INSTRUCCIONES PARA LA CUMPLIMENTACIÓN DE LA MEMORIA FINAL

La memoria final se compone de la memoria técnica y la memoria económica. Todos los campos son obligatorios.

Dentro de la memoria técnica, el título del proyecto, la fecha de inicio y de fin, los participantes, la institución, el nombre del grupo de investigación, el resumen y objetivos y la descripción de las acciones de difusión y sostenibilidad del proyecto, aspectos a cumplimentar en los datos descriptivos, serán publicados de forma permanente en la página web http://www.fundacionhergar.org/ y en el archivo del histórico de proyectos subvencionados de la Fundación Hergar.

Por ello deben contener la extensión óptima para su publicación. Los puntos del 1 al 6 han de tener una **extensión mínima de 900 palabras**. Así mismo, ha de respetar los siguientes indicadores de estilo: **interlineado de 1.5** y letra **Times New Roman 11**.

MEMORIA FINAL

MEMORIA TÉCNICA

DATOS DESCRIPTIVOS

- Título del Proyecto: El turismo 2.0 y los Social media como medios de influencia en las decisiones de compra
- Fecha de inicio: 20/03/2014 Fecha fin: 19/03/2015
- Nombre del Grupo de Investigación: Big Data and Business Intelligence in Social Media
  - Nombre Investigador Principal: Mª del Rocío Martínez Torres
  - DNI:
  - Dirección personal: Facultad de Turismo y Finanzas, Avda. San Francisco Javier s/n 41018 Sevilla (España)
  - Teléfono: 954554310
  - E-mail: rmtorres@us.es
- Institución/entidad a la que se asocia el Investigador Principal: Universidad de Sevilla
- Subvención concedida por la Fundación HERGAR (en euros): 500€

Indicar las personas que han participado en el proyecto subvencionado, así como la entidad a la que pertenecen

Apellidos, nombre	DNI	Institución a la que pertenecen
Martínez Torres, Mª del Rocío		Universidad de Sevilla
Toral Marín, Sergio		Universidad de Sevilla
Castellanos Verdugo, Mario		Universidad de Sevilla
Díaz Fernández, Mª Carmen		Universidad de Sevilla
Espasandín Bustelo, Francisco		Universidad de Sevilla
Oviedo García, Mª de los Ángeles		Universidad de Sevilla
González Rodríguez, Mª Rosario		Universidad de Sevilla
Bessis, Nik		University of Derby
Lawless, Aileen		Liverpool John Moores University

1. Resumen y objetivos del proyecto

El turismo 2.0 representa un nuevo paradigma basado en el uso de los medios electrónicos y social media en el que las empresas turísticas deben adaptarse continuamente a escenarios cambiantes. La rápida proliferación de la web 2.0 y el *e-commerce* ha propiciado un rápido auge de los portales de opinión o boca a boca electrónico, donde los usuarios pueden expresar opiniones y preferencias sobre una gran diversidad de productos y/o servicios. Portales como *TripAdvisor, Ciao, Epinions* representan hoy en día unos influyentes canales a través de los cuales los usuarios pueden aprender sobre la oferta turística y la calidad de los productos ofertados, basando muchas veces en ellos sus decisiones de compra. A su vez, las compañías turísticas pueden utilizar estas opiniones para recopilar las preferencias de los usuarios y modificar consecuentemente sus productos o servicios. No obstante, la información contenida en estos portales se caracteriza por ser dispersa, masiva y no estructurada, y se encuadra dentro del campo denominado *Big Data*. Estas características dificultan la extracción de información relevante para las compañías y hace necesario la utilización de nuevas técnicas metodológicas.

Este proyecto propone la utilización de técnicas basadas en el análisis de redes sociales y el análisis semántico para la identificación de los líderes de opinión (también llamados en la literatura "*influencers*"), que son aquellos usuarios que gozan de una elevada reputación y cuyas opiniones o revisiones ejercen una gran influencia sobre las decisiones de compra de otros usuarios. La identificación de estos usuarios es clave para las empresas turísticas. No sólo porque permite concentrarse en aquellas opiniones verdaderamente interesantes dentro de los cientos o miles de opiniones compartidas, sino porque las revisiones de estos usuarios apuntan las mejores innovaciones que podrían realizarse dentro de los productos o servicios ofertados, así como posibles campañas de marketing.

Las técnicas metodológicas propuestas contemplan la extracción de la información masiva, dispersa y no estructurada, para generar modelos de redes sociales y modelos semánticos que permitan extraer datos sobre la participación de los usuarios y los contenidos compartidos. También se incluyen algoritmos estadísticos de identificación y clasificación de los *influencers* y sus opiniones a partir de los datos generados.

Como resultado del proyecto se generará un software en R que permita la extracción automatizada de los modelos sociales y semánticos, y su procesamiento estadístico para identificar los líderes de opinión. La propuesta tiene un claro carácter multidisciplinar, propio del campo de los *Big Data*, que involucra áreas como la informática, los sistemas de información, la estadística y las ciencias sociales.

Los grupos de interés a los que afecta la propuesta son, principalmente, todas las compañías turísticas interesadas en hacer un seguimiento de su marca y sus productos a través de los portales de opinión para futuras mejoras. También aquellas compañías que aplican técnicas de marketing viral, que pueden hacer uso de los líderes de opinión para mejorar la eficiencia de los procesos virales. Asimismo,

los resultados de este proyecto también tienen implicaciones importantes sobre los propios gestores de portales de opinión, interesados en medir la reputación de los usuarios y la credibilidad de sus opiniones.

La metodología propuesta es genérica y aplicable a cualquier organización, producto o servicio. Es importante reseñar que todo el proceso de extracción de información y generación de los modelos sociales y semánticos estará automatizada mediante la herramienta software resultado del proyecto, por lo que las compañías no necesitarán dedicar ingentes recursos humanos para la monitorización de la información, siendo únicamente necesario un analista especializado de datos.

El **objetivo general** del proyecto es analizar el impacto del turismo 2.0 sobre las decisiones de compra a través del grupo de usuarios líderes de opinión.

Los principales **objetivos específicos** de impacto que se persiguen con esta propuesta son:

- Identificar los líderes de opinión o *influencers* en las comunidades de opinión turísticas, como grupo clave para conocer las preferencias de los usuarios o como grupo sobre el que desarrollar estrategias de marketing viral
- Definir metodologías para la identificación de los *influencers* alternativas a la reputación y popularidad, que resulten más objetivas y más difícilmente manipulables
- Proporcionar un conjunto de herramientas desarrolladas en R a los gestores de las comunidades de opinión turísticas que permitan clasificar los patrones de comportamientos de los diferentes grupos de usuarios
- Aplicar metodologías alternativas para la gestión y procesamiento de los *Big Data* en el contexto de las comunidades de opinión en turismo

2. Descripción de la relación de actividades desarrolladas durante la ejecución del proyecto

El plan de trabajo se estructuró en 4 tareas básicas:

Tarea 1: Marco teórico y estudios previos. Esta tarea se centró en recopilar bibliografía relacionada con las comunidades de opinión, así como en el análisis de comunidades de opinión con relevancia a nivel mundial. La recopilación bibliográfica sirvió para establecer el cuerpo teórico necesario para definir las aportaciones de los artículos de investigación. Esencialmente, la bibliografía previa se centra en la identificación de los usuarios líderes de acuerdo a la teoría definida por von Hippel. En cuanto a las estrategias de implantación de las comunidades de opinión, éstas se basan fundamentalmente en portales web que funcionan como comunidades, donde los usuarios pueden escribir revisiones y opiniones, comentar las ideas o revisiones de otros usuarios o incluso puntuarlas según una escala. Típicamente, existen comunidades implantadas bien por terceras compañías, donde se pueden revisar y comentar una amplia variedad de productos pertenecientes a muchas temáticas, o bien comunidades patrocinadas por una compañía concreta para que los usuarios puedan comentar sus productos o servicios.

Dentro de las primeras se eligió el portal ciao.com, que cuenta en la actualidad con más de 7 millones de revisiones sobre 1.4 millones de productos y cerca de 1.3 millones de usuarios registrados. Dentro de las comunidades patrocinadas por una compañía se eligió el portal MyStarbucksIdea, donde usuarios de Starbucks de todo el mundo pueden proponer mejoras sobre los productos y servicios de esta compañía.

Tarea 2: Extracción de datos. La extracción de datos se centró por una parte en la extracción de las actividades de participación de los usuarios y por otra en la extracción del contenido compartido.

Para la extracción de los datos se desarrolló un programa crawler en R. Para ello se utilizó la función *readLines* del paquete básico de R, que se encarga de leer el contenido de una URL. No obstante, las páginas web están en código HTML y, por tanto, contiene muchos tags pertenecientes al propio código. Por este motivo, el archivo HTML se parseó mediante la función *htmlParse*, que genera una estructura en R que representa el árbol HTML. A partir de esta estructura y utilizando las expresiones regulares soportadas por R se extrajo la información relevante. En el caso de las actividades de participación, dicha información se refiere al alias que identifica a cada usuario, su experiencia como revisor dado por el número de revisiones previas, la puntuación numérica otorgada con su revisión y la valoración de su revisión por el resto de miembros de la comunidad. Para el análisis del contenido compartido se extrajeron las cabeceras y cuerpos de los mensajes enviados.

En total, dentro de la plataforma Ciao se han analizado más de 17.000 revisiones realizadas por más de 13.000 usuarios diferentes. La plataforma Ciao distingue 28 categorías principales, establecidas

desde el propio portal que, a su vez, se subdividen en multitud de categorías secundarias definidas por los propios usuarios.

Por otra parte, en el caso de la plataforma MyStarbucksIdea existen de 15 categorías. El número de revisiones extraídas por cada una de las categorías se detalla en la Tabla 1.

	Category	Nº
Product ideas	Coffee & Espresso Drinks	**9500**
	Frappuccino & Beverages	**2687**
	Tea & other drinks	**7405**
	Food	**9500**
	Merchandise & Music	**7113**
	Starbucks Card	**9500**
	New Technology	2633
	Other Product Ideas	**8508**
Experience ideas	Ordering or Payment & Pick-Up	6338
	Atmosphere & Locations	9500
	Other Experience Ideas	**9500**
Involvement ideas	Buiding Community	4453
	Social Responsability	**6984**
	Other Involvement Ideas	4699
	Outside USA	**1208**

Tabla 1. Ideas extraídas por categoría en el portal MyStarbucksIdea.

Tarea 3: Análisis de la actividad participativa y contenido compartido. Como resultado del análisis de la actividad participativa se validó una primera hipótesis referente a que tanto las revisiones de productos como la actividad de los usuarios siguen una distribución según una ley de potencias data por la ecuación $Cx^{-\alpha}$, siendo C una constante y α el exponente de la distribución de potencias. La Figura 1 detalla la frecuencia de revisión de los productos y servicios incluidos en Ciao en tanto que la Figura 2 muestra la frecuencia de participación de los usuarios. En ambos caso se muestra superpuesta el ajuste de una distribución de ley de potencias, siguiendo para ello el método propuesto por Clauset *et al.* (2007). El valor del exponente obtenido es de α =2.29 para el caso de la frecuencia de revisiones de productos y servicios, y de α =3.5 para la participación de los usuarios. El test de Kolmogorov-Smirnov en ambos casos se encuentra por debajo del valor umbral dado $1.63/N^{0.5}$. Por tanto, la hipótesis nula se cumple y el ajuste es adecuado en ambos casos.

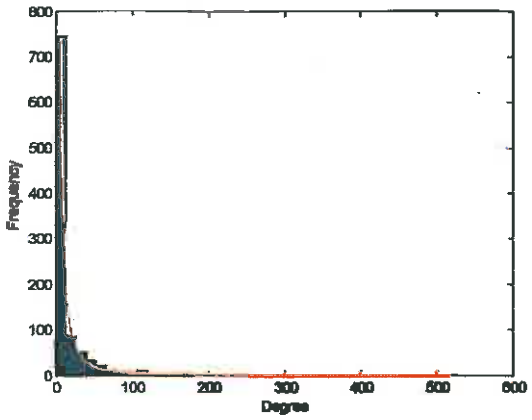

Figura 1. Frecuencia de revisión de los productos y servicios incluidos en Ciao.

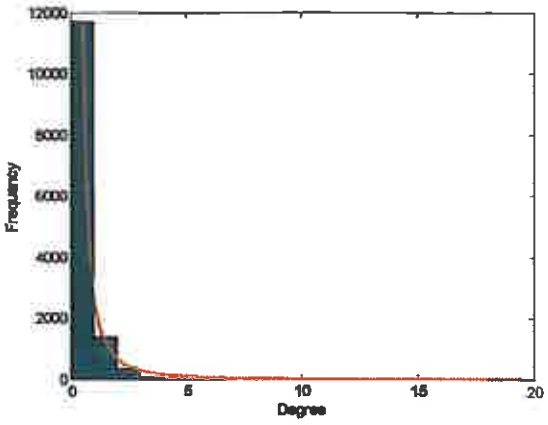

Figura 2. Frecuencia de participación de los usuarios registrados en Ciao.

El siguiente análisis que se llevó a cabo fue intentar identificar los usuarios líderes o influencers a partir de la red social que forman los miembros de la comunidad. Para ello es necesario en primer lugar definir quiénes son los usuarios líderes a partir de sus características definidas por la teoría de usuarios líderes de Von Hippel (1998). Estudios previos muestran que la actividad y la reputación son las dos características esenciales de los usuarios líderes. La actividad puede medirse por el número de revisiones enviadas por cada usuario en tanto que la reputación viene dada por la puntuación o rating recibidas por las revisiones enviadas. Ambos datos fueron capturados en la tarea anterior. En el trabajo se consideraron diferentes umbrales para ambas variables, distinguiendo de este modo lo casos de (a) a (h) detallados en la Tabla 2.

Cases	Posted reviews	% users
(a)	≥2	15.5
(b)	≥3	4.4
	Rev_rating	**% users**
(c)	≥50	12.3
(d)	≥60	9.1
(e)	≥70	6.0
(f)	≥80	3.8
(g)	≥90	3.0
(h)	≥100	2.5

Tabla 2. Valores umbrales para la selección de los usuarios líderes.

A partir de los ocho casos de la variable dependiente de la **Tabla 2** se llevaron a cabo 8 regresiones logísticas binarias, utilizando las siguientes variables independientes:

- Grado: es el número de líneas que inciden (grado de entrada) o salen (grado de salida) de él (Toral et al., 2009a). Mayores grados de nodos producen redes más densas, porque los nodos involucran más arcos y el valor medio del grado de los nodos de una red no es una medida dependiente del tamaño de la red.

- Centralidad de cercanía: es un índice de centralidad basado en el concepto de distancia. La centralidad cercana de un nodo se calcula considerando el total de distancias entre un nodo y todos los demás nodos, donde la distancia más larga ofrece una menor puntuación de centralidad cercana. La centralidad cercana es un índice definido para toda la red y se calcula como la variación en la centralidad cercana de los vértices dividida por la variación máxima posible en la puntuación de centralidad cercana en una red del mismo tamaño (Toral et al., 2009b).

- Centralidad de intermediación: es una medida de la centralidad que reside en la idea de que un nodo es más central en la medida en que actúe como intermediario en una red de comunicación (Nooy et al., 2005). Es decir, la centralidad de un nodo depende de la medida en la que es necesario como enlace para facilitar la conexión de otros nodos dentro de la red. Si se define una geodésica como el camino más corto entre dos nodos, la centralidad de intermediación de un vértice es la proporción de todas las geodésicas entre pares de nodos que incluyen este nodo, y la centralidad en la intermediación de una red es la variación en la centralidad de intermediación de los nodos dividida por la máxima variación posible en la centralidad de intermediación en una red del mismo tamaño.

- Coeficiente de clustering: mide si los vecinos de primer nivel de un nodo interactúan entre ellos. Es una medida de cohesión local a partir de las interacciones entre los nodos vecinos de un nodo.

La * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Tabla 3 muestra los resultados de las ocho regresiones logísticas binarias, mientras que la Tabla 4 detalla las tablas de clasificación de los usuarios líderes y no líderes, así como los porcentajes de usuarios correctamente clasificados. Se puede observar que se ha obtenido un porcentaje de clasificación

satisfactorio en todos los casos. En el caso de los usuarios líderes el porcentaje es siempre sensiblemente más bajo toda vez que sólo representan un pequeño porcentaje dentro de la comunidad de usuarios. El test de Hosmer y Lemeshow que evalúa la bondad del ajuste fue no significativo en todos los casos, lo que significa una buena concordancia entre los resultados predichos y observados.

Case	Variables	Logistic coefficient (standard error)	Wald
(a)	**degree**	-0.003 (0.000)	32.803***
	Closeness	30.628 (1.932)	251.255***
	betweeness	7113.895 (461.478)	237.637***
	CC	-12.600 (0.413)	929.461***
	Nagelkerke R^2	0.840	
(b)	**degree**	0.001 (0.000)	5.445*
	Closeness	16.200 (2.439)	44.130***
	betweeness	1369.163 (145.300)	88.794***
	CC	-9.597 (0.459)	436.375***
	Nagelkerke R^2	0.721	
(c)	**degree**	-0.002 (0.000)	19.756***
	Closeness	22.756 (1.716)	175.895***
	betweeness	2204.314 (195.039)	127.733***
	CC	-9.329 (0.255)	1342.689***
	Nagelkerke R^2	0.702	
(d)	**degree**	-0.001 (0.000)	3.901*
	Closeness	17.231 (1.517)	129.000***
	betweeness	1100.862 (137.632)	63.977***
	CC	-8.127 (.235)	1198.723***
	Nagelkerke R^2	0.620	
(e)	**degree**	0.000 (0.000)	0.189
	Closeness	14.010 (1.650)	72.127***
	betweeness	1020.512 (122.133)	69.818***

Case	Variables	Logistic coefficient (standard error)	Wald
	CC	-7.480 (0.264)	805.633***
	Nagelkerke R^2	0.597	
(f)	degree	0.000 (0.000)	0.012
	Closeness	16.664 (2.611)	40.732***
	betweeness	1043.436 (123.785)	71.055***
	CC	-8.562 (0.398)	462.481***
	Nagelkerke R^2	0.665	
(g)	degree	0.000 (0.000)	0.848
	Closeness	17.177 (3.140)	29.924***
	betweeness	933.355 (117.366)	63.243***
	CC	-8.768 (0.471)	346.476***
	Nagelkerke R^2	0.675	
(h)	degree	0.000 (0.000)	0.360
	Closeness	23.394 (3.812)	37.651***
	betweeness	339.915 (80.079)	18.018***
	CC	-9.103 (0.496)	336.668***
	Nagelkerke R^2	0.630	

* p<0.05 ** p<0.01 *** p<0.001

Tabla 3. Resultados de la regresión logística para los 8 casos considerados.

Case	Observed	Estimated		Percentage correct
		Non-influencers	Influencers	
(a)	Non-influencers	12037	67	99.4
	Influencers	185	208	52.9

Case	Observed	Estimated		Percentage correct
		Non-influencers	Influencers	
	Total percentage correct			98.0
(b)	**Non-influencers**	11880	72	99.4
	Influencers	207	338	62.0
	Total percentage correct			97.8
(c)	**Non-influencers**	10619	313	97.1
	Influencers	421	1144	73.1
	Total percentage correct			94.1
(d)	**Non-influencers**	11007	316	97.2
	Influencers	529	645	54.9
	Total percentage correct			93.2
(e)	**Non-influencers**	11583	122	99.0
	Influencers	400	392	49.5
	Total percentage correct			95.8
(f)	**Non-influencers**	11922	79	99.3
	Influencers	230	266	53.6
	Total percentage correct			97.5
(g)	**Non-influencers**	12037	67	99.4
	Influencers	185	208	52.9
	Total percentage correct			98.0
(h)	**Non-influencers**	12123	55	99.5
	Influencers	180	139	43.6
	Total percentage correct			98.1

Tabla 4. Matrices de clasificación.

Los resultados de la regresión logística muestran que existe una dependencia nula o en muchos casos no significativa del grado de los nodos. Un valor alto del grado significa que el nodo se encuentra conectado a muchos otros nodos, que son usuarios de la comunidad. Esto significa que el usuario está realizando revisiones sobre productos muy populares que, a su vez, reciben revisiones de otros muchos consumidores. El resultado obtenido se puede justificar si se tiene en cuenta que una gran mayoría de usuarios solamente realiza una revisión, y que esta trata sobre algún producto popular evaluado también por muchos usuarios similares. Por otra parte, los resultados muestran también una dependencia significativa y positiva con la centralidad de cercanía e intermediación. Cuanto mayor es la centralidad de los nodos en términos de distancia e intermediación respecto a otros nodos, mayor es la probabilidad de ser un usuario líder. Por último, existe una dependencia significativa pero negativa con el coeficiente de clustering. El coeficiente de clustering mide el ratio entre las interacciones de los nodos en una vecindad de un nodo dado y su grado. Dado que el grado no es significativo, este resultado se puede interpretar como un valor bajo del numerador, que significa que los usuarios líderes normalmente revisan una amplia variedad de productos. Por ejemplo, diferentes destinos turísticos, o varios hoteles dentro de una misma ciudad.

Para el análisis del contenido compartido se trabajó con las opiniones de la plataforma MyStarbucksIdea. La técnica a aplicar es la conocida como Latent Semantic Indexing (LSI) o Análisis Semántico Latente. Se basa en crear una matriz término-documentos a partir de un conjunto de palabras clave relacionadas con el tema de interés, y aplicar una descomposición en valores singulares para reducir la alta dimensionalidad del espacio original. LSI considera que aquellos documentos con muchas palabras en común en el nuevo espacio originado se encuentran semánticamente cercanos, en tanto que aquellos que tienen pocas en común están semánticamente lejanos (Deerwester et al., 1990).

Desde un punto de vista formal, dado n documentos que contienen m palabras clave, la matriz términos-documentos es una matriz A de mxn donde el elemento (i,j) representa el número de veces que la palabra i aparece en el documento j, siendo $n<<m$. Dos documentos o columnas de la matriz A son similares si comparten muchas palabras clave en común. En definiva, el coseno de los vectores columna define la similitud de los documentos. El principal problema de usar la matriz de términos-documentos es que los resultados son muy dependientes del conjunto de palabras clave elegido. Por este motivo, LSI aplica una descomposición en valores singulares que permite proyectar la matriz A sobre un conjunto reducido k de dimensiones, utilizando los k mayores valores singulares de la matriz A, con $k<<n$. LSI recalcula el número de ocurrencias en este nuevo espacio de k dimensiones. En este espacio reducido, el significado de las palabras clave se infiere del contexto en el que ocurren. De este modo LSI evita el

problema de la sinonimia, dado que las palabras sinónimas suelen usarse en el mismo contexto (Kawaguchi *et al.*, 2006).

La Figura 3 muestra los resultados obtenidos en forma de dendograma. Lo primero que se observa es que la organización semántica de las subcategorías es diferente de la propuesta en la plataforma Web.

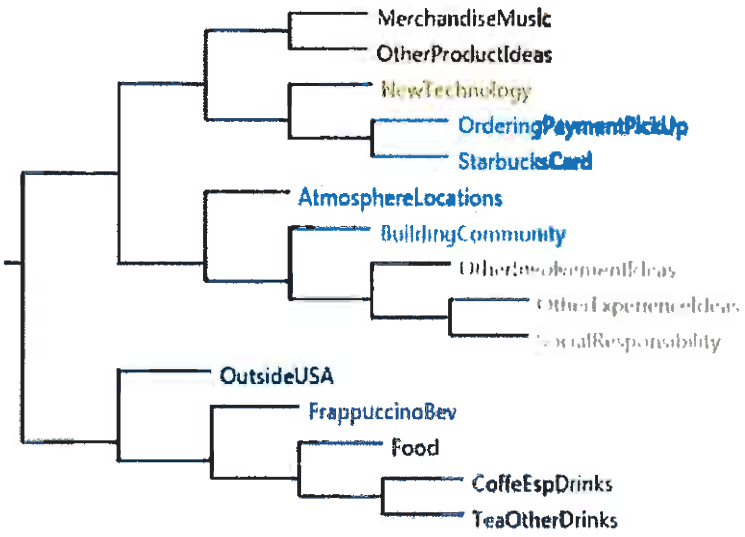

Figura 3. Agrupación de categorías mediante LSI.

En primer lugar, existe una clara relación entre las bebidas y comidas ofrecidas por Starbucks, como puede apreciarse en la parte inferior del dendograma. La parte superior muestra la otra mitad de las ideas de producto, que se refieren a Merchandise & Music y Other Product Ideas. No obstante, Starbucks Card y New Technology aparecen semánticamente más cercanos a las Experience Ideas. Estas dos subcategorías junto con Ordering o Payment & Pick-Up constituyen el núcleo de las experiencias de los clientes en Starbucks. Esto significa que los clientes consideran Starbucks Card y Technology no tanto como productos sino como facilitadores de su experiencia en Starbucks. La otra mitad de las Experience Ideas es Atmosphere & Locations, que aparecen en el dendograma cerca de las Involvement Ideas. El último grupo que se encuentra en la mitad del dendograma es el de las Involvement Ideas, que incluye Other Experience Ideas.

Dentro de la plataforma MyStarbucksIdea se denominan Ideas in Action a aquellas ideas proporcionadas por los usuarios que han sido finalmente adoptadas por la compañía. La Figura 4 muestra la distribución de las Ideas in Action por cada una de las categorías de ideas.

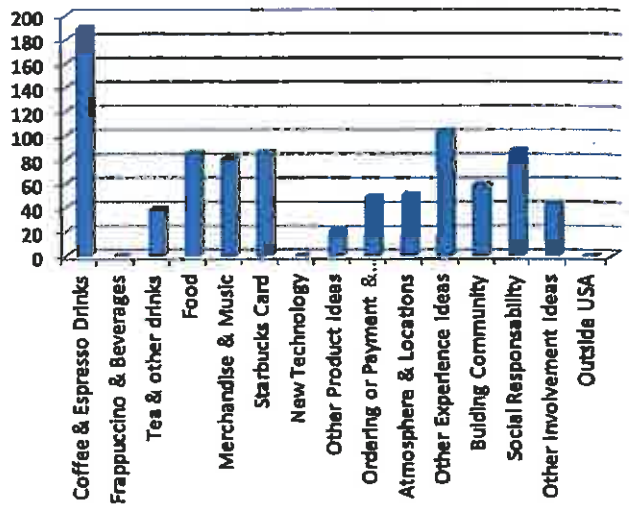

Figura 4. Distribución de Ideas en Acción por categoría.

Coffee & Espresso Drinks, con 190 ideas, es claramente la categoría en la que un mayor número de ideas han sido seleccionadas. Este resultado está en línea con el producto principal ofrecido por Starbucks y que se encuentra más íntimamente ligado a su imagen de marca. El segundo y tercer lugar lo ocupan Other Experience Ideas y Social Responsibility, respectivamente.

La categoría Other Experience Ideas es el lugar para todos aquellos comentarios no directamente clasificables en las otras categorías, tales como partners, otros tipos de premios de fidelidad o cambios en la decoración del local. Esta categoría incluye las percepciones y emociones de los usuarios con respecto a Starbucks, que resulta prioritario para una compañía que considera que la experiencia de tomar un café en Starbucks es un sello distintivo respecto a la competencia. Lo mismo puede decirse de la responsabilidad social. Starbucks se posiciona como una compañía amigable con el medio ambiente, preocupada por los problemas sociales tanto a nivel global como en las comunidades locales en las que se encuentra presente. En general, puede concluirse que las tres categorías en las que más ideas se han implementado están muy asociadas a la imagen de marca de la compañía.

Tarea 4: Identificación de usuarios líderes. Este estudio ha proporcionado algunas implicaciones gerenciales importantes. Por una parte, las revisiones online generadas por los consumidores permiten que las compañías se acerquen a los consumidores y usuarios (Lee, 2007), anticipando tendencias futuras que pueden ser relevantes (Gamon et al., 2005). En este contexto, la identificación de usuarios líderes resulta de gran interés para muchas compañías, dada la influencia que sus revisiones pueden tener en las decisiones de compra de otros consumidores (Ku et al., 2012). Por otra parte, la identificación de los usuarios líderes también permite la aplicación de técnicas de marketing viral para incrementar el interés

sobre determinados productos o mejorar la fidelidad a una marca. Estas técnicas funcionan de manera más eficiente cuando se aplican sobre los usuarios líderes que, según se ha visto, ocupan posiciones relevantes dentro de la red (centralidad). Esta posición les permite realizar la difusión de información de forma creíble y eficiente a través de portales de opinión como los descritos en este este estudio (Kiss and Bichler, 2008). Por último, y desde la perspectiva de los gestores de las comunidades de opinión, también es interesante la identificación de los usuarios líderes para reconocerlos explícitamente como usuarios con buena reputación e intentar retenerlos.

El análisis de los contenidos también resulta de interés para contrastar las preferencias de los consumidores frente a las prioridades de las compañías. Del análisis del portal MyStarbucksIdea se desprende que los usuarios tienden a centrarse en el núcleo de actividad de la compañía, fundamentalmente en aquellas ideas que les proporcionan mayores beneficios y comodidad. En general, los servicios proporcionan valor a los clientes cuando son usados (Vargo and Lusch, 2004). En consecuencia, este valor está íntimamente ligado a la experiencia previa de los clientes y esta experiencia incluye no solamente tomar los productos que ofrecen en Starcbucks, sino también el resto de actividades que rodea a consumir en el establecimiento, tales como pedir, pago, o tarjetas de fidelidad. El análisis de las denominadas Ideas in Action revela que la compañía se centra en las características distintivas de la imagen de marca. En este sentido, hay estudios que señalan una relación positiva entre la imagen de marca, el valor percibido por los consumidores y sus decisiones de compra (Wu, 2008; Cretu & Brodie, 2007)

Tarea 5: Difusión de resultados. Se elaboró un plan de difusión de resultados que incluye publicaciones en congresos y revistas, detalladas en el punto 5 de este documento.

Referencias:

- Clauset, A., Shalizi, C. R., Newman, M. E. J. (2007), Power-law distributions in empirical data, SIAM Review, 51:661-703.
- Cretu, A. E, Brodie, R. J. 2007. The influence of brand image and company reputation where manufacturers market to small firms: A customer value perspective, Industrial Marketing Management, 36, 2, 230-240.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis, Journal of the American Society of Information Science, 41 (6), 391-407.
- Gamon, M., Aue, A., Corston-Oliver, S. and Ringger, E. (2005), Pulse: Mining customer opinions from free text, Advances in intelligent data analysis, 6, 121−132.

- Kawaguchi, S., Garg, P. K., Matsushita, M., Inoue, K. (2006). MUDABlue: An automatic categorization system for Open Source repositories, The Journal of Systems and Software 79, 939-953.
- Kiss, C. and Bichler, M. (2008), Identification of influencers — Measuring influence in customer networks, Decision Support Systems, 46, 233–253.
- Ku, Y.C., Wei, C.P. and Hsiao, H.W. (2012), To whom should I listen? Finding reputable reviewers in opinion-sharing communities, Decision Support Systems, 53, 534–542.
- Lee, T.Y. and Bradlow, E. T. (2007), Automatic construction of conjoint attributes and levels from online customer reviews. The Wharton School, University of Pennsylvania.
- Nooy, W.; Mrvar, A., y Batagelj, V. (2005). Exploratory Network Analysis with Pajek, Cambridge University Press, New York.
- Toral, S. L.; Martínez-Torres, M. R., y Barrero, F. (2009a). Virtual Communities as a resource for the development of OSS projects: the case of Linux ports to embedded processors, Behavior and Information Technology, 28 (5), 405-419.
- Toral, S. L.; Martínez-Torres, M. R.; Barrero, F., y Cortés, F. (2009b). An empirical study of the driving forces behind online communities, Internet Research, 19 (4), 378-392.
- Vargo, S.L. and Lusch, R.F. 2004. Evolving to a new dominant logic for marketing, Journal of Marketing, 68, 1-17.
- Von Hippel, E. (1988). The Sources of Innovation, New York, Oxford University Press.
- Wu, W. C., 2008. The Study of Influence of Brand Equity, Customer Value, Customer Satisfactions and Customer Loyalty ?A Case Study of Wretch, MA Taiwan: Department of Communications Management, Ming Chuan University.

3. Describir las acciones de difusión y sostenibilidad del proyecto, aportando las evidencias oportunas

Publicaciones:

- M.R. Gonzalez-Rodriguez, M. R. Martinez-Torres, S. L. Toral, "Monitoring travel-related information on Social Media through sentiment analysis", 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, UCC 2014, London (UK), pp. 636-6441.
- M. R. Martinez-Torres, S. L. Toral, F. Rodriguez-Piñero Royo, "Company and user preferences in Open Innovation Communities through content analysis", VII International Meeting in Dynamics of the Socio-Economic System, Dyses 2014, Seville (Spain).
- M. Olmedilla, M. R. Martinez-Torres, S. L. Toral, "A long tail study of eWOM communities", International Conference on Communities and Technologies, ICCT 2015, New York (USA), aceptado.

Artículos enviados:

- M. R. Martínez-Torres, F. Rodriguez-Piñero, S. L. Toral, "Customer preferences versus managerial decision making in Open Innovation Communities: the case of Starbucks", Technology Analysis and Strategic management.

Página web:

La página web del grupo de investigación se encuentra disponible en:
http://grupo.us.es/gbdosdata/

WEBDATANET:

La investigadora principal participa en el comité ejecutivo de la Acción Europea WEBDATANET. Dentro de esta Acción se ha organizado una sesión especial titulada 'Big Data analysis and its applied áreas' (http://webdatanet.cbs.dk/images/SESSIONSKATJA/rocomartneztorres.pdf), dentro del meeting final de la Acción a celebrar en Salamanca en mayo de 2015 (http://webdatanet.cbs.dk/index.php/component/content/article/176).

4. Indicar si se ha seguido la metodología de trabajo propuesta o ha habido alguna variación

Las cuatro tareas básicas en las que se dividió el trabajo se realizaron en las fechas previstas. Para una duración total de 12 meses (48 semanas), el resultado temporal fue el siguiente:

- Tarea 1: Marco teórico y estudios previos: Semanas 1-9
  - Revisión bibliográfica: Semanas 1-5.
  - Comunidades de opinión: Semanas 5-9
- Tarea 2: Extracción de datos: Semanas 10-18
  - Redes de usuarios: Semanas 10-14.
  - Contenido compartido: Semanas 13-18
- Tarea 3: Análisis: Semanas 19-27
  - Análisis de redes sociales: Semanas 19-24
  - Análisis semántico: Semanas 24-27
- Tarea 4: Identificación de usuarios líderes: Semanas 28-40
  - Parámetros topológicos: Semanas 28-36
  - Análisis de contenido: Semanas 32-40
- Tarea 5: Difusión: Semanas 39-48

5. Indicar los órganos de evaluación y seguimiento para la consecución de los objetivos fijados. Periodicidad prevista para el seguimiento e indicadores.

El seguimiento para la consecución de los objetivos fijado se hizo en base al grado de cumplimiento del plan establecido.

Indicador: Grado de cumplimiento

6. Indicar si se han identificado nuevas necesidades para alcanzar los objetivos del proyecto

Contribuciones teóricas:

La principal contribución teórica se refiere la identificación de los usuarios líderes usando parámetros topológicos de la red social de la que forman parte. Para ello se han utilizado algunas de las características definidas por Von Hippel sobre los usuarios líderes y se ha avanzado en su caracterización. La identificación de este perfil de usuario también permite una alternativa a los esquemas de evaluación colectiva, que no distinguen entre usuarios experimentados o con buena reputación.

Contribuciones metodológicas:

Desde un punto de vista metodológico se han aportado nuevas técnicas de extracción y procesado de la información visible en los Social Media.

Por lo que respecta a la extracción de datos, se ha utilizado R para acceder de manera masiva a los contenidos e interacciones de los usuarios a través de varios portales de opinión. Esto permite trabajar con un gran volumen de usuarios y contar con diversidad y variedad de opiniones. En la parte de procesamiento, se ha trabajado con técnicas de análisis de redes sociales y de análisis de textos, que permiten transformar conceptos intangibles como la participación o las opiniones en datos medibles y procesables.

Contribuciones prácticas:

La propuesta ofrece mejoras significativas a los administradores y a las compañías interesadas en hacer un seguimiento sobre los portales de opinión online. Por un lado, somete a crítica los esquemas de evaluación colectiva implantados en muchas comunidades de opinión y mide la reputación de los usuarios mediante técnicas más difícilmente manipulables. Por otra parte, ofrece un método alternativo para captar las revisiones susceptibles de ser implementadas mediante la identificación de los usuarios líderes. Un mejor rendimiento en el proceso de evaluación de las ideas compartidas se traduce en una mejora de la capacidad de absorción de conocimiento por parte de la organización, evitando dedicar grandes recursos humanos a la labor de evaluación de las ideas y que estas puedan ocultarse o identificarse tardíamente.

MEMORIA ECONÓMICA

JUSTIFICACIONES

Se deberán presentar los justificantes (copia del pago bancario o factura con recibí) y/o facturas de las actuaciones realizadas y los correspondientes recibos de pagos.

La justificación económica se realizará mediante la aportación de original y fotocopia o copia compulsada.

- Costes Directos:

- Material Inventariable:
En esta partida se contabiliza aquel material que no es susceptible de un rápido deterioro debido a su uso. Algunos de los materiales que incluiríamos en esta partida son el mobiliario, equipos de investigación, ordenadores, impresoras, etc. Es necesario adjuntar factura proforma.

	Descripción	Justificación	Número de Artículos	Coste Unitario	Total Coste
			a	b	a x b
1	TOSHIBA PORTEGE R30-17C (PT341E-077041CE)	Adquisición de un ordenador dedicado a la extracción y análisis de grandes volúmenes de información. Este dispositivo fue co-financiado con dinero de otro proyecto del que la IP de este proyecto también es responsable	1	1486.78	410
2					

- Material Fungible:
En esta partida se contabiliza el material que sufre un rápido deterioro y que requiere reposición. Algunos de los materiales que incluiríamos en esta partida son las grapadoras, folios, bolígrafos, correspondientes al material de oficina.

	Descripción	Justificación	Número de Artículos	Coste Unitario	Total Coste
			a	b	a x b
1	Unidad NAS D-Link DNS-320LW	Adquisición de dispositivo de almacenamiento masivo para guardar la información capturada con el programa crawler. Este dispositivo fue co-financiado con dinero de otro proyecto del que la IP de este proyecto también es responsable	1	106.61€	90€
2					

Cantidad obtenida / prevista y Fuentes de financiación :

Costes Directos					Total Costes Directo	Costes Indirectos	Total Gastos Proyecto
C. Personal	C. Operacional					Total Costes Indirectos (No > 7%)	
G.Pers	Inventariable	Fungible	Viajes y Dietas	Otros Gastos			
	1486,78	106,61	0	0	1593,39	0	1593,39
Total							

Ingresos	Otras Fuentes		Dotación Fundación Hergar	Total Ingresos
Financiación Propia	Subv. Publica	Privada		
	1093,39	0	500	1593,39
Total	1093,39		500	1593,39

Firma de la Investigadora Principal

Sello de la entidad

Mª del Rocío Martínez Torres

UNIVERSIDAD DE SEVILLA

Fdo. Mª del Rocío Martínez Torres

En Sevilla, a 12 de marzo de 2015

J. RODRIGO JIMENEZ DIAZ
UPI - FERIA
FERIA Nº135, LOCAL UPI
41002 SEVILLA
Tel. +34 954 962 911

e-Mail info@upiferia.com
C.I.F 45659910B

FACTURA

U.SEVILLA DPTO. INGENIERIA ELECTRONICA

41092 SEVILLA
SEVILLA
Q-4118001-I

Número factura	Fecha	Fecha Valor	Referencia
A/1388	24/09/2014	24/09/2014	

Descripción

Cantidad	Código	Artículo	Precio		IVA	Subtotal
1,00	880898768775	TOSHIBA PORTEGE R30-17C (PT341E-077041CE)	1.486,78		21,00	1.486,78
1,00					Subtotal	1.486,78

U.P.I.
UNIÓN PROFESIONAL INFORMÁTICA
J. RODRIGO JIMÉNEZ DIAZ - N.I.F. 45.659.910-B
C/ [illegible] Sevilla
Tel.: 954 [illegible] Fax 954 90 84 38

UNIVERSIDAD DE SEVILLA

	Descuento	%	Dto P.Pago	%	IVA	Base Imponible	Importe IVA	Importe R.E.
					21,00%	1.486,78	312,22	

TOTAL FACTURA
1.799,00 €

Forma de Pago
TRANSFERENCIA
BANCARIA
Vencimientos :

Nº de Cta. J. Rodrigo Jimenez
ES60 0049 5739 87 2795020135

UNIVERSIDAD DE SEVILLA
ÁREA DE CONTRATACIÓN Y PATRIMONIO
UNIDAD DE INVENTARIO

IMPRESO PARA DAR DE ALTA EN EL INVENTARIO GENERAL | **CÓDIGO:**

A rellenar por la Unidad de Inventario

Nº INVENTARIO GENERAL: .....

A rellenar por el Centro/Departamento/Servicio

Nº JUSTIFICANTE DEL GASTO: .....

DESCRIPCIÓN DEL ELEMENTO

Nº DE INVENTARIO DE CENTRO: 2006

DESCRIPCIÓN DEL BIEN: UN ORDENADOR PORTÁTIL

ELEMENTOS QUE LO COMPONEN:

MARCA: TOSHIBA ..... **MODELO:** PORTEGE ..... **Nº SERIE:** .....

MATRÍCULA(vehículos): ..... **Nº DE BASTIDOR (vehículos):** .....

SITUACIÓN DEL BIEN: .....

UBICACIÓN DEL ELEMENTO

UBICACIÓN ECONÓMICA (Indicar orgánica): 18.09.01.01.01 y 18.09.01.27.09 ..... 2014 - 211 2013 - 1506 13,9 % 86,1 %

UBICACIÓN ORGANIZATIVA (centro de coste): Dpto. Administración de Empresas y Comercialización e Investigación de Mercados (Marketing) .....

UBICACIÓN GEOGRÁFICA

CAMPUS: Ramón y Cajal .....

EDIFICIO: Facultad de Turismo y Finanzas .....

PLANTA: 1ª .....

LOCAL: Dpto. Administración de Empresas y Comercialización e Investigación de Mercados (Marketing) .....
SUBLOCAL: 1. El turismo 2.0 y los social media como medios de influencia en las decisiones de compra. 2. Análisis del paradigma del Software de código abierto desde la perspectiva del análisis semántico y del análisis de redes sociales. .....

VALORACIÓN DEL ELEMENTO

PROVEEDOR/TRANSMITENTE: JUAN RODRIGO JIMENEZ DIAZ UPI - FERIA .....

NÚM. DE FACTURA: A/1388 ..... **FECHA DE FACTURA/DOCUMENTO:** 24/09/2014 .....

PRECIO DE ADQUISICIÓN (Unitario con IVA incluido aplicando prorrata): 1.486,78 € .....

VALOR DECLARADO EN DOCUMENTO OFICIAL (Adquisiciones sin factura) .....

VALOR ESTIMADO (Otras altas) .....

Firma y sello del responsable funcional

Fdo.: D. Antonio José [illegible]

LA UNIDAD DE INVENTARIO

UNIVERSIDAD DE SEVILLA