



PERIODISMO COMPUTACIONAL: EVOLUCIÓN, CASOS Y HERRAMIENTAS

Computational journalism: Evolution, cases and tools



Mari Vázquez y Lluís Codina



✉ **Mari Vázquez** es profesora asociada de la *Universitat Pompeu Fabra* de Barcelona desde 2006 y bibliotecaria en la *Universitat Oberta de Catalunya*. Es doctora por la *Universitat Pompeu Fabra* de Barcelona desde 2015. Imparte docencia en los grados de Periodismo y de Comunicación Audiovisual, y en el *Master universitario online en Documentación Digital* del *Instituto de Educación Continua*. Es miembro del grupo de investigación *DigiDoc* y forma parte de la *Unidad de Investigación en Periodismo*, grupo consolidado reconocido por la *Generalitat de Catalunya*.
<https://orcid.org/0000-0002-3284-2590>

Universitat Oberta de Catalunya, Biblioteca
Rambla del Poblenou, 158. 08018 Barcelona, España
mari.vallez@upf.edu

Universidad Pompeu Fabra, Departamento de Comunicación
Roc Boronat, 138. 08018 Barcelona, España
mvallez@uoc.edu



Lluís Codina es profesor de la *Universitat Pompeu Fabra* de Barcelona. Imparte docencia en la *Facultad de Comunicación*, en los grados de Periodismo y de Comunicación Audiovisual. Es coordinador del *Master Universitario en Comunicación Social (MUCS)* de esa universidad e imparte docencia también en los *Masters universitarios online en Documentación Digital* y en *Buscadores*, del *Instituto de Educación Continua*. Es miembro promotor del grupo de investigación *DigiDoc*, y coordinador de su Seminario de investigación, así como del *Observatorio de Cibermedios*. Forma parte de la *Unidad de Investigación en Periodismo*, grupo consolidado reconocido por la *Generalitat de Catalunya*.
<http://orcid.org/0000-0001-7020-1631>

Universidad Pompeu Fabra, Departamento de Comunicación
Roc Boronat, 138. 08018 Barcelona, España
lluis.codina@upf.edu

Resumen

El periodismo computacional ha irrumpido con fuerza en el intenso panorama de las innovaciones periodísticas, debido a su gran potencial de impacto en las redacciones. Dada la reciente implantación de esta especialidad, todavía en fase experimental en las redacciones que lo han incorporado, en este trabajo presentamos un resultado de tipo exploratorio sobre el periodismo computacional, a la vez que lo relacionamos con variaciones bien implantadas del periodismo, como el periodismo de datos. El trabajo se ha realizado a partir de la revisión bibliográfica y de los principales informes sobre esta materia, así como a partir de una serie de estudios de caso. Los resultados pueden ayudar tanto a enfocar nuevas investigaciones en este campo, como a completar el panorama de la innovación para profesionales e investigadores del sector que necesiten disponer de una visión global del impacto que puede alcanzar el periodismo computacional.

Palabras clave

Periodismo computacional; Periodismo de datos; Periodismo algorítmico; Periodismo aumentado; Periodismo automatizado; Robot periodista; Periodismo estructurado; Innovación periodística.

Abstract

Computational journalism has burst into the intense panorama of journalistic innovations, due to its great impact potential on newsrooms. Given the recent implantation of this specialty, still in the experimental phase in the newsrooms that have incorporated it, in this work we present an exploratory result on computational journalism, at the same time that we relate it with well-established variations of journalism, such as data journalism. The work has been carried out from the review of the main reports on this matter, as well as from a series of case studies. The results can help both to focus new research in this field, and to complete the panorama of innovation for professionals and researchers in the sector who need to have a global vision of the impact that computer journalism can have.

Keywords

Computational journalism; Data journalism; Algorithm journalism; Automated journalism; Robot journalism; Structured journalism; Journalism innovation.

Artículo recibido el 11-03-2018
Aceptación definitiva: 09-07-2018

1. Introducción

El análisis de la implantación y el impacto de las tecnologías en los medios de comunicación es un tema recurrente que ha sido tratado desde numerosas perspectivas. Está plenamente implantado el concepto de periodismo digital, incluso en su denominación más académica, cibermedios. Sin embargo se queda corto para hacer referencia a las nuevas tendencias que comienzan a instaurarse en los medios internacionales de mayor prestigio.

Se ha empezado a utilizar el concepto de periodismo computacional en referencia a los sistemas que automatizan parte de los procesos de creación y de personalización de contenidos periodísticos. Este concepto puede tener dos lecturas:

- Continuidista: Utilización de la tecnología para agilizar y optimizar las tareas realizadas por los periodistas. Supondría el proceso natural de evolución y consolidación de la tecnología en las redacciones;
- Disruptiva: Implica nuevas formas de hacer periodismo, cuando las redacciones incorporan el uso intensivo de la tecnología en fases de sus procesos que hasta hace poco eran impensables. Su utilización conlleva en este caso el uso de técnicas y herramientas que permiten abordar el periodismo desde nuevos enfoques que de otra forma no serían viables.

En el modelo continuista tendría cabida el periodismo de investigación que a través del procesamiento masivo de datos ha logrado adquirir una dimensión inédita. Un excelente ejemplo sería el caso de *The Panama Papers*. Este modelo se puede identificar también con el así llamado periodismo de datos.

El concepto de periodismo computacional puede tener dos lecturas, una continuista y otra disruptiva

El modelo disruptivo no está tan extendido y se caracteriza principalmente por la creación automatizada de noticias para ámbitos temáticos concretos, como pueden ser los resultados deportivos y la información financiera de las empresas. En ambos casos se trabaja con datos estructurados que se integran en un pequeño relato o noticia. Algunos medios llevan ya algún tiempo experimentando con estos sistemas, por ejemplo *The Washington Post* y *Los Angeles Times* entre otros.

A nivel económico la incorporación intensiva de la tecnología en los medios de comunicación también tiene un impacto directo:

- proporciona una diferenciación respecto a la competencia que puede resultar una ventaja competitiva muy valiosa para el medio;

- puede ayudar a utilizar los recursos de forma más eficiente y reducir costes en un sentido amplio. Al hablar de este aspecto, casi siempre se asocia exclusivamente al impacto que puede tener en la reducción de puestos de trabajo, aunque tiene repercusiones en muchos otros contextos. Nuestro artículo no se centra en este aspecto, aunque hay que tener en cuenta que todos los cambios implican reajustar el sistema con unas cuestiones positivas y otras negativas.

Los análisis de tendencias demuestran que el uso de la tecnología es uno de los elementos clave a la hora de plantearse la estrategia y evolución de los medios. La evolución tecnológica ya no afecta sólo a los procesos y a los canales de difusión, sino también a la creación y conceptualización de las noticias. Por tanto el periodismo computacional no compite con el mejor periodismo tradicional, sino que es un aliado de éste, en especial al liberar recursos que hasta ahora se dedican a tareas repetitivas (como la compilación de resultados deportivos o datos financieros) que pueden dedicarse por ejemplo a periodismo de investigación.

El objetivo de este artículo es realizar una exploración del concepto de periodismo computacional y mostrar sus usos y su alcance. Para ello se han establecido dos objetivos específicos:

- i) analizar las dimensiones que impactan en el periodismo computacional;
- ii) mostrar su vertiente práctica: casos de uso en los medios de comunicación y herramientas utilizadas.

2. Metodología

Se ha utilizado una combinación de metodologías, basada en la revisión bibliográfica sistematizada y en el análisis de casos. Para llevar a cabo la revisión bibliográfica se ha recogido la producción científica indexada en las bases de datos *Web of Science* y *Scopus* desde los últimos 7 años utilizando las siguientes palabras clave: periodismo computacional – *computational journalism*–, periodismo robot –*robo or robot journalism*–, periodismo automático –*automated journalism*–, periodismo algorítmico –*algorithm journalism*–.

También se han identificado los informes emitidos por las organizaciones internacionales de más prestigio (*Reuters Institute, Nieman Journalism Lab, Poynter Institute, The Tow Center for Digital Journalism*...) y publicados en los últimos tres años. La selección de los documentos se basa en el interés de cubrir las diferentes dimensiones del periodismo computacional y en el análisis tanto desde su vertiente profesional como académica.

Así mismo hemos llevado a cabo el análisis de casos que se muestran en el apartado cinco del estudio. Adicionalmente hemos revisado diversas herramientas relacionadas de las que también damos cuenta en el apartado correspondiente de este trabajo.

El artículo se organiza de la siguiente forma: en la próxima sección se presenta la revisión bibliográfica, detectando tendencias en los medios y la evolución del concepto de periodismo computacional desde sus inicios. En la sección cuarta se muestra su transdisciplinariedad, mostrando las disciplinas que convergen en él. En el apartado quinto se realiza un análisis del impacto que tiene en las redacciones y en el sexto se recopila una selección de aplicaciones que a pesar de su heterogeneidad pueden agruparse bajo la etiqueta de periodismo computacional. A continuación se presenta la discusión y por último las conclusiones.

3. Marco teórico

La terminología utilizada para el concepto de periodismo computacional es muy variada porque no existe una línea clara que limite su alcance. **Coddington** (2015) intenta desentrañar la complejidad del concepto bajo la idea del 'giro cuantitativo del periodismo' que destaca por promover el trabajo en red, usar de forma intensiva los *big data*, y fomentar la participación del público. **Hamilton y Turner** (2009) lo entienden como el conjunto de herramientas que los periodistas utilizan para descubrir, explicar y distribuir historias, que también utilizan los algoritmos para crearlas.

A continuación se recogen las denominaciones utilizadas:

- periodismo robot (*robot journalism*) (**Carlson**, 2015; **Montal**; **Reich**, 2016);
- periodismo automático (*automated journalism*) (**Graefe**, 2016; **Lindén**, 2017a);
- periodismo algorítmico (*algorithm journalism*) (**Diakopoulos**, 2015; **Dörr**; **Hollnbuchner**, 2017);
- periodismo aumentado (*augmented journalism*) (**Marco**; **Siegman**, 2017);
- periodismo de datos o periodismo basado en los datos (*data journalism, data-driven journalism*) (**Parasie**; **Dagiral**, 2013);
- incluso el concepto pionero de periodismo asistido por ordenador (*computer-assisted reporting*) (**Houston**, 2014; **Meyer**, 1999).

Algunas de estas denominaciones son anteriores a la emergencia del concepto de periodismo computacional, pero muchos autores (como los que hemos ido citando) los relacionan ahora entre sí.

La bibliografía académica y profesional generada sobre esta materia ha ido *in crescendo* durante la última década. Los temas abordados son muy variados:

- opciones tecnológicas e innovadoras utilizadas (**Newman**, 2017; **Shearer**; **Gottfried**, 2017; **Gynnild**, 2014);
- desafíos éticos y legales que comporta la creación de noticias con algoritmos (**Weeks**, 2014);
- conceptos de autoría y copyright (**Montal**; **Reich**, 2016; **Carlson**, 2015);
- necesidad de transparencia de los algoritmos para no cuestionar su manipulación (**Diakopoulos**, 2015; **Hamilton**; **Turner**, 2009) y asegurar la integridad de los datos (**Dörr**; **Hollnbuchner**, 2017);
- percepción del lector sobre las noticias generadas

(**Wölker**; **Powell**, 2018; **Graefe et al.**, 2018);

- opinión de los periodistas sobre el uso de robots en las redacciones (**Thurman**; **Dörr**; **Kunert**, 2017; **Jung et al.**, 2017; **Lindén**, 2017b).

En este panorama el periodismo de datos ocupa un papel destacado, puesto que ya es una actividad muy consolidada. No obstante se requiere una nueva fase en la que es necesario centrarse en los modelos de datos para lograr la implementación y expansión del periodismo automatizado (**Caswell**; **Dörr**, 2018). Por ejemplo, en el informe *El impacto de los datos en la era de la postverdad* (**Albert**, 2017), profesionales de la comunicación analizan cómo afronta el sector la revolución de los datos y consideran que su uso es una de las formas de conseguir un periodismo de calidad, ya que permite luchar contra la postverdad. **Cohen**, **Hamilton** y **Turner** (2011) consideran que los avances que la tecnología pueda comportar al periodismo revertirán directamente en los ciudadanos y en la gobernanza de los estados.

Por su parte el informe *Scenarios for the future of journalism* (**Kasem**; **Van-Waes**; **Wannet**, 2015) dibuja cuatro posibles escenarios realistas y plausibles para el periodismo en 2025. Los cuatro escenarios permiten reflexionar sobre la evolución del periodismo en los próximos años y las estrategias a llevar a cabo. Dibujan dos ejes: el grado de disrupción tecnológica y la procedencia de los contenidos.

La tecnología es uno de los elementos clave a la hora de plantearse la estrategia y evolución de los medios de comunicación

Por último pero no menos importante, podemos destacar el informe elaborado por *Reuters Institute, Putting Europe's robots on the map: Automated journalism in news agencies* (**Fanta**, 2017), donde se analiza el uso que hacen las agencias de noticias europeas de la automatización de noticias en las redacciones. Este informe detecta que las agencias han empezado tímidamente a utilizar estas técnicas aunque aún no todas ellas. Centrándose en España y en el entorno del periodismo de datos, el trabajo de **Blanco-Castilla** y **Quesada** (2016) recopila varios capítulos que analizan la implantación del periodismo de datos en este país.

4. Transdisciplinariedad del periodismo computacional

El periodismo computacional se caracteriza por su transdisciplinariedad. Los límites de las disciplinas individuales se trascienden y se aplican perspectivas múltiples para generar conocimiento emergente (**Nicolescu**, 2002). Las dos principales disciplinas que convergen son periodismo e informática, aunque también es interesante estudiar los vínculos con otros campos: documentación, estadística y lingüística. Su transversalidad permite abrir un nuevo espacio de conocimiento que integra diferentes disciplinas y comporta una nueva forma de interactuar con la información y de difusión en los medios.

A continuación se examinan especialidades y actividades que tienen una vinculación estrecha con el periodismo y que utilizan la tecnología de forma intensiva. Por ello se pueden agrupar bajo la etiqueta de periodismo computacional, entendiendo en este caso el concepto de la forma más amplia posible.

Una primera especialidad es el periodismo de datos, que se basa en la explotación e interpretación de grandes colecciones de datos estructurados para descubrir patrones o tendencias a partir del análisis estadístico, y así poder extraer historias periodísticas. Para trabajar con los datos se requieren programas que permitan limpiarlos, procesarlos y visualizarlos de forma atractiva. Los datos suelen proceder de las administraciones públicas, aunque cada vez hay más diversidad de fuentes de datos. También se pueden implementar mecanismos para obtener los datos directamente, por ejemplo rastrear la Web para obtener información que pueda convertirse en datos.

Cuando la información es desestructurada se pueden aplicar técnicas de procesamiento del lenguaje natural (Vázquez; Pedraza-Jiménez, 2007) para extraer información de forma automática. Por ejemplo se identifican en documentos las entidades que aparecen, como pueden ser nombres de personas, organizaciones o lugares. También pueden obtenerse las relaciones semánticas que se establecen entre los términos de los documentos a partir del análisis lingüístico. Todo ello con el objetivo de comprender y/o aprehender el lenguaje natural para obtener nueva información. Estas técnicas se pueden aplicar junto a la minería de datos para realizar un análisis de los sentimientos de los documentos (Arcila-Calderón; Barbosa-Caro; Cabezuelo-Lorenzo, 2016), entre otros ejemplos. Sin embargo el procesamiento de información desestructurada aún se encuentra en fase incipiente. La aplicación de la inteligencia artificial basada en técnicas de aprendizaje automático o de ingeniería del conocimiento para procesar la información empieza a ofrecer resultados que van más allá de la experimentación.

Vinculada a este aspecto ha surgido una nueva corriente denominada periodismo estructurado, que prima el acceso directo a los datos y a los documentos para permitir la construcción del relato narrativo según la interactividad que cada usuario realiza cuando consulta la pieza informativa, y que además puede ir evolucionando con el paso del tiempo (Freixa; Pérez-Montoro; Codina, 2017). Esto implica un cambio en el concepto de creación periodística, ya que se requiere una estructura que permita la integración y reutilización de contenidos. Se trata de crear procedimientos que permitan maximizar la vida de las noticias, con la utilización de metadatos que ayudan a almacenar la información de forma estructurada, técnicas propias de la documentación.

En este contexto la difusión de la información implica nuevas formas para transmitirla que van más allá del texto y la imagen, por ello la visualización de la información adquiere un papel muy relevante. En esta disciplina se usan diferentes técnicas (diagramas, gráficas, esquemas, etc.) para facilitar la aprehensión, la interpretación y la comunicación de los contenidos (Pérez-Montoro, 2009). La información visual cada vez ocupa un papel más destacado en la comunicación

por el gran impacto que consigue, sin embargo hay que tener en cuenta que la vertiente estética siempre debe estar supeditada a la informativa en el periodismo (Cairo, 2011).

Los programas informáticos de generación de contenidos [proceso conocido en inglés como *robot journalism* (Carlson, 2015; Lindén, 2017b)], han empezado a ser utilizados en las redacciones de algunos importantes medios de comunicación. Esto es posible en temas en los que la base principal de la noticia está constituida por datos que se pueden vincular entre sí por unas estructuras lingüísticas muy concretas. Las crónicas deportivas de competiciones, las noticias sobre la actividad económica de las empresas, o la descripción de catástrofes naturales son ejemplos de contenidos generados automáticamente. Además estas técnicas permiten personalizar contenidos y cubrir noticias locales.

“ Su transversalidad permite abrir un nuevo espacio de conocimiento que integra varias disciplinas y que comporta una nueva forma de interactuar con la información y de difusión en los medios ”

Las especialidades recogidas aquí muestran que el periodismo computacional es una nueva área que se caracteriza por su transdisciplinariedad, que permite aproximarse al periodismo desde diferentes perspectivas integrando técnicas y metodologías de otras disciplinas. También hay que destacar que se trata de un entorno muy dinámico en el que constantemente se están produciendo innovaciones.

A continuación se recogen otros aspectos del periodismo en los que la tecnología tiene un papel destacado y que resulta un instrumento de ayuda a los periodistas en el proceso de crear noticias. Algunos son:

- optimización de las noticias para mejorar su posicionamiento web (SEO);
- analítica web de las noticias;
- curación de contenidos.

En el primer caso se trata de redactar noticias que cumplan con las pautas recomendadas para favorecer su aparición en las primeras posiciones de la lista de resultados de un buscador cuando los usuarios realizan una búsqueda a partir de las palabras clave principales para las que se desea posicionar. Existe una variedad de pautas a considerar que pueden ofrecer buenos resultados (García-Carretero *et al.*, 2016; Richmond, 2008). El uso de programas que ayuden a optimizar los contenidos para favorecer SEO *on page* son cada vez más imprescindibles para los periodistas (Wang; Han; Rush, 2016).

Una vez la noticia ha sido publicada, interesa aplicar la analítica web para recopilar información de cómo interactúan los usuarios. De este modo el análisis de los datos puede ayudar a optimizar la experiencia de navegación de los usuarios (Wang, 2018) e incluso ser un elemento a la hora de planificar la estrategia de contenidos del medio, aunque pueda resultar peligroso y tendencioso (Tandoc; Thomas, 2015).

La curación de contenidos consiste en buscar, filtrar y seleccionar información de calidad, para después analizarla, interpretarla y difundirla (Guallar, 2014). El concepto es relativamente nuevo y se asocia a la documentación, aunque se ha vinculando rápidamente al periodismo y a las nuevas formas de generar contenidos (Guerrini, 2014). Algunos ejemplos concretos de cómo se está aplicando en la prensa pueden extraerse de Guallar (2017).

5. Aplicación del periodismo computacional en las redacciones

El periodismo de datos, basado en el procesamiento masivo de datos, ya se encuentra totalmente instaurado en gran parte de los medios de comunicación y se pueden citar innumerables excelentes trabajos que han alcanza-

do un gran impacto. El otro gran eje del periodismo computacional, la creación de noticias de forma automática, no cuenta con una trayectoria tan consolidada, pero sí que hay numerosos ejemplos que muestran que es una tendencia que empieza a imponerse en determinados temas.

A continuación se analizan cuatro casos de estudio que muestran el uso del periodismo computacional en las redacciones. En el primer bloque se muestran dos casos de éxito de la aplicación de periodismo de datos, uno nacional y otro norteamericano. Se trata de proyectos que han tenido gran repercusión internacional y que se han mantenido a lo largo del tiempo y aún continúan vigentes. En el segundo bloque se recogen dos casos que utilizan la tecnología para crear de forma automática noticias.

| | |
|---------------------------------|--|
| Nombre del caso 1 | Medicamentalia https://medicamentalia.org/es |
| Medio de comunicación | Fundación Ciudadana Cívico, España |
| Descripción | Investigación periodística internacional que analiza la brecha global en el acceso a la salud. Consta de tres investigaciones: precios de medicamentos (2015), inmunización y precios de vacunas (2017), y anticonceptivos (2018). |
| Especialidad | Periodismo de datos |
| Año | Primera investigación: 2015; última investigación: 2018 |
| Principales aportaciones | El proyecto tiene como objetivo acabar con la opacidad de las empresas farmacéuticas en lo referente a los asuntos públicos y las relaciones de poder en la salud pública. Es un proyecto vivo, que acepta la participación ciudadana para recabar información. Para su desarrollo ha contado con la financiación de becas de innovación periodística, además de una amplia red de colaboradores y medios de primer nivel para amplificar el alcance de la investigación. Las investigaciones cuentan con un detallado apartado de metodología, que explica de dónde se han obtenido los datos, cómo se han procesado, los problemas que se han sorteado y las limitaciones que presentan. Un ejemplo de la información recopilada es la figura 1, que muestra el precio de 14 medicinas en 61 países. Se comparan los días que un trabajador con un salario medio de cada país debe invertir para comprar un medicamento. |
| Limitaciones | No se cuenta en todos los casos con información actualizada de la OMS. Se combinan fuentes de datos que en algunos casos presentan pequeñas diferencias metodológicas. |

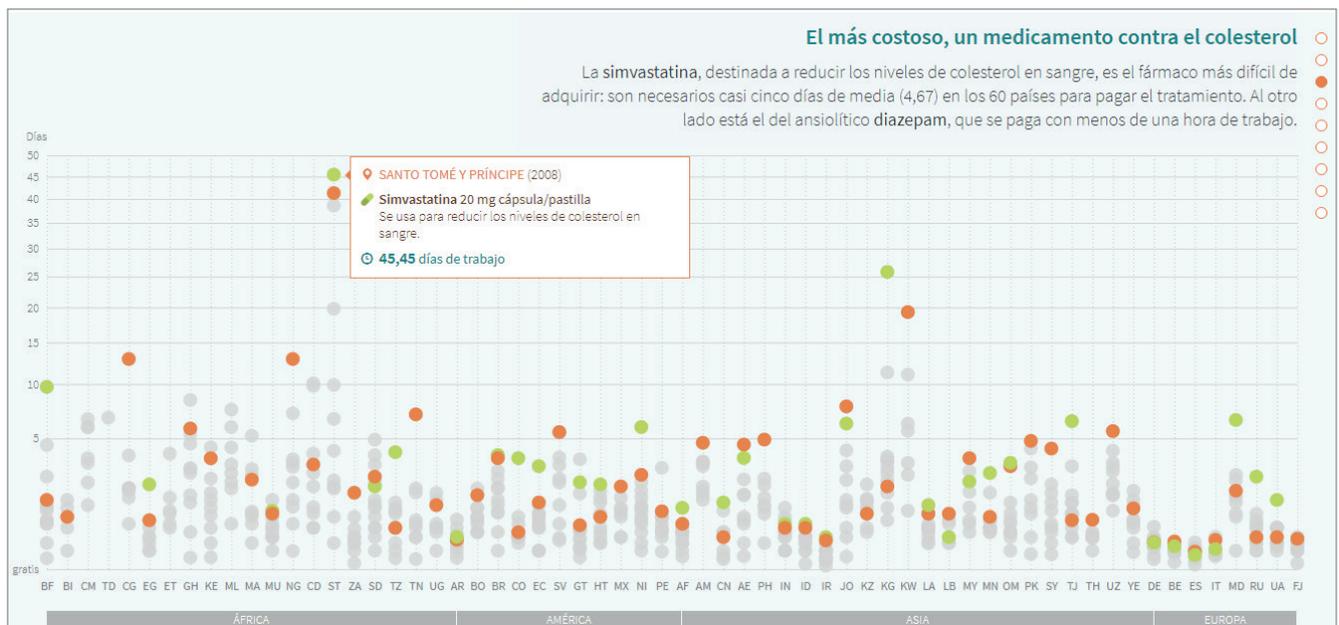


Figura 1. Días de trabajo necesario en cada país para comprar un medicamento. Proyecto de investigación periodística Medicamentalia de Fundación Ciudadana Cívico.

| | |
|---------------------------------|---|
| Nombre del caso 2 | The homicide report http://homicide.latimes.com |
| Medio de comunicación | Los Angeles Times, EUA |
| Descripción | Investigación periodística local, el condado de Los Ángeles, que recoge los datos personales y estadísticos de las muertes por violencia en la zona. |
| Especialidad | Periodismo de datos |
| Año | Proyecto iniciado en 2007 que aún continúa vigente recabando información de víctimas de la violencia. |
| Principales aportaciones | Durante su larga trayectoria el proyecto ha pasado por varias fases. Actualmente proporciona un mapa interactivo, una base de datos y un blog que narra los homicidios acaecidos en el condado. Se recopila cualquier muerte que sea considerada un homicidio por un forense; es decir, la muerte de un ser humano por parte de otro. Se dan los detalles básicos de cada víctima y su raza. Además, se solicita la colaboración ciudadana para ampliar la información. Cada caso se actualiza cuando se realizan arrestos y se juzga a los sospechosos en los tribunales. Al ser un proyecto que abarca un gran período de tiempo ofrece gran cantidad de información retrospectiva que permite entrelazar y vincular la información actual. La figura 2 recoge un ejemplo de la información facilitada de cada víctima. |
| Limitaciones | Inconsistencias al considerar una muerte como homicidio, ya que a veces no coincide el dictamen del forense con el de la policía, ya que no todos los homicidios son actos criminales. Algunos casos de discrepancia son las víctimas de accidentes de tráfico y las muertes por agentes de policía. |

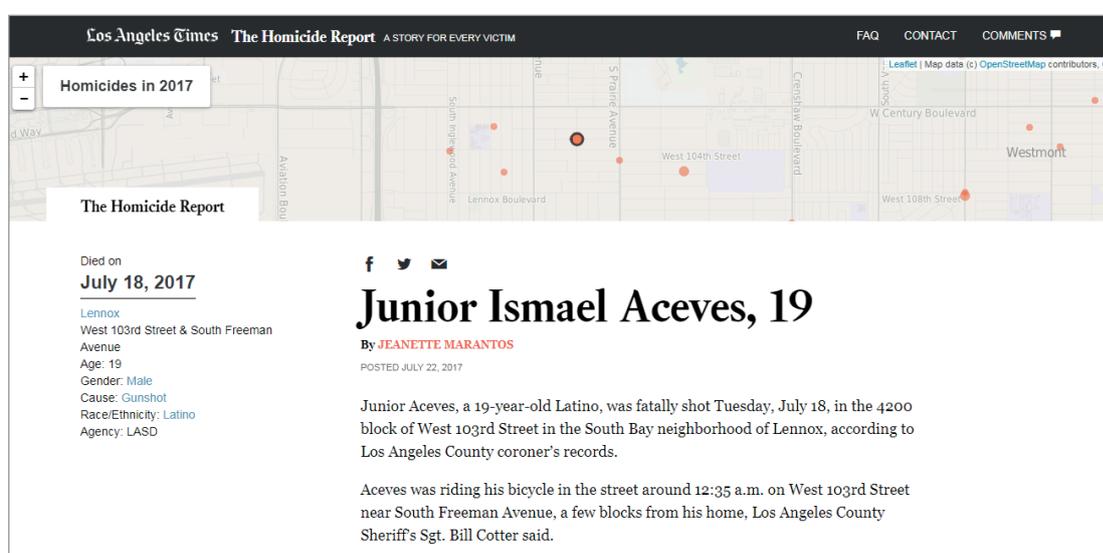


Figura 2. Reportaje en *The homicide report* publicado por *Los Angeles Times*.

Los ejemplos de generación de contenidos con sistemas automáticos, no son tan vistosos como los anteriores, pero sí que son ya una realidad. En EUA llevan años experimentando con estos sistemas, utilizando principalmente la tecnología de las compañías *Automated Insights* y *Narrative*

Science, aunque también algunos medios de comunicación han creado sus propios algoritmos, como *Heliograf* de *The Washington Post* o *Quakebot* de *Los Angeles Times*. En España los ejemplos son más escasos, pero algunos medios ya han empezado a experimentar con esta tecnología.

| | |
|---------------------------------|--|
| Nombre del caso 3 | Contaminación en Barcelona en tiempo real https://www.ara.cat/meteo/contaminacio/noticia |
| Medio de comunicación | Ara.cat, España |
| Descripción | Creación de piezas periodísticas sobre los niveles de contaminación en Barcelona a partir de los datos facilitados por la red de vigilancia y previsión de la contaminación atmosférica. Las noticias son actualizadas constantemente. |
| Especialidad | Creación automática de noticias |
| Año | Proyecto iniciado a principios de 2017 |
| Principales aportaciones | En función de los valores registrados por las estaciones de la <i>Generalitat</i> , un algoritmo procesa los datos en tiempo real de contaminación y redacta automáticamente la historia. El robot es capaz de elaborar una noticia compleja con un titular informativo, un subtítulo y diversos párrafos. La figura 3 muestra una de las noticias generadas con este sistema. |
| Limitaciones | El algoritmo se basa en estructuras lingüísticas previamente definidas y en los límites establecidos por la <i>CE</i> y la <i>OMS</i> de concentración de dióxido de nitrógeno y de partículas en suspensión. |



Figura 3. Noticia de contaminación atmosférica generada automáticamente en Ara.cat.

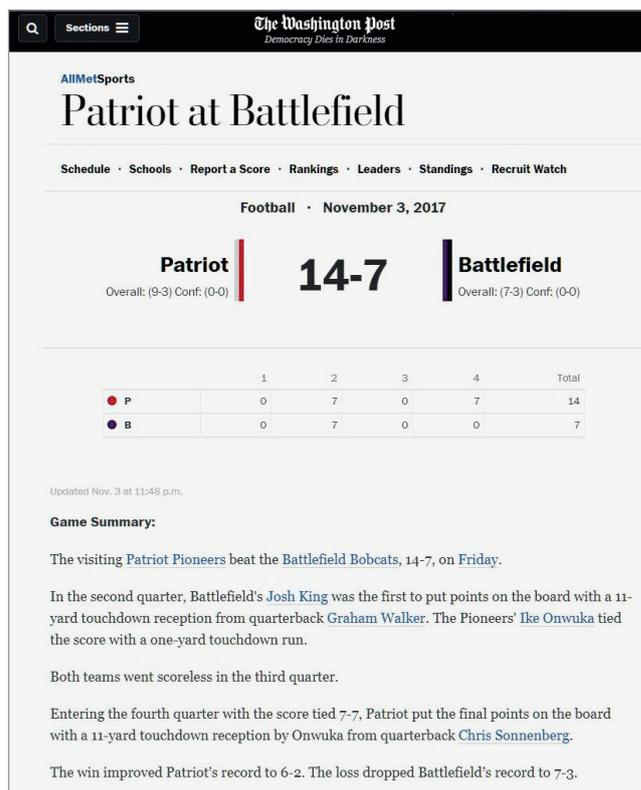


Figura 4. Noticia de la liga de fútbol escolar generada automáticamente en The Washington Post.

| | |
|---------------------------------|---|
| Nombre del caso 4 | Automated storytelling to cover high school football https://www.washingtonpost.com/allmetsports/2017-fall/games/football/87087 |
| Medio de comunicación | The Washington Post, EUA |
| Descripción | Creación de piezas periodísticas sobre los partidos de fútbol de las escuelas secundarias del área de Washington D.C. |
| Especialidad | Creación automática de noticias |
| Año | Proyecto iniciado a mediados del 2017 |
| Principales aportaciones | Cada noticia se basa en los resultados obtenidos por los equipos, principales jugadas, estadísticas de los jugadores y comparativa respecto a otros trimestres, junto con los propios rankings regionales semanales elaborados por The Post. Las historias se actualizan automáticamente cada semana usando los datos presentados por los entrenadores de fútbol de las escuelas de secundaria. La figura 4 muestra una de las noticias generadas con esta aplicación. |
| Limitaciones | Los datos tienen que obtenerse manualmente. |

6. Herramientas utilizadas en periodismo computacional

A continuación se recopilan algunas aplicaciones que tienen un papel destacado en periodismo computacional. No es una lista exhaustiva, pero sí representativa de las últimas tendencias. Se han clasificado en tres grandes grupos:

Gestión y visualización de datos

- *Open Refine*: aplicación gratuita para trabajar con datos sin procesar, que permite limpiarlos, transformarlos en diferentes formatos y organizarlos de una forma rápida.
<http://openrefine.org>
- *Overview*: plataforma de código abierto que ayuda a leer y analizar miles de documentos de manera rápida. Incluye búsqueda de texto completo, visualizaciones, detección de entidades y agrupamiento de temas, en un entorno de

trabajo visual fácil de usar.
<https://www.overviewdocs.com>

- *Tableau Public*: herramienta gratuita que permite crear y compartir gráficos interactivos, mapas y paneles con actualización en tiempo real. Los contenidos creados se publican en la plataforma y pueden ser compartidos a través de la web.

<https://public.tableau.com/s>

- *Flourish*: a partir de hojas de cálculo permite crear gráficos de gran impacto visual y generar historias interactivas. Es gratuito.

<https://flourish.studio>

- *Opta*: recopila, analiza y categoriza eventos en directo de gran variedad de deportes. Los datos se recogen utilizando un estándar independientemente del país o continente y se incluyen en una base de datos que puede consultarse

para ofrecer información en vivo.
<https://www.optasports.com>

Generación de contenidos y procesamiento de lenguaje natural

- *Wordsmith* de *Automated Insights*: plataforma de generación de lenguaje natural que convierte los datos en narraciones que pueden personalizarse.
<https://automatedinsights.com/wordsmith>
- *Quill* de *Narrative science*: sistema basado en la generación de lenguaje natural que transforma automáticamente los datos en narrativas inteligentes a escala, en un lenguaje conversacional, fácil de entender, y destacando para cada caso la información más relevante.
<https://narrativescience.com/#quill-platform>
- *OpenCalais*: desarrollada por *Thomson Reuters*, procesa el texto identificando entidades, relaciones, eventos, temas y etiquetas sociales. Facilita el fichero en formato RDF con toda la información extraída. Permite estructurar la información desestructurada.
<http://www.opencalais.com>
- *API Natural Language* de *Cloud*: aplicación de *Google* que extrae información de textos no estructurados utilizando los modelos de aprendizaje automático de *Google*. También puede utilizarse para captar el sentimiento transmitido en un texto o en las redes sociales.
<https://cloud.google.com/natural-language>

Compartir contenidos

- *DocumentCloud*: herramienta de almacenamiento que permite a los periodistas compartir, analizar, anotar, organizar y publicar documentos originales en la Web.
<https://www.documentcloud.org>
- *Ushahidi*: permite recoger información/datos directamente de los testigos a través de *crowdsourcing*. Se desarrolló para ayudar a la sociedad civil a mejorar el flujo de información de abajo hacia arriba.
<https://www.ushahidi.com>
- *Project Shield*: servicio gratuito que utiliza tecnología de *Google* para proteger webs de noticias de ataques *distributed denial-of-service* (DDoS) en internet. Se trata de un proyecto de *Jigsaw* (anteriormente *Google Ideas*) que tiene como objetivo ayudar a luchar contra la censura en internet, reducir las amenazas de ataques digitales y favorecer la libertad de expresión.
<https://projectshield.withgoogle.com/public>

7. Discusión

Gracias a la revisión sistematizada, a los casos considerados y a las herramientas de las que hemos dado cuenta, hemos podido mostrar que el concepto de periodismo computacional es amplio y presenta diversas dimensiones.

Algunas de ellas tienen un carácter claramente continuista, como es el caso de las nuevas tendencias en el periodismo de datos, muy orientado al periodismo de investigación, sobre todo entendido como periodismo de denuncia.

También se ha visto que hay otras derivaciones que pueden

acabar afectando de manera más profunda el trabajo en las redacciones, como es el caso del llamado coloquialmente *robot periodista*, a cargo del cual es posible que en el futuro quede la cobertura de ciertas noticias. Concretamente aquellas que se benefician de la síntesis y la agregación de datos, como son curiosamente temas aparentemente tan alejados como resultados deportivos y financieros.

El periodismo estructurado, fuertemente basado en el uso de bases de datos, es otra especialidad cuyo planteamiento conceptual es altamente prometedora por los resultados que presenta y la indudable utilidad ciudadana. No obstante éste como otros casos de innovación debe superar aún la validación final del favor del público.

Se trata de un entorno muy dinámico en el que constantemente se están produciendo innovaciones

8. Conclusiones

El periodismo computacional se caracteriza por su transdisciplinariedad y multidimensionalidad, donde la transgresión de las disciplinas configura una nueva realidad que otorga al periodismo un carácter dinámico que permite su evolución y adaptación a los nuevos entornos.

Asimismo se caracteriza por un uso intensivo de la tecnología, no sólo como pieza clave para el proceso de creación de las noticias sino también como aliada para la redacción de las noticias, su optimización para diferentes entornos, y su facilidad para personalizar contenidos.

Se requieren periodistas con conocimientos técnicos, que sean capaces de configurar herramientas, procesar datos, crear visualizaciones de la información con diferentes soportes multimedia, e incluso de analítica web para planificar la estrategia de contenidos. Las universidades tienen que proporcionar profesionales que puedan cubrir las necesidades que demanda el mercado laboral.

Para finalizar y a modo de resumen, el periodismo computacional se focaliza en dar respuesta a las siguientes cuestiones: ¿cómo descubrir nuevas historias usando datos y algoritmos?, ¿cómo contar historias atractivas visualmente?, ¿cómo personalizar contenidos para diferentes tipos de lectores?, ¿cómo cubrir noticias de eventos locales? Por tanto, se trata de un sector en evolución que cuenta con un gran recorrido.

9. Agradecimiento

Este trabajo forma parte del proyecto *Creación y contenido interactivo en la comunicación de información audiovisual: audiencias, diseño, sistemas y formatos* CSO2015-64955-C4-2-R *Ministerio de Economía y Competitividad (Mineco/Feder)*, (España).

10. Referencias

Albert, Sílvia (2017). *El impacto de los datos en la era de la postverdad*. Madrid: Perspectivas Wellcomm.
<http://perspectivas2017.well-comm.es>

- Arcila-Calderón, Carlos; Barbosa-Caro, Eduar; Cabezero-Lorenzo, Francisco** (2016). "Técnicas *big data*: análisis de textos a gran escala para la investigación científica y periodística". *El profesional de la información*, v. 25, n. 4, pp. 623-631.
<https://doi.org/10.3145/epi.2016.jul.12>
- Blanco-Castilla, Elena; Quesada, Montse** (2016). *Periodismo de datos*. La Laguna (Tenerife): Cuadernos Artesanos de Comunicación. CAC 112. ISBN: 978 84 16458 56 1
<http://www.revistalatinacs.org/068/cuadernos/2016/cac112.pdf>
- Cairo, Alberto** (2011). *El arte funcional: infografía y visualización de información*. Madrid: Alamut. ISBN: 978 84 9889 067 9
- Carlson, Matt** (2015). "The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority". *Digital journalism*, v. 3, n. 3, pp. 416-431.
<https://doi.org/10.1080/21670811.2014.976412>
- Caswell, David; Dörr, Konstantin** (2018). "Automated journalism 2.0: Event-driven narratives". *Journalism practice*, v. 12, n. 4, pp. 477-496.
http://www.zora.uzh.ch/id/eprint/137060/1/Automated_Journalism_2.0_EventDriven_Narratives.pdf
<https://doi.org/10.1080/17512786.2017.1320773>
- Coddington, Mark** (2015). "Clarifying journalism's quantitative turn". *Digital journalism*, v. 3, n. 3, pp. 331-348.
<https://doi.org/10.1080/21670811.2014.976400>
- Cohen, Sarah; Hamilton, James; Turner, Fred** (2011). "Computational journalism". *Communications of the ACM*, v. 54, n. 10, pp. 66-71.
<https://doi.org/10.1145/2001269.2001288>
- Diakopoulos, Nicholas** (2015). "Algorithmic accountability. Journalistic investigation of computational power structures". *Digital journalism*, v. 3, n. 3, pp. 398-415.
http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/algorithmic_accountability_final.pdf
<https://doi.org/10.1080/21670811.2014.976411>
- Dörr, Konstantin; Hollnbuchner, Katharina** (2017). "Ethical challenges of algorithmic journalism". *Digital journalism*, v. 5, n. 4, pp. 404-419.
<https://doi.org/10.1080/21670811.2016.1167612>
- Fanta, Alexander** (2017). *Putting Europe's robots on the map: automated journalism in news agencies*. Oxford: Reuters Institute for the Study of Journalism.
<http://reutersinstitute.politics.ox.ac.uk/our-research/putting-europes-robots-map-automated-journalism-news-agencies>
- Freixa, Pere; Pérez-Montoro, Mario; Codina, Lluís** (2017). "Interacción y visualización de datos en el periodismo estructurado". *El profesional de la información*, v. 26, n. 6, pp. 1076-1090.
<https://doi.org/10.3145/epi.2017.nov.07>
- García-Carretero, Lucía; Codina, Lluís; Díaz-Noci, Javier; Iglesias-García, Mar** (2016). "Herramientas e indicadores SEO: características y aplicación para análisis de cibermedios". *El profesional de la información*, v. 25, n. 3, pp. 497-504.
<https://doi.org/10.3145/epi.2016.may.19>
- Graefe, Andreas** (2016). *Guide to automated journalism*. New York: Tow Center for Digital Journalism, Columbia University.
<https://doi.org/10.7916/D80G3XDJ>
- Graefe, Andreas; Haim, Mario; Haarmann, Bastian; Brosius, Hans-Bernd** (2018). "Readers' perception of computer-generated news: Credibility, expertise, and readability". *Journalism*, v. 19, n. 5.
<https://doi.org/10.1177/1464884916641269>
- Guallar, Javier** (2014). Content curation en periodismo (y en documentación periodística). *Hipertext.net*, n. 12.
<https://doi.org/10.2436/20.8050.01.15>
- Guallar, Javier** (2017). "Content curation in digital media: Between retrospective and real-time information". En: Campos-Freire, Francisco; Rúas-Araújo, Xosé; Martínez-Fernández, Valentín A.; López-García, Xosé (eds.). *Media and metamedia management*. Springer, pp. 37-46. ISBN: 978 3 319 46066 6
https://doi.org/10.1007/978-3-319-46068-0_6
- Guerrini, Federico** (2014). *Newsroom curators and independent storytellers: Content curation as a new form of journalism*. Oxford: Reuters Institute for the Study of Journalism.
<http://reutersinstitute.politics.ox.ac.uk/our-research/newsroom-curators-and-independent-storytellers-content-curation-new-form-journalism>
- Gynnild, Astrid** (2014). "Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets". *Journalism: Theory, practice & criticism*, v. 15, n. 6, pp. 713-730.
<https://doi.org/10.1177/1464884913486393>
- Hamilton, James; Turner, Fred** (2009). *Accountability through algorithm: Developing the field of computational journalism*. California: Center for Advanced Study in the Behavioral Sciences.
<https://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>
- Houston, Brant** (2015). *Computer-assisted reporting: a practical guide*. Londres: Routledge. ISBN: 978 1 317 51943 0
- Jung, Jaemin; Song, Haeyeop; Kim, Youngju; Im, Hyunsuk; Oh, Sewook** (2017). "Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists". *Computers in human behavior*, v. 71, pp. 291-298.
<https://doi.org/10.1016/j.chb.2017.02.022>
- Kasem, Ila; Van-Waes, Mark J. F.; Wannet, Kim C. M. E.** (2015). *What's new (s)? Scenarios for the future of journalism (Journalism 2025)*. Amsterdam: Netherlands Press Fund.
<http://www.journalism2025.com/bundles/svdjui/documents/Scenarios-for-the-future-of-journalism.pdf>
- Lindén, Carl-Gustav** (2017a). "Algorithms for journalism: The future of news work". *The journal of media innovations*, v. 4, n. 1, pp. 60-76.

<https://doi.org/10.5617/jmi.v4i1.2420>

Lindén, Carl-Gustav (2017b). "Decades of automation in the newsroom". *Digital journalism*, v. 5, n. 2, pp. 123-140.
<https://doi.org/10.1080/21670811.2016.1160791>

Marconi, Francesco; Siegman, Alex (2017). *The future of augmented journalism. A guide for newsrooms in the age of smart machine*. New York: AP Insights.
https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf

Meyer, Philip (1999). "The future of CAR: declare victory and get out". En: Poynter. *When nerds and world collide: Reflections on the development of computer assisted reporting*. Florida: The Poynter Institute for Media Studies, pp. 4-5.
<http://www.unc.edu/~pmeyer/carfuture.doc>

Montal, Tal; Reich, Zvi (2017). "I, robot. You, journalist. Who is the author?". *Digital journalism*, v. 5, n. 7, pp. 829-849.
<https://doi.org/10.1080/21670811.2016.1209083>

Newman, Nic (2017). *Journalism, media and technology predictions 2017*. Oxford: Reuters Institute for the Study of Journalism.
<https://goo.gl/njQdCt>

Nicolescu, Basarab (2002). *Manifesto of transdisciplinarity*. Albany: State University of New York Press. ISBN: 978 0 7914 5262 2

Parasie, Sylvain; Dagiral, Eric (2013). "Data-driven journalism and the public good: 'Computer-assisted-reporters' and 'programmer-journalists' in Chicago". *New media & society*, v. 15, n. 6, pp. 853-871.
<https://doi.org/10.1177/1461444812463345>

Pérez-Montoro, Mario (2009). "Visualización de la información". *Glossarium-BITri*.
<http://glossarium.bitrum.unileon.es/Home/visualizacion-de-la-informacion>

Richmond, Shane (2008). "How SEO is changing journalism". *British journalism review*, v. 19, n. 4, pp. 51-55.

<https://doi.org/10.1177/0956474808100865>

Shearer, Elisa; Gottfried, Jeffrey (2017). *News use across social media platforms 2017*. Washington: Pew Research Center.
<http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017>

Tandoc, Edson; Thomas, Ryan (2015). "The ethics of web analytics. Implications of using audience metrics in news construction". *Digital journalism*, v. 3, n. 2, pp. 243-258.
<https://goo.gl/8mqfYg>
<https://doi.org/10.1080/21670811.2014.909122>

Thurman, Neil; Dörr, Konstantin; Kunert, Jessica (2017). "When reporters get hands-on with robo-writing". *Digital journalism*, v. 5, n. 10, pp. 1240-1259.
<http://openaccess.city.ac.uk/16350>
<https://doi.org/10.1080/21670811.2017.1289819>

Vázquez, Mari; Pedraza-Jiménez, Rafael (2007). "El procesamiento del lenguaje natural en la recuperación de información textual y áreas afines". *Hipertext.net*, n. 5.
<http://www.upf.edu/hipertextnet/numero-5/pln.html>

Wang, Qun (2018). "Dimensional field theory. The adoption of audience metrics in the journalistic field and cross-field influences". *Digital journalism*, v. 6, n. 4, pp. 472-491.
<https://doi.org/10.1080/21670811.2017.1397526>

Wang, Shuguang; Han, Eui-Hong; Rush, Alexander (2016). "Headliner: An integrated headline suggestion system". En: *Computation+Journalism Symposium California*.
<https://goo.gl/9fvdx4>

Weeks, Lin (2014). "Media law and copyright implications of automated journalism". *Journal of intellectual property and entertainment law*, v. 4, n. 1, pp. 67-94.
<https://jipel.law.nyu.edu/vol-4-no-1-3-weeks>

Wölker, Anja; Powell, Thomas E. (2018). "Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism". *Journalism* (in press).
<https://doi.org/10.1177/1464884918757072>

Cronología de la Documentación Española

