciones respecto a la protección de datos a nivel internacional.

EL NEGOCIO DEL CONOCIMIENTO

Las compañías nos ofrecen tras un simple clic miles de servicios, cuyas ventajas de uso y, sobre todo, su aparente gratuidad, hace que los consumamos sin reparar en las consecuencias. Llegados a este punto, planteo dos preguntas que, de nuevo, corresponde analizar a otras disciplinas pero cuya respuesta viene condicionada por aspectos técnicos tangentes, o yo diría que incluso secantes: primero, ¿realmente son gratuitos estos servicios?; y segundo, ¿somos libres a la hora de elegir si usarlos o no? A estas alturas, la respuesta a la primera de estas cuestiones parece clara. La respuesta a la segunda es más compleja. La deriva de nuestro entorno nos arrastra cada vez más a consumirlos si no queremos convertirnos en marginados digitales. Para ilustrar estas reflexiones voy a poner un ejemplo que creo les resultará cercano.

El usuario tipo es una persona que suele tener al menos: una cuenta de correo electrónico personal en un servidor gratuito, por ejemplo en Gmail; un perfil en una red social generalista, como Facebook; y, dando por indiscutible la prácticamente obligada tenencia de un teléfono móvil, por ejemplo con sistema operativo Android, usa una aplicación móvil de mensajería instantánea, pongamos por caso WhatsApp. Como supondrán, estos ejemplos no están escogidos al azar.

Gmail superó hace años a Outlook como plataforma de correo electrónico más utilizada en el mundo, no por su calidad como gestor de correo, sino más bien por la cantidad de servicios a los que podemos acceder con un única cuenta, la de Google. A buen seguro, todos los aquí presentes, tengamos o no con una cuenta en Google, hemos utilizado alguno de sus servicios: Google Maps, Google Calendar, YouTube, o simplemente el buscador original. ¿Sabían que si los utilizamos con nuestra sesión abierta en el navegador, toda la actividad queda registrada y asociada a nuestra cuenta de usuario?

Pero hay mucho más. Android, que como todos saben es propiedad de Google desde 2005, domina las plataformas móviles con una cuota de mercado superior al 90% en España. El aprovechamiento de nuestro terminal Android se ve seriamente mermado si no lo tenemos asociado a una cuenta de Google, principalmente porque sin ella no podemos descargar aplicaciones de Google Play ni podemos sincronizar la información que tenemos almacenada en los otros servicios. Ello hace que prácticamente nadie se resista a asociar su cuenta al teléfono o crear una si no la tiene. Esta cuenta de Google vinculada a nuestro terminal está constantemente activa, lo que le permite, de igual manera que ocurre en con el navegador Web, registrar toda nuestra actividad y asociarla a nuestra cuenta. Además, como ya comenté anteriormente, esta actividad esta localizada geográficamente y marcada temporalmente.

Les invito a que hagan la siguiente prueba –por favor, una vez concluido el acto-. Desde su cuenta de Gmail, abran la aplicación *My Account*¹, busquen cualquier fecha y utilicen el servicio ‘historial de ubicaciones’ para situar la información. El resultado les sorprenderá. Gracias a esto puedo saber, por ejemplo, dónde estuve y qué hice el día que nació mi segundo hijo –para mí una sensación entrañable, pero a la vez inquietante, al pensar que Google también lo sabe-. Siento decirles a los usuarios de iPhone que pueden dejar de congratularse. Apple hace exactamente lo mismo.

El modelo de ecosistema de aplicaciones interconectadas no sólo ha reportado a Google un éxito indiscutible, sino que está consiguiendo lo que comenté anteriormente, es decir, la posibilidad de relacionar nuestra información dispersa en la red, haciéndola accesible para ellos desde un único punto: nuestra cuenta de usuario. Sospecho además que esto es sólo la punta del iceberg, y que hay mucho más detrás otros dispositivos como Google Glass, Google Car o los *smart watches*.

¹ Artículo publicado en El País del 7 de agosto de 2016:

http://verne.elpais.com/verne/2016/08/07/articulo/1470564011_821392.html

Por otro lado, Facebook prácticamente monopoliza el mercado de las relaciones personales en la red, superando los 1500 millones de usuarios activos que representan más de 70% de los usuarios adultos online. En este aspecto relega a Google al segundo puesto del ranking, con algo más de 1000 millones de usuarios. La principal fuente de ingresos de Facebook es la publicidad segmentada, la cual implementa basándose en las preferencias personales de cada usuario según su interacción con amigos o páginas de empresas u organizaciones. El algoritmo de Facebook aplica diversas técnicas de análisis inteligente de datos para modelar nuestro perfil como consumidor a partir de las historias que escribimos y que leemos, así como de nuestras reacciones ante las mismas: si nos gustan o no, si las compartimos, si las comentamos y cómo lo hacemos o si las etiquetamos. No sé si coincidirán conmigo, pero después de lo visto, me parece cuanto menos ingenuo pensar que la utilización de esta información se limita sólo a la segmentación de usuarios con fines comerciales. Recordemos que hace unos años salió a la luz que Facebook manipuló las cuentas de 700.000 usuarios para hacer un experimento psicológico tratando de modelar el comportamiento de las personas cuando observaban noticias positivas o negativas. Me inquieta pensar que se pueda cambiar a la opinión de una persona mostrándole determinados mensajes en determinados momentos. Si lo pienso desde un punto de vista político o ideológico, me aterra.

Facebook adquiere WhatsApp en 2015, plataforma de mensajería instantánea de sobra conocida que ha alcanzado la cifra de los 1000 millones de usuarios en el presente año, de los que más del 70% son activos diariamente. Con este número de usuarios, esta cuota de uso y sus expectativas de crecimiento, a priori parece obvio dónde está el negocio. Sin embargo, aunque el modelo de negocio estaba basado inicialmente en el pago anual, concretamente 1 dólar al año por usuario, pronto advirtieron que esta decisión mermaría significativamente su expansión, lo que llevó a la compañía a confirmar en enero de este mismo año que el servicio sería gratuito de por vida. WhatsApp tampoco ofrece publicidad, al menos de momento, luego esa rama de ingresos también queda podada. Hace

unos meses se habilitó el cifrado de extremo a extremo, el cual evita que cualquier mensaje pueda ser leído por alguien que no sean el emisor y el receptor, incluida la propia compañía. Entonces, si no se obtienen ingresos por uso ni por publicidad y tampoco pueden acceder a la información que los usuarios intercambian, ¿qué empuja a Facebook a invertir en esta compra 19.000¹ millones de dólares? La respuesta es sencilla: no compraban la plataforma, sino sus usuarios y lo que es más importante, la red de contactos de cada uno de ellos.

Robin Dunbar², profesor de Psicología Experimental de la Universidad de Oxford, ha demostrado recientemente que la media de amigos de un usuario en Facebook está en torno a 150, de los cuales sólo 4 pueden considerarse amistades verdaderas. Sin embargo, los contactos de WhatsApp son mucho más cercanos por dos razones. La primera que la comunicación es privada, más interactiva y más rápida, lo que nos permite tener abierto constantemente un canal personal, el cual solemos utilizar únicamente con nuestros contactos más allegados. La segunda es que la cuenta va vinculada al número de teléfono, y solemos ser muchos más celosos a la hora de facilitar nuestro teléfono que nuestro usuario de Facebook. Podemos decir entonces que la red de contactos es más pequeña y de mejor calidad, y por tanto más fiable. Ese es, a mi entender, el verdadero valor detrás de esa compra aparentemente fuera de mercado. Sólo hace unos días⁴, WhatsApp ha anunciado un cambio en sus términos de uso y política de privacidad que contempla el intercambio de información con Facebook, además de otras filiales como Instagram. Parece que el círculo se ha cerrado.

¹ Según la información enviada a la Comisión del Mercado de Valores de EE.UU., Facebook compra WhatsApp por esta cantidad, la cual supone más de cinco veces sus ingresos netos anuales.

² Dunbar R.I.M. 2016 *Do online social media cut through the constraints that limit the size of offline social networks?* R.Soc. opensci. 3:150292. DOI: 10.1098/rsos.150292

⁴ El anuncio se produjo el 25 de agosto de 2016, lo que me obligó a cambiar la última frase de este párrafo (“Si hacen uso de esa información o no aún no lo sabemos”) por la actual.
http://tecnologia.elpais.com/tecnologia/2016/08/25/actualidad/1472130602_996229.html

En definitiva, estas dos empresas, Google y Facebook, junto con sus filiales, se reparten gran parte del pastel de la información digital que transita por Internet. Saben de nosotros más que nosotros mismos. Evidentemente, el resto de plataformas no son desechables, pero estimo que lo expuesto es lo suficientemente ilustrador.

En el ámbito personal, los peligros que derivan de la gestión de los datos son, principalmente, los relacionados con la privacidad. La propiedad de los datos depende de varios factores como el proveedor del servicio, la legislación del país en el que se almacenan y la forma en la que fueron generados. Con el avance del Internet de las Cosas, la situación se complicará. Cuando, por ejemplo, los vehículos estén conectados en red e intercambien información, ¿serán los conductores, los propietarios o los fabricantes los dueños de los datos que se generan? La legislación en cuestión de propiedad intelectual y protección de datos quizá no ha avanzado al ritmo requerido por la evolución de la tecnología. El panorama que empieza a vislumbrarse abre la posibilidad de nuevos escenarios, en algunos casos inimaginables aún, que supondrán un desafío para los legisladores. No podemos ni debemos pretender frenar el avance de la tecnología, pero sí regular su uso.

EL RETO DE LA UNIVERSIDAD

Como mencioné al principio de esta lección, la capacidad para transformar las amenazas derivadas de la inminente llegada de la cuarta revolución industrial en oportunidades dependerá de la preparación de la sociedad para hacer frente a los retos que está nos planteará. Es éste un aspecto en el que la Universidad, como protectora del saber, generadora de conocimiento e ilustradora del ser humano, debe asumir su responsabilidad y exigir al resto de instituciones que se comprometan con la suya. Ya lo dicen nuestros estatutos en su artículo 3 sobre la Misión de la Universidad: “[...] la Universidad Pablo de Olavide está al servicio de la sociedad y se define como un lugar de reflexión y pensamiento crítico comprometido con la contribución al progreso, [...] y con la respuesta a las necesidades y problemas de la sociedad contemporánea [...]”

En España, y concretamente en nuestra universidad, el compromiso que viene del lado de la investigación, el desarrollo y la innovación, está siendo razonablemente cubierto, si tenemos en cuenta los escasos recursos de los que disponemos y soslayando, a mis luces, la escasa transferencia tecnológica existente. Pese a ello, no podemos competir en este aspecto fuera de nuestras fronteras más que en áreas muy concretas. La posibilidad de colocarnos siquiera cerca de la vanguardia investigadora europea pasa por un incremento significativo de la inversión en I+D+i, incremento que hoy por hoy nos resignamos a ver como una utopía.

Sin embargo, la otra parte de ese compromiso, la académica, presenta quizá expectativas diferentes. Los estudiantes que precisamente hoy comienzan su andadura en esta casa, esperan terminar sus estudios de grado en el año 2020, ese horizonte que aparece recurrentemente. Mirando un poco más allá, la mayoría de los

niños y niñas –se estima que un 65%¹– que hoy cursan educación primaria tendrán que dedicarse a profesiones que aún no existen, muchas de las cuales ni siquiera las podemos imaginar. Las universidades deben preocuparse desde este mismo momento por la formación de estos estudiantes, presentes y futuros, adaptando los programas formativos de los títulos que ya están en marcha y diseñando nuevos títulos acordes con la demanda competencial esperada. Si no lo hacen, si no lo hacemos, podría darse la paradoja de pasar de un paro juvenil casi generalizado a no poder cubrir los nuevos puestos de trabajo por falta de cualificación.

No podemos olvidar que serán las empresas las que exigirán esas nuevas competencias a nuestros egresados. Por ello, en mi opinión, las universidades públicas, porque las privadas ya lo hacen, tendrán que intensificar la participación de las empresas tanto en el diseño de los programas formativos de los nuevos títulos como en la impartición de los mismos. Pero para ello, la oferta a las empresas debe resultarles más atractiva y rentable de lo que es hoy en día. Las empresas perciben la colaboración con las universidades como un gasto, en vez de una inversión. Mientras no revirtamos esta opinión, seguiremos formando a nuestros estudiantes en competencias diferentes a las demandadas en el mercado laboral. Quizá haya llegado el momento de adoptar y adaptar alguna de las fórmulas que tanto éxito han tenido en otras partes del mundo, en los que el binomio Universidad-Empresa tiene una permeabilidad recíproca y constante en ambos sentidos y en todos los aspectos.

Finalmente, la calidad, ese valor tan intangible como necesario, pero cuyo significado genérico se ha degradado y envilecido últimamente por la insistencia desmesurada de incluirla en todas partes. La mejora de la calidad docente está cada vez más enfocada hacia el “cómo se enseña”, aspecto crucial pero no exclusivo, ya que este sesgo está eclipsando cada vez más lo que para mí es

¹ Dato extraído del informe del Foro Económico Mundial de Davos 2016:
http://www3.weforum.org/docs/WEF_AMNC16_Report.pdf

fundamental: el “qué se enseña” en nuestras universidades. Observo con preocupación una degradación paulatina de la calidad en este último sentido, provocada por la inclusión de programas formativos más propios de escuelas profesionales que de centros universitarios, así como por la aquiescencia de modas e innovaciones desprovistas de fondo. De nada servirá la personalización de la enseñanza con los medios tecnológicos que hoy he esbozado aquí si las competencias adquiridas carecen de contenido y no proporcionan a los profesionales del futuro una actitud de respeto y admiración por el conocimiento.

EPÍLOGO

Avanzamos inexorablemente hacia una época que cambiará por completo nuestra manera de vivir, de pensar y de actuar. La economía, el empleo, la política, la enseñanza, la sanidad... absolutamente todos los ámbitos de la sociedad se verán alcanzados por ese trascendental desafío que debemos afrontar con cauteloso optimismo. En el corazón de la transformación está el desvanecimiento de las fronteras entre lo físico y lo virtual. Este hecho elevará el crecimiento de la inteligencia colectiva¹ a cotas difíciles de imaginar, pues su fuente de información ya no se limitará a las personas, sino que incluirá también a las máquinas.

Como la mayoría de las cosas en la vida, no es la tecnología en sí misma, sino la finalidad con la que se usa lo que podemos calificar de buena o mala. En este sentido, la mala utilización de la información provoca que los usuarios pierdan la confianza en las nuevas tecnologías y obvien, por tanto, los beneficios que pueden desprenderse de las mismas. Las ventajas del análisis inteligente de datos masivos son innumerables y no deben ser eclipsadas por un potencial uso comercial o incluso perverso de las tecnologías que lo implementan.

El temor a que las tecnologías inteligentes puedan desnudar nuestra privacidad y hacernos rehenes de quienes las controlen no es infundado. Sin embargo, haciendo una lectura senequista, la capacidad de convertir ese temor en confianza pasa por concienciar a la sociedad de las consecuencias del uso de las nuevas tecnologías para no dejarla caer en la *ciberinopia*. Conocer y hacer conocer el presente tecnológico sobre el que se construirá el futuro y transmitirlo a las generaciones venideras se revela como un aspecto capital.

¹ La inteligencia colectiva es una forma de inteligencia que surge de la colaboración de muchos individuos, generalmente de una misma especie.

Me ilusiona pensar que la educación universitaria puede ser un espacio en el que alcancemos un alto grado de competitividad dentro y fuera de nuestras fronteras, entendiendo la misma en términos de preparación de nuestros egresados. Sin embargo, mientras la *secundarización* de las aulas universitarias siga generalizándose, la burocratización siga entorpeciendo las labores académicas y las instituciones gubernamentales sigan alejando la educación del centro del desarrollo económico, esta posibilidad se irá enterrando hasta quedar dilapidada para siempre.

Por último, en el contexto doméstico, donde aparentemente sí tenemos más capacidad de influencia, nuestros hijos aprenden a usar las nuevas tecnologías de manera autónoma y natural, sin miedo, pues las perciben como una parte más de la vida. Al igual que participamos en otras facetas de su desarrollo personal, tenemos que hacerlo asimismo en el plano que hoy abordamos para inculcarles la necesidad de hacer un uso responsable y mesurado de las nuevas tecnologías.

Como en todo, el tiempo me dará o me quitará la razón. Afortunadamente no tendré que esperar mucho para resolver la incógnita. Espero haber conseguido despertar su curiosidad, o al menos que algunos de los pasajes de esta lección les haya sorprendido o extrañado. Como dijo Ortega y Gasset: “Sorprenderse, extrañarse, es comenzar a entender”.

Desde el corazón, muchas gracias por acompañarme.



UNIVERSIDAD

**PABLO[®]
OLAVIDE**

SEVILLA