

# CULTURA ALGORÍTMICA Y TECNOLOGÍAS DEL LENGUAJE HUMANO

Jesús Octavio ELIZONDO y Esaú VILLATORO

## INTRODUCCIÓN

**E**n este trabajo analizaremos, desde una perspectiva crítica, el fenómeno de los algoritmos y las tecnologías del lenguaje humano (TLH) con el fin de identificar el papel que aquéllos tienen en la conformación de una nueva forma de producir bienes y servicios para diferentes plataformas informacionales y que, de seguir el camino aquí descrito, eventualmente constituirían parte de una “cultura algorítmica”. Para alcanzar este objetivo, revisaremos la dualidad cultura y tecnología; con ello, buscamos señalar el papel relevante que el almacenaje, la organización y la recuperación de la información mediante algoritmos tiene en la producción cultural. Si los algoritmos computacionales realizan ya tareas asignadas tradicionalmente a la cultura entonces, ¿cabría inferir un escenario donde los algoritmos generen información para otros algoritmos a la par que lo hagan para los seres humanos? De ser así, ¿qué implicaciones tendría esto para la cultura?

## DE CULTURA Y TECNOLOGÍA A CULTURA ALGORÍTMICA

El fenómeno que nos interesa analizar aquí es el uso de los algoritmos computacionales, ya que esta práctica cultural está construyendo, en cierta medida, una nueva forma de entender la relación entre cultura y la tecnología. Los algoritmos computacionales tienen como característica la capacidad de interpretar y proponer contenidos informacionales de acuerdo a las prácticas del sujeto interpretante (usuario). Es decir, el uso generalizado de los algoritmos en los motores de búsqueda como los que emplean Google, Facebook y Netflix está haciendo algo más que brindar el servicio de búsqueda ofrecido. Estas compañías realizan una importante tarea que era exclusiva de la cultura: seleccionar, procesar y difundir la información pertinente para una comunidad de hablantes. Por esta razón es pertinente preguntarse: ¿Qué lugar ocupan los algoritmos computacionales en la vida diaria? ¿El uso generalizado de algoritmos incide en las prácticas culturales? De ser así, ¿en qué casos y de qué manera lo hacen? Con el fin de dar respuesta a estas preguntas y analizar la posibilidad de que estemos ante la conformación de una “cultura del algoritmo”, emplearemos los principios de la disciplina conocida como *lingüística computacional o tecnologías del lenguaje humano*.

Algunos autores contemporáneos (Striphas 2014) sostienen que nos encontramos en un momento histórico en el que es posible observar un reencuentro entre los conceptos de cultura y tecnología. Para la década de 1980, Williams detectó en los diccionarios de la lengua (inglesa) que los conceptos cultura y tecnología se habían vuelto menos antagónicos que lo que había estado teóricamente y en la práctica (Striphas 2014). Esta intuición se debe en parte al acercamiento de la cultura a los procesos propios de la computación y la incipiente digitalización de la época, así como al hecho de que ya se veía llegar al naciente mercado de las computadoras personales. Este fenómeno vio su explosión durante las décadas siguientes, cuando la computadora personal

se instalara como una tecnología disruptiva. Sucedió en un corto periodo de tiempo, seguido únicamente por el fenómeno de la telefonía móvil. Dichos fenómenos corrieron a la par de importantes cambios en la ciencia, producto –en parte– de la llegada de la cibernética y las teorías de la información computarizada durante el último tercio del siglo XX. Hoy experimentamos un cambio importante respecto al sentido de la tecnología y la cultura. Si bien no se trata de un regreso a la concepción pre-moderna, resulta evidente que ambos conceptos se están fusionando en formas diversas: *Digital Humanities*, *Big data*, *Crowd Computing* así como Cibercultura y Tecnologías del lenguaje humano son sólo algunos ejemplos comunes de este fenómeno.

Tomemos el caso de Google. Esta compañía usa como insumos para la producción de sus servicios electricidad, silicón y plástico, que bajo la dirección de sus ingenieros se emplean en la tarea de almacenar, organizar, buscar y recuperar en forma jerarquizada (*rankear*) la importancia que tienen personas, lugares, objetos e ideas. De hecho, el sistema de *ranking* PageRank (Brin 1998) fue la punta de lanza de esta compañía. Si bien los medios y los fines difieren, estas tareas no son muy diferentes del sentido y propósito último de la cultura: abocarse a la tarea de determinar lo mejor de lo que se ha pensado y dicho. Si los algoritmos computacionales proveen de este tipo de información jerarquizada de manera automática a otros sistemas de la misma manera que lo hacen para los sujetos interpretantes (usuarios) entonces, en este punto, cabe preguntarse ¿Qué dice de la cultura humana que aspectos clave sean inteligibles únicamente para los sistemas informacionales?

El ejemplo sobre Google, anteriormente mencionado, ilustra lo dicho hasta ahora, pero es posible encontrar casos similares en internet. Facebook dedica buena parte de sus esfuerzos en realizar tareas similares, a saber: determinar quiénes de tus amigos y cuáles de sus comentarios aparecerán en tu perfil de noticias. Lo mismo vale para los sitios de compras en línea, de *streaming* de

música y videos, y otros servicios donde hayamos dejado huella de nuestros gustos y preferencias. Particularmente importantes son aquéllos sitios donde hemos realizado alguna compra. Pero lo que es relevante señalar aquí es la manera en que la cultura del algoritmo se retroalimenta (creando el *feedback*) para producir nuevos hábitos de pensamiento, conducta y expresiones que de otra manera no se darían sin la intervención de la cultura algorítmica en el contexto de las ciencias computacionales. El problema aquí es que este tipo de cultura tiende más a reforzar lo existente que a propiciar nuevas formas de conocer y crear, nuevas formas de hacer cosas. Esto es lo que se ha llamado “personalización” de las preferencias o lo que algunos autores (Pariser 2011) llaman *Your Loop*, es decir, tu círculo autorreferente, recursivo (que es tanto virtuoso o vicioso). De la misma manera, los sistemas algorítmicos tienen la capacidad de presentar productos culturales –bienes y servicios– que de otra manera habrían sido muy tardados de encontrar por uno mismo. De aquí que la cultura algorítmica sea equivalente a lo eficiente y pertinente de sus algoritmos.

En este punto cabe preguntarse: ¿Quién está detrás de las fuerzas que mueven estas macrotendencias y que impulsan el cambio? Evidentemente los actores bien posicionados en el mercado son los de mayor visibilidad: Google, Apple, Facebook, Amazon, Netflix y proveedores de servicios *triple-play* entre otros. Pero lo interesante del caso es que ninguno de éstos, de manera individual, es responsable de los cambios que presenciamos: son elementos de una macro tendencia que los contiene y los superan. Si pensamos en la cultura de los algoritmos, tenemos que regresar a la década de 1940 para comenzar con la creación de la *información* como valor y el subsecuente desarrollo de la informática. Y es que si aceptamos la noción de que todo es potencialmente información, y por lo tanto computable, entonces podemos ver un elemento clave para entender los cambios que estamos viviendo.

## TECNOLOGÍAS DEL LENGUAJE HUMANO

Las tecnologías del lenguaje humano tratan de buscar mecanismos computacionales que permitan almacenar, organizar, recuperar y presentar lenguaje natural. Para ello, realizan un tratamiento automático de éste; por tanto, intentan trasladar e integrar el conocimiento que las personas tenemos de la lengua en los agentes, para que puedan emular las acciones que podemos realizar de forma innata. Para lograr este objetivo, incorporan modelos teóricos, métodos y técnicas de diferentes disciplinas: lingüística, filosofía, psicología e ingeniería, ya que todas ellas están implicadas o pueden resultar útiles para tratar los diferentes procesos que envuelven el lenguaje natural. Cada una de ellas estudia la lengua desde puntos de vista y objetivos distintos, lo cual ha conllevado también el uso de terminología diferente para hacer referencia a la misma idea. La lingüística utiliza el término *lingüística computacional*, la ingeniería informática usa la expresión *ingeniería del lenguaje natural*. Sin embargo, el concepto más utilizado tradicionalmente por la comunidad científica es *procesamiento del lenguaje natural*, aunque actualmente está muy extendida la expresión *tecnologías del lenguaje humano*.

### *Recuperación de información: el algoritmo de recomendación*

Uno de los problemas principales dentro del área de Tecnologías del lenguaje humano es la Recuperación de Información (IR, por sus siglas en inglés). La tarea principal de los sistemas de recuperación de información consiste en: una vez dada una consulta, formulada en lenguaje natural por algún usuario, obtener documentos relevantes que satisfagan las necesidades de información del usuario (Grossman 2004).

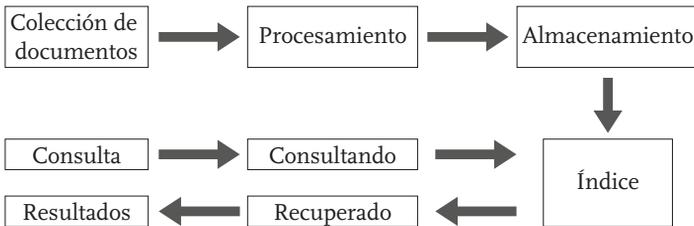


Diagrama 1. Sistema típico de recuperación de información

En este contexto, el término se refiere a que los documentos recuperados deberán estar semánticamente relacionados a la necesidad de información del usuario. Además, los documentos deberán estar ordenados/presentados de acuerdo a la relevancia que cada uno de estos tenga con respecto a la consulta dada. Es importante señalar que la relevancia sólo puede ser evaluada por el usuario que formula la consulta, por lo que la evaluación de la efectividad de un sistema de recuperación de información no puede ser del todo objetiva;<sup>1</sup> aunque actualmente se han creado colecciones de documentos estándar y desarrollado protocolos para la evaluación y comparación de este tipo de sistemas (como el *Cross-lingual evaluation forum* y el *Text retrieval conference*) (Baeza-Yates 1999; Grossman 2004). Nótese también que no se especifica el tipo de colección ni la modalidad de los documentos, por lo que por documento nos referiremos indistintamente a documentos de texto, imágenes, audio, video, etcétera.

Para solucionar la tarea de recuperación de información se han propuesto muchos modelos. Un modelo de recuperación se define como el conjunto de métodos y estrategias que nos permiten representar y organizar la colección de documentos, definir consultas y compararlas con los documentos en la colección. Uno de los esquemas que ha mostrado ser muy exitoso es el modelo de espacio vectorial, inicialmente propuesto por Salton (1990).

<sup>1</sup> Para realizar una evaluación objetiva sería necesario conocer *a priori* todos los documentos relevantes a las consultas de cada usuario.

La idea intuitiva de este modelo de recuperación establece que el significado de un documento está dado por las palabras que este contiene. Entonces, propone llevar los documentos y la consulta a una representación vectorial, donde cada componente del vector representa a las palabras del documento y de la consulta así, al hacer una comparación vectorial entre documentos y consulta se podrán identificar los documentos más similares (relevantes) a la consulta. Este modelo, a pesar de su simplicidad, ha mostrado ser un algoritmo eficiente y efectivo al grado de que sigue siendo ampliamente utilizado por máquinas de recuperación de información actuales como Google, Amazon, Yahoo!, etcétera.

Retomemos el caso Google. Ésta es una empresa estadounidense multinacional que se especializa en el almacenamiento, búsqueda y recomendación de productos y servicios en internet. Algunos de los servicios que actualmente ofrece Google son la búsqueda de información y cómputo distribuido así como el desarrollo de *software* y generación de publicidad dirigida. En su esquema original, Google nació como una máquina de recuperación de información que implementaba un modelo de espacio vectorial para hacer la búsqueda de documentos. Sin embargo, lo que fue un éxito rotundo fue su algoritmo de recomendación. En otras palabras, el algoritmo propuesto para ordenar y presentar la información al usuario fue lo que lo hizo atractivo para el público. Este algoritmo de recomendación, mejor conocido como PageRank (Brin 1998) es un algoritmo especializado en el análisis de conexiones entre objetos interconectados dentro de una red, el cual permite asignar un valor de importancia a cada elemento dentro de la red, indicando hasta cierto punto la “relevancia” de dicho objeto dentro de la red.

El algoritmo PageRank (Brin 1998), fue propuesto por Serge Brin y Lawrence Page en 1996 como parte de un proyecto de curso, el cual tenía como objetivo principal formular e implementar una máquina de búsqueda novedosa. Dos años después, en 1998, Page y Brin fundaron Google Inc.

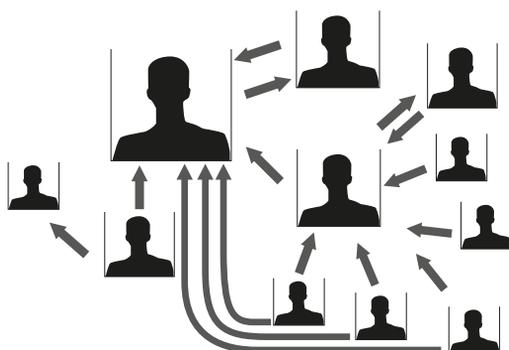


Diagrama 2. Page Rank

En 2006 la compañía Netflix anunció el Premio Netflix:<sup>2</sup> una competencia en la cual se evaluarían algoritmos que fueran capaces de mejorar de manera sustancial la precisión en las predicciones sobre cuánto disfrutaría una persona/usuario una película, tomando como referencia sus preferencias y hábitos. La realidad es que Netflix, siendo una empresa que se dedicó durante algunos años a la venta/compra de DVD, actualmente es una empresa multimillonaria que ofrece contenidos visuales a través de transmisiones en línea. Su éxito se debe en gran medida a la evolución a una empresa de *software*, de tal manera, que ahora funciona montada sobre algoritmos computacionales que constantemente están en desarrollo y mejoramiento. En 2014, Netflix cuenta ya con 48 millones de subscriptores alrededor del mundo (Bulygo 2013). Esto le permite a los desarrolladores de Netflix recuperar diariamente una cantidad enorme de datos. Estos son los insumos que les permiten tomar decisiones importantes y estar mejor informados sobre su audiencia. Tomemos otro ejemplo: decidir si vale la pena invertir cien millones de dólares en una producción original como *House of Cards* (las cadenas de televisión tradicionales no pueden gozar de estos privilegios). Los

<sup>2</sup> El premio consistió en un millón de dólares.

“ratings” son meras aproximaciones y la decisión sobre el lanzamiento de nuevas series se basa en intuiciones y/o experiencias previas de los involucrados en el proceso. En este sentido, compañías de software como Netflix tienen una enorme ventaja sobre la competencia al poder conocer muy bien a sus clientes, en otras palabras; son capaces de determinar con una precisión muy alta el grado de éxito que tienen sus nuevos contenidos, gracias a toda la retroalimentación producida por los mismos usuarios. Pero, ¿Qué tipo de información es empleada por Netflix? ¿Cómo obtienen la retroalimentación? La respuesta es simple y tiene que ver con los siguientes factores:

- a) Conocer cuándo se hacen pausas, retrocesos y/o avances
- b) ¿Qué día de la semana viste contenido?
- c) La fechas en que se utilizó el sistema
- d) Durante cuánto tiempo viste contenido
- e) Dónde estuviste (ubicación geográfica)
- f) Dónde lo utilizaste (tipo de dispositivo)
- g) Cuándo se pausa y cuándo se abandona o se regresa
- h) Las calificaciones asignadas (estrellas)
- i) Las peticiones/consultas realizadas
- j) La búsquedas/navegación en la interfaz: ¿Cuánto tarda el usuario en encontrar contenido atractivo?

Entonces, a través de proponer algoritmos (un ecosistema de algoritmos) que sean capaces de interpretar y aprender de toda la información generada por los usuarios, Netflix es capaz de recomendar contenido a los usuarios, mismo que tiene altas posibilidades de ser visto por las audiencias.

Los factores mencionados anteriormente tienen además, una función adicional: la de proporcionar información al algoritmo de recomendación de Netflix de tal forma que sea posible hacer sugerencias personalizadas a los usuarios (actualmente el 75% de la actividad de los usuarios se realiza gracias a las recomendaciones).

El algoritmo de recomendación emplea la noción de similitud para poder hacer una personalización más adecuada. En este contexto, la similitud no solo se mide en términos de palabras, sino que también se toman en cuenta información multi-modal (meta-datos de las películas, puntuaciones, visitas, etcétera).

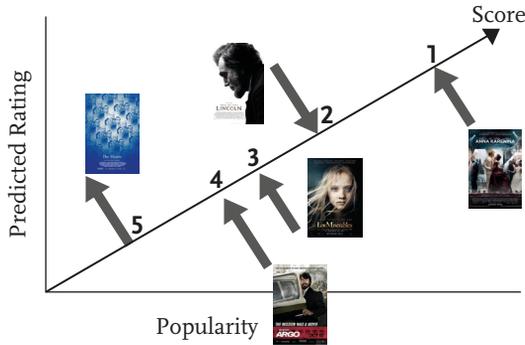


Diagrama 3. Recomendación personalizada

Finalmente, podemos decir que Netflix enfrenta el problema de cómo presentar la información que se determinó como similar a los gustos/necesidades de cada usuario, es decir, cómo *rankear* los contenidos de forma que los primeros en la lista sean los que tienen mayores posibilidades de ser vistos por la audiencia. El esquema más simple para presentar los contenidos sería por medio de considerar las puntuaciones de cada película. Sin embargo, esta forma de organizar y presentar la información va en contra de la personalización, pues resultaría en una lista con los mismos objetos en el mismo orden. En su lugar, un ejemplo simple sería un esquema de *ranking* de dos dimensiones. Dado que la estrategia que permite determinar similitudes nos genera un valor, indicando que tanto le podría gustar al usuario un determinado título, se construye una función lineal que combina la popularidad de los títulos con el valor predicho de “gusto”, de tal forma que la función de recomendación genera un orden personalizado para cada usuario.

## CONCLUSIONES

Es de esperar que los algoritmos y la computación a gran escala continúen por el camino que siguen ahora, es decir, cada vez más presentes dentro de la toma de decisiones sobre los asuntos del día de los sujetos generadores de información sistematizable. En este orden de ideas, imaginemos que para la industria del entretenimiento como en el ejemplo anterior, se vuelve un hábito cruzar historias en apariencia muy distintas y crear con ellas nuevas narrativas. Esto es lo que hace la compañía *Narrative Science*. Este tipo de Narrativas transmediáticas serán cada vez más comunes en nuestro ecosistema algorítmico y serán evidentemente generadas por computadoras. No habrá ya más guionistas o escritores que sigan desarrollando su actividad al margen de estos criterios.

Por otro lado, los algoritmos comienzan a jugar un papel importante a la hora de decidir qué aspectos de la cultura tienen luz verde para la producción. Esto es un hecho en el caso de guiones cinematográficos que son puestos a prueba con algoritmos que los analizan en busca de suficientes elementos característicos de un éxito en taquilla. Como hemos visto, Netflix analiza información de manera similar con sus subscriptores antes de dar luz verde a su producción y programación, con el fin de asegurarse de que los gustos y hábitos de sus espectadores encajen bien con la selección de los personajes, género, estructura, etcétera, de una nueva serie televisiva. Podemos decir que la cultura tiene ahora dos audiencias: la gente y los algoritmos computacionales. Ambos juegan un papel significativo en la construcción de los materiales audiovisuales que llegan a las pantallas del público en general (Striphas 2014). Considerando todo lo dicho hasta ahora, podemos preguntarnos ¿Cómo se percibe el futuro de las industrias culturales? Nos aventuramos a dibujar al menos dos escenarios: aquél que implica el uso de algoritmos en la producción cultural y otro que no lo hace al menos de momento. De ser así, esto implicaría que podría haber una recursividad, un círculo

recurrente (*loop*) en la práctica cultural, de la cultura algorítmica en general, lo que a largo plazo podría acarrear contradicciones intrínsecas a toda la sociedad.

Los algoritmos conllevan un sesgo por lo predictivo; el “si tal cosa, entonces tal otra”: la noción de que los asuntos humanos son mejores si se pueden predecir, poner a prueba con todas sus variables posibles antes de que ocurran. Esto, por desgracia, es un modelo que no difiere del de la automatización del trabajo, el censo y el registro cuantitativo, conceptos éstos característicos de las teorías gerenciales, que han propiciado políticas burocráticas. ¿Cuáles son los peligros de una era donde se puede predecir casi todo? De entrada podemos decir que habría un riesgo muy grande de caer en un “determinismo informacional” (Pariser 2011) que sería el resultado ineludible del exceso de información personalizada en la Web y en los sistemas de banda ancha. En este escenario, todo aquello a lo que se le dé “click” determina lo que se verá a continuación: una especie de bagaje o historia escrita en la red que estaríamos condenados a repetir actualizada cada vez, en nuevas formas culturales. Podríamos vernos estancados en un contaste *loop* recursivo de uno mismo y sus preferencias. Manuel Castells (2000) habló ya del tiempo atemporal que es aquél en el que todo existe ahí, en la red permanentemente, sin un destino que alcanzar, sino de permanecer. ¿Qué hacer ante este escenario? No se gana nada con armarnos contra los algoritmos computacionales y repetir la historia del Ludismo de la Revolución industrial contra las máquinas en el siglo XIX. Si se desea mantenerse al margen de los localizadores, uno se puede dar de baja de Facebook y Google incluso cancelar su suscripción a Netflix como muchos están haciendo ya (sobre todo en el caso de Facebook). Una mejor solución es apelar al derecho a privacidad en internet también llamada Ley del Derecho al Olvido (o a ser olvidado) recientemente aprobada en la Unión Europea (De Terwangne 2012). Pero lo que es esencial es armarnos con el conocimiento de cómo funcionan estas computadoras y cómo

pueden ser empleadas legítimamente al servicio de las mayorías propiciando el empoderamiento de los ciudadanos, mejorando la calidad de vida, del aprendizaje y de nuestras relaciones sociales. Es esencial familiarizarse con la manera en que las compañías pueden abusar de estos sistemas informacionales y estar alerta.

## BIBLIOGRAFÍA

- BAEZA-YATES, Ricardo; Berthier RIBEIRO-NETO. 1999. *Modern Information Retrieval*. Nueva York: Addison Wesley.
- BERNERS-LEE, Tim; James HENDLER; Ora LASSILA. 2001. "The semantic web", *Scientific America*, núm. 501 (mayo): 29-37.
- BRIN, Sergey; Lawrence PAGE. 1998. "The Anatomy of a Large-scale Hypertextual Web Search Engine", en *Proceedings of the Seventh International Conference on World Wide Web 7*, 107-117. Stanford: Universidad de Stanford.
- BULYGO, Zach. 2013. *How Netflix Uses Analytics To Select Movies, Create Content, and Make Multimillion Dollar Decisions*. Página web: <https://blog.kissmetrics.com/how-netflix-uses-analytics/>, (consultado el 25 de junio de 2014).
- CASTELLS, Manuel. 2000. "Virtual time" en *Rise of the network society*, Vol. I. The Information Age: economy, society and culture. West Sussex; Malden: WilleyBlackwell.
- COLE, Ron. 1995. *Survey of the state of the art in human language technology*. Cambridge: Cambridge University Press, <http://www.dfki.de/~hansu/hlt-survey.pdf>, (consultado el 25 de junio de 2014).
- DE TERGANGNE, Cécil. 2012. "Privacidad en internet y el derecho a ser olvidado/ derecho al olvido" en *IDP revista internet, derecho y política*. España: Universidad Oberta de Catalunya. <http://www.redalyc.org/articulo.oa?id=78824460006> (consultado el 25 de junio de 2014).

- GROSSMAN, David A.; Ophir FRIEDER. 2004. *Information Retrieval, Algorithms and Heuristics*, segunda edición. Berlín; Heidelberg; Nueva York: Springer.
- PARISER, ELI. 2011. *The filter bubble*. Nueva York: The Penguin Press.
- RUSLAN, Mitkov. 2003. *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.
- SALTON Gerard; Chris BUCKLEY. 1990. "Improving retrieval performance by relevance feedback", en *Journal of the American Society for Information Science*, núm. 41 volumen 4: 288-297.
- STRIPHAS, Ted. 2014 "Algoritmic Culture. 'Culture now has two audiences: people and machines.'" a conversation with Ted Striphas" en *Medium*. <https://medium.com/futurists-views/2bdaa404f643>, (consultado el 25 de mayo, 2014).
- TUKER, Patric. 2012. *The naked future*. Nueva York: Penguin.
- VALLEZ, Mari. 2009. "La web semántica y el procesamiento del lenguaje natural", en *Web Semántica y Sistemas de Información Documental*, Luis Codina; Mari Carmen Marcos; Rafael Pedraza-Jiménez (Coordinadores), 155-180. Trea: Gijón.

*Cultura visual y sistemas  
de significación. Dando sentido a los algoritmos, los medios  
y la creatividad en el espacio de la comunicación*, se terminó  
de imprimir en octubre de 2015  
en los talleres de Tipos Futura S. A. de C. V.,  
Francisco González Bocanegra #47B, Col. Peralvillo,  
Del. Cuauhtémoc, CP 06220.  
*150 ejemplares*

# CULTURA VISUAL Y SISTEMAS DE SIGNIFICACIÓN

Dando sentido a los algoritmos,  
los medios y la creatividad  
en el espacio de la comunicación

---

Esta obra colectiva debate sobre el estado actual de la cultura visual, en un momento en el que la relación entre las tecnologías de la información, el diseño y la comunicación es innegable.

Este libro está dirigido a estudiosos e investigadores de estos tres campos... pero también es de interés para creadores audiovisuales, curadores de exposiciones en museos y cinéfilos.

Se abordan temas como la cultura algorítmica, los retos del análisis cinematográfico, la relación de la creatividad con los medios, así como el uso de tecnologías en los museos y el sentido de las formas arquitectónicas de estos recintos.

Se ofrece una visión amplia del fenómeno de la cultura visual desde los campos epistemológicos de la ecología de los medios, la semiótica y la retórica visual.

**Jesús Octavio Elizondo (Editor)**

