

## ANÁLISIS DE NAVEGACIÓN DE GEOPORTALES

ESTHER HOCHSZTAIN<sup>1</sup>, CARLOS LÓPEZ VÁZQUEZ<sup>1,2</sup>, MIGUEL ANGEL BERNABÉ<sup>2</sup>

estherhochsztain@gmail.com carlos.lopez@ieee.org ma.bernabe@upm.es

1 Universidad Ort Uruguay

2 Universidad Politécnica de Madrid

### RESUMEN

En este artículo se presenta un marco conceptual para el análisis de navegación en geoportales acompañado de un estudio de caso. Un geoportal es un sitio web especializado considerado como una forma de depositar y acceder a contenido geográfico en la web. Actualmente, se entiende que los geoportales constituyen un elemento clave para las Infraestructuras de Datos Espaciales (IDE) porque permiten difundir contenidos y capacidades.

Dado que un geoportal es un sitio web, para contar con información que apoye su desarrollo y contribuir a que tanto usuarios como propietarios del sitio satisfagan sus objetivos, es necesario identificar sus demandas y evaluar en qué medida éstas se logran satisfacer. Este artículo tiene como objetivos principales proponer un marco conceptual para el análisis de la navegación de los usuarios en un geoportal y presentar un caso de estudio de análisis de *web server logs* en base a técnicas de *Web Usage Mining*.

La metodología aplicada comprende la determinación de los módulos del marco conceptual, el estudio de las características de los *web server logs* y su procesamiento en base a herramientas de *Web Usage Mining* y a una aplicación en R desarrollada específicamente para este trabajo.

Los principales resultados obtenidos permiten identificar las particularidades de los geoportales, y se enmarcan en el modelo de desarrollo de infraestructura de datos geográficos de tercera generación centrado en el usuario.

**PALABRAS CLAVE:** Geoportal, Infraestructura de Datos Espaciales (IDE), *Web Usage Mining*, *web server logs*.

## 1. INTRODUCCIÓN

Un geoportal es un sitio web especializado considerado como una forma de depositar y acceder a contenido geográfico en la web o, alternativamente, un sitio web donde se puede encontrar contenido geográfico. Los geoportales constituyen un elemento clave para las Infraestructuras de Datos Espaciales (IDE) porque permiten difundir contenidos y capacidades (Tait 2005) .

En este sentido, puede plantearse que la oferta y demanda de aplicaciones web asociadas a datos espaciales está aumentando rápidamente. Los geoportales permiten que los usuarios finales consulten, visualicen y descarguen datos espaciales. Constituyen la forma que toma la oferta para satisfacer la demanda creciente de datos espaciales por parte de los usuarios finales. Si bien el intercambio de información espacial siempre existió, el mismo ha aumentando enormemente como consecuencia de la popularización de Internet y del desarrollo de nuevas tecnologías asociadas a información espacial.

Algunos de los beneficios de compartir información espacial se asocian, entre otros, con las economías de escala en la producción de información geográfica y la intención de evitar la duplicación de información, tal y como se señala por parte de (Goodchild, Fu, and Rich 2007).

Las IDE han experimentado cambios que permiten reconocer tres generaciones, centradas en productos, procesos y usuarios respectivamente, dependiendo del peso relativo de los mismos (Hennig and Belgui 2011) . La primera generación tiene como principio rector la integración y gestión de los datos existentes en las agencias gubernamentales; la segunda generación se focaliza en establecer un vínculo entre personas y datos mediante el desarrollo de aplicaciones de datos espaciales; la tercera generación está guiada por los usuarios (del sector privado, organizaciones e individuos) e integra iniciativas gubernamentales y privadas. Las tres generaciones coexisten actualmente, y no hay una frontera nítida entre ellas. En la tercer generación las necesidades de los usuarios constituyen un elemento central, y sus requerimientos determinan la naturaleza de la IDE, componentes básicos, metadatos y herramientas de acceso (web services y geoportales). [(Man 2011); (Sadeghi-Niaraki et al. 2010)]

Los geoportales permiten que los usuarios finales consulten, visualicen y descarguen datos espaciales. Constituyen la forma que toma la oferta para satisfacer la demanda creciente de datos espaciales por parte de los usuarios finales. Aunque el intercambio de información espacial

siempre existió, el mismo ha aumentando enormemente como consecuencia de la popularización de Internet y del desarrollo de nuevas tecnologías asociadas a información espacial.

Tres componentes se requieren para implementar un geoportal: (a) un sitio web que presente la aplicación geográfica o portal, (b) web services para publicar funcionalidades geográficas, y (c) el software de gestión de datos que brinda un entorno relacional para gestionar contenido geográfico. La Figura 1 muestra los componentes de un geoportal, los elementos que los integran, los entornos tecnológicos y estándares así como sus principales funciones. Los geoportales brindan aplicaciones GIS distribuidas con capacidades para buscar, generar y visualizar mapas, publicar y administrar información geográfica.

Components	Elements	Environments	Functions
Web Portal	<b>Web Site</b>	HTML, HTTP, XSL, XML, JSP, ASP	Search, Map Viewer, Publish, Adminstrate
	<b>Web Controls</b>	Java Beans, .NET	Query, Gazetteer, Mapping, Edit, Geocoding,
Web Services	<b>Geographic Web Services</b>	XML, SOAP, WSDL, WMS, WFS, GML	Query, Map render/feature, Transaction, Geocode
Data Management	<b>DBMS Geographic &amp; Tabular Data</b>	SQL	Raster, Vector, Tabular

Figura 1: Componentes de un Geoportal. Fuente: (Tait 2005)

Muchos países y gobiernos locales han desarrollado o están desarrollando portales de información geográfica (denominados geoportales), que permiten consultar, visualizar y descargar información espacial. El proceso de visualización de un mapa por parte del usuario de un geoportal planteado por (Harrie, Mustière, and Stigmar 2011) se presenta en la Figura 2. El usuario requiere cuál información (niveles) desea que se presente de diferentes servicios web; la respuesta del servicio de visualización es un mapa en cierto formato. El servicio de visualización de mapas carece de las características de legibilidad y completitud que tienen los mapas creados por un cartógrafo. Por consiguiente, se desarrollan e implementan métodos automatizados para reemplazar el trabajo cartográfico, lo que constituye una complicada tarea dado que requiere formalizar el trabajo del cartógrafo. Para resolver el problema (Harrie, Mustière, and Stigmar 2011) consideran que se requieren cinco categorías de métodos cartográficos: semántico, fusión (*conflation*), generalización, ubicación de etiquetas y

simbología.

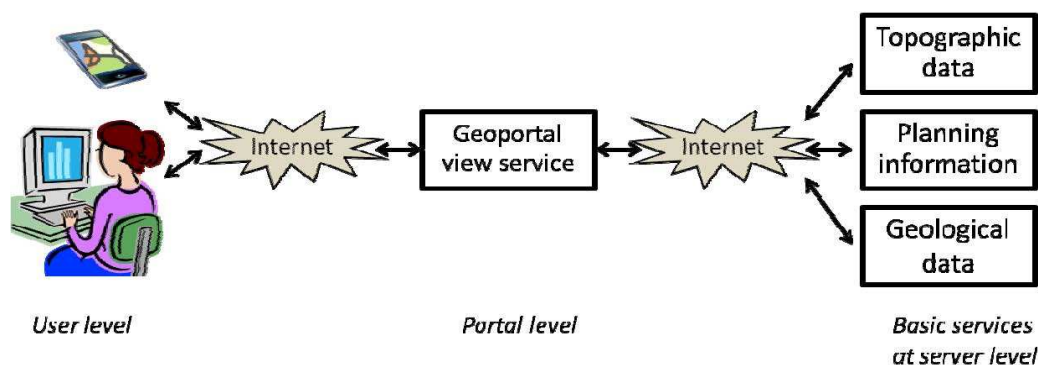


Figura 2: Visualización en un Geoportal. Fuente: (Harrie, Mustière, and Stigmar 2011)

La correcta operación de una IDE requiere, entre otras cosas, describir el suministro de acceso a información geográfica estandarizada. Se considera una forma para generar capacidades e incentivar la participación entre ofertantes y demandantes de información espacial. Como consecuencia de los avances tecnológicos y el hecho de que muchas actividades poseen un componente espacial, actualmente se dispone de grandes volúmenes de datos georreferenciados. A través de las IDE se busca tanto catalogarlos como dejarlos a disposición del público.

Otro enfoque para analizar un geoportal, puede plantearse en base a criterios basados en la economía. Como se señala por parte de (Goodchild, Fu, and Rich 2007) un argumento importante para compartir información geográfica deriva de las economías de escala en la producción, por lo que trata de maximizar la cantidad de usuarios y la validez en el tiempo. Sin embargo, los mismos motivos dan origen a deseconomías de escala, dado que deben contemplarse diferentes comunidades de información con diferencias en el idioma, cultura, terminología y acepciones de las palabras. Se requieren importantes inversiones en documentación (metadatos), software de conversión de formatos, y traducción de documentación (entre otros) para lograr que la información geográfica sea interoperable por diferentes comunidades de información. Si tales costos recaen en el proveedor de la información geográfica se generan deseconomías de escala que redundan desestimulan la expansión a múltiples comunidades de usuarios. Si por el contrario los costos deben ser asumidos por el usuario, se desalienta el uso de la información y por tanto el tamaño del mercado con la obvia consecuencia de reducir las economías de escala.

Si nos ubicamos en el tema de evaluación de geoportales, encontramos varios enfoques el análisis. En este sentido, según (Mansourian et al. 2011) pocos *spatial data clearinghouses* nacionales funcionan bien, tanto porque no se puede acceder satisfactoriamente a los recursos de datos espaciales disponibles o usarlos en forma óptima. Proponen un sistema experto como solución al problema.

Como ya se ha planteado, el proceso de creación y operación de una Infraestructura de Datos Espaciales, y de un geoportal en particular, habitualmente se basa en beneficios para la sociedad, la economía y su medio ambiente. Sin embargo, en la práctica la bibliografía señala que no resulta fácil evaluar el logro de tales objetivos. Un marco conceptual multi-visión fue propuesto para enfrentar este problema ((Grus et al. 2011) ; (Crompvoets et al. 2008) ). El aspecto central del marco conceptual multi-visión se basa en poner de relieve las múltiples ópticas de análisis pertinentes, las que se clasifican tanto en base a los aspectos evaluados (organizacionales, técnicos, etc.) como en función de diferentes propósitos (conocimiento, desarrollo, etc.). La evaluación se desarrolla en cuatro fases: valoración del propósito, valoración del enfoque, aplicación y evaluación.

Dado que un geoportal es un sitio web, para contar con información que apoye su desarrollo y contribuir a que tanto usuarios como propietarios del sitio satisfagan sus objetivos, es necesario identificar sus demandas y evaluar en qué medida éstas se logran satisfacer. Es necesario capturar el conocimiento disponible, no explícito del comportamiento de los usuarios y expectativas de los propietarios, para optimizar el geoportal y la relación con los usuarios.

A continuación en la sección 2 se presenta la metodología del trabajo, seguida por los resultados y discusión en la sección 3 y finalizando con las conclusiones en la sección 4.

## 2. METODOLOGÍA

La metodología aplicada comprende la determinación de los módulos del marco conceptual (*framework*), el estudio de las características de los *web server logs*, su procesamiento tanto en base a herramientas de Web Usage Mining como a una aplicación en R (R Development Core Team 2008) desarrollada específicamente para este trabajo.

Web Mining es la aplicación de metodologías y técnicas de Data Mining a información y

datos web. Web Usage Mining es una de las áreas que la componen y consiste en el proceso de extracción de patrones de uso de la web. Se basa esencialmente en el procesamiento y análisis de web server logs. Conociendo los patrones de uso es posible personalizar los sitios web, mejorar el diseño de los sitios, conocer las necesidades de los usuarios, contar con información para tomar decisiones con el objetivo de aumentar la satisfacción de los usuarios, gestionar mejor las relaciones con los clientes, entre otras aplicaciones ( [(Spiliopoulou et al. 2012); (Berendt 2012) ; (Devi et al. 2012); (Liu, Mobasher, and Nasraoui 2011); (Le and Zhou 2011); (Prasad, Reddy, and Acharya 2010); (Tiwari et al. 2011); (Adda et al. 2010)]).

## 2.1. MARCO CONCEPTUAL

El marco conceptual propuesto, que se presenta en la Figura 3, se basa en la metodología estándar para proyectos de Minería de Datos CRISP-DM (Azevedo and Santos 2008) (Chapman et al. 2000) (Wirth and Hipp 2000), comprende los siguientes módulos:

- Características del geoportal
- Comprensión del problema
- Comprensión de los datos
- Preparación de los datos
- Análisis de los datos
- Evaluación de los resultados

El proceso es iterativo, lo que justifica flechas siempre en ambos sentidos al conectar las etapas, pero teniendo en cuenta que lo indispensable y más frecuente son los vínculos en sentido horario, puede ocurrir que sea necesario retornar a una etapa anterior (marcado en la figura por flechas más pequeñas y de color más claro). Merece un análisis especial el vínculo entre las características del geoportal y la evaluación de los resultados, dado que puede considerarse que la evaluación de los resultados realimenta el proceso, dando lugar a nuevas interpretaciones y semántica asociada a las características del geoportal, que podrían estar desencadenando nuevamente el proceso.

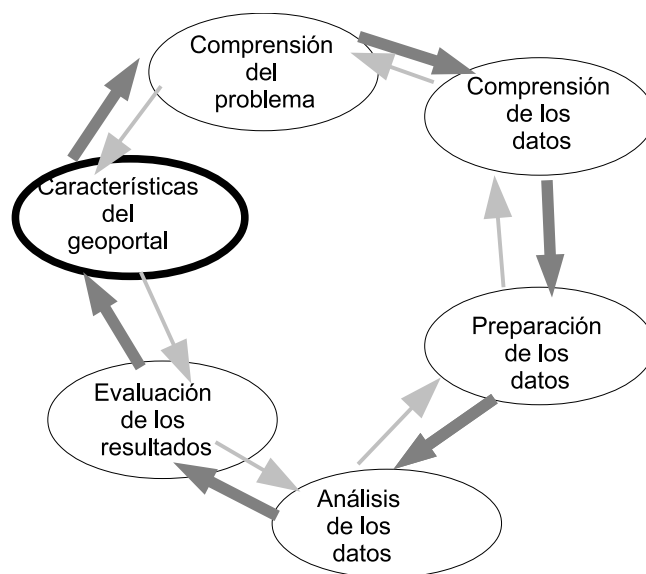


Figura 3: Marco Conceptual

### 2.1.1. Características del geoportal

Esta etapa se basa en estudiar todos los aspectos que rodean la creación y desarrollo del geoportal, desde el punto de vista de los antecedentes institucionales y las normativa vigente (leyes, decretos, etc.), los aspectos asociados al geoportal en sí en cuanto a su arquitectura, herramientas y decisiones técnicas, recursos humanos vinculados con el geoportal, así como a los usuarios del geoportal. Estos conceptos se muestran en la Figura 4.



Figura 4: Características del geoportal. Nota: Las imágenes utilizadas fueron obtenidas en Google Imágenes

- Caracterización de la institución responsable de la Infraestructura de Datos Espaciales (IDE)
- Presentación de los antecedentes de creación y desarrollo del geoportal, objetivos del geoportal, arquitectura y componentes del geoportal así como la normativa aplicable al geoportal
- Recursos humanos asociados al geoportal
- Actividad asociados a los usuarios del geoportal.

### **2.1.2. Comprensión del problema**

Esta etapa abarca la identificación de los objetivos del proyecto desde el punto de vista 'de negocio'. En base a los datos recabados en la etapa anterior, es posible delinear un proyecto de investigación. Esta etapa requiere identificar los principales problemas que se presentan para el geoportal e identificar las posibilidades de explorar datos para buscar soluciones al mismo. La comprensión del problema abarca aspectos tales como:

- identificación de los promotores del proyecto
- participantes del proyecto
- recursos humanos asignados al proyecto
- dominio principal (gubernamental, diversos socios, etc.)
- recursos asignados (infraestructura informática, recursos financieros, etc.)
- métricas de éxito

El resultado de esta etapa es generar un plan de proyecto que busca resolver el problema planteado, acorde a las características de los participantes, recursos disponibles y formas de cuantificar el logro de los objetivos. Esta etapa comprende asimismo establecer, en forma preliminar las técnicas y herramientas para apropiadas.



### 2.1.3. Comprensión de los datos

La etapa de comprensión de los datos tiene por objetivo obtener un conjunto inicial de datos, para luego describirlos y explorarlos. En particular, debe tenerse en cuenta que la principal fuente de datos son *web server logs*.

El estándar de formato más usado en los ficheros de registros (*logs*) es el formato común de registros (*Common Log Format CLF*), que describe los siete campos esenciales que debe contener un fichero de traza y que se describen seguidamente.

- **Host.** El *host* es la dirección Internet de la máquina cliente que realiza las peticiones, y es también la dirección a la cual será enviada la respuesta *HTTP*. Generalmente es una dirección IP .
- **Ident.** Información de identificación suministrada por el cliente.
- **Authuser.** En el caso de que la comunicación se realice a través del protocolo seguro SSL, el usuario está obligado a identificarse y suministrar un login y una password para establecer una comunicación segura. Tras la identificación, se genera un identificador arbitrario (*Authuser*) que acompaña al resto de las comunicaciones intercambiadas entre el navegador (*browser*) cliente y el servidor de *web*. Este identificador es muy útil para posteriormente asociar todos los registros (*logs*) asociados a la sesión de un usuario .
- **Time.** La fecha y hora en la que el servidor de finaliza de procesar la petición HTTP del usuario.
- **Request.** Porción de la dirección URL que especifica explícitamente la petición del usuario. Generalmente en la solicitud (*request*) se especifica una acción solicitada por un usuario y un identificador del objeto sobre el cual se solicita dicha acción.
- **Status.** Código numérico que especifica el procesamiento con o sin éxito de la petición del usuario.
- **Bytes.** Cuenta de bytes devueltos desde el servidor de al cliente.

#### 2.1.4. Análisis de los datos

El análisis de los datos abarca su estudio en base a técnicas de estadística descriptiva, así como la construcción de modelos. Las técnicas y herramientas utilizadas pueden ser diversas, tales como:

- herramientas de Web Usage Mining disponibles (AlterWind, Analog, entre otras)
- adaptación de herramientas de Data Mining a las especificidades del análisis de *web server logs* (SPSS Modeler, Weka, entre otras)
- análisis basados en los resultados obtenidos mediante herramientas SEO (Search Engine Optimization)
- desarrollo de aplicaciones específicas, para las cuales generalmente se utilizan lenguajes o sistemas ampliamente usados en el área estadística: R, SPSS, SAS, entre otros

#### 2.1.5. Evaluación de los resultados

La evaluación de los resultados debe vincularse al plan de proyecto realizado y a los objetivos desde la perspectiva de negocio establecidos en las etapas previas. Se espera que los resultados obtenidos arrojen luz acerca de los problemas planteados, y sean una realimentación del proceso de análisis. Asimismo, los resultados comprenden la revisión del proceso y la identificación de características que permitan guiar los pasos siguientes, determinando acciones futuras.

La evaluación de los resultados abarca, entre otros, los siguientes aspectos:

- confrontación de los resultados objetivos con las hipótesis previas
- generación de hipótesis a testear en futuros proyectos
- explicación de los resultados obtenidos, a la luz de las características del geoportal, la institución promotora, datos de entorno <sup>1</sup>
- conocimiento de las personas involucradas y forma de trabajar en relación al geoportal en las distintas organizaciones

---

<sup>1</sup>Por ejemplo que en el período de estudio hubiera surgido un interés especial del público que conduzca a aumentar cierto tipo de consultas

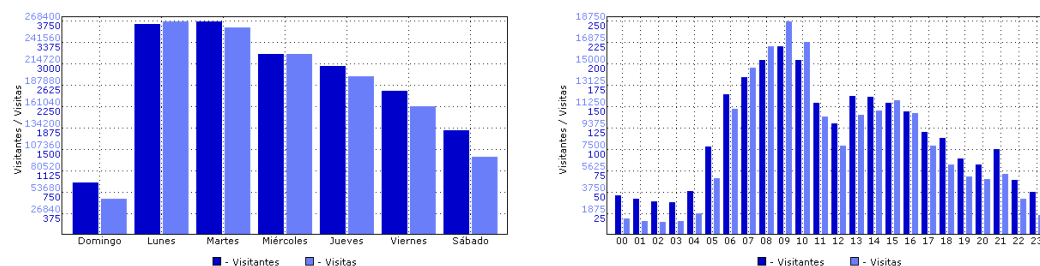
Luego de haber analizado los diversos módulos o etapas que conforman el marco conceptual propuesto, a continuación se presentan los resultados obtenidos.

### 3. RESULTADOS Y DISCUSIÓN

Los resultados obtenidos se basan en tres tipos de análisis:

- uso de la herramienta de análisis de web server logs AlterWind Log Analyzer
- desarrollo de una aplicación para análisis de web server logs, en base a R y SPSS.
- el uso de una herramienta SEO

Utilizando la herramienta gratuita AlterWind LogAnalyzer Lite ([www.alterwind.com](http://www.alterwind.com)), se muestran en la Figura 5 los resultados que se obtuvieron al analizar los *web server logs* de una semana de un geoportal.



(a) Actividad diaria

(b) Actividad por hora

Figura 5: Análisis de web server logs con la herramienta AlterWind

En lo referente a la actividad diaria, en la Figura 5(a) se observa que la mayor actividad se registra lunes y martes, fueron los días de mayor actividad, descendiendo paulatinamente en los restantes días. Al observar la actividad por hora, en la Figura 5(b) se observa que las horas de mayor actividad se presentan en la mañana, entre las 8 y las 10. Excepto el factor de escala, puede apreciarse que visitantes y visitas presentan un comportamiento muy similar, como era esperable <sup>2</sup>.

Utilizando R como herramienta para preprocesamiento de los *web server logs* y SPSS para la visualización de los resultados, se obtuvieron resultados que son acordes a lo que ocurre en

---

<sup>2</sup>En el análisis de las Figuras 5(a) y 5(b) debe tenerse que se manejan escalas diferentes para visitantes y visitas

otros sitios web analizados. La Figura 6 presenta gráficos obtenidos del análisis de los ad hoc de los *web server logs* del geoportal.

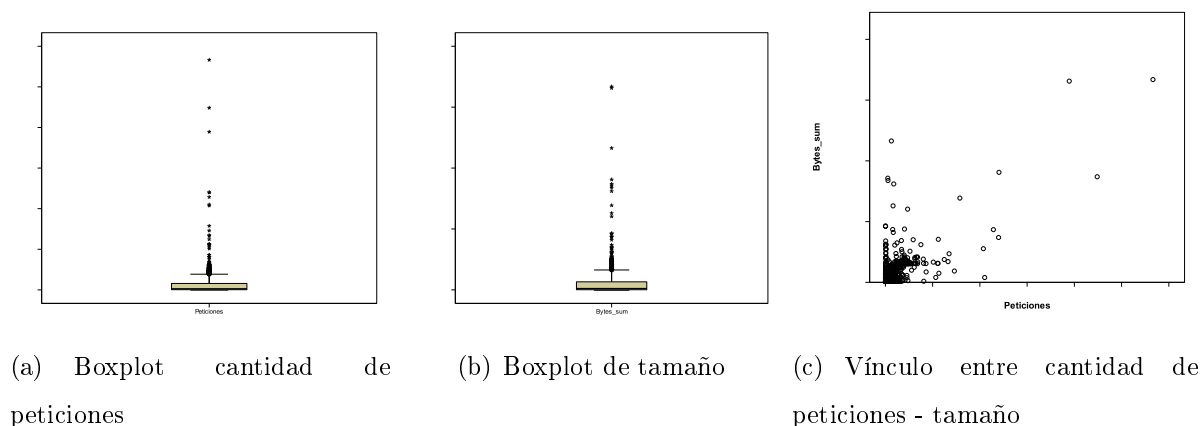


Figura 6: Análisis de *web server logs* de un geoportal

La cantidad de peticiones por sesión se concentra en valores pequeños, registrándose algunas sesiones atípicas con muchas peticiones, como muestra la Figura 6(a). La Figura 6(b) muestra un comportamiento similar para el tamaño (cantidad de bytes) al anteriormente reseñado para la cantidad de peticiones, lo que no sorprende por ser un patrón identificado generalmente en el análisis de *web server logs*. Por otra parte, la Figura 6(c) muestra el vínculo entre cantidad de peticiones y el tamaño, observándose que la gran mayoría de los valores se concentran en valores reducidos en ambas variables. El coeficiente de correlación de 0,7 deja en evidencia el efecto de los valores atípicos.

Por último, se obtuvo información del geoportal utilizando la herramienta SEO Alexa ([www.alexa.com](http://www.alexa.com)). La mayoría de los visitantes se encuentran en España (el 93.7%). El promedio (de tres meses) de vistas de página por usuario es 1.6. Los usuarios (en relación a los usuarios generales de Internet) pueden caracterizarse por su mayor porcentaje de: edades entre 35 y 54 años, sexo masculino, nivel de educación elevado (graduado), que consulta desde el trabajo como se muestra en la Figura 7

De la discusión de los resultados obtenidos, se concluye que no cabe duda de la necesidad de integrar diferentes puntos de vista para el análisis de un geoportal, y que los resultados obtenidos resultan de gran interés para la optimización y adecuación de los geoportales a las finalidades por ellos perseguidas. Sin embargo, son múltiples las posibilidades de cometer errores, en particular dado que la gran cantidad de datos a procesar requiere contar con algoritmos sumamente

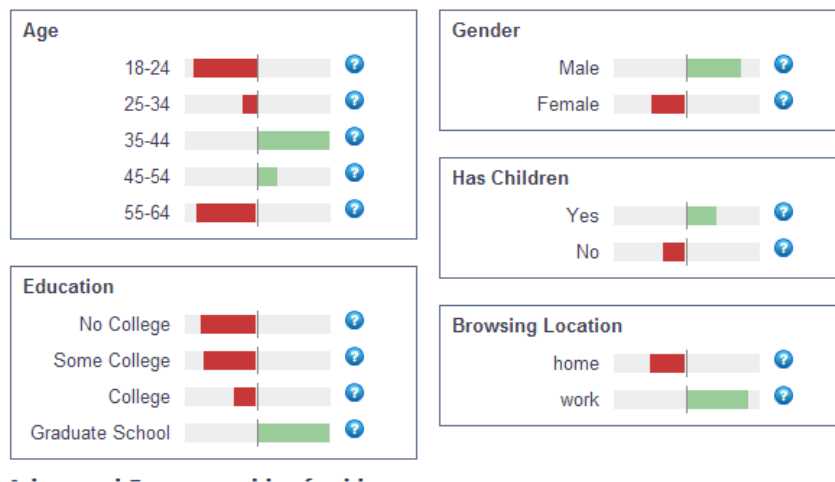


Figura 7: Datos demográficos de visitantes

eficientes.

## 4. CONCLUSIONES

Este estudio se enmarcan en el modelo de desarrollo de infraestructura de datos geográficos de tercera generación. En el mismo se propone un marco conceptual para análisis de navegación en geoportales, y se presenta un extracto de los resultados obtenidos en la etapa de análisis de los datos. El marco conceptual propuesto se basa en el estándar CRISP-DM, el que fue adaptado para contemplar los requerimientos específicos para análisis de geoportales. Cada una de las etapas del mismo requiere contemplar las particularidades de las IDE, así como del geoportal en estudio en particular. Sin embargo, a pesar de las especificidades es posible identificar un marco de análisis común.

Al instanciar el marco conceptual propuesto al caso de un geoportal específico se pudieron identificar particularidades en los datos que deberán ser estudiados posteriormente con mayor detalle. Los principales resultados obtenidos permiten identificar y confirmar las particularidades de navegación en geoportales.

Este trabajo forma parte de un proyecto cuyo objetivo general consiste en el establecimiento de un marco conceptual que permita evaluar cuantitativamente el logro de los objetivos de los geoportales IDE en base al análisis de la navegación de los usuarios. El mismo se desglosa en dos tareas: a) evaluar cuantitativamente la actividad de los usuarios al navegar en el geoportal

IDE y b) comparar los resultados obtenidos del análisis de la navegación contra los objetivos planteados para el geoportal IDE. En particular, en futuros trabajos se espera responder a las siguientes preguntas: ¿cuales son las características específicas de los geoportales IDE desde el punto de vista del análisis de la navegación que los distingue de otros portales?, ¿cómo debe ser el proceso de análisis de la navegación en geoportales IDE?, ¿cuales son los objetivos de los geoportales IDE?, ¿pueden identificarse patrones de navegación en los geoportales IDE? y ¿cómo pueden contrastarse los objetivos de los geoportales IDE a través del análisis de la navegación?

## 5. REFERENCIAS

- Adda, M., P. Valtchev, R. Missaoui, and C. Djeraba. 2010. "A framework for mining meaningful usage patterns within a semantically enhanced web portal." *Proceedings of the Third C\* Conference on Computer Science and Software Engineering*. ACM, 138–147.
- Azevedo, A., and M.F. Santos. 2008. "KDD, SEMMA and CRISP-DM: a parallel overview." *Proceedings of the IADIS European conf. data mining*. 182–185.
- Berendt, B. 2012. "More than modelling and hiding: towards a comprehensive view of Web mining and privacy." *Data Mining and Knowledge Discovery*, pp. 1–41.
- Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. 2000. "CRISP-DM 1.0 Step-by-step data mining guide."
- Crompvoets, J., A. Rajabifard, B. Loenen, and T. Delgado Fernandez. 2008. "A multi-view framework to assess spatial data infrastructures."
- Devi, B.N., Y.R. Devi, B.P. Rani, and R.R. Rao. 2012. "Design and Implementation of Web Usage Mining Intelligent System in the Field of e-commerce." *Procedia Engineering* 30:20–27.
- Goodchild, M.F., P. Fu, and P. Rich. 2007. "Sharing geographic information: an assessment of the Geospatial One-Stop." *Annals of the Association of American Geographers* 97 (2): 250–266.

- Grus, Lukasz, Watse Castelein, Joep Cromptoets, Theo Overduin, Bastiaan van Loenen, Annemarie van Groenestijn, Abbas Rajabifard, and Arnold Bregt. 2011. "An assessment view to evaluate whether Spatial Data Infrastructures meet their goals." *Computers, Environment and Urban Systems* 35 (3): 217–229.
- Harrie, L., S. Mustière, and H. Stigmar. 2011. "Cartographic Quality Issues for View Services in Geoportals." *Cartographica: The International Journal for Geographic Information and Geovisualization* 46 (2): 92–100.
- Hennig, S., and M. Belgui. 2011. "User-centric SDI: Addressing Users Requirements in Third-Generation SDI. The Example of Nature-SDIplus." *Geoforum Perspektiv* 10, no. 20.
- Le, Z., and P. Zhou. 2011. "A Framework for Web Usage Mining in Electronic Government." *IFIP Advances in Information and Communication Technology (AICT)* 252 (252): 487–496.
- Liu, B., B. Mobasher, and O. Nasraoui. 2011. "Web Usage Mining." *Web Data Mining*, pp. 527–603.
- Man, E. 2011. "Spatial Data Infrastructuring: praxis between dilemmas." *International Journal of Spatial Data Infrastructures Research*, vol. 6.
- Mansourian, A., E. Omid, A. Toomanian, and L. Harrie. 2011. "Expert system to enhance the functionality of clearinghouse services." *Computers, Environment and Urban Systems* 35 (2): 159–172.
- Prasad, G.S., N.V.S. Reddy, and U.D. Acharya. 2010. "Knowledge Discovery from Web Usage Data: A Survey of Web Usage Pre-processing Techniques." *Information Processing and Management*, pp. 505–507.
- R Development Core Team. 2008. "R: A Language and Environment for Statistical Computing." ISBN 3-900051-07-0.
- Sadeghi-Niaraki, A., A. Rajabifard, K. Kim, and J. Seo. 2010. "Ontology Based SDI to Facilitate Spatially Enabled Society." *Proceedings of GSDI*, vol. 12.
- Spiliopoulou, M., B. Mobasher, O. Nasraoui, and O. Zaiane. 2012. "Guest editorial: special issue on a decade of mining the Web." *Data Mining and Knowledge Discovery*, pp. 1–5.

- Tait, M.G. 2005. “Implementing geoportals: applications of distributed GIS.” *Computers, Environment and Urban Systems* 29 (1): 33–47.
- Tiwari, S., D. Razdan, P. Richariya, and S. Tomar. 2011. “A web usage mining framework for business intelligence.” *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*. IEEE, 731–734.
- Wirth, R., and J. Hipp. 2000. “CRISP-DM: Towards a standard process model for data mining.” *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*. Citeseer, 29–39.