

La cultura científica más allá de las encuestas: Wikipedia, la prensa y los libros de texto ¹

Carlos G. Figuerola, Tamar Groves, Miguel Ángel Quintanilla
Instituto Universitario de Estudios de la Ciencia y Tecnología, Universidad de Salamanca

Resumen

Debido a la influencia que los factores culturales tienen en la articulación de políticas de ciencia, tecnología e innovación, es importante comprender cómo la sociedad percibe la ciencia. Las encuestas son la herramienta más común para la evaluación de la percepción social de la ciencia, pero la información que nos proporcionan puede ser comparada y complementada con otras fuentes. En este trabajo describimos un proyecto de investigación, actualmente en marcha, que aborda la percepción de la ciencia e base a otros soportes de información: la prensa, los libros de texto y la Wikipedia. Para el análisis de los contenidos en los diferentes soportes utilizamos un marco teórico único basado en el concepto de cultura científica.

Uno de los objetivos básicos de nuestro proyecto es probar la validez del modelo de cultura científica para organizar una multiplicidad de contenidos y soportes de información. La cantidad de información científica que contienen estos tres soportes es enorme. Para poder analizarla es necesario emplear procedimientos automáticos de localización y almacenaje de información. Otro objetivo de nuestro proyecto es por lo tanto crear una base documental de contenidos de cultura científica en español. El tercer objetivo de nuestro proyecto es clasificar el contenido científico de la base documental según nuestro modelo de cultura científica con la idea de extraer unos indicadores de cultura científica que se podrían utilizar en distintos contextos.

Dada la importancia que los factores culturales tienen en la articulación de políticas de ciencia, tecnología e innovación, es importante comprender cómo la sociedad percibe la ciencia. Las encuestas son la herramienta más común para la evaluación de la percepción social de la ciencia. En nuestro proyecto de investigación pretendemos comparar y complementar el uso de las encuestas con otros tres soportes de información: la prensa, los libros de texto y la Wikipedia. Para el análisis de los contenidos en los diferentes soportes utilizamos un marco teórico único basado en el concepto de cultura científica.

El primer objetivo de nuestro proyecto es probar la validez del modelo de cultura científica para organizar una multiplicidad de contenidos y soportes de información. La cantidad de información científica que contienen estos tres soportes es enorme. Para poder analizarla es necesario emplear procedimientos automáticos de localización y almacenaje de información. El segundo objetivo de nuestro proyecto es por lo tanto crear una base documental de contenidos de cultura científica en español. El tercer objetivo de nuestro proyecto es clasificar el contenido científico de la base documental según nuestro modelo de cultura científica con la idea de extraer unos indicadores de cultura científica que se podrían utilizar en distintos contextos.

¹ Este trabajo ha sido financiado por el Ministerio español, de Ciencia e Innovación, Subprograma de Proyectos de Investigación Fundamental, ref. FFI2011-27763

Para cada uno de los soportes hemos desarrollado una metodología que permite clasificar su contenido según nuestro modelo. En el caso de la prensa hemos usado análisis de contenido. En el caso de Wikipedia utilizamos procesos automáticos de recuperación y clasificación y en el caso de los libros de texto, manejamos el programa *atlas-ti* que permite codificar a los párrafos según la presencia y ausencia de palabras claves.

Todavía estamos en las primeras etapas de la investigación; sin embargo ya hemos detectado que mientras en los soportes tradicionales la cultura científica está caracterizada por una división disciplinar tradicional, en Wikipedia la distribución de contenidos no parece depender solamente de las disciplinas académicas sino también de preocupaciones sociales actuales. Además hemos visto que un fenómeno parecido pasa con los libros de texto de la asignatura *Ciencias para el Mundo Contemporánea* que aspira superar el aislamiento de la enseñanza de la ciencia de sus implicaciones sociales.

El modelo de Cultura Científica

Nadie discute la importancia de la difusión pública de la ciencia y la tecnología y de la correlativa participación del público informado en cuestiones y debates de interés en relación con la ciencia y la tecnología. Muchos de los grandes debates de nuestra época (cambio climático, seguridad, salud, capacidad innovadora, etc.) son debates en torno a cuestiones de carácter científico. Para participar en ellos de forma racional y efectiva, se requiere que la población tenga acceso a información científica y técnica compleja, además de que existan cauces e instituciones adecuadas para canalizar la participación.

Por esta razón han adquirido cada vez más importancia los estudios de comprensión pública de la ciencia y las iniciativas de potenciación de la comunicación pública de la ciencia. A pesar de lo mucho que se ha avanzado en este terreno, no existen aún resultados concluyentes que nos permitan entender mejor los procesos de comunicación pública, su incidencia en las actitudes de la población hacia la ciencia y la tecnología (Montañés, 2010).

En este contexto consideramos interesante desarrollar el concepto de cultura científica que hemos utilizado de forma tentativa en algunos estudios previos (Quintanilla, et al., 2009). Este concepto es una extrapolación del de cultura tecnológica utilizado en trabajos anteriores (Quintanilla et al., 2004). El núcleo fundamental del modelo es la distinción entre cultura tecnológica incorporada a sistemas técnicos (las reglas de operación de un sistema técnico, por ejemplo) y cultura tecnológica no incorporada a sistemas técnicos (los fines sociales para los que se usa o se desarrolla una tecnología, por ejemplo). El primer componente de la cultura tecnológica es inseparable de la propia tecnología, pero el segundo es mucho más flexible y puede adoptar diversas configuraciones, en función de otros elementos culturales de la sociedad de referencia, dando así pie a diferentes procesos de desarrollo tecnológico y de apropiación social de la tecnología. Esta distinción es muy relevante para entender los procesos de transferencia tecnológica entre sistemas culturales muy diferentes (J. C. Martínez, 2010), así como para mejorar nuestra comprensión de cómo los factores culturales inciden en la capacidad y la propensión a innovar de una sociedad (Quintanilla, 2012). El propósito actual es ampliar el modelo a la cultura científica en general (no solo tecnológica).

La cultura científica es aquella parte de la cultura de un grupo social que consiste en información relacionada y compatible con la actividad científica. Se pueden distinguir dos tipos de componentes en la cultura científica: la ciencia propiamente dicha, por una parte, y el resto de la información, representacional, práctica o valorativa que

forma parte de la cultura general del grupo, y tiene que ver con la ciencia aunque no forme parte de la actividad científica como tal. En la cultura científica así entendida siempre se pueden distinguir dos tipos de componentes: las creencias, reglas de actuación y valores científicos propiamente dichos (podríamos llamarlos a estos cultura científica en sentido estricto) y las creencias, reglas de actuación y valores referidos a la ciencia y compatibles con ella, aunque que no forman parte de la ciencia (cultura científica en sentido lato).

Una consecuencia interesante de todo esto es que, en principio, para un mismo nivel del desarrollo de la ciencia, puede haber diferentes configuraciones de la cultura científica en sentido lato. La otra es que si aplicamos reglas de consistencia para la integración de la información cultural en conjuntos coherentes de información, tenemos que aceptar que no cualquier configuración cultural es compatible con el estado real de la ciencia. Esto significa algo muy simple: en función del estado real de la ciencia podemos establecer criterios efectivos para definir qué es una cultura científica adecuada (es decir compatible con el conocimiento científico) en cada momento.

La base conceptual es el modelo teórico desarrollado en Quintanilla, 2000 y Aibar y Quintanilla, 2002. En este modelo, cultura es definida como la información representacional, práctica y evaluativa transmitida a través del aprendizaje social. Así, en nuestro modelo distinguimos tres tipos de cultura: información representacional, información operacional e información evaluativa. La cultura científica y tecnológica de un país se caracteriza en los siguientes términos:

Cultura científica: la información cultural relacionada con actividades científicas, métodos, resultados y su relación con cualquier otra actividad social. La expresión "cultura científica" la usamos aquí no para referirnos a la cultura profesional de los científicos, sino a la parte de la cultura de un individuo o de un colectivo relativa al conocimiento y actividades científicos.

El modelo distingue dos clases o niveles de cultura científica que podemos denominar como intrínseco y extrínseco. El primero se refiere a componentes culturales inmanentes a las actividades científicas: el conocimiento científico en cada área, teorías debatidas por los científicos, hechos descubiertos a través de la investigación científica, explicaciones e interpretaciones científicas de fenómenos naturales o sociales. etc. También se incluyen las normas del método científico, las normas de la investigación empírica, la comunicación científica de los resultados de esa investigación, así como los valores que se supone que guían la actividad científica (la objetividad, coherencia, precisión, etc).

El nivel extrínseco se refiere a todo lo representacional, práctico y evaluativo que está relacionado con las actividades científicas, así como con las instituciones y personas que la llevan a cabo, pero que no forman parte de la cultura científica intrínseca. La imagen de la ciencia, en el sentido de cómo la gente la percibe, la regulación jurídica de las instituciones científicas, la evaluación de la ciencia desde un punto de vista cultural, moral, político, religioso ... Todo esto es parte de lo que entendemos como cultura científica extrínseca.

Cultura tecnológica: es la información cultural sobre sistemas técnicos, su funcionamiento, diseño y producción y cualquier otra clase de actividad relacionada con la tecnología. Nuevamente, "cultura tecnológica" significa aquí no sólo el conocimiento profesional de tecnólogos e ingenieros, sino también se refiere a componentes relativos a la tecnología, su invención, producción, difusión y uso. Otra vez podemos distinguir dos clases de cultura tecnológica, intrínseca y extrínseca. La

cultura tecnológica intrínseca podría explicitarse como toda la información cultural necesaria para diseñar, producir y usar adecuadamente un dispositivo técnico. La extrínseca, por su parte, sería toda la información cultural (creencias, normas, actitudes y valores) que están relacionados con los sistemas técnicos así como con las actividades, conocimiento, personas e instituciones que diseñan, producen y difunden tecnologías; pero que no forman parte de la tecnología o del sistema técnico. Por ejemplo, las teorías y opiniones sobre la influencia de la tecnología de la información en la reducción de puestos de trabajo en la industria.

La distinción entre cultura intrínseca y extrínseca es importante porque, como se ha sugerido en otro lugar (Quintanilla 2005, Quintanilla, 2012), el nivel de innovación de un país o de una empresa depende, además de otros factores, de dos características: la capacidad de innovar y la propensión a innovar. La primera, la capacidad para innovar, está directamente relacionada con el acceso lo que hemos llamado cultura científica o tecnológica intrínseca, mientras que la segunda, la tendencia a la innovación depende más de elementos de cultura científica o tecnológica extrínseca.

Los libros de texto

Este trabajo se lleva a cabo en dos fases y su propósito es estudiar la presencia de cultura científica y tecnológica, así como factores de cultura innovadora y emprendedora en los libros de texto de la Enseñanza Secundaria Obligatoria en España, estudiantes entre 12-16 años de edad. En una primera fase se analizaron 81 libros de las materias obligatorias y de algunas opcionales durante los años 2005-2006.

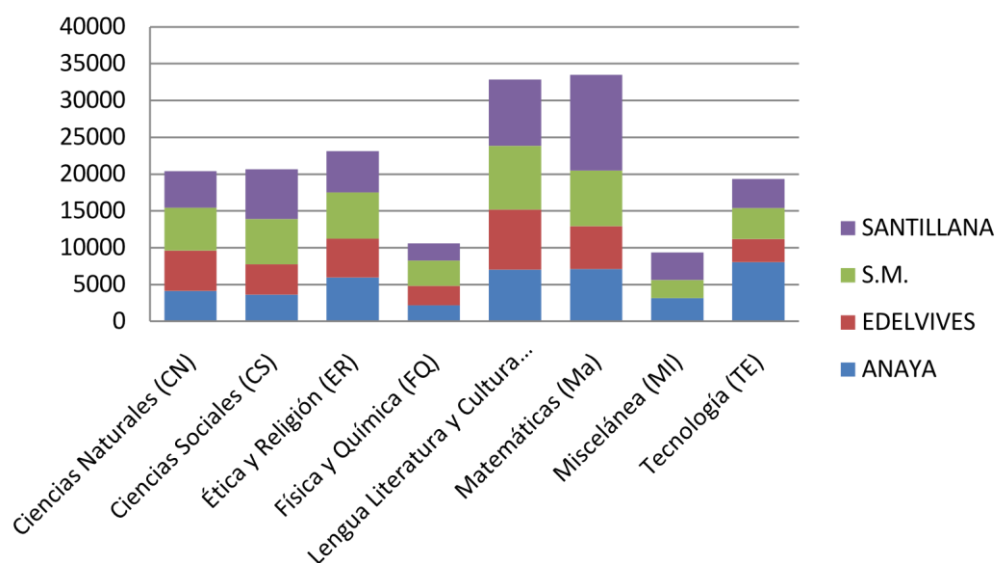


Figura 1. Párrafos de cada materia y editorial

El análisis de contenido de la base documental se realizó a través de varias fases:

1. Elaboración de un thesaurus de vocabulario científico de las palabras clave que caracterizan cada disciplina. Este procedimiento se realizó obteniendo de forma automatizada la lista de palabras presentes en cada texto y comparándolas con un corpus de referencia. Esto permitió obtener una serie de listas de palabras (aproximadamente setenta para cada asignatura) que caracterizan a los textos de cada materia.

2. Análisis automático de la presencia del vocabulario científico con cada texto de la base documental a través del software Atlas ti codificando de manera automática cada

párrafo según se presente o no alguna de las palabras clave asignadas para cada disciplina. Con lo cual se cubre el primero de los objetivos planteados.

3. Se agruparon textos de diversas disciplinas en los que aparecen contenidos de cultura científica, tecnológica e innovación a partir de la codificación automática y se llevo a cabo un análisis manual de los contenidos de cultura científica, tecnológica y de innovación que aparecen en este grupo de textos.

4. Análisis de metáforas de promesas y amenazas presentes en los textos seleccionados, es decir valoraciones ante la ciencia ya sean positivas o negativas.

Las conclusiones principales fueron:

- Hay un equilibrio claro entre los contenidos científicos y tecnológicos, de un lado y humanísticos y sociales, de otro. En cuanto a materias individuales, Matemáticas es la más importante.

- El análisis de las redes semánticas entre los diferentes contenidos científicos muestra una densa red de relaciones entre ellos que sobrepasa los límites de las disciplinas académicas. Es decir, todas las disciplinas están conectadas con contenidos genéricos que pueden ser relevantes desde el punto de vista de la cultura de la innovación y el emprendimiento. Hay, al menos aparentemente, una excepción: Matemáticas carece de relaciones significativas con el resto de las materias.

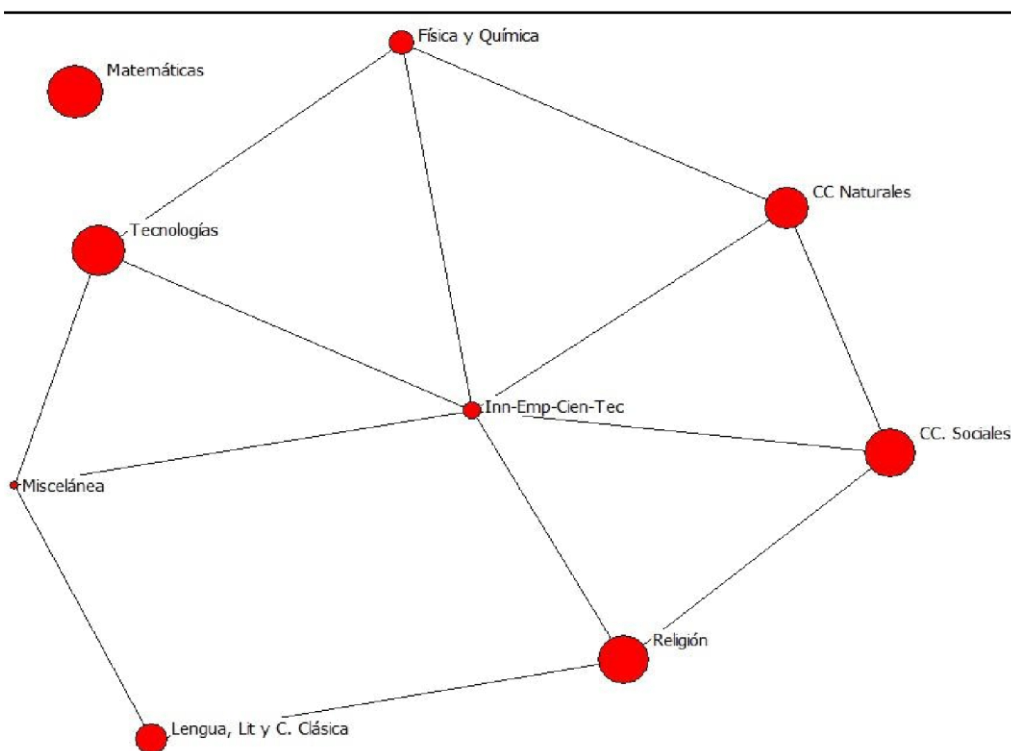


Figura 2. Red semántica de las diferentes materias

- Otra característica importante es la clara predominancia de las Tecnologías de la Información, desde el punto de vista de usuario, en el área de Educación Tecnológica.

- Sin embargo, hay suficientes elementos en los libros de texto de la ESO que permiten conectar los contenidos científicos y tecnológicos con conceptos relevantes de la cultura del emprendimiento y la innovación como, por ejemplo, los conceptos genéricos de Conocimiento, Competitividad, Industria, Producción, Economía, etc.

- El análisis manual de los contenidos identificados como Innovación y Emprendimiento, Ciencia y Tecnología refuerza la conclusión de que los libros de texto de la Enseñanza Obligatoria Secundaria se caracterizan por una visión sesgada de la

cultura científica. De un lado hay muy poca relación entre los contenidos científicos y los tecnológicos; de otro lado, hay poca relación entre elementos intrínsecos y extrínsecos de la cultura científica. Los contenidos científicos no están relacionados de forma significativa con su aplicación práctica. Y la tecnología es percibida más como un instrumento que uno tiene que aprender a usar que como un conjunto de soluciones prácticas basadas en el conocimiento científico. Además, parece haber una división en lo referente a la transmisión de la cultura científica: los libros de ciencia proporcionan conocimiento científico, mientras los libros humanísticos y de ciencias sociales transmiten representaciones de la ciencia y actitudes sobre ella.

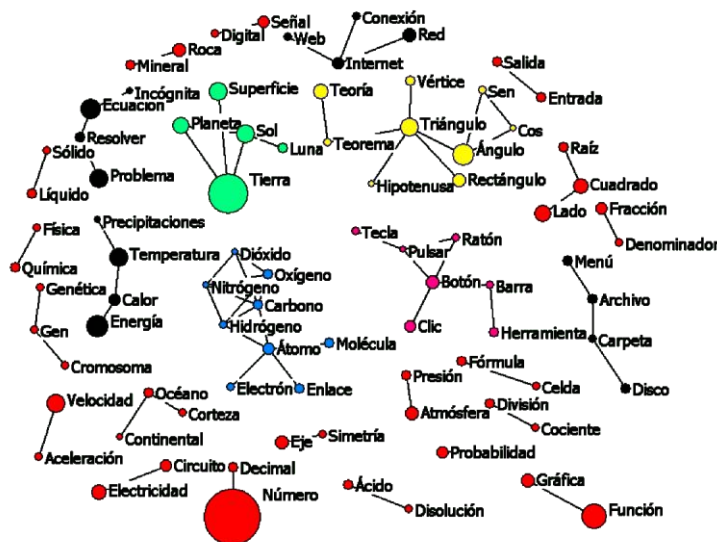


Figura 3. Conceptos de Tecnología

- A pesar de todo esto, la percepción de la ciencia transmitida a los estudiantes no es pesimista ni negativa; en realidad es optimista, o polarizada entre ambas posiciones (visión positiva de la ciencia y visión pesimista).

En la fase actual se está completando esta parte con el análisis de los libros de texto de la asignatura *Ciencias para el mundo contemporáneo*, del Bachillerato, así como con nuevos libros de Ciencias y Tecnología de primer y cuarto curso de la Enseñanza Secundaria Obligatoria.

Noticias de Prensa

De otro lado, se ha iniciado un estudio basado en noticias de prensa relacionadas con la Ciencia y la Tecnología. Por su mayor facilidad para operaciones de tratamiento automático se decidió trabajar con las versiones digitales de periódicos de tirada nacional; así, se seleccionaron los sitios web de El País, El Mundo y Público. El elevado número de documentos de prensa disponibles hacía inviable una selección manual, por lo que se decidió aplicar un clasificador que seleccionase de forma automática las noticias o documentos que tratasen, de una u otra forma, sobre temas relacionados con la Ciencia y la Tecnología.

Se trata de analizar un período amplio, el comprendido entre 2002 y 2011. No se han considerado todos los días de estos años, sino una muestra amplia construida ad-hoc. En el diseño de esta muestra se ha procurado cubrir los diferentes días de la semana a lo largo de todos los meses del año, de manera que para cada diario se han considerado unos 840 días; salvo para Público, que apareció en 2007.

La obtención de las noticias para cada uno de esos 840 días ha sido todo lo exhaustiva posible. Pero no ha sido sencilla, dada la estructura de los respectivos portales web, poco favorable a ello. En efecto, ha sido preciso emplear técnicas de *web scraping* (Moody et al, 2003), siempre dentro de la legalidad, limitándose a las partes de acceso público y abierto de dichos portales. La cantidad global de noticias de todas las secciones obtenidas para cada periódico está reflejada en la tabla 1.

	Noticias recogidas
El Mundo	86.139
El País	74.368
Público	27.357
TOTAL	187.864

Tabla 1. Cantidad de noticias recogidas

Para la selección de noticias enfocadas a Ciencia y/o Tecnología se ha utilizado un clasificador automático basado en algoritmos *Naive Bayes*, (Langley et al., 1992) con algunas especificidades añadidas propias del idioma español. Se trata de un algoritmo relativamente sencillo, basado en estimación de probabilidades a partir de las palabras presentes en una serie de documentos de muestra o entrenamiento, pero que ha demostrado alcanzar niveles de precisión tan elevados como otros algoritmos más complejos (McCallum y Nigam, 1998).

El principal problema de la aplicación de clasificadores automáticos es que éstos necesitan de una colección de documentos pre-clasificados, de los cuales el programa informático aprende a diferenciar unas clases de documentos de otras. El tamaño de esta colección de documentos de aprendizaje o entrenamiento es importante, y el resultado final del clasificador depende en buena medida del número de documentos de entrenamiento disponibles. Para construir dicha colección y reducir el proceso de pre-clasificación al máximo se aplicó un procedimiento de *bootstrapping* consistente en utilizar los documentos de la sección de Ciencia y Tecnología de El Mundo como entrenamiento inicial, y aplicarla al resto de las secciones de ese mismo diario, así como a los otros dos periódicos. Una muestra aleatoria de las noticias así obtenidas se revisó manualmente y con ella se constituyó la colección de entrenamiento aplicada finalmente.

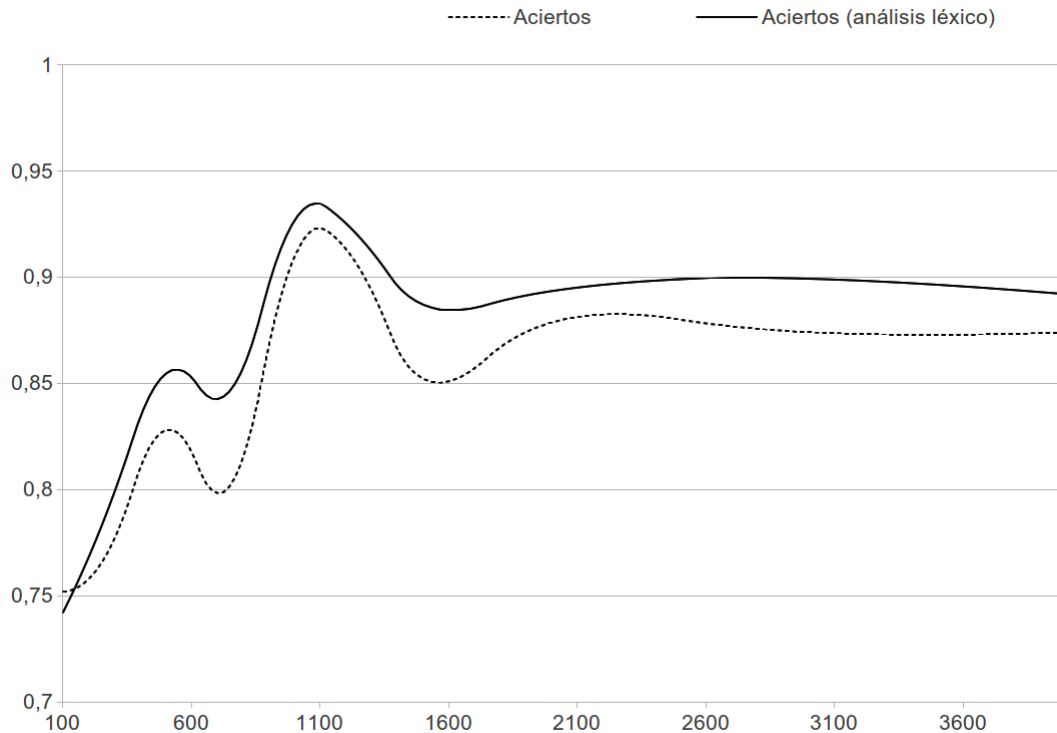


Figura 4. Tamaño de colección de entrenamiento y resultados del clasificador

Como resultado final, 17.573 noticias fueron señaladas por el clasificador automático como enfocadas en Ciencia o Tecnología. Un análisis manual de una muestra aleatoria de ellas mostró una tasa de aciertos del 94.1%. En la actualidad se trabaja con una selección de 1000 noticias clasificadas como de Ciencia y Tecnología, procediendo a su análisis cualitativo.

Adicionalmente, es posible someter los textos de las 17.573 noticias confirmadas como de Ciencia y/o Tecnología a análisis automáticos avanzados. Una de estas posibilidades es lo que se conoce como detección de entidades nombradas (*named entities*). De forma genérica consiste en identificar elementos atómicos en un texto y clasificarlos en diferentes categorías (personas, instituciones, lugares, eventos, acciones ...). La *Named Entity Recognition* (NER) es uno de los elementos claves en el funcionamiento del web semántico y existen en la actualidad diversos instrumentos informáticos que permiten llevarla a cabo.

Uno de tales instrumentos es el servicio conocido *OpenCalais*, impulsado por Thomson Reuters y que ofrece buenos resultados para el idioma español (Castrillo Espina, 2013). La tabla 2 recoge las 15 entidades más frecuentes de algunos de los tipos en los que se clasifican.

Tipo: Country	España Estados Unidos China Francia Japón Alemania Rusia Italia
---------------	--

	<p>Australia Canadá India Brasil Reino Unido Chile Portugal</p>
Tipo: Person	<p>Cristina Narbona Rafael Bachiller George W. Bush Barack Obama Jane Goodall Yuri Gagarin Al Gore Charles Darwin Stephen Hawking Elena Espinosa Albert Einstein José Luis Rodríguez Zapatero Isabel II Cristina Garmendia Carlos Duarte</p>
Tipo: Organization	<p>NASA ONU Unión Europea Greenpeace Ministerio de Medio Ambiente Agencia Espacial Europea Organización Mundial de la Salud UNESCO Consejo Superior de Investigaciones Científicas OMS Universidad de California Gobierno español Casa Blanca Ministerio de Sanidad Consejo de Seguridad Nuclear</p>
Tipo: Company	<p>Endesa Twitter Google Microsoft EDF Iberdrola Boeing Sony Jet Propulsion Laboratory Facebook BBVA Nintendo Nokia IBM Exxon</p>

Tabla 2. Entidades nombradas más frecuentes

La posibilidad de poner en relación de coocurrencia unas entidades y otras, así como la de analizar la evolución temporal de la aparición en prensa de determinadas entidades resulta prometedora.

La detección automática de temas (topic detection) consiste en la aplicación de técnicas automáticas para identificar materiales (fundamentalmente texto, por ahora) temáticamente relacionados (Jo et al., 2007; Jo et al. 2010). El escenario típico es el de un flujo más o menos continuo de documentos textuales, por ejemplo noticias de prensa, o comentarios en redes sociales como Twitter (Berrocal et al., 2013). Es obvio que tales técnicas pueden ser aplicadas a nuestra base de datos de noticias y podrían arrojar resultados útiles.

La Wikipedia

La web es un medio completamente novedoso en el panorama de la cultura científica, al que las encuestas y estudios tradicionales de comunicación pública de la ciencia no han prestado todavía suficiente atención. Su utilización para la elaboración de indicadores de cultura científica ofrece dificultades específicas, pero también resulta prometedor para acceder a aspectos de la cultura científica actual que no es fácil captar por otros medios. Dentro de la web, la Wikipedia es un fenómeno digno de mención, en este contexto. Al tratarse de un diccionario enciclopédico elaborado y mantenido colectivamente, a través de la web social, puede ser muy útil para revelar detalles y características sobre la percepción 'social' del conocimiento científico así como de la participación social en la elaboración, uso y difusión de éste.

Una de las ventajas de la Wikipedia es su acceso abierto, incluso a sus características de organización interna. Entre otras cosas, es posible descargar la base de datos completa, sin restricciones, de manera que constituye una fuente de datos de incalculable valor. Organizada, como es sabido, en idiomas, nos hemos centrado en la Wikipedia en español, trabajando, por el momento, con los datos de enero de 2012. Hay en esa versión unos 860.000 artículos o entradas, cada uno de los cuales está etiquetado con una o varias categorías; éstas son campos temáticos informales, de manera que cada editor es libre de asignar a cada artículo las categorías que estime oportuno, así como de crear categorías nuevas ad-hoc. La versión de la Wikipedia con que trabajamos tiene unas 60.000 categorías efectivas, una vez descartadas las puramente administrativas y las que sólo tienen uno o ningún artículo.

Por el momento, el principal problema con que nos enfrentamos consiste en localizar las entradas relacionadas con Ciencia y Tecnología, presumiblemente una pequeña parte de esos 863.000 artículos mencionados antes. Descartadas selecciones manuales a modo de muestreo, hemos decidido una primera aproximación a partir de las categorías temáticas, cuya cantidad es más reducida.

Hemos recopilado los artículos etiquetados con cada una de las 60.000 categorías y hemos extraído todos los hipervínculos existentes en esos artículos hacia otras entradas de la misma Wikipedia. De forma acumulativa, esto nos ha permitido construir un grafo de las relaciones o hiperenlaces entre todas las categorías (Chernov et al., 2006).

La aplicación de técnicas de Análisis de Redes Sociales nos ha permitido observar aspectos interesantes sobre aspectos del etiquetado social, o sobre el diferente funcionamiento de etiquetas o categorías transversales frente a otras de tipo específico. La figura 5 representa la distribución de categorías según el número de artículos, y claramente se ve que sigue las leyes de la exponenciación (Broder et al., 2000; Voss, 2005) mostrando fuertes semejanzas con las distribuciones analizadas por

Cattuto et al. (2007) para otros casos de etiquetado social. La tabla 3 de otro lado, muestra las 15 categorías con mayor grado absoluto (entendido como número de categorías enlazadas, en un sentido o en otro). El carácter marcadamente transversal es evidente; incluso en aquellas más cercanas la Ciencia es posible apreciar ese carácter (Términos Zoológicos, Botánicos, Categorías Taxonómicas, ...).

El Análisis de Redes Sociales nos ha permitido también aplicar algoritmos de Detección de Comunidades; las comunidades son grupos de categorías fuertemente hiperenlazadas entre sí y presumiblemente, por tanto, afines temáticamente. Así, el algoritmo conocido como *Infomap* (Rosvall y Bergstrom, 2008) nos permitió agrupar las 60.000 categorías en 839 comunidades que, al ser una cantidad pequeña, han podido ser revisadas de forma manual para identificar las categorías relacionadas con la Ciencia y la Tecnología. Éstas alcanzan la cifra de 3.471, y la cantidad de artículos o entradas que contienen es de 216.459; ésta es, en principio, la cantidad de entradas de Ciencia y Tecnología que contiene la Wikipedia en español. Sin embargo, como indican Holloway et al. (2007), una parte numéricamente importante de los artículos están elaborados de forma automática por programas importantes a partir de árboles taxonómicos, listas de asteroides, censos postales, etc.

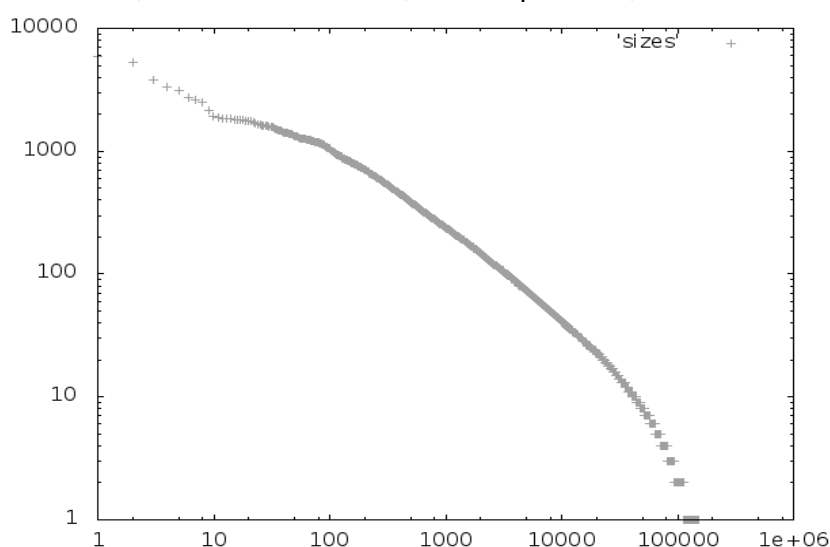


Figura 5. Distribución log-log de categorías y número de artículos

Categoría	Grado absoluto de enlaces
Personas_vivas	43940
Siglo_XX	25185
Años_1990	17641
Años_2000	16821
Siglo_XIX	12021
Años_1980	11783
Grandes_ciudades_de_España	8310
Escritores_en_español	7557
Términos_de_geografía	7551
España	7040
Años_1970	6683

Categoría	Grado absoluto de enlaces
Paseo_de_la_Fama_de_Hollywood	6506
Términos_botánicos	6465
Términos_zoológicos	6441
Actores_de_cine_de_Estados_Unidos	6226
Películas_de_Estados_Unidos	5956
Siglos	5934
Categorías_taxonómicas	5778
Unidades_de_longitud	5498

Tabla 3. Categorías Wikipedia con mayor grado de enlaces

La figura 6 muestra el grafo de todas las categorías de la Wikipedia, con aquellas que son de Ciencia y Tecnología marcadas en color rojo. El grafo agrupa los hiperenlaces entre todos los artículos de cada categoría y está elaborado tendiendo a situar próximas las categorías más interenlazadas entre sí. Claramente se aprecian núcleos compactos de categorías de Ciencia y Tecnología (también otros, correspondientes a otras áreas temáticas) que pueden ser analizados con mayor detalle, a nivel de los artículos o entradas individuales y sus enlaces.

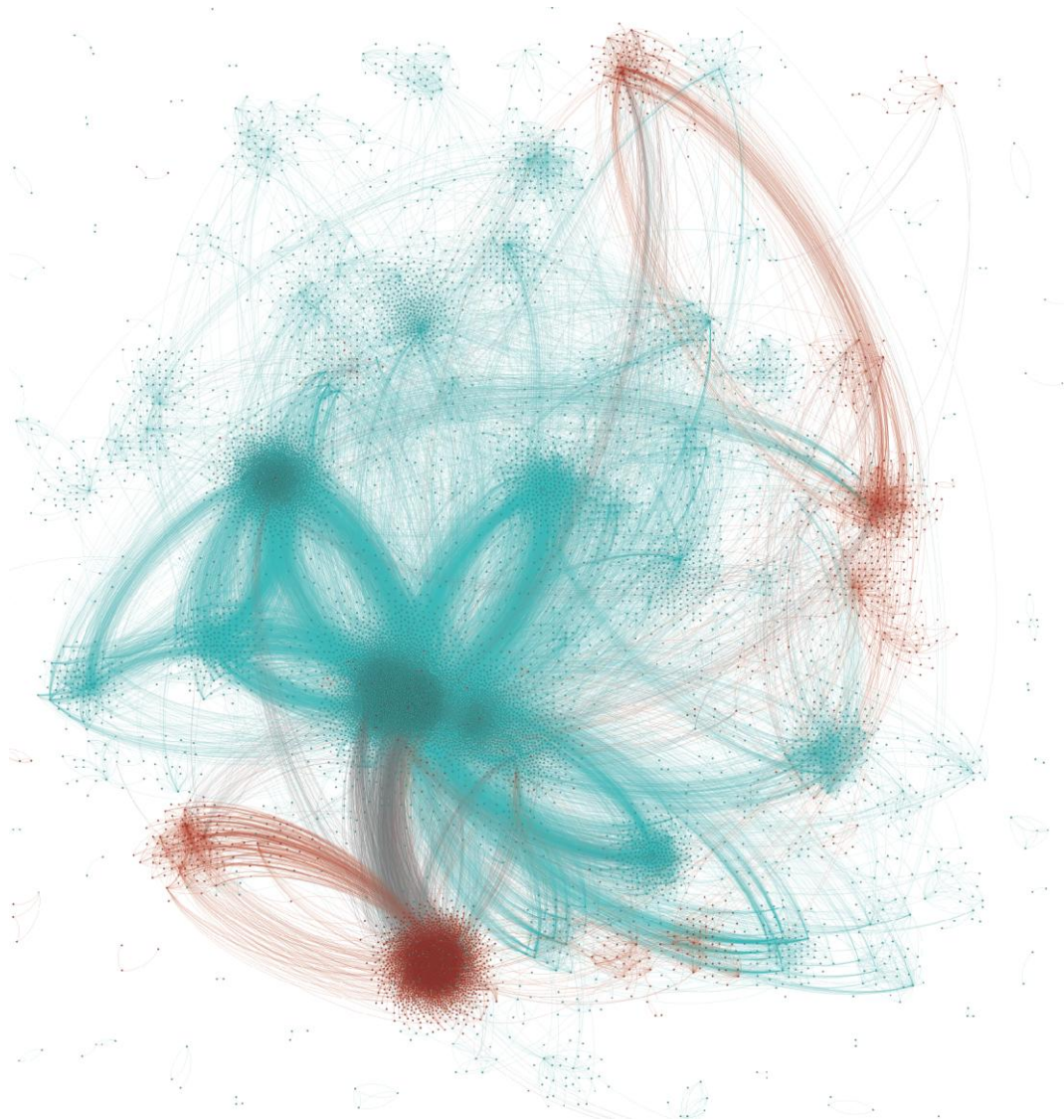


Figura 6. Las categorías de Wikipedia según sus enlaces

En esta fase, precisamente, se encuentra este aspecto de nuestro proyecto; pero parece factible obtener información de interés acerca de aspectos como la amplitud de la cobertura de las diferentes disciplinas o áreas temáticas y de la relación de éstas entre sí. También de las conexiones de temas o áreas científicas con otras zonas temáticas no científicas ni tecnológicas. De otro lado, el mismo tipo de análisis aplicado a datos de fechas posteriores (por ejemplo, de 2014) y su comparación con éstos podría resultar interesante.

Conclusiones y trabajo futuro

En las páginas precedentes hemos descrito los datos recopilados, las fuentes y las metodologías de recopilación y filtrado, en su caso, utilizadas. Algunos de esos datos han sido ya analizados a fondo y hemos podido extraer valiosas conclusiones: éste es el caso de los libros de texto de la Enseñanza Secundaria Obligatoria española, sin perjuicio de nuevos análisis.

Con otros datos, como las noticias de prensa o la Wikipedia, hemos finalizado la fase de recopilación y preparación y está en marcha el análisis de esas informaciones. Las

metodologías a aplicar en dicho análisis parecen claras: análisis cualitativo manual en el caso de las noticias de prensa, así como, desde el punto de vista automatizado, detección de temas, coocurrencia y tratamiento de entidades nombradas. En el caso de la Wikipedia, aplicación de técnicas de Análisis de Redes Sociales, además de la utilización complementaria de técnicas cuantitativas simples (visitas a los artículos, número de ediciones/editores, ...).

Sin embargo, además del estudio por separado de los datos proporcionados por estas fuentes, tal vez existe la posibilidad de vincular entre sí algunas de ellas. En efecto, las noticias de prensa transmiten eventos relacionados con Ciencia y Tecnología que, en la medida en que los medios se hacen eco de ellos, son reveladores acerca de la manera en que la gente percibe la cultura científica y tecnológica. Algunos autores sugieren que determinados eventos pueden también producir eco en la propia Wikipedia (Osborn et al., 2012): un aumento en las visitas o consultas a determinados artículos, pero también un aumento en el número de ediciones de determinadas entradas, o incluso la aparición de nuevos artículos de la Wikipedia ¿es posible confirmar y evaluar el efecto de ciertos eventos y noticias en la Wikipedia? Los datos disponibles invitan a explorar ese camino, y ésta será una de las tareas futuras en nuestro proyecto.

Bibliografía

Aibar, E. y M. A. Quintanilla (2002). *Cultura tecnológica : estudios de ciencia, tecnología y sociedad*. Barcelona: ICE Universidad de Barcelona

Alonso Berrocal, J. L.; Figuerola, C. G. y Zazo Rodríguez, A. F. (2013). REINA at RepLab2013 Topic Detection Task: Community Detection, *CLEF 2013, Lab on Online Reputation Management*, <http://reina.usal.es/biblio/files/berrocal2013reina.pdf>

Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins, A. y Wiener, J. (2000). Graph structure in the web, *Computer networks*, 33 (1), 309-320

Castrillo Espina, M. (2013). OpenCalais. Desarrollo de las tecnologías del lenguaje y utilidad del reconocimiento de entidades nombradas para la labor del documentalista, Universidad de Salamanca (Tesis Fin de Máster). <http://reina.usal.es/biblio/files/castrillo2013opencais.pdf>

Cattuto, C.; Schmitz, C.; Baldassarri, A.; Servedio, V.; Loreto, V.; Hotho, A.; Grahl, M. y Stumme, G. (2007). Network properties of folksonomies, *AI Communications*, 20, 245-262

Chernov, S.; Iofciu, T.; Nejd, W. y Zhou, X. (2006). Extracting Semantics Relationships between Wikipedia Categories. SemWiki, 2006, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.5507&rep=rep1&type=pdf>

Holloway, T.; Bozicevic, M. y Börner, K. (2007). Analyzing and visualizing the semantic coverage of Wikipedia and its authors Complexity, *Wiley Online Library*, 12, 30-40

Jo, Y.; Lagoze, C. y Giles, C. L. (2007). Detecting research topics via the correlation between graphs and texts, *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 370-379

Jo, Y.; Hopcroft, J. E. y Lagoze, C. (2011). The web of topics: discovering the topology of topic evolution in a corpus, *Proceedings of the 20th international conference on World wide web*, 257-266

Langley, P.; Iba, W. y Thompson, K. (1992). An analysis of bayesian classifiers. *Proceedings National Conference on Artificial Intelligence*. San Antonio, CA: AAAI Press and MIT Press, 223–228

Martínez, J.C. (2010). *Los factores culturales y locales en los procesos de transferencia de tecnología en un programa de cooperación al desarrollo en la República Dominicana*, <http://hdl.handle.net/10366/76521>

McCallum, A. y Nigam, K. (1998). A comparison of event models for naive bayes text classification. *AAAI-98 workshop on learning for text categorization*, 41-48. <http://www.kamalnigam.com/papers/multinomial-aaaiws98.pdf>

Montañés, Ó. (2010). *Problemas epistemológicos de la comunicación pública de la ciencia*. Instituto de Estudios de la Ciencia y la Tecnología. Salamanca: Universidad de Salamanca.

Moody, K. & Palomino, M. (2003). SharpSpider: Spidering the Web through Web Services, *First Latin American Web Congress*, 219-221

Osborne, M.; Petrovic, S.; McCreadie, R.; Macdonald, C. y Ounis, I. (2012). Bieber no more: First story detection using Twitter and Wikipedia *SIGIR 2012 Workshop on Time-aware Information Access*

Quintanilla, M. A. (2000) Técnica y cultura. *Teorema*, XVII/3, 49-69.

Quintanilla Fisac, M. A.; Lawler, D.; González Rodríguez, M. D.; Montañés Perales, O.; Montero Becerra, J. A. y Pedrera Gómez, D. J. (2004). Public understanding, scientific culture perception and civic engagement indicators, Scientific Knowledge and cultural diversity. *Public communication of science and technology network*. ISBN: 84-497-0122-8, pg: 433-435.

Quintanilla, M. A. (2005). *Tecnología: Un enfoque filosófico y otros ensayos de filosofía de la tecnología*. México: Fondo de Cultura Económica

Quintanilla Fisac, M. A.; Escobar, M.; Montero, J. A. et al. (2009). *La cultura científica y tecnológica en los libros de texto de la educación secundaria obligatoria* (Documento de Trabajo. Instituto ECYT).

Quintanilla, M.A. (2012). Cultura, Tecnología e innovación. En E. Aibar & M. A. Quintanilla (Eds.), *Ciencia, tecnología y sociedad. Enciclopedia Iberoamericana de Filosofía*. Madrid: Trotta.

Rosvall, M. y Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure, *Proceedings of the National Academy of Sciences*, National Academy of Sciences, 2008, 105, 1118-1123.