

Anclando "La Balsa de Piedra" digital de las IDEs al mundo de la Web Semántica

Co-localización espacial de recursos IDE y la Web de datos

Lopez-Pellicer, Francisco J.; Rentería-Agualimpia, W.; Valiño, Juan; Béjar, Rubén; Noguerras-Iso, Javier; Muro-Medrano, Pedro R.

Resumen

José Saramago en su libro "La Balsa de Piedra" ("A Jangada de Pedra" en Portugués) plantea una inquietante hipótesis: España y Portugal, habiendo perdido todo contacto físico con Europa, se lanzan en un errante viaje en solitario por el Océano Atlántico. De algún modo, esto mismo está ocurriendo a las Infraestructuras de Datos Espaciales (IDEs). Estas se han convertido en balsas digitales de piedra: islas de información en la Web que han perdido todo contacto con la cultura dominante en la Web. Esta cultura está caracterizada por el concepto de "mashup" o aplicación intensiva en el consumo de datos procedentes de diferentes fuentes, y, con cada vez mayor frecuencia, con los conceptos casi sinónimos de Web de Datos, Web Semántica y Linked Data. Un buen ejemplo del aislamiento de las IDEs con la Web real es, a modo de ejemplo, el auge fuera de en las Administraciones Públicas (AA.PP.) de aplicaciones cuyo soporte de mapas y datos geográficos de referencia no es proporcionado por las IDEs auspiciadas por las AA.PP., sino por servicios de mapas y datos propietarios no estándar, con Google Maps a la cabeza.

Una de las causas del aislamiento de las IDEs de la Web radica paradójicamente en una de sus mayores virtudes: sus servicios estándar están pensados por y para la comunidad de agentes participantes de cada IDE, o de otras IDEs asociadas. En consecuencia, es difícil que un usuario ajeno a una IDE concreta, o al mismo concepto de IDE (que es lo habitual), pueda ser consciente del volumen de recursos públicos, oficiales, de calidad asociados a una localización. No podemos culpar a los estándares usados por las IDEs. Los usuarios están dispuestos a utilizar casi cualquier servicio Web, casi sin importarle el proveedor, procedencia o incluso complejidad, con tal que devuelvan datos procesables, relevantes, con enlaces a otros recursos, con razonables políticas de propiedad intelectual, en condiciones económicas competitivas, y en formatos ampliamente soportados. Este rasgo dominante en la cultura actual de la Web.

Esta ponencia plantea un aprovechamiento inteligente de este comportamiento dominante para ofrecer enlaces semánticos (en el sentido dado por Tim Berners-Lee "procesable por máquinas" [1]) a recursos publicados mediante los servicios OGC de una o varias IDEs utilizando el poder expresivo de la localización geográfica. Esta información está complementada con enlaces a recursos de la Web de Datos relacionados, lo que facilita su interpretación tanto por usuarios como por máquinas. Este aprovechamiento se materializa en lo que denominamos como servicio de "co-localización espacial". En su forma más básica, un servicio de co-localización espacial devuelve a partir de una posición espacial enlaces a recursos situados o relacionados con ella, incluyendo peticiones a servicios OGC y recursos de la Web de Datos, en un formato procesable por máquinas.

El consumo de este servicio en aplicaciones aumenta su valor añadido al enriquecer la información geográfica y semántica presentada al usuario, y simplificar la integración de dicha información. Desde el punto de vista de las IDEs, este servicio facilita la diseminación y el uso de sus recursos por terceros, haciendo factible restablecer el contacto la Web dominante.

PALABRAS CLAVE

IDE, mashups, Web semántica, co-localización.

1. INTRODUCCIÓN

Se dice que la primera ley de la geografía es "todo se relaciona con todo lo demás, pero las cosas cercanas están más relacionadas entre sí" [2]. Esta primera ley se puede aprovechar para generar enlaces semánticos entre recursos publicados mediante los servicios OGC de una o varias IDEs utilizando el poder expresivo de la localización geográfica. Esta información puede complementarse con enlaces a otros recursos, por ejemplo, descripciones de las entidades geográficas situadas en la misma localización. La generación de este tipo de enlaces se puede delegar a lo que denominamos en este artículo como *servicio de co-localización espacial* (SCE). En su forma más básica, un SCE devuelve a partir de una representación de una posición espacial (un punto, una caja envolvente, un

polígono) enlaces a recursos situados o relacionados con dicha posición. Estos recursos incluyen peticiones a servicios OGC y a recursos de la Web de Datos.

El consumo de este servicio en aplicaciones aumenta su valor añadido al enriquecer la información geográfica y semántica presentada al usuario, y simplificar la integración de dicha información. Desde el punto de vista de las IDEs, este servicio facilita la diseminación y el uso de sus recursos por terceros, haciendo factible restablecer el contacto la Web dominante.

Este trabajo se organiza como sigue. En la sección 2 presentamos conceptos relacionados con el servicio de co-localización espacial. A continuación en la sección 3, se describe el concepto de SCE. Sigue en la sección 4 la descripción de Balsa, una aplicación que ofrece un servicio de SCE para IDEs. Finalmente, presentamos como conclusión un resumen de las ventajas que aporta un servicio de co-localización espacial a una IDE.

2. CONCEPTOS RELACIONADOS

En esta sección presentamos conceptos relacionados que no forman parte del concepto de IDE que están relacionados con el concepto de servicio de co-localización espacial.

El primero de estos conceptos es el de "mashup". Una "mashup" o aplicación Web híbrida es una aplicación Web que usa y combina datos, contenido generado o funcionalidad de dos o más proveedores [3]. Por ejemplo, una mashup geográfica medioambiental puede combinar mapas procedentes de diferentes administraciones locales con datos de diversidad biológica proporcionados por organizaciones internacionales. El término "mashup" implica que el desarrollo de dicha aplicación no ha requerido un fuerte esfuerzo en la integración de información delegándose la integración a librerías fácilmente configurables. Un buen ejemplo son los clientes de visualización. Los programadores pueden incluir en su "mashup" mapas y datos geográficos utilizando estas librerías, sin requerir un conocimiento previo acerca de Sistemas de Información Geográfica. Es lo que se ha dado en llamar neo-geografía [4]. Consideramos que las "mashups" geográficas pueden aprovechar los servicios de co-localización espacial para acceder a los recursos publicados mediante una IDE.

Otro concepto relevante de la Web actual es la Web Semántica. La Web Semántica o Web de Datos es una colección de buenas prácticas y tecnologías orientadas a facilitar la comprensión de los datos en la Web. Las tecnologías que se suelen considerar parte de la Web Semántica son el modelo de descripción de recursos *Resource Description Framework* (RDF), sus formatos de intercambio (*RDF/XML*, *N3*, *Turtle*), y especificaciones de modelos (*RDF Schema*, *Web Ontology Language*) y reglas (*Rule Interchange Format*). El término "comprensión" en el contexto de la Web Semántica quiere decir facilitar que los metadatos sobre los recursos se basen en RDF [1]. Con ello, entre otros aspectos, se puede mejorar la recuperación de información al hacer explícita la semántica de los recursos.

Una de las buenas prácticas más conocidas es Linked Data (LD) o Datos Enlazados. Este término describe una metodología para publicar, compartir y conectar datos utilizando la Web. En sus primeras versiones, LD estaba enfocada a la publicación de datos en RDF. Con desarrollo de la comunidad LD las reglas han evolucionado hacia una concepción más amplia de LD. Los principios de esta visión amplia de LD han sido resumidos por Tim Berners-Lee [5] a tres simples reglas:

- Dar un nombre que comience por http a cualquier recurso interesante.
- Devuelve información en algún formato Web estándar cuando un se recupere su nombre en la Web.
- La información devuelta debe contener datos, pero es todavía más importante que devuelva relaciones a otros recursos cuyos nombres comiencen por http.

LD es un estilo de publicación de datos en la Web Semántica que simplifica la creación de enlaces entre recursos, facilita su descubrimiento mediante motores de búsqueda que indexan la Web, y su consumo por aplicaciones Web. Existen una serie de herramientas que permiten trabajar

directamente con datos publicados como Datos Enlazados: navegadores de RDF (*Disco*, *OpenLink Browser*), motores de búsqueda de RDF (*SWSE*, *Swoogle*), y librerías (*SemWeb Client Library*, *SWIC*).

Aun más importante que LD es la disponibilidad de información publicada como LD. La iniciativa *Linking Open Data*¹ impulsa la publicación de datos con licencias abiertas como LD. En octubre de 2007 el tamaño de la información publicada por esta iniciativa era de 2.000 millones de sentencias RDF que incluían 2 millones de enlaces entre las diferentes colecciones. En mayo de 2009 el conjunto de sentencias se había duplicado mientras que el número de enlaces se había multiplicado por 70. En septiembre de 2010 el volumen de sentencias había alcanzado los 25.000 millones y el número de enlaces era de 395 millones. La Figura 1 muestra las diferentes colecciones consideradas parte de LOD en septiembre de 2010 y los enlaces entre ellas. Una parte notable de la LOD es geográfica e incluye colecciones como *GeoNames*, *GeoSpecies*, *LinkedGeoData* (*OpenStreetMap*), *US Census*, y *Linked Sensor Data*.

El objetivo del SCE es relacionar en base a la localización recursos de una IDE con los recursos de la Web Semántica publicados como Linked Data. Esta relación “ancla” las IDEs a la Web Semántica. Por ejemplo, un SCE habilita a las “mashups” capaces de acceder a recursos de la Web Semántica el descubrimiento de recursos relacionados en una IDE.

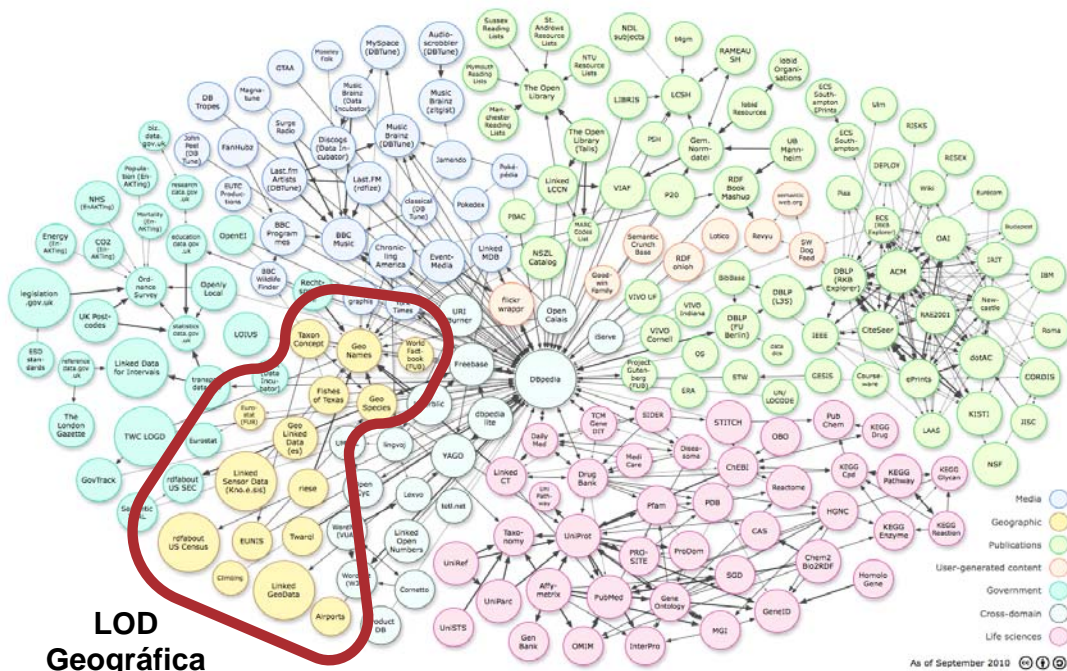


Figura 1 Estado actual de la LOD, clasificada por temáticas, compilado por Richard Cyganiak y Anja Jentzsch y disponible en <http://lod-cloud.net/>

3. SERVICIO DE CO-LOCALIZACIÓN ESPACIAL

Un SCE devuelve a partir de una posición espacial o de un recurso que contiene una descripción espacial enlaces a recursos con una localización espacial similar, incluyendo peticiones a servicios OGC y recursos de la Web de Datos, en un formato procesable por máquinas. Un SCE es capaz de trabajar tanto sobre recursos IDEs como sobre recursos de la Web Semántica gracias a que existen esquemas bien conocidos establecidos para describir las propiedades espaciales de los recursos Web. En las IDEs, estos esquemas están establecidos principalmente por OGC. En la Web Semántica, estos esquemas están definidos por el W3C², o son adaptaciones de los definidos por OGC [6].

El objetivo de un SCE espacial es devolver una colección de recursos que espacialmente estén relacionados. La Figura 2 describe cómo opera el SCE. Primero, tras la petición del usuario, localiza

¹ <http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

² <http://www.w3.org/2003/01/geo/>

una serie de fuentes de recursos potenciales. Estas fuentes pueden ser tanto servidores de una o varias IDEs como servidores de la Web Semántica. Aplicando alguna de las estrategias descritas a continuación, solicita a dichos servidores recursos candidatos. Una vez obtenidas las respuestas, filtra y ordena los recursos devolviendo los enlaces a los recursos en un formato adecuado al usuario. A partir de este punto el usuario puede acceder a los recursos navegando los enlaces. El usuario, una vez recuperadas las representaciones de los recursos, extraerá la información que necesite si sabe procesar el formato.

La estrategia más general parte de la descripción de una geometría espacial (punto, caja envolvente, polígono) con la que se puede preguntar a servidores SPARQL [7] y a servidores de descarga de datos y metadatos por recursos que se solapen con dicha geometría. SPARQL es el protocolo de la Web Semántica que ofrece una funcionalidad similar a los protocolos de descarga de datos y metadatos que existen en las IDEs. El resultado devuelto puede estar ordenado en función del grado de similitud de la extensión espacial con la geometría. Otras estrategias pueden partir de una URL que identifica un recurso de la Web Semántica o de las IDEs. Si la URL apunta a un recurso de la Web Semántica, la geometría se obtiene de sus propiedades espaciales explícitas o se agregan las propiedades espaciales de aquellos recursos que formalmente se pueden considerar *parte de*. Si la URL apunta a un recurso de una IDE accesible en un servidor de descarga de datos y metadatos, la geometría se obtiene su extensión espacial.

El SCE espacial puede actuar sobre una colección de fuentes cerradas o, incrementar la colección de fuentes utilizadas incorporando los de recursos sobre los que el usuario quiere descubrir recursos co-localizados. Por ejemplo, si a un SCE se solicita la obtención de los recursos co-localizados a un fenómeno almacenado en un WFS desconocido para el servicio, este nuevo WFS puede tenerse en cuenta para la resolución de futuras peticiones.

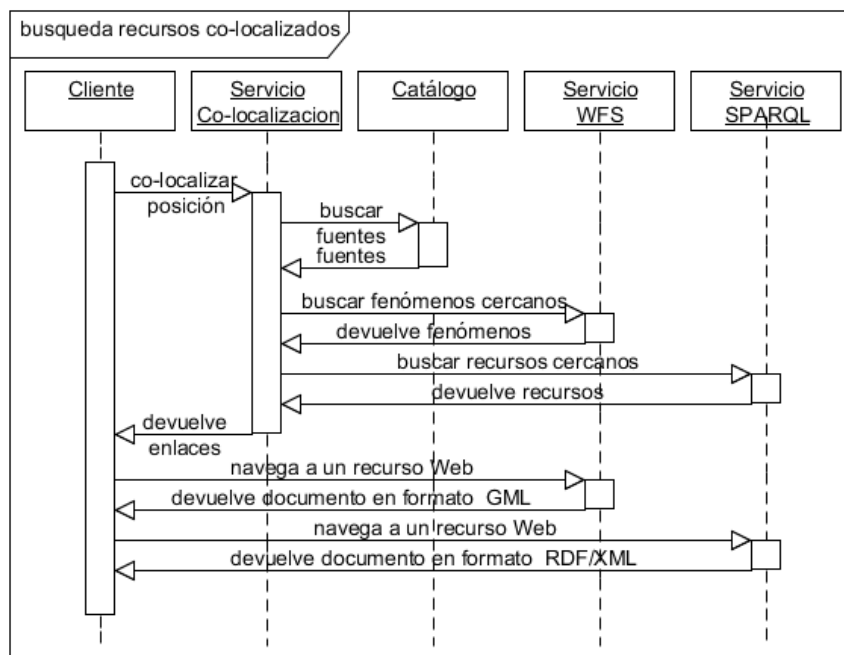


Figura 2. Búsqueda de recursos co-localizados

4. BALSAS: APLICACIÓN A INFRAESTRUCTURAS DE DATOS ESPACIALES

Balsa³ (Figura 2) es una aplicación que puede ser usada para añadir un SCE espacial a una IDEs. Un SCE espacial en el contexto de las IDE es un servicio que a partir de una descripción que esté relacionada con una localización, devuelve enlaces a recursos espacialmente relacionados publicados en la IDE o relacionados con ellos en un formato procesable por máquinas. Gran parte de los datos ofrecidos por las IDEs y la Web Semántica no son descargables directamente. Los datos de las IDEs suelen estar almacenados en bases de datos espaciales y pueden ser accedidos para su descarga enviando peticiones a servicios de descarga que soportan preguntas, tales como los servicios OGC WFS y WCS. Balsa tiene las siguientes características:

- Proporciona un SCE espacial de recursos que devuelve como resultado enlaces a datos y metadatos vía servidores WFS y CSW, y a recursos de la Web Semántica. El servicio es conforme con la especificación *OpenSearch*⁴ para facilitar su descubrimiento por "mashups".
- Proporciona un interfaz Linked Data (LD), o Datos Enlazados, a servicios de acceso a datos y metadatos vía servidores WFS y CSW, y a servicios de la Web Semántica (servidores SPARQL). Es decir, asigna identificadores Web a los recursos co-localizados que apuntan al servidor Balsa, y cuando el cliente solicita el recurso al servidor Balsa le devuelve la localización efectiva del recurso.
- Incluye la capacidad de añadir enlaces a los metadatos de los recursos co-localizados. Estos enlaces pueden apuntar a un meta dato accesible mediante un servicio CSW o a recursos de la Web Semántica.
- Devuelve respuestas en formato de intercambio orientado a usuarios como HTML o RSS, a aplicaciones como RDF/XML, y a aplicaciones geográficas como GML.

Balsa está diseñada para proporcionar un SCE, y, adicionalmente, un servicio LD. Supongamos que Balsa está ejecutándose en la dirección <http://www.ejemplo.es/balsa/>. En la configuración por defecto se pueden realizar preguntas a la siguiente dirección:

`http://www.ejemplo.es/balsa/loc?q={término}`

Donde *término* puede ser:

- Una posición espacial que describe un punto, una caja envolvente o un polígono en un sistema de referencia de coordenadas. Por defecto se asume WGS 84.
- Un nombre de recurso que comienza por http. En el contexto de Balsa consideramos recursos tanto los recursos de la Web Semántica como cualquier petición a un servidor de una IDE.
- Un nombre de una localización.

El formato de salida puede ser especificado mediante parámetros o negociado a nivel de protocolo http. Los formatos admitidos son los habitualmente empleados en las "mashups" (RSS), en las aplicaciones semánticas (basados en RDF) y en las aplicaciones geográficas (GML, KML). También puede devolver la información en formato HTML para su visualización directa por el usuario.

Balsa opera como un SCE espacial y un servicio de publicación LD (Figura 3). El SCE implementado en Balsa aplica diferentes estrategias para obtener recursos co-localizados. Por ejemplo, a partir de una posición espacial crea preguntas a las fuentes de datos registradas para localizar recursos situados en la misma localización. Si el término es una URL, la URL se resuelve y se extrae la información espacial contenida. A partir de ese punto se buscan recursos situados en dicha localización. Finalmente, si es texto se intenta localizar recursos con información espacial relacionada con dicho texto. En la medida de lo posible, las respuestas a una petición de co-localización son almacenadas localmente por Balsa para reducir el tiempo de respuesta de futuras peticiones.

³ El nombre de Balsa es un homenaje a la obra de José Saramago.

⁴ <http://www.opensearch.org/>

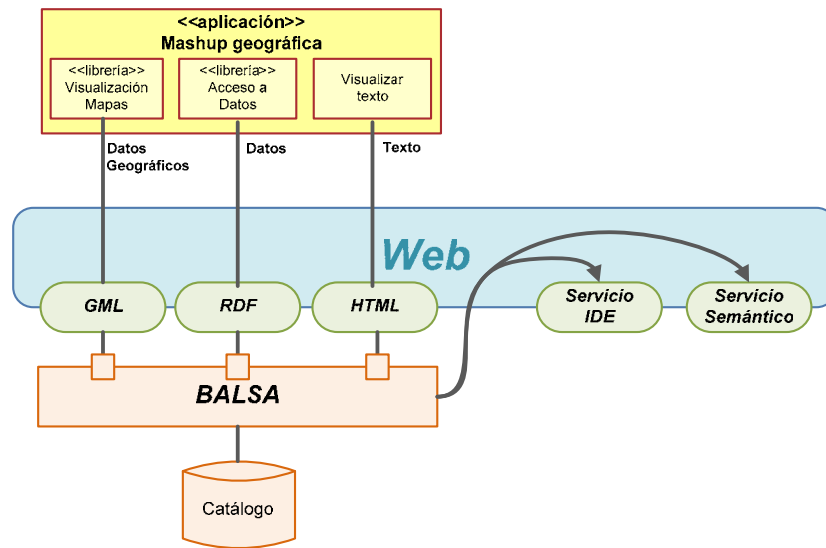


Figura 3. Arquitectura de Balsa

Cuando se configura Balsa para utilizar un servidor WFS de una IDE, se añade una entrada a la configuración del Balsa que especifica la localización del servicio, los tipos de fenómenos relevantes y las propiedades susceptibles de ser exploradas por Balsa. Cuando Balsa reciba una petición de co-localización buscará fenómenos almacenados en el servidor WFS que puedan estar espacialmente co-localizados restringiendo la búsqueda a los tipos indicados. La búsqueda se basará en la localización de los fenómenos y en el valor de las propiedades señaladas en la configuración. La respuesta del SCE contendrá hiperenlaces a cada uno de los recursos del servidor WFS que formen parte de la respuesta. Este hiperenlace será una pregunta *GetFeature* o *GetFeatureById* utilizando el estilo HTTP KVP.

Si se configura Balsa para proporcionar un interfaz LD a los recursos del servidor WFS, los enlaces que devuelva Balsa apuntarán a Balsa, y será Balsa la responsable de redirigir al cliente al servidor WFS. Es lo que se conoce como reescritura de URL. Por ejemplo, si el servidor WFS configurado es <http://www.idee.es/IDEE-WFS/ogcwebservice>, cuando un cliente solicita la URL:

<http://www.ejemplo.es/balsa/resource/idee-wfs/VerticeRedOrdenInferior/2864>

El cliente es redirigido mediante una respuesta HTTP 303 a la siguiente URL:

<http://www.idee.es/IDEE-WFS/ogcwebservice?SERVICE=WFS&VERSION=1.1.0&REQUEST=GetFeature&NAMESPACE=xmlns%28ideewfs=http://www.idee.es/wfs%29&TYPENAME=ideewfs:VerticeRedOrdenInferior&FILTER=%3CFilter%20xmlns:ideewfs=%22http://www.idee.es/wfs%22%3E%3CPropertyIsEqualTo%3E%3CPropertyName%3Eideewfs:numeroROI%3C/PropertyName%3E%3CLiteral%3E2864%3C/Literal%3E%3C/PropertyIsEqualTo%3E%3C/Filter%3E>

El interfaz LD proporciona dos servicios adicionales basados en la negociación del contenido de la respuesta: visualización para humanos y datos para máquina. Si la aplicación cliente solicita que el contenido de la respuesta sea HTML, es decir, para humanos, la aplicación cliente es redirigida mediante una respuesta HTTP 303 a la siguiente URL:

<http://www.ejemplo.es/balsa/page/idee-wfs/VerticeRedOrdenInferior/2864>

Esta dirección devuelve una página HTML con un simple cliente de visualización capaz de solicitar al servidor WFS el recurso y visualizarlo. La página puede contener información adicional almacenada previamente por Balsa sobre el recurso. La aplicación cliente puede solicitar que el contenido de la respuesta sea procesable por máquinas. En ese caso, la aplicación cliente es redirigida a la siguiente URL:

<http://www.ejemplo.es/balsa/data/idee-wfs/VerticeRedOrdenInferior/2864>

Esta dirección devuelve un documento RDF que contiene toda la información que conoce Balsa sobre dicho recurso, además del enlace al recurso en el servidor WFS.

La Figura 4 presenta una visión detallada de las interacciones entre los diferentes servicios cuando Balsa opera simultáneamente como SCE e interfaz LD a los recursos co-localizados.

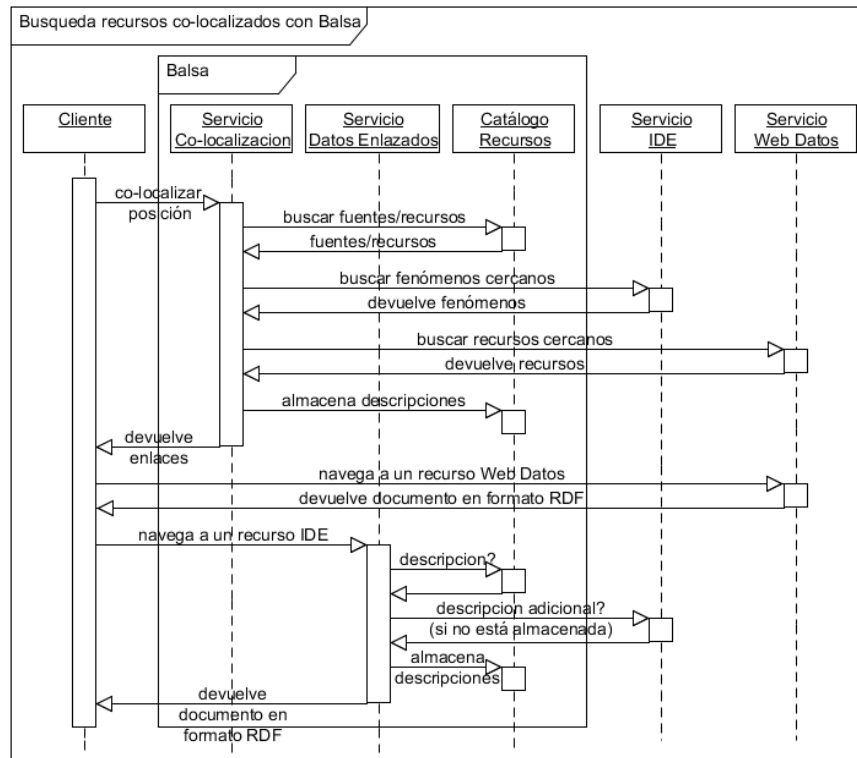


Figura 4. Búsqueda de recursos co-localizados con Balsa y acceso a sus descripciones como LD

5. CONCLUSIONES

Este artículo ha presentado el SCE espacial como un aprovechamiento inteligente de los hábitos de publicación de recursos en la Web. Este servicio, en su forma más básica, devuelve a partir de una posición espacial enlaces a recursos situados o relacionados con ella, incluyendo peticiones a servicios OGC y recursos de la Web de Datos, en un formato procesable por máquinas. Es decir, devolvemos un documento que relaciona a diferentes recursos disponibles en la Web. También presentamos la aplicación Balsa como un ejemplo del SCE espacial. Además, esta aplicación ofrece un servicio LD que permite que las peticiones a servicios OGC puedan ser procesadas como recursos de la Web de Datos.

El uso de servicios de co-localización y de aplicaciones como Balsa permite acceder a los recursos ofrecidos por las IDEs por parte de usuarios que no son conscientes de la existencia de las IDEs y que sólo desean acceder a datos procesables, relevantes, con enlaces a otros recursos, con razonables políticas de propiedad intelectual, en condiciones económicas competitivas, y en formatos ampliamente soportados.

7. AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el Gobierno de España a través de los proyectos "España Virtual" (ref. CENIT 2008-1030), TIN2009-10971 y PET2006_0245; del Gobierno de Aragón a través del proyecto PI075/08; del Instituto Geográfico Nacional (IGN) y de GeoSpatiumLab S.L.

6. REFERENCIAS

- [1] Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284 (5), 28-37 (2001)
- [2] Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. Economic geography 46(3), 234-240 (1970)
- [3] Miller, C.C.: A Beast in the Field: The Google Maps Mashup as GIS/2" Cartographic, 41 (3), 187-199 (2006)
- [4] Turner, A. J. Introduction to Neogeography. O'Reilly Shortcuts. O'Reilly, (2006)
- [5] Berners-Lee, T.: The next Web. In TED Conference, disponible en http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html (2009)
- [6] Lieberman, J.: Geospatial Semantic Web Interoperability Experiment Report 0.5.0, Technical Report, OGC 06-002r1, Open Geospatial Consortium Inc. (2006)
- [7] Prud'hommeaux, E., Seaborne, A.: SPARQL Protocol and RDF Query Language, W3C Recommendation, W3C (2008).

7. CONTACTOS

LOPEZ-PELLICER, Francisco J.
fjlopez@unizar.es
Universidad de Zaragoza
Departamento de Informática e
Ingeniería de Sistemas

BEJAR, Rubén
rbejar@unizar.es
Universidad de Zaragoza
Departamento de Informática e
Ingeniería de Sistemas

RENTERIA-AGUALIMPIA, Walter
walterra@unizar.es
Universidad de Zaragoza
Departamento de Informática e
Ingeniería de Sistemas

NOGUERAS-ISO, Javier
jnog@unizar.es
Universidad de Zaragoza
Departamento de Informática e
Ingeniería de Sistemas

VALIÑO, Juan
juanv@unizar.es
Universidad de Zaragoza
Departamento de Informática e
Ingeniería de Sistemas

MURO-MEDRANO, Pedro R.
jnog@unizar.es
Universidad de Zaragoza
Departamento de Informática e
Ingeniería de Sistemas