

---

# Detectando anomalías en los metadatos de cartotecas

*Detecting outliers in metadata from digital map libraries*

---

**Walter RENTERIA-AGUALIMPIA, Francisco J. LOPEZ-PELLICER, Aneta J. FLORCZYK, Juan LÓPEZ DE LARRINZAR, Javier LACASTA, Pedro R. MURO-MEDRANO, F. Javier ZARAZAGA-SORIA**

Departamento de Informática e Ingeniería de Sistemas Universidad de Zaragoza, C/ María de Luna 1, 50.018 Zaragoza, España  
{walterra, fjlopez, florczyk, jlacasta, prmuro, javy}@unizar.es

## Resumen

La validación de metadatos asociados a mapas digitalizados es uno de los problemas más relevantes relacionados con la preservación de la información cartográfica en las bibliotecas de mapas digitales. Este artículo presenta las primeras fases de investigación de un nuevo método automático para asegurar su validez. Este método está basado en la combinación de técnicas de agrupación espacial y métodos estadísticos de análisis geográfico.

**Palabras clave:** DML. Metadatos. Agrupaciones espaciales. Análisis geográfico estadístico. Detección de anomalías.

## Abstract

The validation of the metadata associated with digitalized maps is one of the main problems related to the preservation of cartographic information in digital map libraries. This article describes the first research phases of a new automatic method to ensure the validity of spatial metadata. The method is based on the combination of spatial clustering techniques and geographical statistical analysis.

**Keywords:** DML. Metadata. Spatial clustering. Geographic statistic analysis. Outliers detection.

## 1. Introducción

En el mundo de las Bibliotecas de Mapas Digitales (*Digital Map Libraries*, DML), también conocidas como cartotecas digitales, los metadatos señalan el camino para acceder a la cartografía. Es decir, los metadatos no solo describen o documentan qué información proporcionan los mapas, sino que además nos dicen dónde se encuentran (Zarazaga-Soria et al., 2003).

Las cartotecas digitales procuran tener metadatos que describan bien la parte espacial de los recursos indexados. Si los metadatos contienen inconsistencias en la información espacial, se crean distorsiones que entorpecen la visualización, localización e interpretación de los recursos de la cartoteca. Por ejemplo un mapa digital cuya extensión espacial corresponde a Alemania, pero su descripción textual se refiere a España, generará resultados de búsqueda incoherentes. Algunos de estos problemas y situaciones de inconsistencias en la descripción de recursos cartográficos son recogidos por Monnier (1991) y Hill (2006, p. 156-161).

De cara al uso y explotación de los recursos de una cartoteca digital, es importante asegurar la consistencia entre la cartografía digitalizada y los metadatos que la describen. Asegurar la validez de los metadatos asociados a los mapas digitalizados es uno de los problemas más im-

portantes de la preservación de la información cartográfica en las bibliotecas de mapas digitales (Anguita et al., 2012). El aseguramiento se puede hacer de forma manual, pero es una tarea ardua dado el volumen de la información gestionada por las cartotecas actuales.

Un tipo de validación automática es el aseguramiento de la conformidad del metadato con un estándar (Capdevila et al., 2012). Una aproximación diferente es la validación semántica que analiza la coherencia y la consistencia conceptual (no solo sintáctica) de la descripción de los recursos en los metadatos de acuerdo a los conceptos de un dominio determinado

En este artículo se analiza el uso combinado de técnicas de *agrupación o clustering espacial y métodos estadísticos de análisis geográfico* para la validación semántica del conocimiento de las cartotecas en el dominio espacial, en particular sus registros de metadatos de recursos cartográficos. Aunque la validación puede abarcar otros aspectos —como el temporal, las escalas o los resúmenes, entre otros—, en este trabajo nos enfocaremos en la validación de la descripción espacial.

Como explicaremos más adelante, las técnicas arriba mencionadas aplicadas a los registros de metadatos permiten detectar incoherencias entre la descripción de la extensión geográfica

cubierta por un recurso (mapa digitalizado) mediante referencias espaciales indirectas (específicamente, los nombres de lugares o topónimos) y la descripción mediante referencias espaciales directas (es decir, las coordenadas de un sistema de referencia espacial).

Más concretamente, estas técnicas detectan si hay correspondencia entre el topónimo descrito en el registro de metadato y el topónimo asociado a la extensión geográfica de los datos en la realidad. El propósito central es responder a la pregunta: ¿Son coherentes las descripciones textuales y las descripciones basadas en coordenadas que encontramos en los registros de metadatos de las cartografías digitales?

## 2. Trabajos relacionados

Dado el gran número de registros históricos de mapas, los volúmenes de información manejados son enormes y cada vez crecen más (Fernandez-Wytenbach et al., 2007). Por esta razón es común tanto en los sistemas de búsqueda tradicionales, como en las DML, que cualquier consulta recupere cientos de resultados representados en largas listas paginadas de ítems.

En este punto surge la necesidad de mecanismos que permitan organizar y presentar un gran número de resultados de manera que el usuario pueda verlos como uno solo. Uno de los mecanismos que suple esta necesidad se denomina agrupación o “*clustering*” el cual permite formar agrupaciones de elementos similares según algún criterio. Hay diferentes tipos de técnicas de clustering:

- *La agrupación o clustering textual* se define como la agrupación de metadatos en función de la similitud de su contenido, ya sea palabras clave, resúmenes o todo el cuerpo del texto (Lacasta, Nogueras-Iso y Zarazaga-Soria, 2010).
- *La agrupación o clustering espacial* es la agrupación de los metadatos bajo el criterio de distancia o distribución espacial, que puede usar diferentes métricas o medidas para determinar la proximidad espacial entre los elementos analizados (Ester et al., 1996).

En la línea de estas definiciones, Bookstein et al. (1998) proponen un sistema que optimiza la recuperación de recursos. La optimización asigna valores de relevancia a las palabras adyacentes y las agrupa usando el *clustering textual*.

Rauber y Merkl (1999) desarrollaron un sistema de exploración de documentos en *Bibliotecas Digitales* basado en *clustering textual* denominado SOMLib. Éste sistema organiza los docu-

mentos de temas similares en función de su cercanía espacial. El objetivo principal de SOMLib es superar la limitación de los motores de búsqueda, que representan los resultados como una lista unidimensional.

La organización y navegación de bibliotecas digitales es estudiada por Liu, Wu y Liu (2011). Su objetivo principal es ayudar al usuario a encontrar información agrupando los documentos más similares. Su enfoque se basa en calcular la similitud entre documentos usando una nueva forma de representar sus características. Este enfoque logra acelerar el *clustering de textos* basado en mapas autoorganizados.

Skupin y Buttenfield (1996) presentan un mecanismo de visualización de documentos de bibliotecas digitales. En el citado trabajo se usan metáforas espaciales para representar como puntos de una gráfica los documentos recuperados. Los puntos son agrupados por cercanía espacial usando el *clustering espacial*.

Desde una perspectiva de acceso a la información por discapacitados visuales, Luk et al. (2000) desarrollaron un sistema de asistencia visual para acceder a bibliotecas digitales que utilizaba un *clustering espacial* en el proceso de conversión de una pregunta en código Braille a consultas al sistema.

MobiTOP es una propuesta que mejora el proceso de búsqueda y recuperación de un sistema basado en la localización (Quach et al., 2010). En dicho sistema, las anotaciones cumplen la función de los metadatos de las bibliotecas digitales, con la diferencia de que estas anotaciones son provistas por los usuarios y organizadas de manera jerárquica. Esta propuesta contempla la posibilidad de incluir los detalles de las ubicaciones de manera agregada, como en un foro de discusión. El enfoque usa técnicas de *clustering espacial* exclusivamente para mejorar el proceso de búsqueda.

En el ámbito de sistemas de búsquedas espaciales sobre metadatos, generalmente se presentan dos tipos de búsquedas: por tema o por extensión geográfica. En la combinación de estos dos tipos de búsqueda emerge la necesidad de diseñar heurísticas que permitan detectar inconsistencias entre tema, nombres de lugares y extensiones geográficas. El uso de análisis estadísticos y heurísticas en el ámbito de las bibliotecas de mapas digitales ya ha sido propuesto antes por Fernandez-Wytenbach et al. (2007). La diferencia principal de nuestra propuesta con el citado trabajo radica en la finalidad con la que se han usado los métodos y heurísticas. Fernandez-Wytenbach las usan para *acceder* a los contenidos históricos, mien-

tras que en nuestro trabajo son usadas para *detectar* incoherencias y *verificar* la consistencia de los metadatos cartográficos de las cartotecas.

En el trabajo de revisión bibliográfica se han encontrado pocos trabajos que analicen espacialmente la información descriptiva de recursos geográficos de las bibliotecas de mapas digitales y en particular de las cartotecas. Adicionalmente no se han encontrado trabajos que usen técnicas de agrupación espacial para analizar las inconsistencias en los metadatos de los recursos cartográficos digitales.

### 3. Clustering de metadatos geográficos

#### 3.1. Metodología del enfoque propuesto

El *clustering* es una técnica propia del área de minería de datos e inteligencia artificial que permite agrupar elementos en función de algún criterio o medida de similitud (Sugihara, Okabe y Satoh, 2011; Ester et al., 1996).

Existen al menos dos formas en las que se pueden agrupar espacialmente los metadatos usando información explícita de su descripción:

- Agruparlos de acuerdo a los *topónimos* asociados a su ubicación.
- Agruparlos aplicando alguna medida de distancia espacial entre su posición geográfica.

En este trabajo se asume que los metadatos que formen un *cluster* en razón de su posición geográfica poseen alguna referencia textual a la localización común donde están situados. Por ejemplo, en el caso de un conjunto de mapas históricos de ciudades españolas, se asume que los topónimos en los metadatos de las cartografías de cada ciudad deberían concordar con los topónimos naturalmente asociados al mismo territorio. Al formar agrupaciones, deberíamos poder generar al menos un *cluster* en la zona geográfica de cada ciudad. Por ejemplo, para el *cluster* asociado al territorio de Barcelona, la descripción textual de los metadatos pertenecientes a ese *cluster* debería contener el topónimo "Barcelona". Si no se cumple esta concordancia entre la información espacial y la textual, es muy posible que haya discrepancia en los metadatos. Por lo tanto sería útil informar de las siguientes situaciones:

- Advertir de la ausencia del término "Barcelona" como información descriptiva relevante que podría mejorar la capacidad de sistemas de recuperación que usan búsquedas de información espacial textual en las cartotecas.

- Señalar como posibles inconsistencias o anomalías aquellos metadatos que, perteneciendo al *cluster* de Barcelona, no posean el *topónimo* "Barcelona" asociado naturalmente a su *cluster*.
- Complementar y enriquecer la descripción con el *topónimo* "Barcelona", que estaba ausente hasta ese momento.

Para poder detectar anomalías en los metadatos, se parte de la siguiente hipótesis: *detrás de una agrupación o cluster geográfico subyace una propiedad, característica o fenómeno que podría estar directamente asociado con el motivo de la agrupación o incluso revelar su origen* (Getis y Ord, 1992). Desde el punto de vista espacial dicha propiedad podría ser el *topónimo* del lugar donde está el *cluster*.

En la línea de este planteamiento, algunas de las preguntas que surgen son las siguientes:

- ¿Qué porcentajes de elementos del *cluster* hacen referencia textual a la ubicación donde se encuentra el *cluster*? ¿Es significativo dicho porcentaje?
- Al analizar espacialmente las divisiones creadas por los *clusters* ¿qué propiedad es resaltada? o ¿Se podría informar cuál es el motivo de la agrupación?
- ¿Bajo qué propiedad toma sentido o "es más significativo" el análisis del grupo?

Con miras a resolver estas preguntas, se plantea la metodología de trabajo que se presenta en la figura 1. Se pueden identificar las siguientes etapas:

- *Agrupar metadatos en clusters*. Partiendo del repositorio de registros de metadatos de mapas digitales, se extrae la información espacial, que es agrupada de acuerdo a su localización.
- *Geocodificación inversa*. A partir de la localización de cada *cluster*, se aplica un subproceso que asocia a cada *cluster* nombres de lugar (*topónimos*). El *topónimo*, o posiblemente conjunto de *topónimos*, pasan a ser los identificadores que caracterizan el *cluster*.
- *Extracción de información espacial textual*. Otra hipótesis se centra en que la co-ocurrencia espacial de metadatos no se da por casualidad, sino que estará reflejada en la información textual o descriptiva de carácter espacial. En otras palabras, los registros de metadatos que concurren en un mismo lugar, deberían tener topónimos en común.

- **Validación estadística.** Esta fase usa métodos de análisis estadísticos de *clustering* geográfico para validar tres aspectos:

- La adecuada agrupación de elementos.
- La correspondencia entre los topónimos del cluster y los topónimos presentes en los metadatos.
- La significancia del topónimo respecto al cluster.

A partir de estos análisis se propondrán correcciones o anotaciones de topónimos ausentes que enriquezcan los registros, a través de un proceso de realimentación con el repositorio, señalado con línea punteada en la figura 1.

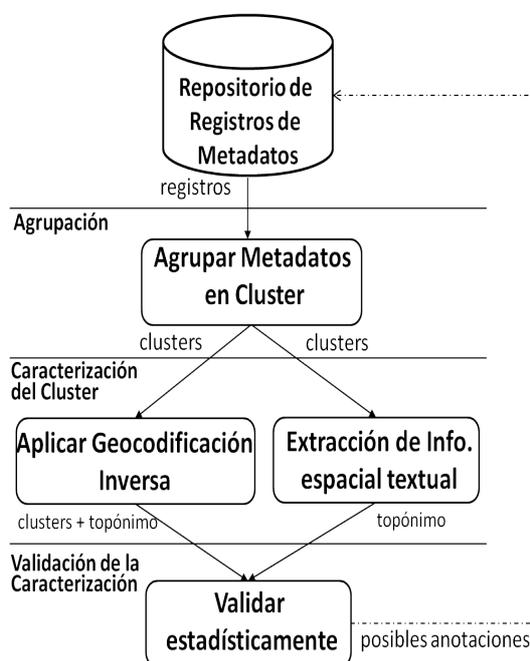


Figura 1. Método de detección de anomalías basado en clustering

### 3.2. Tratamiento previo de los registros

El punto de partida del procedimiento general es el repositorio de registros de metadatos de las cartotecas, de los cuales se extraen todos aquellos que posean explícitamente su localización. La localización, entendida como la extensión geográfica o caja rectangular "*Bounding Box*" es homogeneizada en un mismo sistema de coordenadas y transformada a su centroide. Este proceso de filtrado prepara los datos para poder aplicar el algoritmo de *clustering* espacial.

Cabe mencionar que aquellos metadatos que no poseen la información espacial de su localización son desestimados para el cálculo en el siguiente paso.

### 3.3. Clustering

El proceso de detección de metadatos inconsistentes por medio de *clustering* utiliza el algoritmo DBSCAN (Ester et al., 1996) para identificar agrupaciones de metadatos geográficos. Entre sus principales ventajas están las siguientes:

- Es un método que no requiere supervisión.
- Es capaz de detectar *clusters* de recursos de formas geométricas diversas.
- Permite identificar recursos espacialmente anómalos (*outlier*).
- Se ha comprobado la eficiencia del DBSCAN para gestionar bases de datos espaciales de gran tamaño como es el caso de las bibliotecas de mapas digitales actuales (Sander et al., 1998).

En síntesis, este algoritmo genera agrupaciones de metadatos de recursos cartográficos identificando las zonas donde hay mayor densidad o concentración de recursos. Además, identifica los recursos cartográficos aislados como anomalías desde el punto de vista de la densidad espacial.

### 3.4. Geocodificación inversa

En esta fase del proceso de detección se requiere saber cuáles *topónimos* están asociados a cada cluster. Para suplir esta necesidad, se usa la geocodificación inversa. Este proceso consiste en asignar nombres de lugares (*topónimos*) a unas coordenadas geográficas conocidas (GeoNames, 2008; Florczyk et al., 2009).

Dado que un *cluster* puede tener una forma geométrica compleja, en este punto surge la necesidad de desarrollar un nuevo geocodificador inverso que pueda asociar *topónimos* a geometrías más complejas que las de un punto o una caja rectangular. El requerimiento mencionado no es soportado por los geocodificadores estudiados (GeoNames, 2012; Google Geocoding API, 2012; Geocoder.ca, 2012). Por esta razón se implementó uno propio. Sus principales características son:

- Permite asociar topónimos a clusters de metadatos cartográficos con formas geométricas diversas o complejas.
- Además de los topónimos, proporciona el porcentaje de área del cluster asociada realmente al topónimo. Esta información puede ser usada como ponderación en sistemas de recuperación.

Sumado a estos aspectos, en el geocodificador inverso desarrollado se puede especificar el

nivel de división política en el que se desea establecer la operación de geocodificación (i. e., país, estado o comunidad y provincia). La aclaración hecha, surge dado que un geocodificador inverso podría retornar como respuesta para una misma coordenada o región, tanto una calle como un barrio, una ciudad, un estado o país. El geocodificador inverso toma como datos de entrada la extensión geográfica del *cluster* y regresa los topónimos asociados de unidades administrativas proporcionadas por la base de datos de áreas administrativas globales GADM (Hijmans, García y Wieczorek, 2012).

### 3.5. Extracción de información espacial textual

Partiendo de la hipótesis de que la co-ocurrencia espacial de los datos es reflejada en los metadatos a través de los topónimos allí documentados, este módulo identifica los topónimos de cada metadato para ser comparados en la siguiente fase. Esta identificación se hace comparando los campos de los metadatos contra una base de conocimiento externa de nombres de lugares. Cabe mencionar que esta fase en sí misma presenta otros retos, ya que no es fácil identificar los topónimos que no se encuentran documentados como tales, sino que están como texto libre.

Sumado a esto se ha observado que —posiblemente debido a diversas prácticas de documentación— los topónimos no siempre están asociados a la posición real del recurso cartográfico, sino al proveedor, al publicador o, incluso siendo un único proveedor, poseen servidores en diferentes lugares incluidos en el metadato. La falta de claridad de la proveniencia de estos topónimos y su pobre descripción aumentan el reto de su identificación.

### 3.6. Análisis estadístico y validación

El principal objetivo es determinar si un cluster es consistente desde el punto de vista espacial y textual. Para ello se usan métodos de análisis estadísticos. En este sentido se encontró que para lograr este objetivo de validación el análisis debería permitir cuatro tareas principales:

1. Verificar si la propiedad textual que es objeto de análisis (topónimos) es una de las razones principales que ha ocasionado el cluster.
2. Determinar el grado de relevancia de cada topónimo asociado naturalmente a la zona donde está, que bien podría ser un valor de relevancia en función del área o la población.
3. Detectar clusters y metadatos que difieren significativamente de su mayoría de vecinos.

4. Determinar si hay un alto porcentaje de metadatos con referencia explícita al topónimo asociado geográficamente al *cluster*.

A partir de esta idea, si después de haber corroborado que los topónimos naturalmente asociados al cluster son coherentes con los topónimos descritos en los metadatos, entonces podrían advertirse como posibles incoherencias o anomalías los metadatos que perteneciendo al cluster no hacen referencia a dichos *topónimos*.

En la literatura se han desarrollado métodos estadísticos para determinar si los clusters son de un tamaño geográfico considerable y su concentración no ha ocurrido por casualidad, (Knox, 1989; Kulldor, 2006; Besag y Newell, 1991). Aunque estos métodos y análisis aportan confiabilidad a este enfoque, solo permiten establecer el grado de significancia de los clusters de metadatos desde el punto de vista espacial.

### 3.7. Caso de uso

Para sintetizar la comprensión de estas fases y realiza una evaluación inicial, se presenta el ejemplo de la figura 2. En este ejemplo se han analizado registros de metadatos del área de Barcelona. Estos registros 1, 2, 3, 4 tienen declaradas explícitamente sus extensiones geográficas, indicadas con líneas punteadas en la figura. De igual modo, tienen explícitamente asociados sus topónimos: Barcelona, Barcelona, Valencia y Barcelona respectivamente.

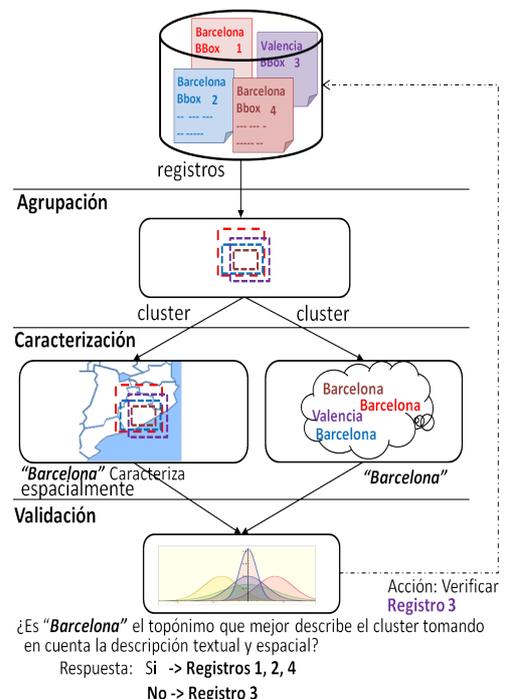


Figura 2. Ejemplo de aplicación del Método de detección de anomalías basado en clustering

Al aplicar el algoritmo de *clustering* se agrupan los metadatos espacialmente cercanos, posteriormente se aplican dos procesos simultáneos sobre el cluster resultante. El primero aplica geocodificación inversa para identificar el topónimo asociado al cluster, en este caso es “Barcelona”. El segundo identifica el topónimo más probable a partir del conjunto de registros que forman el cluster, cuyo resultado es “Barcelona”. El proceso de validación analizaría éstos resultados evaluándolos a través de una pregunta como la siguiente: ¿Es “Barcelona” el topónimo que mejor describe el cluster, tomando en cuenta la descripción textual y espacial? De esta pregunta surgen respuestas que pueden servir como indicador de calidad para realizar anotaciones en los registros bien descritos (1, 2 y 4) y anotaciones que sugieren la necesidad de revisar el registro 3, correspondiente a Valencia.

Lo más relevante de este caso de uso es que contiene tres ejemplos de búsqueda que pueden ser mejorados con las anotaciones obtenidas:

- Si se realiza una búsqueda por la extensión espacial de un recurso en el área geográfica de Barcelona, se encontrará entre los resultados un registro etiquetado como perteneciente a Valencia, lo cual podría ser entendido como una inconsistencia. Al estar ahora anotado como un registro que requiere revisión, puede ser excluido de la respuesta.
- Si se realiza una búsqueda espacial usando el topónimo “Valencia”, se conseguirá que el registro 3 cubre el área geográfica de la ciudad de Barcelona. Esto también podría llegar a ser entendido como una inconsistencia. El criterio de exclusión anterior también se puede aplicar en este caso.
- Existe la posibilidad de que en la ciudad de Barcelona exista una calle, un barrio, parque u otra entidad geográfica susceptible de ser etiquetada con el término “Valencia”. Sin embargo, la ausencia del topónimo “Barcelona” en el registro 3, puede ocasionar la invisibilidad del recurso cuando se realice una búsqueda espacial con una conjunción, por ejemplo, registros de metadatos en el área de Barcelona y que contengan el topónimo Barcelona. Por lo tanto, la anotación nos podría indicar un potencial problema de invisibilidad del recurso para algunas preguntas triviales. Este problema se resolvería tras la revisión del recurso y la corrección de su descripción textual.

#### 4. Sigüientes pasos en la investigación

Al realizar el análisis estadístico de cada cluster, un metadato sería identificado como candidato a inconsistente si, perteneciendo a un cluster específico, carece de la descripción común que aparece en un número significativo de los demás metadatos del cluster.

Los resultados permitirían detectar aquellos metadatos carentes de descriptores útiles para la recuperación de los recursos y que, por carecer de dichos descriptores, pueden ocasionar que los recursos sean menos visibles respecto a sus vecinos espaciales. Esto quiere decir que, en tareas de recuperación de información sobre cartotecas, son más invisibles aquellos recursos cuya descripción no contiene referencia al mismo *topónimo* al que hacen referencia un número significativo de vecinos del mismo cluster. Es decir, es más difícil recuperarlos. Por lo tanto, hay que advertir de la necesidad de mejorar la documentación de dichos metadatos en particular.

Un paso adicional consistirá en aprovechar la capacidad de los resultados del *clustering* para señalar zonas donde hay que examinar cuidadosamente la información descriptiva de los recursos cartográficos. Este análisis permitirá ahorrar tiempo de exploración a los analistas de las bibliotecas de mapas digitales.

#### 5. Conclusiones y trabajo futuro

A lo largo de este trabajo se han señalado algunos de los retos de la preservación de la información cartográfica en las bibliotecas de mapas digitales. Entre estos resultan de interés aquellos que están vinculados con la validación de la información de carácter espacial de los metadatos asociados a los mapas.

Debido a que las aplicaciones y sistemas de búsqueda en parte deben su fiabilidad a la veracidad con la que los metadatos describen los recursos, es importante desarrollar nuevos mecanismos automáticos que asistan las bibliotecas de mapas digitales para la validación, corrección, “curación” y administración de los metadatos asociados a los mapas.

En este trabajo se han presentado las primeras fases de investigación de un método orientado a asegurar la validez de los metadatos asociados a los mapas digitales, detectando automáticamente posibles inconsistencias en los metadatos. Su detección y posterior enmienda puede mejorar la confiabilidad de los análisis y las búsquedas que se realizan sobre la información cartográfica de la biblioteca.

El método propuesto se sustenta en la combinación de técnicas de agrupación espacial con métodos estadísticos para advertir tanto la presencia de anomalías como la ausencia de información espacial descriptiva relevante. Aplicado a los sistemas de búsqueda en cartotecas, este método podría mejorar la precisión de las búsquedas espaciales.

## Agradecimientos

Este trabajo ha sido parcialmente financiado por el Gobierno de España a través del proyecto TIN2012-37826-C02-01, el Instituto Geográfico Nacional (IGN), y GeoSpatiumLab S.L. El trabajo de Walter Rentería Aqualimpia ha sido cofinanciado por el Gobierno de Aragón a través de la beca B181/11. El trabajo de Aneta Jadwiga Florczyk ha sido cofinanciado por el Gobierno de España a través de la beca AP2007-03275. El trabajo de Juan López ha sido cofinanciado por el Gobierno de España a través de la Ayuda INNCORPORA INC-TU-2011-1528.

## Referencias

- Anguita, S.; Montaner, C.; Oller, J.; Roset, R. (2012). Digital preservation at the institut cartogr de catalunya. // 7th International Workshop on Digital Approaches to Cartographic Heritage. Barcelona. 1 (April 2012) 75-82.
- Besag, J.; Newell, J. (1991). The detection of clusters in rare diseases. // Journal of the Royal Statistical Society Series A (Statistics in Society). 143-155.
- Bookstein, A.; Klein, S.; Raita, T. (1998). Clumping properties of content-bearing words. // Journal of the American Society for Information Science. 49(2):102-114.
- Capdevila, J.; Agudo, J.; Zarazaga-Soria, F.; Barrera, J.; Sánchez, A.; Soteres, C.; Criado, M.; Crespo, M. (2012). Gateway MARC21-ISO19115: definition and reference implementation. // 7th International Workshop on Digital Approaches to Cartographic Heritage. Barcelona, (Spain) April, (1):66-73.
- Chao Liu, Y.; Wu, C.; Liu, M. (2011). Research of fast som clustering for text information. // Expert Systems with Applications. 38(8):9325-9333. DOI 10.1016/j.eswa.2011.01.126.
- Ester, M.; Kriegel, H.; Sander, J.; Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. // Proc. of the 2nd Intl Conf. on Knowledge Discovery and Data Mining KDD96, AAAI Press. 226-231.
- Fernandez-Wytenbach, A.; Alvarez, M.; Bernabe-Poveda, M.; Borbinha, J. (2007). Digital map library services in the spatial data infrastructure (sdi) framework: The dig-map project. // Proc. of the 23th International Conference in Cartography. International Cartographic Association (ICA-ACI). Moscow, (Russia) August. 4-10.
- Florczyk, A.; López-Pellicer, F.; Gayán-Asensio, D.; Rodrigo-Cardiel, P.; Latre, M.; Noguera-Iso, J. (2009). Compound Geocoder: get the right position, GSDI 11 World Conference and the 3rd INSPIRE Conference 2009, Rotterdam, 15-19 June 2009.
- GeoNames (2012). GeoNames geographical database. <http://www.geonames.org/>, (2012-05-05).
- Geocoder.ca (2012). <http://geocoder.ca/?reverse=1> (2012-05-05)
- Getis, A.; Ord, J.K. (1992) The analysis of spatial association by use of distance statistics. // Geogr. Anal. 24:3, 189-206
- Google Geocoding API (2012). <https://developers.google.com/maps/documentation/geocoding/?hl=en> (2012-03-25).
- Hijmans, R.; Garcia, N.; Weiczorek, J. (2012). GADM database of Global Administrative Areas. GIS Shapefile. <http://www.gadm.org/> (2012-04-15).
- Hill, L.L. (2006), Georeferencing: The Geographic Associations of Information (Digital Libraries & Electronic Publishing). MIT Press.
- Knox, E. (1989). Detection of clusters. // Methodology of enquiries into disease clustering London: Small Area Health Statistics Unit. 17:17-22
- Kohonen, T. (1990). The self-organizing map. // Proceedings of the IEEE. 78:9, 1464-1480.
- Kulldor, M. (2006). Tests of spatial randomness adjusted for an inhomogeneity. // Journal of the American Statistical Association. 101:475, 1289-1305.
- Lacasta, J.; Noguera-Iso, J.; Zarazaga-Soria, F. (2010). Terminological Ontologies: Design, Management and Practical Applications. // Semantic Web and Beyond / Ed. Ramesh Jain and Amit Sheth, Springer US, 2010. 9, 1-24. ISBN 978-1-4419-6981-1.
- Luk, R.; Yeung, D.; Lu, Q.; Leung, E.; Li, S.; Leung, F. (2000). Digital library access for chinese visually impaired. // Proceedings of the fifth ACM conference on Digital libraries. 244-245.
- Monmonier M, (1991) How to Lie with Maps. Chicago, IL.: University of Chicago Press.
- Quach, H.; Goh, D.; Razikin, K.; Pham, T.; Kim, T.; Theng, Y.; Lim, E.; Chang, C.; Chatterjea, K.; Sun, A. (2010). Enhancing search in a geospatial multimedia annotation system. // Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services. 617-624.
- Rauber, A.; Merkl, D. (1999). The somlib digital library system. // Abiteboul S.; Vercoustre, A. M. (eds). Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science. Berlin: Springer. vol 1696, 852-852. Doi 10.1007/3-540-48155-9-21
- Sander, J.; Ester, M.; Kriegel, H.; Xu, X. (1998). Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications. // Data Mining and Knowledge Discovery, Springer Netherlands. 2:2 (1999) 169-194. doi 10.1023/A:1009745219419, 1998.
- Skupin, A.; Buttenfield, B. (1996). Spatial Metaphors for Visualizing Very Large Data Archives. // Proceedings GIS/LIS'96. Bethesda: American Society for Photogrammetry and Remote Sensing. 607-617.
- Sugihara, K.; Okabe, A.; Satoh, T. (2011). Computational method for the point cluster analysis on networks. // Geoinformatica. 15, 167-189. DOI 10.1007/s10707-009-0092-5
- Zarazaga-Soria, F. J.; Lacasta, J.; Noguera-Iso, J.; Torres, M. P.; Muro-Medrano, P. R. (2003). A Java Tool for Creating ISO/FGDC Geographic Metadata. // Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung. Beiträge zu den Münsteraner GI-Tagen. IfGI prints. 18 (2003) 17-30.

Fecha de envío: 2012-05-16.

Fecha de aceptación: 2012-08-21.

