



Facultad de Turismo  
Universidad de Málaga



# IX CONGRESO NACIONAL TURISMO Y TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

Málaga, 25 y 26 de octubre de 2012  
Palacio de Ferias y Congresos

## **Turistic RobSoC: Collective Intelligence y Robotic Social Curation aplicada al Turismo**

Eneko Astigarraga<sup>1</sup>

Josu Azpillaga<sup>2</sup>

Luis Fernandez<sup>2</sup>

Aitzol Naberan<sup>2</sup>

Daniel Reguera<sup>3</sup>

Urko Zurutuza<sup>3</sup>

<sup>1</sup> Universidad de Deusto

<sup>2</sup> CodeSyntax. Internet Solutions

<sup>3</sup> Universidad de Mondragon – Escuela Politecnica Superior

CodeSyntax – Poligono Azitain 3-K; E-20600 Eibar, Gipuzkoa

[eastigarraga@codesyntax.com](mailto:eastigarraga@codesyntax.com)

## **Resumen**

RobSoC Turístico, por Robotic Social Curation, es una extensión ampliada y aplicada al sector turismo de las investigaciones y experiencias que en CodeSyntax se comenzaron a desarrollar con el proyecto Umap, un intento práctico de obtener inteligencia colectiva del flujo de las redes sociales centrado en el análisis de información intercambiada en el seno de comunidades lingüísticas.

Si en Umap el objeto social compartido era el idioma mismo, en Turístico RobSoC el objeto social es el Turismo, y se avanza un paso más en la denominada content curation aplicada al turismo. Así, mediante la aplicación Turístico RobSoC se analizan, filtran y valoran los enlaces que los usuarios comparten, de manera que por un lado la inteligencia colectiva se convierte en información estructurada y por otro de manera automatizada se valoran, seleccionan, editan y comparten los contenidos turísticos que son más pertinentes y relevantes a juicio de la comunidad de expertos en turismo que comunican información sobre el tema a través de las redes sociales.

Así, con Turístico RobSoC el equipo investigador de CodeSyntax, una pyme dedicada al desarrollo de aplicaciones de Internet, ha pretendido centrar su investigación así como la aplicación resultante, en el ámbito de la inteligencia colectiva y la social curation automatizada en el sector del turismo. El proyecto ha contado con el apoyo del equipo investigador de la Escuela Politécnica Superior de la Universidad de Mondragón, Mondragón Unibertsitatea, en el proyecto ICCU- y el de la Universidad de Deusto, que han permitido poner en marcha con éxito una aplicación que es ya operativa.

**Palabras clave:** collective intelligence, social curation, microblogging, social network, turismo, umap

## **Introducción: social media, inteligencia colectiva y Social curation**

Numerosos autores (Zeng, Chen, Lusch, & Shu-Hsing, 2010) han demostrado y defendido que en los medios sociales o Social Media se genera, disemina y comunica contenido entre las comunidades de forma conversacional y distribuida. Las fronteras entre los autores y los lectores están desapareciendo a la vez que el consumo y diseminación de la información se mezclan con el proceso de generación y compartición. En este sentido, Internet y las tecnologías móviles están siendo las plataformas clave para el florecimiento y la intensificación del Social Media.

En la denominada sociedad en red (Castells, 2005), cualquier nodo de la misma es tan capaz como cualquier otro de transmitir un mensaje. Como subrayaba el filósofo y sociólogo francés Pierre Levy nadie sabe todo, todos sabemos algo, pero todo el conocimiento reside en las redes. En definitiva, en la sociedad en red el conocimiento se genera de forma colectiva (Levy, 1994; 1997).

Pero pese a la aparente fusión entre autores y lectores, entre emisores y receptores, como sugieren Bhattacharyya et al. (2011), la opinión de aquellos que son considerados por la comunidad como expertos convenientemente informatizada, analizada y ponderada, permite obtener contenido interesante y de mucho valor para la propia comunidad, para las empresas y organizaciones del sector y también para expertos, generadores de contenido y el público en general.

En cualquier ámbito el Social Media es ya una parte importante del ecosistema de la información, lo que está generando nuevas líneas de investigación y una avalancha de aplicaciones orientadas al mismo. Como día a día podemos observar, desde un punto de vista comercial o de negocio, las redes y los medios sociales son una nueva fuente de información para la innovación de productos, una nueva forma de relacionarse con clientes y proveedores, un poderoso medio de marketing, ... Al mismo tiempo, son un medio inigualable para informar a ciudadanos, recabar sus preocupaciones o formar grupos de apoyo o de rechazo que, por ejemplo, tanta importancia tienen en el ámbito de los proyectos turísticos.

En Turistic RobSoC el equipo investigador de CodeSyntax, una pyme dedicada al desarrollo de aplicaciones de Internet, ha pretendido avanzar sobre los resultados obtenidos en proyecto Umap (Astigarraga, Azpillaga, Fernandez & Naberan, 2011) y centrar su

investigación así como la aplicación resultante, en el ámbito de la inteligencia colectiva y la social curation automatizada en el sector del turismo, y con el apoyo del equipo investigador de la Escuela Politécnica Superior de la Universidad de Mondragón, Mondragón Unibertsitatea, en el proyecto ICCU- y la Universidad de Deusto se ha puesto en marcha con éxito una aplicación que es ya operativa (CodeSyntax, 2011)

*ICCU, "Inteligencia Colectiva a partir del flujo Comunidades de Usuarios",* es un proyecto desarrollado por la empresa CodeSyntax en colaboración con la Escuela de Ingeniería, MGEP de Mondragón Unibertsitatea y enfocado al estudio, definición y desarrollo de un sistema flexible de seguimiento y filtrado de información de redes sociales, con objeto de extraer la inteligencia colectiva contenida en las participaciones de los usuarios, en el sector turístico. Ha sido desarrollado en el marco del programa GAITEK, un programa de apoyo a la I+D del Gobierno Vasco, financiado parcialmente con fondos europeos.

Aunque desde el punto de vista de las TICs la investigación en torno al social media ha tendido a centrarse en lo que se denominan Social Media Analytics y Social Media Intelligence, la misma no puede enfocarse desde una sola disciplina. El Social Media Analytics se ha preocupado de crear las herramientas para recoger, monitorizar, analizar, resumir, y visualizar información de los medios sociales. Y si bien desde el punto de vista técnico, hay numerosos retos por resolver, la analítica de los medios sociales persigue facilitar la conversación e interacción entre las comunidades por un lado y por otro extraer patrones e inteligencia útiles que sean utilizables en la práctica.

Por su parte, el Social Media Intelligence contiene un conjunto de datos enriquecidos o metadatos que no ha sido tratado sistemáticamente (etiquetas, opiniones, apreciaciones, evaluaciones, valoraciones, perfiles...).

En Turistic RobSoC se ha tratado de conjugar ambos enfoques y los problemas planteados, al igual que la naturaleza de las aplicaciones sociales, son muy diversos: ¿cómo se puede capturar de forma eficiente la variación del interés sobre un tema en los cortos y crípticos mensajes de Twitter?, o, por ejemplo, ¿cómo puede afectar una opinión negativa sobre un hotel de un cliente concreto en las reservas de turistas ingleses?

Turistic RobSoC trata de avanzar, y en buena medida lo logra, en la búsqueda de soluciones automáticas, aunque probablemente todavía, como señalaba B. Liu en su artículo de 2010, se necesite de asistencia humana para la interpretación y aprovechamiento de

resultados.

La dificultad de gestionar la ingente cantidad de información turística, o de otros ámbitos, disponible es un hecho constatado. En el sector turístico existe la necesidad de filtrar la información relevante debido al exceso de información. En este contexto, los *Content Curators* identifican, organizan y comparten la información que será más relevante para sus clientes potenciales. Son capaces de proporcionar la información pertinente, aunque también detallan las relaciones entre esa información para que los clientes entiendan cómo todo ese contenido se reúne. Como defiende Bhargava en su manifiesto (Bhargava, 2009) la figura del *Content Curator* es necesaria.

La *Content Curation* es una nueva tendencia que consiste en seleccionar, editar y compartir los contenidos más pertinentes respecto a un tema específico, utilizando una combinación de máquina y recursos humanos y que puede inscribirse dentro de la denominada web semántica.

La "content curation" suscita controversia entre los profesionales de los medios de comunicación (Bhargava, 2009), con profesionales destacados a favor y en contra de la práctica. Sin embargo, responsables de nuevos medios y editores como Arianna Huffington de The Huffington Post y Michael Arrington de Tech Crunch defienden apasionadamente el content curation, dado que entienden que los modelos de negocio de los medios de comunicación tienen que evolucionar.

En este contexto, para ejercer eficientemente su labor los *Content Curator* necesitarán de ayudas automatizadas para consumir y gestionar la información con el fin de interpretar y entender mejor cómo hacer frente a las necesidades de información de sus usuarios. Turistic RobSoC permite eliminar incluso, o reducir al mínimo la labor de los *Content Curators*, que deben de procesar rápidamente la información diaria, usando herramientas (como la que ofrece Turistic RobSoC) que organizan y automáticamente clasifican contenidos, entregan resúmenes, lo evalúan y clasifican según sea necesario. Turistic RobSoC pretende también ofrecer algunas de esas herramientas a estos nuevos profesionales del sector turístico.

Turistic RobSoC permite recolocar el contenido relevante de forma rápida y fácil. Materializar esta potencialidad puede ayudar a colocar al sector turístico de cualquier territorio, destino u oferta, en una posición ventajosa respecto a otros destinos u ofertas.

## 2. Metodología: extracción de información turística desde las redes sociales

Como se ha señalado, extraer información relevante de los inputs generados por los distintos usuarios de las redes sociales requiere definir, en un primer momento, la comunidad o colectivo a estudiar, que estará caracterizado por compartir algún interés o rasgo. Éste puede ser, como en el caso de UMAP (CodeSyntax, 2010), el uso de un determinado idioma (catalán, gaélico, euskara...), tener un mismo interés (coches, motos, perfumes...), ser usuario de una marca o empresa, algo social o político, pertenecer a una comunidad geográfica... , o como en Turistic RobSoc, el sector turístico.

Figura 1: Noticiero automatizado para el sector turístico en Niagarank. Septiembre 2012

El primer paso del proceso consiste en definir un primer colectivo inicial de expertos en el tema a tratar, en este caso turismo: agencias, profesionales, empresas, organizaciones, comunicadores, mayoristas, destinos, promotores, ... RobSOc analiza las redes sociales de los mismos y va ampliando y creando de forma automatizada, dinámica y jerarquizada el colectivo de expertos que permitirá extraer la información relevante. En concreto, Turistic RobSOc lleva a cabo un seguimiento de más de 3.000 expertos y de forma dinámica va actualizando un ranking de los mismos. Por ejemplo, en la figura que sigue a continuación podemos apreciar el ranking de expertos generado el 7 de septiembre de 2012 a las 13:00 horas:



Usuario	Cu
Ruralworker	htt
promoturismo	htt
TravelReportMX	htt
MundoTuristico	htt
turismoas	htt
turiskopio	htt
fanturistic	htt
Turijobs	htt
Vsinlimite	htt
Crisnaveda	htt
SEGITTUR	htt
turisticate	htt
lara_soler	htt
andalucianet	htt
David_Mora	htt
guiasviajar	htt
gersonbeltran	htt
hacemosturismo	htt
viajesnet	htt
internetyturism	htt

Figura 2: ranking dinámico de expertos (2012/09/07, 13:00)

A partir de la información suministrada por los expertos, se trata de leer, interpretar y extraer información relevante compartida por todas esas personas que hablan sobre un tema - en este caso escriben sobre él- en un determinado idioma (español, inglés, francés, alemán), y analizar únicamente los tweets que cumplan con esos requisitos. Es la tarea que desarrolla Turistic RobSOc

A continuación pasamos a exponer una aplicación concreta de Turistic RobSOc aplicada al sector turismo. Es decir, el proceso seguido es el siguiente:

- Recoger en tiempo real todos los inputs en redes sociales (tweets) de todos los expertos seleccionados
- Discriminar los mensajes de los expertos según un parámetro concreto. Sólo nos interesan algunos de ellos, los referidos a temáticas turísticas
- Analizar los tweets y extraer trends o palabras clave, así como establecer un ranking de links compartidos por esa comunidad de expertos.

Los robots de Turistic RobSOc siguen a varios centenares de cuentas de Twitter, y valoran esas cuentas a partir de un ranking de interés específico relacionado con el turismo en España. El ranking es dinámico y automático, una vez definidos unos parámetros, el sistema recaba información de usuarios potenciales y los valora en relación al tema; no dependen, pues, de ningún registro, ni el mismo es necesario.

Una vez definidos los usuarios más interesantes en el sentido que poseen un mayor grado de expertise y son referenciados por la comunidad, el sistema recoge todos sus tweets y analizamos los enlaces que intercambian. Sobre esos enlaces también se aplica otro algoritmo que intenta detectar el contenido asociado al tema en cuestión. Y cuando una noticia dada sobrepasa un umbral de publicación, es seleccionado y va a portada del noticiero automatizado y su contenido es objeto de análisis.

Actualmente RobSOc publica las noticias con unos snippets de texto mínimo, un thumbnail si detecta imágenes, y en algunos casos en los que los sitios web de origen permiten el embebido de contenidos (vídeos de Youtube, por ejemplo), también se integran en los noticieros automatizados. Además, se integran como comentarios a la noticia los tweets que han mencionado el enlace en cuestión y también, si existe, la conversación agregada.

### *Minería de datos para la explotación de los datos proporcionados por Turistic RobSoc*

En el proceso de investigación se han analizado los diferentes algoritmos de explotación de datos y algoritmos de aprendizaje para automatizar al máximo el proceso de minería de datos.

Analizando las diferentes posibilidades y evaluándolas se procedió a utilizar técnicas *clustering mediante kmeans o k-medias* (MacQueen, 1967) para la parte de minería de datos y para el aprendizaje de modelos en la técnica probabilísticas como el muestreo de Gibbs. De hecho, fueron las dos herramientas seleccionadas para el primer desarrollo que está en funcionamiento.

### *Visualización de los datos*

Al igual que en UMAP, la visualización de datos se basa en un herramienta de monitorización donde el cliente puede visualizar, analizar y estudiar la información para mejorar su estrategia, ver que se está comentando y que le sirva como una herramienta eficaz a la hora de la toma de decisiones.

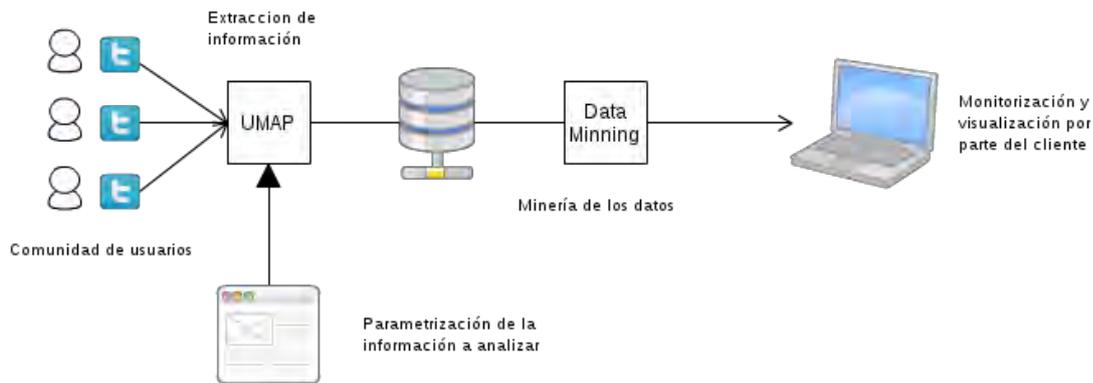


Figura 3. Arquitectura de enlace entre los diferentes componentes del producto final

### Características

Las características que presenta Turistic RobSOc definidas en función de los requisitos funcionales del producto final, son los siguientes:

Sistema configurable por parte del cliente: El cliente ha de poder configurar el ámbito a analizar.

Sistema capaz de extraer el los flujos de información de las diferentes redes sociales: En este momento es operativo con la red social Twitter.

Sistema capaz de procesar información y mostrarla de una manera clara: visualización de datos en web. pública

Sistema capaz de añadir usuarios a la comunidad a analizar: El cliente puede añadir diferentes expertos turisticos potenciales para analizar las informaciones que dan en las redes sociales

Sistema capaz de generar información extraída de las redes sociales: inteligencia colectiva y social curation turístico



Figura 4. Caso de uso para la herramienta del cliente

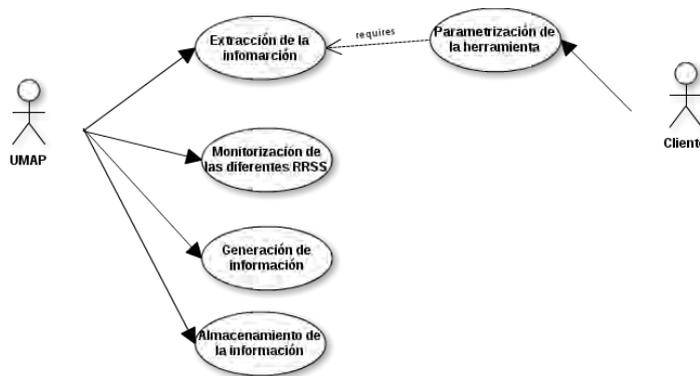


Figura 5: Caso de uso para la extracción y almacenamiento de la información

### Sistema de Análisis de datos

Para el apartado del análisis de datos, básicamente se han utilizado técnicas de clustering y aprendizaje de modelos. En el análisis de datos el input es el flujo de información y el output es la clasificación de la información en función de la polaridad negativa/positiva y en la polaridad objetiva/subjetiva. Para la consecución de éste fin, se ha subdividido en dos funciones principales, primero detectar la polaridad negativa/positiva utilizando técnicas de “*sentiment analysis*” en los que se trata de filtrar la información y segundo en detectar el grado de objetividad y subjetividad que tienen los mismos con técnicas de “*opinion minning*”.

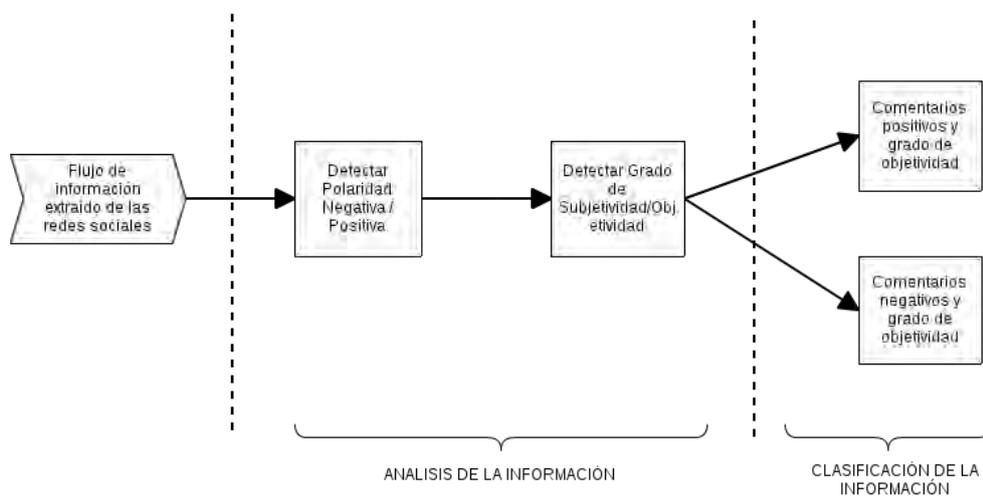


Figura 6: Diseño del sistema de análisis de datos

### Producto final

Como producto final, Turistic RobSoc se ha desvelado como una importante herramienta para la extracción de inteligencia colectiva de las redes sociales aplicada al turismo y que puede personalizarse en función de oferta turística, destino, y otros parámetros del clientes.

Además, como herramienta automatizada de social curation ofrece información relevante en tiempo real que también puede ser parametrizada por el cliente.

Por otra parte, la información recogida y almacenada por Turistic RobSoc puede ser analizada y extraer datos referidos a, por ejemplo:

- Marcas más comentadas en las redes sociales en los últimos meses
- Destinos más mencionados
- Regiones o áreas geográficas más mencionadas
- Publicaciones/periodicos/fuentes de información más mencionados
- Polaridad negativa/positiva / polaridad objetiva/subjetiva obtenida
- ...

El análisis de datos también es parametrizable y podría permitir mediante su uso a largo plazo, entre otras funciones, el análisis histórico de tendencias, que posibilite conocer la evolución de destinos, de marcas, hoteles, ...

### **3. Conclusiones y líneas futuras de investigación**

En las líneas que preceden hemos expuesto Turistic RobSoc aplicado al sector turístico. Los resultados obtenidos son alentadores, el sistema es capaz de obtener información relevante para y referente al sector turístico y que puede ser extrapolable a otros subsectores como turismo rural, de negocios, de invierno etc o a intereses compartidos: comerciales, marcas, regiones, productos, etc.

El sistema aplicado por CodeSyntax en Turistic RobSoc y derivado de Umap permite extraer, almacenar y tratar de forma automatizada y en función de diferentes parámetros, los temas de interés para esas comunidades de usuarios y las redes sociales que generan. Se han obtenido noticieros automatizados para el sector turístico que seleccionan y ponderan las informaciones relevantes, ofreciendo un resumen de la inteligencia colectiva extraída desde las redes sociales que se conforman a partir de las mismas.

Se ha validado la utilidad de la herramienta y se han abierto posibilidades de aplicación en otros campos como los arriba mencionados: análisis y extracción de información a partir de redes sociales en torno a productos turísticos, marcas, destinos, ... Su aplicación en otro tipo de redes sociales además del microblogging constituye otra de las líneas de investigación futura.

Así, Turistic RobSoc abre el camino para futuras aplicaciones en la extracción de información e inteligencia colectiva manera automatizada de aplicadas al sector turístico que compartan intereses a través de redes sociales.

### **4. Referencias**

*Astigarraga, E., Azpillaga, J., Fernandez, L., et al. (2011) “Umap, inteligencia colectiva extraída de las redes sociales, 2011. Revista El profesional de la información. EPI SCP, Barcelona, Spain. pp. 542-547.*

*Bhargava, Rohit (2009) “Manifesto For The Content Curator: The Next Big Social Media Job Of The Future ?” Recuperado de*

- <http://rohitbhargava.typepad.com/weblog/2009/09/manifesto-for-the-content-curator-the-next-big-social-media-job-of-the-future-.html> Septiembre 2009*
- Bhattacharyya Prantik, Ankush Garg, Shyhtsun Felix Wu; (2011) "Analysis of user keyword similarity in online social networks" en *Soc. Netw. Anal. Min.* (2011) 1:143–158
- Castells Manuel (2005) *La sociedad en red*. Alianza Editorial, Madrid
- CodeSyntax (2011) *Proyecto UMAP*. Referencias en <http://www.umap.eu> y <http://www.codesyntax.com/taller-cs/blog/topics/umap>
- Levy, Pierre (1994). "L'intelligence collective. Pour une anthropologie du cyberspace". La Découverte, Paris
- Levy, Pierre (1997) "Cyberculture". Éditions Jacob, Paris
- Liu, B. (2010) "Sentiment Analysis: A multifaced problem", *IEEE Intelligent Systems*, May/June 2010.
- MacQueen, J. B. (1967). "Some Methods for classification and Analysis of Multivariate Observations". 1. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press. pp. 281–297.
- Zeng, D., Chen, H., Lusch, R., Shu-Hsing, L. (2010) "Social Media Analytics and Intelligence", *IEEE Intelligent Systems*, Nov/Dec 2010

## 5. CV autores

Eneko Astigarraga. MBA por la UPV/EHU y Licenciado en CC.PP y Sociología por la Universidad de Deusto. Consultor Senior para diversas organizaciones y empresas. Ha participado en proyectos de i+d relacionados con las redes sociales e Internet. Actualmente es profesor de Prospectiva e Internet en la Empresa en la Deusto Bussines School, Universidad de Deusto y gerente de CodeSyntax.

Josu Azpillaga. Ingeniero Industrial por la Universidad de Navarra. Colaboró como investigador en el Centro Tecnológico IKERLAN, así como en MIK (Mondragon Innovation & Knowledge). Ha dirigido proyectos relacionados con tecnologías web 2.0. Actualmente es responsable de I+D+i en CodeSyntax

Luis Fernandez. Licenciado en Ciencias de la Información por la UPV/EHU. Es autor de varias publicaciones en torno a Internet, las redes sociales y la generación, búsqueda y

organización de contenidos digitales. Ha dirigido proyectos relacionados con las tecnologías de la lengua. Actualmente es responsable de proyectos y co-fundador en CodeSyntax.

Aitzol Naberan. Ingeniero Informático por la UPV/EHU. Ha participado en el desarrollo de diferentes proyectos tanto de Investigación y desarrollo como proyectos comerciales en el ámbito de las redes sociales , el comercio electrónico y los contenidos digitales. Actualmente es desarrollador senior en CodeSyntax.

Daniel Reguera Ingeniero Superior en Informática, e Ingeniero Técnico en Informática de Sistemas por la Escuela Politécnica Superior de Mondragon Unibertsitatea. En la actualidad lleva a cabo labores de investigación en el área de telemática de la citada Universidad.

Urko Zurutuza. Doctor en Informática, Ingeniero Superior en Informática, e Ingeniero Técnico en Informática de Sistemas por la Escuela Politécnica Superior de Mondragon Unibertsitatea. Compagina labores de docencia en ingeniería con formación a empresas y con labores de investigación en el área de telemática.