



GEORREFERENCIACIÓN MASIVA Y VISUALIZACIÓN DE DATOS DE BIODIVERSIDAD

Master en Tecnolgies de la Informaci3n Geogr3fica

Proyecto Final de Master – Febrero'16

Autor: Alberto 3rpez Mil3n

Tutores: Francesc Uribe Porta (MCNB)

Jos3 Quir3s Jim3nez (LIGIT)

17mtig 2015

Professionals per a la Societat de la Informaci3n

UAB

Universitat Aut3noma de Barcelona

Departament de Geografia



UAB

Universitat Aut3noma
de Barcelona

Resum

Davant la veloç generació de nombroses fonts d'informació sorgeix la necessitat de crear eines que puguin gestionar , emmagatzemar i divulgar l'enorme flux d'informació que produeix l'aparició del Big Data. El present projecte proposa un mètode a seguir per poder georeferenciar de forma massiva els bancs de mostres que posseeix el Museum de Ciències Naturals de Barcelona, de manera que per a cada registre s'ha calculat les coordenades geogràfiques, així com la incertesa associada a la posició. Aquest es un mètode més eficient, ràpid i de fàcil divulgació per emmagatzemar la informació.

També es desenvolupa un visor web on es mostra les diferents propostes cartogràfiques, que s'han elaborat , per a representar la informació generada després de la georeferenciació del banc de dades.

Resumen

Ante la veloz generación de numerosas fuentes de información surge la necesidad de crear herramientas que puedan gestionar, almacenar y divulgar el enorme flujo de información que produce la aparición del Big Data. El presente proyecto propone un método a seguir para poder georreferenciar de forma masiva los bancos de muestras que posee el Museum de Ciencias Naturales de Barcelona, de manera que para cada registro se ha calculado sus coordenadas geográficas, así como la incertidumbre asociada a la posición. Siendo esta una manera más eficiente, rápida y de fácil divulgación para almacenar la información.

También se desarrolla un visor web donde se muestra las diferentes propuestas cartográficas, que se han elaborado para representar la información generada tras la georreferenciación del banco de datos.

Abstract

Given the rapid production of multiple sources of information comes the need for tools that can manage, store and disseminate the enormous flow of information produced by the emergence of Big Data. This project proposes a method to georreferenciate masive sample banks owned by the Museum of Natural Sciences in Barcelona , so that each record has calculated their geographical coordinates, and the uncertainty associated with the position. Since this is a more efficient, fast and easy to store information dissemination way .

A web viewer where different map proposals, which have been developed to represent the information generated after georeferencing databank shown is also developed .

Agradecimientos

A Francesc Uribe por su amabilidad, profesionalidad y predisposición durante mi estancia en el Museum de Ciencias Naturales de Barcelona, así como al resto de personal del museum por su hospitalidad.

A Martí Pericay por su altruista ayuda y apoyo constante en las tareas de desarrollo informático, allanándome cualquier obstáculo que fue apareciendo.

También me gustaría acordarme de todo el personal del LIGIT, desde profesores hasta alumnos, los cuales me han ayudado a una fácil adaptación y, sobretodo, han generado en mí un gran recuerdo de mis dos primeros años en Cataluña, llevándome conmigo a buenos amigos y grandes aprendizajes.

A mis hermanos por el apoyo constante, a mis padres por su impagable fe en mí, a mis abuelos y familiares por los ánimos recibidos, y a Rocío por sus horas de atención, comprensión y cariño.

ÍNDICE

Introducción

Marco institucional	Pág. 2
---------------------	--------

Objetivos del proyecto	Pág. 3
------------------------	--------

Desarrollo del proyecto

Georreferenciación masiva y la incertidumbre asociada

Planteamiento del problema	Pág. 4
Metodología	Pág. 5
Pretratamiento de la información	Pág. 5
Funcionamiento del proceso y conexión con los servicios de geocodificación	Pág. 6
Análisis de los servicios	Pág. 7
Comparativa	Pág. 11
Incertidumbre	Pág. 14
Elección del servicio más favorable	Pág. 15
Implantación	Pág. 16
Resultados	Pág. 21

Representación cartográfica de una base de datos biológicos

Planteamiento del problema	Pág. 22
Metodología	Pág. 23
Implantación	Pág. 23
Resultados	Pág. 26

Conclusiones	Pág. 28
--------------	---------

Referencias bibliográficas/web	Pág. 29
--------------------------------	---------

Anexos

JSON generado por OpenCage	Pág. 29
Código comentado del visor cartográfico	Pág. 31

Introducción

Marco institucional

El origen del Museu de Ciències Naturals de Barcelona se remonta al año 1882, cuando se inaugura un nuevo equipamiento científico en los terrenos del Parc de la Ciutadella. El Museu Martorell aloja colecciones de historia natural y etnología. El centro abierto al público es el resultado de la donación hecha al ayuntamiento de la Ciudad por parte de Francesc Martorell y Peña, legado que viene acompañado de una entrega en metálico suficiente para levantar el que será primer edificio dedicado a museo en Barcelona.

Hasta finales del siglo XX no se abre un nuevo edificio para albergar un museo en la Ciudad. El paso del tiempo ha permitido acrecentar el patrimonio cultural del primer establecimiento en la Ciutadella, expandirse hasta crear nuevos museos temáticos y ocupar al fin diversos espacios urbanos para acoger colecciones de botánica, geología y zoología. En la actualidad se calcula que el fondo patrimonial del Museo reúne más de tres millones de registros.

No cabe duda de que se trata de un centro de investigación y de servicios científicos de envergadura. Ello se traduce de forma radical en la gestión de los datos relacionados con las muestras de colecciones. El valor de la información custodiada se corresponde con la conversión de la misma en recursos de investigación. Los criterios de rigor, coherencia, perdurabilidad, trazabilidad, difusión, estandarización y reúso se aplican para una óptima explotación en entornos esencialmente virtuales que conviven de forma dinámica con metodologías directamente aplicadas a las muestras materiales de colección.

La calidad de la información que el Museo sirve a través de sus colecciones puede ser mejorada gracias a una mejor contextualización de los elementos naturales que las conforman. En cualquier aproximación científica a los sistemas naturales el componente geográfico es imprescindible para alinear los datos con las teorías e hipótesis puestas en juego. La información espacial sirve para establecer tendencias que se benefician del tamaño de las matrices de datos disponibles. Sólo así es posible discernir interpretaciones en sistemas tan complejos como los biológicos, por ejemplo.

Ante la tendencia a agregar fuentes de datos para incrementar el tamaño muestral resulta indispensable recurrir a parámetros tan estándar como sea posible. La información espacial dispone de estos estándares pero las colecciones previas a las tecnologías de geolocalización no pueden entrar en análisis espaciales si no se procede a su georreferenciación retrospectiva para asignar coordenadas e incertidumbres a los datos textuales de origen. Así es como información anterior a nuestro tiempo es posible que participe de análisis de series temporales para las que el retroceso en el tiempo sólo se consigue por medio de este patrimonio de museos.

Señalada la importancia de georreferenciar bases de datos de colecciones de museo, surgen inmediatamente las restricciones a este procedimiento, siendo las principales:

- Una labor ingente si se trata de centenares de miles de registros.
- Considerar una medida de la calidad de la georreferenciación obtenida que permita ajustar análisis estadísticos posteriores.

En ambos temas la propuesta de proyecto de prácticas sugerido para el Master de Tecnologías en Información Geográfica (MTIG), del departamento de geografía de la Universidad Autónoma

de Barcelona, centra su acción en el tratamiento de la incertidumbre asociada a la georreferenciación, siendo esta una medida del espacio en que muy probablemente hubiera sido recolectada la muestra. La inclusión de los datos de colecciones en proyecciones geográficas digitales encuentra su correlato en la posibilidad de contrastar estos datos de distribución con otras capas de información para dar respuesta a las incógnitas de investigación que se puedan plantear.

Sin duda esta perspectiva tecnológica de la información permite derivadas de todo tipo y por ello el Museo se plantea usos y aplicaciones diversas para un mejor servicio a la comunidad científica. El proyecto de prácticas que aquí se presenta no puede por ello verse como un principio y final sino como un eslabón más aunque muy potente para seguir con futuros desarrollos.

En realidad, el Museo ha convertido la distribución de datos científicos en uno de los motores del programa patrimonial. Aunque no basta con satisfacer las necesidades científico-técnicas sino que se deben atender también los usos de la sociedad en general. Con mayor motivo en un ámbito como las ciencias naturales que conllevan una gran carga de participación amateur. La vía de distribución de datos se combina y retroalimenta con la participación de los usuarios de los servicios web que el Museo alimenta.

Concretando de nuevo el marco de objetivos del proyecto MTIG17 se debe tener en cuenta que el Museo dispone de:

- Hábito en la agregación de datos en plataformas web de ámbito internacional.
- Aplicaciones desarrolladas para la estimación de georreferenciación retrospectiva.
- Evidente experiencia en el manejo de bases de datos científicos.
- Por último, de proyectos ya tangibles que demandan y donde se ubicarían los desarrollos que se propaguen a partir del proyecto de prácticas propuesto.

El proyecto MTIG17 contiene en su germen las capacidades de análisis de datos y de visualización idóneos para desviar la atención desde proyectos complejos de investigación a conjuntos de datos mucho más numerosos, las colecciones, aunque menos dispersas temáticamente.

Objetivos del proyecto

En el presente informe se pretenden alcanzar tres objetivos aparentemente diferentes, aunque existe una gran relación entre cada uno de ellos, siendo necesaria la información que aporta cada uno para la consecución de los otros.

Estos objetivos son:

- Elaborar un diseño para representar la incertidumbre asociada a la posición geográfica de cada registro de la base de datos dada por el museum.
- Llevar a cabo diferentes representaciones cartográficas de la información que el museum posee. Estas representaciones se muestran como propuestas cartográficas para, posteriormente, ser utilizadas en futuros visores de información de la entidad.
- Realizar un procedimiento para llevar a cabo una georreferenciación masiva de los registros, cuyo origen es anterior a la aparición de las herramientas de geolocalización,

para poder otorgarle a éstos una componente geográfica de posición e incertidumbre adecuada a las tecnologías actuales.

Desarrollo del proyecto

El desarrollo del presente proyecto viene marcado por la interacción entre los diferentes objetivos que se proponen. Por ello, se exponen de manera conjunta, es decir, el objetivo referente a la incertidumbre está implementado dentro del objetivo de georreferenciación. Posteriormente se desarrollará la representación cartográfica y el visor web para mostrar los resultados.

Georreferenciación masiva y la incertidumbre asociada a la posición geográfica exacta.

Planteamiento del problema

Los avances de tecnología en el ámbito de la información geográfica han supuesto numerosos cambios y novedades a la hora de trabajar con datos geográficos, facilitando su manejo y gestión por parte de las entidades, tales como el Museum de Ciencias Naturales de Barcelona (MCNB). El cual posee una extensa base de datos compuesta por registros, que representan una gran colección de muestras científicas del medio natural. Esta colección está compuesta por muestras que tienen su origen geográfico repartido por todo el mundo.

Debido a que el origen de la colección natural del MCNB es muy anterior a la irrupción de las nuevas tecnologías de la información geográfica, aparece el problema de cómo georreferenciar o cómo mostrar los datos antiguos en los medios actuales, tanto para el público general como para la comunidad científica.

Por ello, es interesante realizar una georreferenciación retrospectiva de los registros antiguos para poder mostrarlos de manera conjunta a los actuales. Con georreferenciar nos estamos refiriendo a extraer la posición en coordenadas geográficas de cada registro. El problema es que los registros antiguos están geocodificados por topónimos, es decir, cada registro tiene asociado una localidad, municipio, provincia, país...etc.

Para poder tratar de manera homogénea estos registros junto a los más modernos, hay que transformar sus topónimos a coordenadas geográficas. Añadiendo la dificultad de que al ser una base de datos tan extensa hay que hacerlo de forma masiva.

Para solventar este problema se propone utilizar los servicios de geocoding que proporcionan empresas como Google, OpenCage o MapQuest, para poder conocer las coordenadas geográficas de posición para cada registro. Estos servicios transforman una cadena de texto (topónimos) en coordenadas geográficas.

Al problema de georreferenciación masiva hay que añadir un requerimiento propio del MCNB, puesto que trabajan con incertidumbres en la determinación de la posición exacta de los

registros. Esto proporciona mayor calidad a la información geográfica de sus muestras, siendo muy útil en procesos de investigación.

Existen diversas formas de representar la incertidumbre asociada a una posición geográfica en el debate científico. Por lo que tras una revisión bibliográfica se determina que el estudio de John Wiecek, Qinghua Guo & Robert Hijmans (2004) *The pointradius method for georeferencing locality descriptions and calculating associated uncertainty*, satisface los requerimientos demandados por el MCNB.

Este método describe una localidad geográfica con un par de coordenadas y una distancia desde el punto que forman dichas coordenadas, con estos atributos se forma un círculo. La longitud del radio determina una incertidumbre, por lo que con este método se visualiza, fácilmente, el área donde se encuentra la situación exacta de dicha localidad geográfica. Este método resulta muy práctico y eficiente a la hora de manejar los datos necesarios en la georreferenciación.

Metodología

La metodología llevada a cabo en este proyecto se organiza en tres bloques diferentes. El primero marca la necesidad de un pretratamiento de los datos antes de llevarlos a su explotación, después se hará un análisis descriptivo de los diferentes servicios que pueden ser utilizados en la georreferenciación masiva y, finalmente, se propondrá uno de estos servicios como el más adecuado para cumplir satisfactoriamente los objetivos que se pretenden alcanzar en este proyecto.

Pretratamiento de la información

Para empezar, cabe destacar la importancia de tener en cuenta algunas consideraciones en el tratamiento previo de la información.

Para una correcta obtención de resultados es importante proporcionar los datos de manera adecuada al proceso. Dichos datos se encuentran en tablas en formato csv, con una información extensa para cada registro que se pretende georreferenciar.

Los datos que se utilizan en este proyecto han sido facilitados en un formato de datos *Darwin Core Archive*, este tipo de formato es un estándar de datos informáticos sobre biodiversidad. En él se agrupa un contenido de datos sobre la ocurrencia de especies. Dicho formato es el preferido para realizar publicaciones de datos en la red de trabajo GBIF.

Los registros de la base de datos con la que se trabaja hacen referencia a la ocurrencia o aparición de determinadas especies en una escala espacio-temporal.

Es importante que a la hora de visualizar, almacenar y manipular los campos o celdas de dichas tablas se tomen en cuenta algunas consideraciones:

- El tipo de todas las celdas de la tabla ha de ser un **tipo texto**. Esto puede prevenir problemas con el formato de las coordenadas, debido a que algunos programas reconocen que son números y, automáticamente, los ordena como millares, decimales...

Camps

Típus de columna: Text

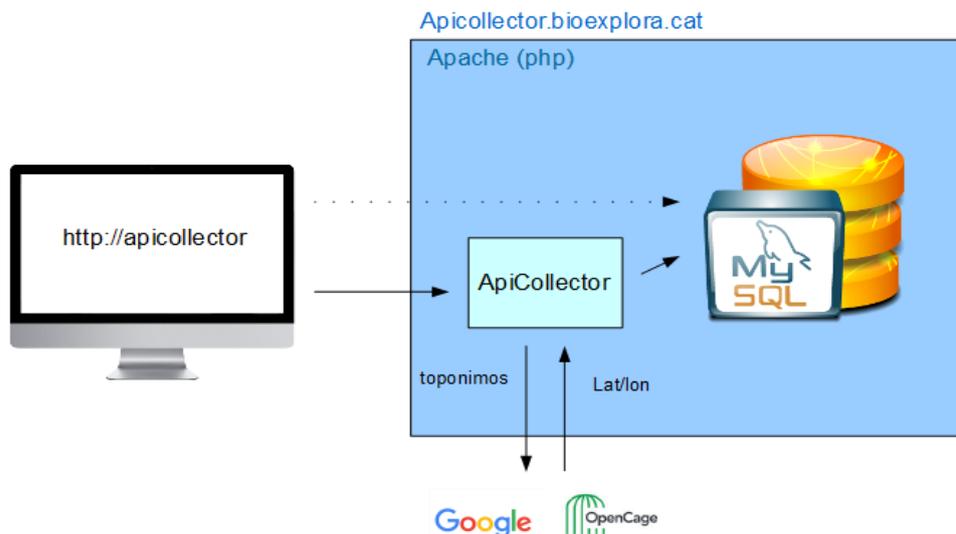
	Text	Text	Text	Text	Text	Text
1	MINIMUMDEP	MAXIMUMDEP	DECIMALLAT	DECIMALLON	GEODETICDA	COORDINATE
2			-0.2182	-78.09	WGS84	62696.89
3			-0.6333	-72.25	WGS84	
4			-0.73101	42.51813	WGS84	
5			-1.494	-78.4332	WGS84	1489305.4
6			-12.0559	-77.0605	WGS84	7643.73
7			-14.3414	-71.4789	WGS84	11623.17
8			-14.8411	-64.8873	WGS84	1172.67

- La codificación que se propone es **UTF8**, para poder visualizar de manera correcta los topónimos (acentos, apóstrofes...etc). Preferible que siempre sea la misma.
- En cuanto al **formato de escritura de los topónimos**, se ha observado un aspecto importante. Por ejemplo, si buscamos un registro que en el valor del campo *locality* tiene: “de la pileta, cueva”, y posteriormente realizamos una búsqueda a través de las APIs, se ha detectado que en la mayoría de los casos genera errores en las coordenadas obtenidas. Para poder georreferenciar de manera correcta y rápida habría que haber introducido: “cueva de la pileta”.
- En algunos servicios se generan errores y resultados inexactos al introducir el nombre de los países en catalán. Lo más recomendable es, que al menos el nombre del país, esté en inglés o castellano.

Funcionamiento del proceso y conexión con los servicios de geocodificación

Una vez que los datos están adecuados y ordenados de manera óptima para ser utilizados, resulta interesante hacer un análisis de los diferentes servicios de geocodificación, para poder determinar cuál es el que mejor se adecua a las necesidades de este proyecto. Aunque todos los servicios se implementan de la misma manera en nuestra propuesta, solo cambia que para cada API de geocodificación hay que elaborar una formulación diferente. La cual se explicará más adelante, en el análisis individual de cada servicio.

Estas formulaciones se encuentran implementadas dentro de ApiCollector y es aplicado a cada registro. Lo cual hace que sea un proceso sencillo. ApiCollector es una aplicación externa de código libre, de desarrollo propio del MCNB creada por Martí Pericay. Esta aplicación resulta muy importante en el proceso porque su funcionamiento interno posee el diseño de actuación que se propone en este proyecto.



Apicollector es una aplicación que se encuentra situada en el servidor, por ejemplo, *apicollector.bioexplora.cat*, y tiene una función de intermediario entre la base de datos, donde se encuentran nuestras tablas con la información de cada registro, y los servicios de geocodificación. Esta aplicación coge la información correspondiente a los topónimos de la base de datos y construye las direcciones url para realizar las peticiones a los servicios de geocodificación. Luego, estos servicios generan un archivo *json* donde se muestra toda la información asociada a una determinada cadena de topónimos. Dentro de este archivo se encuentran gran cantidad de datos.

Debido a que la cantidad de información que devuelven los servicios de geocodificación es muy elevada, Apicollector extrae los datos que se necesitan en este proyecto, para posteriormente, almacenarlos en la base de datos.

Análisis de los servicios de geocodificación propuestos

En este apartado se pretende realizar un análisis de los principales servicios de geocodificación en el mercado, para poder elegir cual es el servicio que más se ajusta a las necesidades demandadas en los requerimientos del proyecto.

La comunicación con estos servicios se realiza a través de peticiones HTTP, por lo que ofrecen un servicio de geocodificación estático. La respuesta que generan dichos servicios puede estar en formato JSON o XML.

Cada servicio tiene una forma concreta en cuanto a la organización de los datos de entrada, es decir la organización de los topónimos. Y a su vez, la respuesta que se genera está organizada de diferente forma en cada servicio.

Se ha observado que el orden de los campos (topónimos) y la separación de estos, a la hora de generar la petición url para cada API, puede afectar en el resultado. Esta afección será diferente según el servicio que se escoja. A continuación, para cada servicio de georreferenciación se muestran las formulaciones correctas de cada uno, según las pruebas realizadas.

¿Qué es geocodificar? Hay diferentes tipos de geocodificación:

- Geocodificación normal: transformación de las direcciones postales en coordenadas geográficas, organizadas en longitud y latitud.
- Geocodificación inversa: es el proceso de transformar las coordenadas geográficas en direcciones postales.

En el presente proyecto se va a utilizar la geocodificación normal, puesto que se tienen cadenas de topónimos que hay que transformar a coordenadas geográficas.

Google Maps Gecoding

La empresa Google tiene una API para proporcionar un servicio de geocodificación, el cual tiene una gran comunidad de usuarios, por lo que existe una abundante cantidad de información acerca de cómo desarrollarlo.

El servicio de Google Maps Geocoding es muy apto para los requerimientos de este proyecto. Tiene un alto grado de respuestas correctas en cuanto a la posición, y además, proporciona información necesaria para poder calcular la incertidumbre en la exactitud de la búsqueda por topónimos.

Para utilizar dicho servicio, al igual que todas las demás APIs que se exponen, hace falta registrar una clave personal para la API. Esto es un proceso sencillo que está muy bien explicado en las direcciones web de cada API. Cada clave tiene un límite de uso, el cual puede incrementarse contratando los servicios de la empresa.

A continuación se muestra la formulación para el servicio de Google:

<https://maps.googleapis.com/maps/api/geocode/output?parameters>

Donde en el apartado *output*, se puede elegir el formato en se quiere obtener la respuesta, los dos valores entre los que se puede escoger son: JSON y XML.

La información de entrada, es decir los topónimos que se encuentran en la base de datos del MCNB, han de introducirse a través del parámetro *parameters*. Este parámetro está compuesto por diversos componentes, con los que se puede acotar la búsqueda o solicitar mayor o menor cantidad de información. En este proyecto basta utilizar el componente *adress*, en él se organizan los topónimos de la siguiente manera:

[https://maps.googleapis.com/maps/api/geocode/json?address=\[locality\],\[municipality\],\[county\],\[stateProvince\],\[country\]&key=YOUR_API_KEY](https://maps.googleapis.com/maps/api/geocode/json?address=[locality],[municipality],[county],[stateProvince],[country]&key=YOUR_API_KEY)

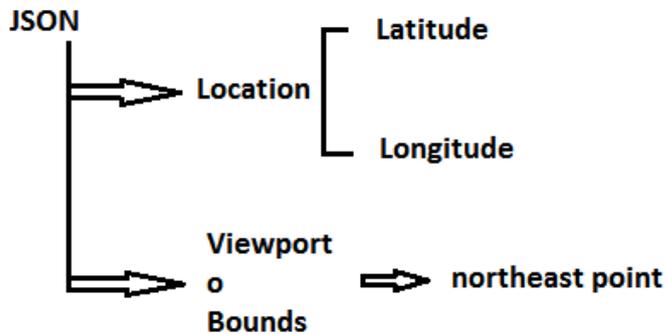
Se ha observado que para este servicio, los caracteres de separación no tienen tanta afección en el resultado de la búsqueda pero es recomendable que estos sean: **,+**

Por ejemplo, esta sería una formulación correcta para obtener las coordenadas geográficas de la Cueva de la Pileta, la cual estaría geolocalizada por los topónimos:

- Locality: Cueva de la Pileta
- Municipality: Benaoján
- County: Málaga
- StateProvince: Andalucía
- Country: España

https://maps.googleapis.com/maps/api/geocode/json?address=cueva%20de%20la%20pileta,+benaojan,+malaga,+espa%C3%B1a&key=AlzaSyD3uhPjQ8b_xgRyJ40tKTa9VaCEIFxU8ZA

De todo el JSON creado por la API de Google, solo necesitamos los siguientes apartados:



Esta información se extraerá del JSON mediante un parseo y, posteriormente, será almacenada en la tabla de registros en formato Darwin Core que se encuentra alojada en la base de datos de nuestro servidor.

OpenCage

El servicio de OpenCage es uno de los mejores servicios para satisfacer los requerimientos de este proyecto. Tiene un alto grado de respuestas correctas en cuanto a la posición, incertidumbre, selección de una confianza mínima en el resultado, y además, se puede acotar la búsqueda mediante las siglas de cada país. A través del componente *countryCode*. De esta manera la búsqueda será acotada al país que se indique en dicho componente, y el resultado será más exacto al lugar que se pretende obtener sus coordenadas geográficas.

El funcionamiento de las demás API, incluido la de OpenCage, será muy similar a la del servicio de Google, aunque la formulación de la petición url cambiará.

La formulación url para la API de OpenCage tiene la siguiente estructura:

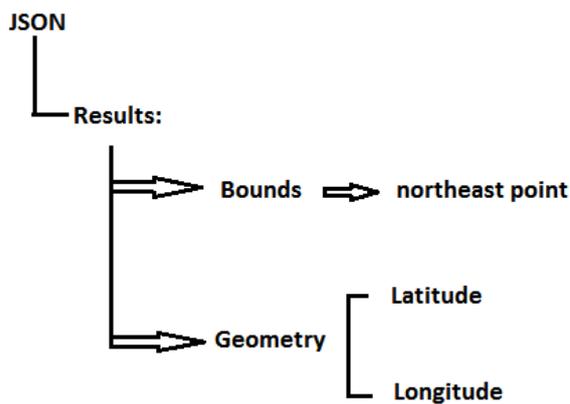
[http://api.opencagedata.com/geocode/v1/json?key=YOUR_KEY&q=\[locality\],\[municipality\],\[county\],\[stateProvince\]&countrycode=\[countryCode\]](http://api.opencagedata.com/geocode/v1/json?key=YOUR_KEY&q=[locality],[municipality],[county],[stateProvince]&countrycode=[countryCode])

A diferencia de Google Maps, para OpenCage es importante separar los campos con “,” para obtener resultados más precisos. Puesto que se observan resultados dispares al cambiar dichos caracteres que separan los topónimos.

Ejemplo para la búsqueda anterior de la Cueva de la Pileta pero con el formato de OpenCage:

<https://api.opencagedata.com/geocode/v1/json?q=cueva%20de%20la%20pileta,+malaga,+spain&key=793e8c738614d7b4ce5c4e887792f612&pretty=1>

OpenCage genera una gran cantidad de información en el JSON de respuesta. Para la realización de este proyecto solo se va a necesitar la información similar que se ha extraído en la anterior API. Por lo tanto el esquema quedaría de la siguiente manera:



MapQuest

Otra de las alternativas como servicio de geocodificación es MapQuest. Muestra una amplia posibilidad de combinar componentes para personalizar la búsqueda, aunque si es cierto que está muy enfocada al formato de direcciones de América del Norte. Esto puede ser un problema para codificar zonas del resto del planeta. Este problema no ha sido detectado en el proceso de pruebas para este proyecto, ya que el ámbito geográfico de trabajo es muy extenso y general.

La formulación de la petición url es la más sencilla comparada con el resto de servicios que se proponen en este estudio. Tan solo es necesario separar los topónimos con un espacio en blanco, aunque esto puede generar resultados algo ambiguos.

[http://www.mapquestapi.com/geocoding/v1/address?callback=renderOptions&inFormat=json&outFormat=json&location=\[locality\]%20\[municipality\]%20\[county\]%20\[stateProvince\]%20\[country\]](http://www.mapquestapi.com/geocoding/v1/address?callback=renderOptions&inFormat=json&outFormat=json&location=[locality]%20[municipality]%20[county]%20[stateProvince]%20[country])

Realizar búsquedas de provincias en este servicio requiere un proceso de programación extra en las aplicaciones utilizadas, por lo que habría que generar protocolos más detallados en MapQuest. Esto resulta ser un hándicap añadido.

Cartodb

El servicio de elaboración cartográfica y análisis espacial Cartodb también posee servicios de geocodificación. Tiene una forma metodológica diferente a los anteriores servicios, mucho más directa y con menor posibilidad de gestionar la búsqueda. Esto se presenta como un aspecto negativo para las necesidades de este proyecto.

La manera de obtener los resultados se basa en subir a la plataforma de Cartodb un archivo con la información ordenada en tablas, y en el gestor de la plataforma se seleccionan las columnas que contienen la información referente a los topónimos.

Posteriormente, cartodb procesa la información y la georreferencia en un elemento cartográfico.

Este proceso no se realiza a través de peticiones url, por lo que no existe una manera autónoma de introducir los datos en la base de datos que se encuentra alojada en el servidor del MCNB. Por lo que para ello, habría que extraer las tablas generadas en cartodb y luego mediante uniones de campo adherirlas a la base de datos.

El servicio de geocodificación de Cartodb no proporciona información para poder calcular la incertidumbre, por lo que no cumpliría con uno de los requisitos importantes del proyecto.

A priori, no parece el método más óptimo, aunque su proceso es más sencillo que los anteriores.

Comparativa de los servicios de geocodificación

Aunque no se ha comentado anteriormente hay que señalar que cada servicio tiene unos límites de uso por día y un coste adicional cuando se sobrepasa dicho límite. Esto puede ser importante a la hora de elegir cual es el servicio más adecuado para el proyecto.

Por lo tanto, el gasto económico que hay que realizar dependerá del número de registros que se pretenden georreferenciar y en cuanto tiempo queremos conseguirlo.

Para poder decidir cuál es el servicio más óptimo hay que realizar una comparación entre estos atendiendo a:

- Calidad en la información que proporcionan. Precisión, exactitud...
- Numero de resultados obtenidos
- ¿Posibilidad para calcular la incertidumbre?
- Coste

Por todo lo anterior, hay que realizar unas pruebas de cada servicio y, posteriormente, comparar dichos resultados.

Pruebas de los resultados de cada servicio de geocodificación

Para realizar las pruebas se ha utilizado el programa de escritorio Qgis. Siguiendo el siguiente esquema:



Siendo representados los mismos registros, tanto con coordenadas conocidas como georreferenciados por la vía masiva. Para cada registro se ha realizado la georreferenciación a través de los diferentes servicios para poder compararlos.

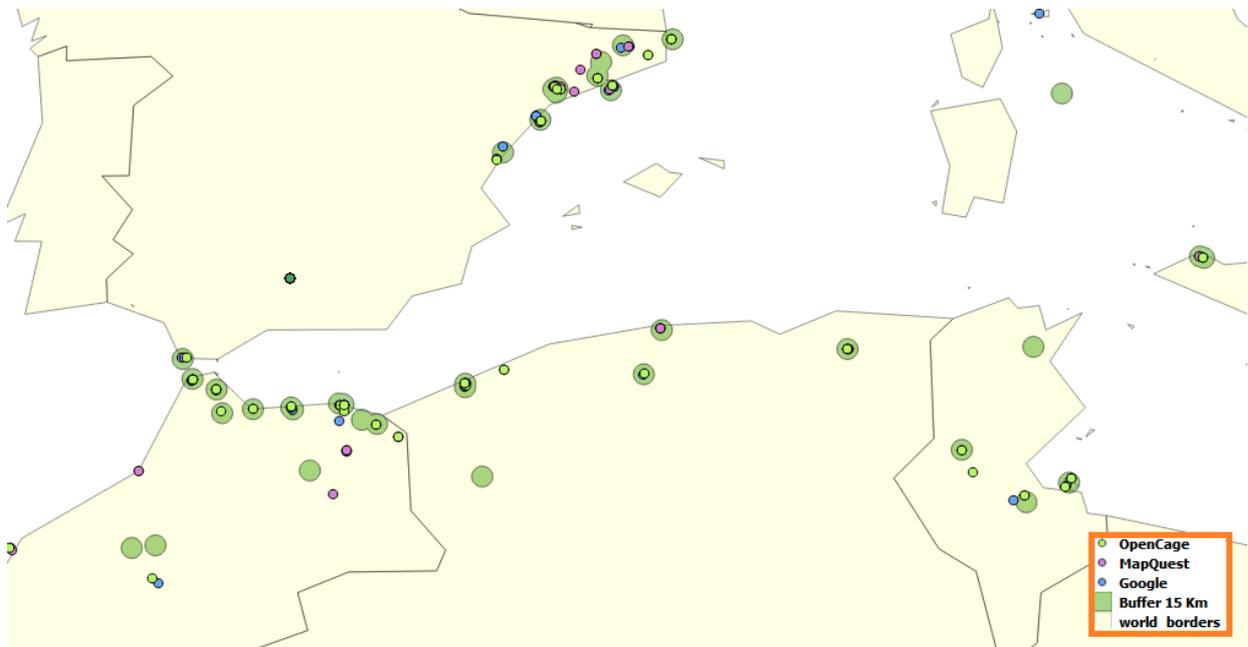
Quedando así una representación para cada registro de:

- Coordenadas medidas con precisión y conocidas.
- Georreferenciación en OpenCage.
- Georreferenciación en Google.
- Georreferenciación en MapQuest.
- Georreferenciación en Cartodb.

De esta manera pueden compararse la precisión y los resultados de cada servicio, respecto a las coordenadas medidas con exactitud para cada topónimo que se ha buscado a través de los diferentes servicios de geocodificación.

De manera, que después de esta georreferenciación, se aplica un buffer a los puntos que representan las coordenadas medidas con exactitud, y se cuentan cuáles de los puntos georreferenciados por los diferentes servicios se encuentran dentro del límite de aceptación de 15 km que se ha propuesto.

A continuación se muestra un extracto con zoom para ver el detalle del proceso de pruebas:



En la figura se puede observar que hay resultados que están dentro del buffer, siendo estos resultados válidos, y otros resultados se encuentran fuera, siendo datos, a priori, erróneos.

A simple vista se pone de manifiesto que OpenCage y Google son los servicios que tienen un mayor grado de exactitud.

Para observar los datos de manera interactiva se pueden realizar a través de:

https://mcnb.cartodb.com/viz/a4e78f02-820f-11e5-8a24-0ef24382571b/public_map

Una vez analizados los resultados de las pruebas se pueden poner en común los resultados de cada servicio de manera normalizada, y así poder realizar una comparativa justa entre cada uno de ellos. Para ello se muestra la siguiente tabla que recoge de manera sintética toda la información generada en este proceso de pruebas.

Comparación entre los diferentes servicios

COMPARATIVA DE APIs						
	Nº resultados	Precisión	¿Devuelve siempre incertidumbre?	Precio	Coste	Descripción
Google	93/100	Alta	Si, límites	2500/día. Posibilidad diarios 100.000 0.50\$/1000 request	43,98 \$	Acepta cualquier formato de entrada de datos. Resultado exacto --> No Bounds
OpenCage	96/100	Buena	Siempre, límites	2500/día. Presupuesto, posiblemente gratis	Pedir ???	Sensible al formato de entrada de los datos. Mejor servicio para representar incertidumbre. Posibilidad buscar provincias
MapQuest	99/100	Buena	Si, rangos de confianza	5000/día	Gratis si es OpenData, pero limitado a 5000 request/day	Rangos: Excelente, bueno, aprox. No es posible generar un radio exacto de la incertidumbre
Cartodb	78/100	Buena	No	250/mes. request	15\$/1000	Servicio más facil y comodo de implementar
					90479 registros aprox.	

Incertidumbre

La metodología llevada a cabo para mostrar los valores de incertidumbre está relacionada con el estudio anteriormente citado, *The pointradius method for georeferencing locality descriptions and calculating associated uncertainty* (Wieczorek, J. et al.2004). Por tanto, se utiliza el método del punto-radio.

Las APIs de geocodificación usadas proporcionan una gran cantidad de información útil para calcular la incertidumbre en la posición exacta del topónimo que se busca, además de sus coordenadas geográficas.

Por lo que se proponen dos maneras para representar la incertidumbre de acuerdo con dicho método de representación:

- Analizando el json devuelto por OpenCage se puede observar que proporciona una información llamada *Confidence*, este parámetro está relacionado con la precisión en la posición del topónimo. Tiene una escala de valores del 0 al 10, a los que relaciona con un tamaño de radio específico a cada uno.
 - 10 - less than 0.25 km distance
 - 9 - less than 0.5 km distance
 - 8 - less than 1 km distance
 - 7 - less than 5 km distance
 - 6 - less than 7.5 km distance
 - 5 - less than 10 km distance
 - 4 - less than 15 km distance
 - 3 - less than 20 km distance
 - 2 - less than 25 km distance
 - 1 - 25 km or greater distance
 - 0 - unable to determine a bounding box

Cada valor de confidence tiene un radio asociado, con el que se puede trazar un círculo de incertidumbre alrededor del punto geográfico proporcionado para el topónimo buscado.

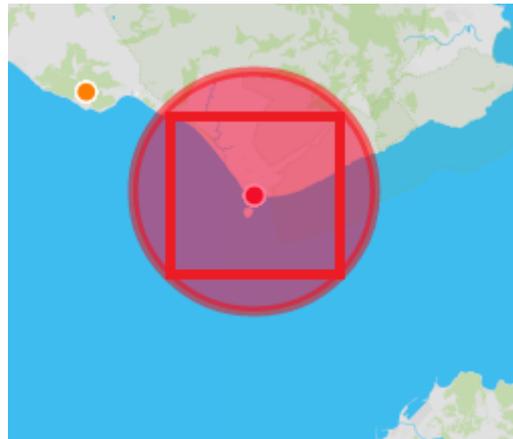
Este parámetro se puede acotar en la petición a OpenCage, con esto se puede tener un control de la confianza deseada en la generación de los datos.

Esta propuesta de hallar la incertidumbre tiene el problema de que para extensiones amplias no proporciona un radio exacto de la búsqueda. Genera resultados aproximados.

- El segundo método puede ser implementado en cualquier servicio de geocodificación propuesto anteriormente. Esto es posible ya que estos servicios proporcionan un

parámetro llamado *Bounds*, en los que aseguran que la búsqueda que se ha realizado se encuentra dentro del marco que forman estos límites.

Cada servicio proporciona estos límites de una manera particular, pero pueden ser fácilmente recogidos y almacenados. Por defecto, los representa mediante un rectángulo, posicionando la búsqueda en el centro de éste. Esta representación no cumple con los requerimientos del MCNB para representar la incertidumbre, por lo que se propone utilizar la distancia entre el punto geográfico dado y la esquina NE del cuadrado generado por los límites.



Este método es más exacto que el anterior, puesto que recoge la distancia del radio de una manera más exacta, sin rangos de aproximación. Creando para cada registro un radio de incertidumbre más fidedigno a el resultado que se genera en la búsqueda de determinados topónimos.

Se ha llevado a cabo una implementación en Apicollector que recoge los datos necesarios para calcular la longitud existente entre el punto que forman las coordenadas geográficas resultantes y el punto NE que determina el parámetro *bounds*, el cual establece los límites del resultado de la búsqueda por topónimos. Esta longitud calculada es el radio de incertidumbre.

Elección del servicio más favorable

Tras un análisis de los resultados obtenido mediante las pruebas anteriormente descritas, es el momento de elegir cuál de los servicios es el más adecuado para garantizar el cumplimiento de los objetivos del presente proyecto. Para ello el servicio ha de cumplir con los siguientes requisitos:

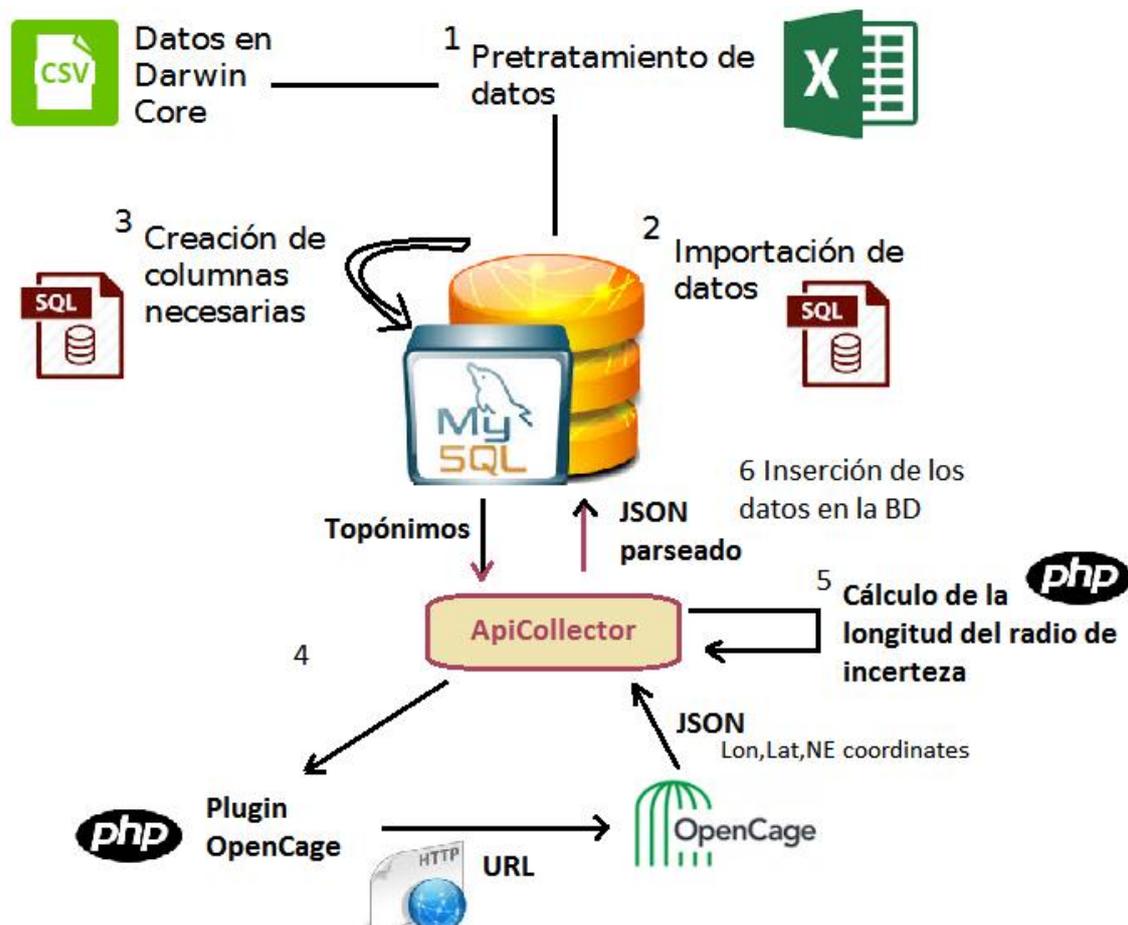
- Ha de facilitar el cálculo del radio de la circunferencia que muestra la incertidumbre.
- Posibilidad de georreferenciar provincias enteras si así se requiere, es decir, si se busca el topónimo Málaga, que se pueda elegir entre obtener los datos para georreferenciar la ciudad de Málaga como los datos para la provincia entera de Málaga. Puesto que es una petición del MCNB.
- Buena precisión en los resultados.
- Tener un coste moderado

Analizando todos estos requisitos, y tras un proceso experimental realizado, se propone utilizar el servicio de geocodificación proveído por OpenCage. Debido a que:

- Genera resultados buenos para geocodificar provincias.
- Alto número de resultados con la información *Bounds*, solo cuando el resultado es muy exacto no lo genera. Mientras que google en este aspecto, cuando el resultado es más o menos exacto genera otro componente llamado *Viewport*, que puede generar confusiones y pérdida de la calidad de la información, puesto que esto no tiene relación con la incertidumbre.
- Las búsquedas pueden acotarse fácilmente por el componente *countryCode*, de manera que se puede acotar las búsquedas para mayor fidelidad en el resultado.

Implementación

El proceso de georreferenciación está implementado en varias tecnologías y lenguajes de programación diferentes. Se propone seguir unas pautas ordenadas para organizar la arquitectura necesaria en este proyecto.



- 1. Pretratamiento de los datos.

Los registros están en un archivo .csv en formato DarwinCore, siempre que se abra este archivo en algún programa tipo hoja de cálculos (LibreOffice, por ejemplo), habrá que darle un tratamiento de texto a todas las celdas de nuestro archivo, a parte se propone darle una codificación UTF8.

Es decir:

Importa

Joc de caràcters: Unicode (UTF-8)

Llengua: espanyol (Espanya)

Des de la fïla: 1

Opcions del separador

Amplada fixa Separat per

Tabulació Coma Punt i coma Espai Altre

Fusiona els delimitadors Delimitador de text: "

Altres opcions

Camp entre cometes com a text Detecta els noms especials

Camps

Típus de columna: Text

	Text	Text	Text	Text	Text	Tex
17		Asia	Sri Lanka LK		western	
18		Asia	Sri Lanka LK		western	
19		Asia	Sri Lanka LK		western	
20		Asia	Sri Lanka LK		central	
21		Asia	Sri Lanka LK		western	
22		Asia	Sri Lanka LK		Sabaragamuwa	
23		Asia	Sri Lanka LK		Sabaragamuwa	
24		Asia	Sri Lanka LK		Sabaragamuwa	

Ajuda D'acord Cancel·la

Es necesario eliminar las “,” de los campos Locality, Municipality, stateProvince, Country...y posteriormente reorganizar la celda. Puesto que esto genera errores y resultados no deseados. De modo que:

Granadella, la ---> la Granadella

Para conseguir esto se propone realizar la siguiente concatenación de fórmulas en LibreOffice calc:

```
=RETALLA(CONCATENA(DRETA(A1;LONG(A1)-TROBA(";",A1));"dejar un espacio blanco aquí";ESQUERRA(A1;TROBA(";",A1)-1)))
```

```
=REDUCIR(CONCATENAR(DERECHA(AC8;LARGO(AC8)-ENCONTRAR(";",AC8));" ";IZQUIERDA(AC8;ENCONTRAR(";",AC8)-1)))
```

Se ha hecho como ejemplo para la celda A1, este campo variará según el formato en el que están estructuradas las tablas.

- 2. Conexión al software de gestión de la base de datos e Importación de los datos.

Para esta ocasión se utiliza como ejemplo el software MySQL Workbench 6.3, aunque se puede llevar a cabo con cualquier otro programa. Es necesario poseer un servidor y una contraseña para poder realizar la conexión. Por ejemplo:

Servidor: **apicollector.bioexplora.cat**

Username: **apicollector**

Contraseña:

Una vez conectados podremos ver el contenido de la base de datos y gestionar las tablas que aparecen.

Importación de la información a la base de datos.

A continuación se van a importar los datos con el tratamiento previo realizado, esta tabla que se va a importar tiene que estar nombrada como “dwc”, puesto que ApiCollector está configurado para trabajar con este nombre de tabla. Los formatos en los que pueden subirse son diversos: csv, dbf, excel... Se recomienda trabajar en csv.

- 3. Creación de campos necesarios antes de empezar a georreferenciar

Es necesario crear una columna nueva en la tabla que se acaba de importar a la base de datos. Este nuevo campo es importante para que ApiCollector pueda funcionar de manera correcta y así tener un control de los posibles fallos o informaciones del proceso de georreferenciación.

Las propiedades de este campo han de ser las siguientes:

Nombre: **opencage_hits**

Extensión: **int (1)**

Este proceso se realiza mediante una línea de código en SQL:

```
ALTER TABLE dwc ADD opencage_hits int(1)
```

- 4. Realizar las peticiones a OpenCage a través de ApiCollector y almacenar el resultado.

Este proceso se ha implementado en ApiCollector. Genera peticiones para cada uno de los registros que se encuentren en la tabla dwc de la base de datos. Estas peticiones son realizadas al servicio de geocoding de OpenCage, con la estructura óptima de los datos para que este servicio funcione correctamente.

Para realizar esto hay que ejecutar la siguiente url:

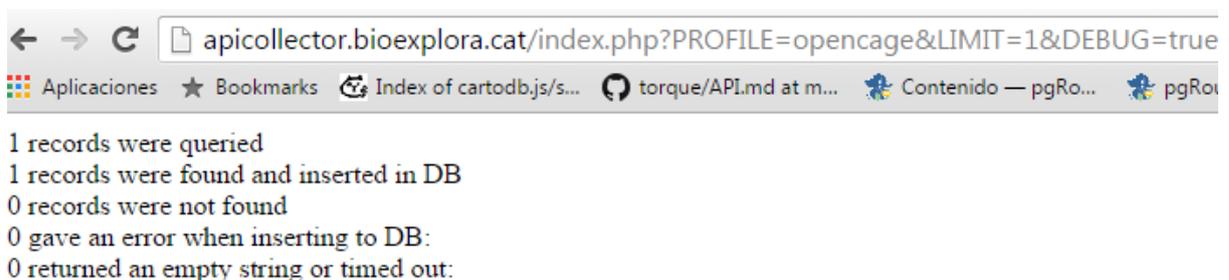
<http://apicollector.bioexplora.cat/index.php?PROFILE=opencage&LIMIT=5&SLEEP=100&DEBUG=true>

ApiCollector posee unos parámetros de configuración que se detallan a continuación:

- PROFILE: ApiCollector trabaja con diferentes servicios de geocoding, en este caso vamos a utilizar OpenCage.

- LIMIT: mediante este parámetro se pueden limitar el número de peticiones que queremos lanzar, si no lo establecemos hará todas las peticiones para todos los registros.
- SLEEP: Milisegundos de espera entre una petición y otra. Por defecto está programado para 1 segundo.
- DEBUG: Si está activado (true) generará un texto de control del proceso almacenado en el archivo log de apicollector.
- ONLYNULL: (boolean). Si está en true, realizará las peticiones de los registros que aún no han sido lanzadas, si false, empezará desde el principio indistintamente de si han sido ya geocodificados o no.

Al mandar dicha url debe aparecer en el navegador la información del proceso (esto puede tardar unos minutos dependiendo del número de registros)



```

1 records were queried
1 records were found and inserted in DB
0 records were not found
0 gave an error when inserting to DB:
0 returned an empty string or timed out:

```

Debido a que algunos registros solo poseen el campo country, hay que lanzar dos veces esta url, efectuando un sencillo cambio en el archivo de configuración de Apicollector:

Primero se lanza por primera vez con la configuración por defecto, esto georreferenciará todos los registros que tienen *locality*, *municipality...etc*, por su localidad.

A continuación hay que georreferenciar los registros que solo tienen countryCode, Opencage no lo hará bien si no modificamos la configuración de la instrucción que se le manda. Para ello antes hay que borrar los datos de la columna *opencage_hits* en los registros que solo tienen countryCode, esto se hace con la siguiente instrucción:

```
UPDATE dbapicollector.dwc SET opencage_hits=null WHERE locality="" AND municipality="" AND stateProvince="" AND county="" AND countryCode<>"
```

Una vez hecho este paso, solo queda modificar la configuración de la instrucción que hay por defecto y se lanza. El archivo a modificar está en la carpeta *conf* (*//datos/web/lib/conf*), y es el archivo *opencage.plugin.php*. Hay que introducir estas líneas de código sustituyendo a las anteriores:

```

$this->config["urlpattern"] =
"http://api.opencagedata.com/geocode/v1/json?key=793e8c738614d7b4ce5c4e887792f612&q
=[country]&countrycode=[countryCode]";

$this->config["queryfield"] =
array("countryCode","country","locality","municipality","county","stateProvince");

```

Para un mayor detalle y comprensión de cómo trabaja ApiCollector visitar:

<https://github.com/mpericay/apicollector>

- 5. Calculo de la longitud del radio de incertidumbre

Este proceso se realiza cuando ya se han obtenido los datos desde el JSON que devuelve OpenCage. Posteriormente, se calcula la distancia existente entre el punto geográfico dado y el punto NE del componente *Bounds*, es decir el punto noreste del límite que establece a la búsqueda, Opencage.

Para poder realizar este proceso de cálculo, dentro de apiCollector esta implementado este código en lenguaje PHP:

```
public function getDistanceBetweenPoints($latitude1, $longitude1, $latitude2, $longitude2, $unit = 'Mi') {
    $theta = $longitude1 - $longitude2;

    $distance = (sin(deg2rad($latitude1)) * sin(deg2rad($latitude2))) + (cos(deg2rad($latitude1))
    * cos(deg2rad($latitude2)) * cos(deg2rad($theta)));

    $distance = acos($distance);

    $distance = rad2deg($distance);

    $distance = $distance * 60 * 1.1515; switch($unit) {
        case 'Mi': break; case 'Km' : $distance = $distance * 1.609344;
    }

    return (round($distance,2));
}
```

- 6. Inserción de los resultados en la base de datos.

Una vez obtenidos la información geográfica y calculado el radio de incertidumbre, apiCollector inserta los datos en la base de datos conectada, de manera automática.

	SUBGENUS	SPECIFICEP	INFRASPECI	TAXONRANK	opencage_hits	opencage_lat	opencage_lon	opencage_radius_km	opencage_confidence
s		rufa		Species	1	41.3338964	1.9216696	4.11	7
s		glandarius		Species	1	41.3281334	1.9117486	2.78	7
pod	Dolichopoda	linderii		Species	1	41.995547	2.4182542	5.44	7
				Genus	1	14.4750607	-14.4529612	414.1	1
		nilotica		Species	1	26.2540493	29.2675469	982.69	1
don		grandis		Species	1	39.7837304	-100.4458825	5489.51	1
ta		californiensis		Species	1	39.7837304	-100.4458825	5489.51	1
nopsis		hirundo		Species	1	-2.9814344	23.8222636	1248.06	1
				Genus	1	13.5066394	104.869423	325.26	1
ta		oregonensis		Species	1	39.7837304	-100.4458825	5489.51	1
eramus		lineatus		Species	1	19.1399952	-72.3570972	142.45	1
		nigrescens		Species	1	18.1850507	-77.3947693	183.18	1
emma		lateralis		Species	1	23.0131338	-80.8328748	708.18	1

Posteriormente se pueden extraer los datos de la tabla eligiendo el formato de salida. Tras este último paso se obtienen los registros georreferenciados, listos para ser utilizados en la representación cartográfica.

Resultados

Después de este proceso de implementación se obtiene como resultado un total de 895623145462145 registros georreferenciados. Estos resultados pueden ser consultados a través de la base de datos MySQL donde han sido almacenados, están organizados en una tabla de formato DARWIN CORE como la siguiente:

	AV	AW	AX	AY	AZ	BA	BB	BC	BD
1	genus	subgenus	specificEpithet	infraspecificD	taxonRank	opencage_hits	opencage_lat	opencage_lon	opencage_r
2	Regulus		ignicapilla		Species	1	41.11667	1.25	41.1468620
3	Sylvia		melanocephala		Species	1	41.1172364	1.2546057	41.1655218
4	Sylvia		atricapilla		Species	1	41.11667	1.25	41.1468620
5	Erithacus		rubecula		Species	1	41.11667	1.25	41.1468620
6	Myiopsitta		monachus		Species	1	41.3825596	2.1771353	41.4679428
7	Riparia		riparia		Species	1	41.4636665	0.4118677	41.4898426
8	Sylvia		melanocephala		Species	1	41.11667	1.25	41.1468620
9	Phaedusa				Genus	1	19.531932	76.0554568	22.0309689
10	Glessula				Genus	1	19.531932	76.0554568	22.0309689
11	Glessula				Genus	1	7.2955	80.6356	7.35420894
12	Glessula				Genus	1	7.2955	80.6356	7.35420894
13	Glessula				Genus	1	6.934287	79.8532704	6.9811557
14	Glessula				Genus	1	7.2955	80.6356	7.35420894
15	Glessula				Genus	1	6.934287	79.8532704	6.9811557
16	Glessula				Genus	1	7.2955	80.6356	7.35420894
17	Glessula				Genus	1	7.2955	80.6356	7.35420894
18	Glessula				Genus	1	7.2955	80.6356	7.35420894
19	Glessula				Genus	1	7.2955	80.6356	7.35420894
20	Subulina		octona		Species	1	6.934287	79.8532704	6.9811557
21	Opeas				Genus	1	21.89338375	106.58861194596	22.4614663
22	Opeas				Genus	1	6.934287	79.8532704	6.9811557
23	Opeas				Genus	1	6.934287	79.8532704	6.9811557
24	Opeas				Genus	1	6.934287	79.8532704	6.9811557
25	Opeas				Genus	1	6.934287	79.8532704	6.9811557
26	Opeas				Genus	1	6.934287	79.8532704	6.9811557
27	Opeas				Genus	1	7.2955	80.6356	7.35420894
28	Zootecus		insularis		Species	1	6.934287	79.8532704	6.9811557
29	Achatina		fulica		Species	1	7.2531997	80.3451354	7.2931997
30	Achatina		fulica		Species	1	7.2531997	80.3451354	7.2931997

Para cada registro se han obtenido sus coordenadas geográficas y la incertidumbre asociada a su posición.

En el anexo electrónico que se adjunta en CD-ROM se muestra el archivo completo resultante del proceso de georreferenciación, con el nombre: *georreferenciació.csv*

Representación cartográfica de una base de datos biológicos

Planteamiento del problema

Uno de los principales problemas a la hora de representar de manera cartográfica es la cantidad de elementos que se quieren mostrar. Puesto que si existe un alto número de elementos, la información que puede proporcionar un mapa se puede ver afectada negativamente por el exceso de información mostrado. Debido a que la base de datos del MCNB es muy extensa la representación cartográfica de esta se presenta como un complicado problema a resolver.

A esto se le suma que la información puede representarse de diferentes formas, que éstas a su vez proporcionan informaciones diferentes. Para solventar estos problemas se plantean diferentes representaciones para que el MCNB decida qué forma de representar la información es más adecuada para sus propósitos. Puesto que está en desarrollo un visor cartográfico llamado TaxoMap 3.0. Este visor tomará los resultados cartográficos que resulten de este proceso de representación cartográfica.

Otro de los problemas a la hora de representar la información que se posee en las bases de datos del MCNB es, ¿con qué capas ambientales se puede complementar la información? Para ello habrá que realizar un análisis de cuáles son las variables ambientales más interesantes para complementar la información que se posee. En este proyecto se proponen las variables de temperatura global, cobertura del suelo global y precipitación global. Debido que son variables que tienen influencia en muchos otros parámetros ambientales, y a partir de ellas se pueden sacar una gran diversidad de valoraciones, siendo a su vez, fáciles de representar junto a los numerosos registros que se pretenden cartografiar.

Resulta importante que la información que se quiere mostrar tenga una lectura fácil y clara, para que pueda ser entendida por usuarios generales hasta usuarios avanzados, estos últimos podrán sacar deducciones propias alternando la información que da la clase de registros, con su posición geográfica más las variables de temperatura y precipitación.

Además, se pretende añadir una representación de la incertidumbre asociada a la posición de cada registro, para así dar mayor cantidad de información en el servicio. El principal problema que aporta este objetivo es tanto a la hora de implementarlo como de representarlo, puesto que su forma de mostrarlo ha de ser en un segundo plano y solo cuando el usuario lo solicite. De esta manera se intentará solucionar a través del evento *click* en el marcador de cada registro. Es decir que cuando se haga click en un marcador, aparezca tanto la información específica de dicho registro, así como, se dibuje un círculo que represente la incertidumbre de la posición.

Para mostrar todo lo anterior, se propone crear un visor web interactivo, que sirva como muestra de toda la información generada en este apartado. Para ello, se han utilizado los lenguajes de programación web: CSS, HTML y JavaScript, junto con las librerías de Leaflet y Cartodb, las cuales son librerías para la representación y creación de mapas, que facilitan su implementación en las tecnologías web a través de sencillos archivos JavaScript.

Metodología

La metodología desarrollada en el apartado de representación cartográfica se puede resumir a nivel general en los siguientes apartados:

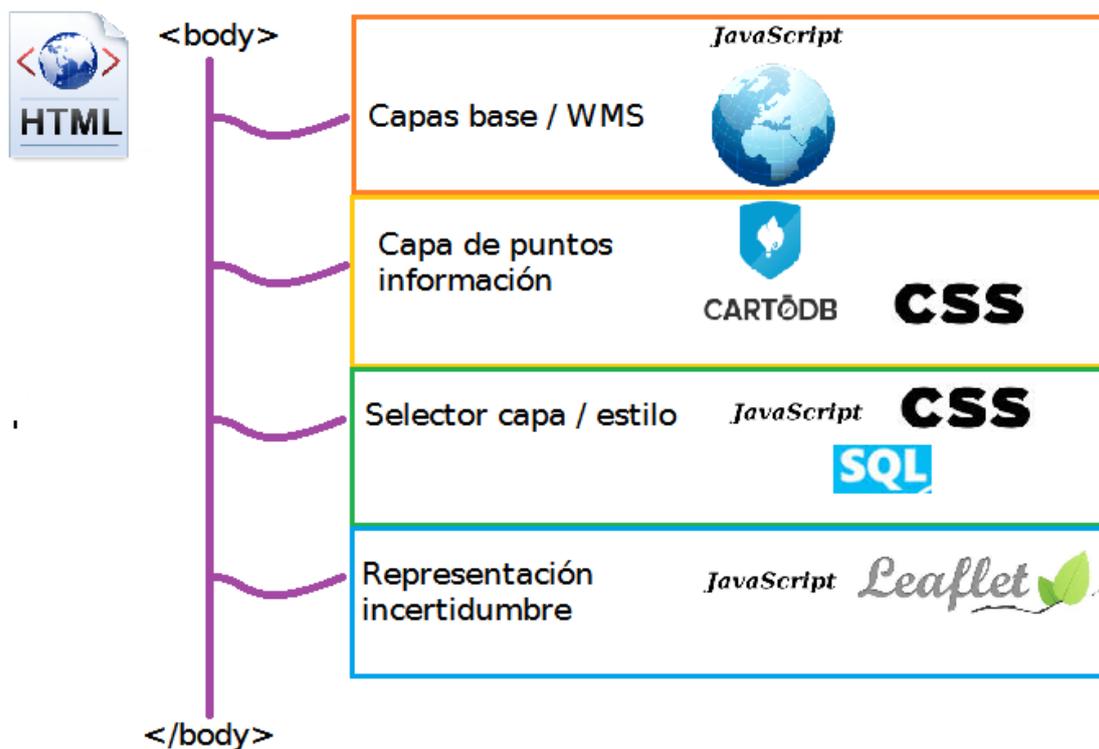
- Desarrollo del visor basándose en las librerías de las APIs de Leaflet y Cartodb javaScript.
- Diseño de la representación de los registros en la plataforma de Cartodb.
- Elección de capas temáticas que complementen la información (servicios WMS).
- Elección de las capas base cartográficas.
- Desarrollo de selectores de información programando en javaScript.
- Desarrollo de la herramienta para mostrar la incertidumbre.

En el siguiente resumen se muestra como está organizada la metodología empleada en el desarrollo de este apartado de representación.



Implementación

En la implementación del apartado de representación cartográfica se han usado lenguajes como JavaScript, CSS y HTML. A parte se han utilizado principalmente dos librerías para la creación del visor, Leaflet API y Cartodb JavaScript API. A esta herramienta creada para mostrar las cartografías propuestas se le han añadido funciones como: controlador de capas a mostrar, representación de la incertidumbre, obtener información de las capas WMS, zoom...entre otras funcionalidades.

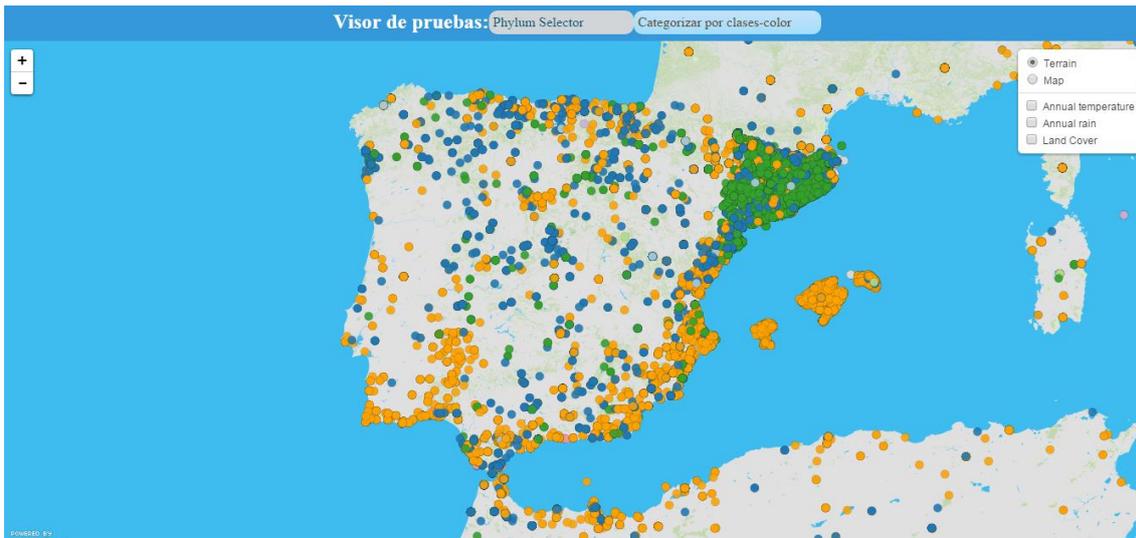


El visor cartográfico elaborado en este proyecto consta de varios elementos, los cuales cada uno ha sido implementado en diferentes lenguajes de programación. El lenguaje HTML ha sido el conector entre los diferentes lenguajes.

El mapa mostrado en el visor ha sido diseñado en lenguaje JavaScript, bajo la librería de Leaflet y Cartodb. Las capas bases están proveídas por el servicio web de OpenStreetMaps, éstas han sido relacionadas con capas representantes de variables ambientales, a través de servicios web WMS.

```
var b2 = L.tileLayer('http://{s}.tile.openstreetmap.org/{z}/{x}/{y}.png', {
  attribution: '&copy; <a href="http://www.openstreetmap.org/copyright" target="_blank">OpenStreetMap</a> contributors'
});
```

La capa de puntos, que representa los registros de muestras naturales del MCNB, está alojada en Cartodb. A través de este servicio ha sido generado el estilo de representación, además, Cartodb proporciona los recursos para poder servir toda la información que se genera en su plataforma. Gracias a esto, a través de la API de JavaScript de Cartodb ha sido posible insertar la capa en el mapa creado anteriormente con la librería de Leaflet.



Para darle mayor funcionalidad e interacción al visor se han implementado funciones de zoom, funciones de obtención de información de las capas WMS y controladores de estilo y contenido cartográfico. Estos selectores de contenido y estilo han sido implementados en JavaScript con conexión a las funciones de Cartodb mediante lenguaje SQL y CSS, para los estilos. Por ejemplo:

```

<div id="navegador" ><h2>Visor de pruebas:</h2>
  <div id="sql" class="layer_selector">
    <dt><a href="#"><span>Phylum Selector</span></a></dt>
    <dd>
      <ul class="hide">
        <li data=" WHERE phylum='Arthropoda'" data-type="sql">Arthropoda</li>
        <li data=" WHERE phylum='Annelida'" data-type="sql">Annelida</li>
        <li data=" WHERE phylum='Brachiopoda'" data-type="sql">Brachiopoda</li>
        <li data=" WHERE phylum='Chordata'" data-type="sql">Chordata</li>
        <li data=" WHERE phylum='Cnidaria'" data-type="sql">Cnidaria</li>
        <li data=" WHERE phylum='Echinodermata'" data-type="sql">Echinodermata</li>
        <li data=" WHERE phylum='Mollusca'" data-type="sql">Mollusca</li>
        <li data=" WHERE phylum='Nematoda'" data-type="sql">Nematoda</li>
        <li data=" WHERE phylum='Platyhelmintha'" data-type="sql">Platyhelmintha</li>
        <li data=" WHERE phylum='Tracheophyta'" data-type="sql">Tracheophyta</li>
        <li data=" WHERE phylum='OTHERS'" data-type="sql">OTHERS</li>
        <li data="" data-type="sql">Reset</li>
      </ul>
    </dd>
  </div>
</div>

```

Parte del código SQL se almacena en data, de manera que al hacer click en cada entrada de la lista se genera un código de SQL diferente, para cada filo animal.

Posteriormente, en la función createSelector se construyen y se realizan las peticiones SQL a la librería de Cartodb para JavaScript, obteniendo la información almacenada en data.

```

function createSelector(layer, incertidumbre) {
  incertidumbre = "";
  var condition = ""; // SQL or CartoCSS string
  var $options2 = $(".layer_selector").find("li");
  $options2.click(function(e) {
    var $li = $(e.target);
    var selected = $li.attr('data');
    var type = $li.data('type'); // 'sql' or 'cartocss'

    if (type === "cartocss") { // if a CartoCSS selector is chosen
      $options2.removeClass('cartocss_selected');
      if (selected !== "simple") {
        $li.addClass('cartocss_selected');
      }
    }

    condition = $('#'+selected).text();
    layer.setCartoCSS(condition);
  } else { // if a SQL selector is chosen, re-query the data
    $options2.removeClass('sql_selected');
    if (selected !== "") {
      $li.addClass('sql_selected');
    }

    layer.setSQL("SELECT * FROM " + tableName + selected);
  }
});

```

En cuanto a la incertidumbre, su implementación se ha desarrollado en JavaScript con la librería de Leaflet. Para representarla se dibuja un círculo cuyo centro es la posición en coordenadas geográficas y cuyo radio es la longitud de incertidumbre calculada. Para ello se ha desarrollado las siguientes líneas de código:

```
sublayer.on('featureClick', function(e, latlng, pos, data) {  
  
    if(selectedCircle) map.removeLayer(selectedCircle);  
    //Dibujo un círculo para la incertidumbre al pinchar sobre el punto  
    selectedCircle = L.circle(latlng, data.coordinateuncertaintyinmeters, {  
        color: 'red',  
        fillColor: '#f03',  
        fillOpacity: 0.5  
    }).addTo(map);  
});
```

Donde *latlng* es el centro del círculo (las coordenadas geográficas del registro) y *data.coordinateuncertaintyinmeters* es el radio calculado en el proceso de implementación de la incertidumbre anterior. Este último dato está almacenado en la base de datos, alojada en Cartodb, y mediante una conexión a través de su API JavaScript podemos obtener dicho valor.

En el anexo electrónico (CD-ROM) se muestra el código comentado paso a paso del visor cartográfico.

Resultados

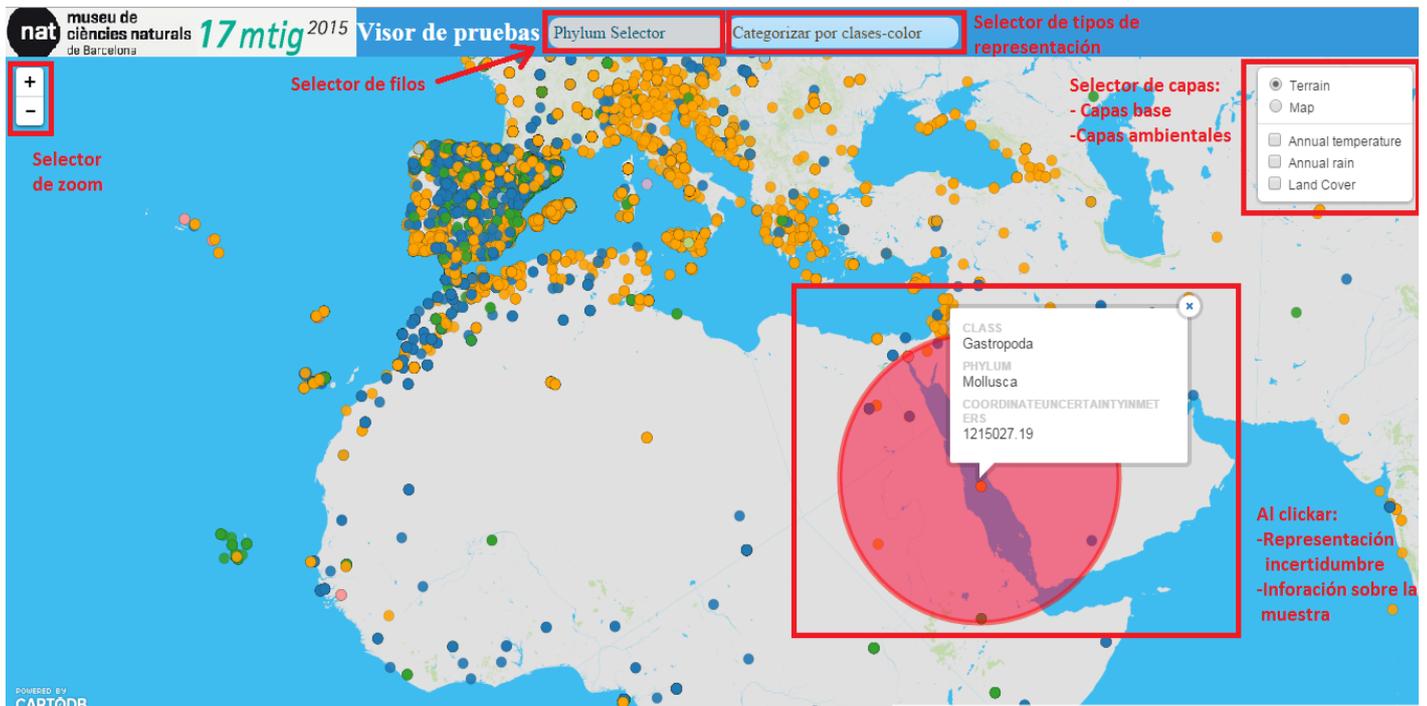
Como resultado se obtiene una cartografía que muestra puntos categorizados por tipos de filo, intercalando información con capas temáticas ambientales, como pueden ser capas de precipitación, cobertura del suelo y temperatura, y que al hacer click en cada punto se proporciona información específica del registro al que representa el punto. Para finalizar, simultáneamente, se dibuja un círculo que representa la incertidumbre asociada a la posición exacta del punto representado.

En resumen, el visor web muestra las propuestas de representación cartográfica para los datos que se han propuesto en este proyecto. De manera que muestra la información en diferentes representaciones:

- Categorizada por tipo de filo.
- Representación de la intensidad de datos en el área geográfica.
- Representación de la densidad de datos en el área geográfica.
- Cluster de datos.

Y proporciona herramientas de:

- Selección de datos.
- Representación de la incertidumbre.
- Información sobre las capas temáticas WMS.



Para la representación de los tipos de filos se ha optado por una diferenciación de colores, otorgándole un color diferente para cada filo. Además el visor muestra otra representación que va en función de la intensidad de muestras existentes en el ámbito geográfico.

Es posible ver la intensidad de muestras que hay para un solo tipo de filo, para ello hay que seleccionar el filo que se quiere representar a partir del selector de filos, y posteriormente, se elige la representación por intensidad.

Los resultados se pueden observar de manera interactiva en el siguiente enlace:

http://develtaxomap.bioexplora.cat/_mtig/index.html (también en anexo electrónico, CD-ROM)

Conclusiones

La irrupción del BIG DATA ha generado que los grandes volúmenes de datos sean una realidad. Han surgido desde numerosas fuentes, desde el uso masivo de las redes sociales a la eclosión de dispositivos que proporcionan grandes cantidades de información de todo tipo. Como pueden ser sensores (temperatura, humedad, caudal...), post en redes sociales, imágenes satélite digitales, posicionamientos GPS de los móviles...etc.

Detrás del fenómeno del BIG DATA se encuentra una respuesta tecnológica. Las herramientas tradicionales no eran capaces de absorber este nuevo escenario, por lo que resulta interesante un desarrollo de las herramientas en las Tecnologías de la Información y de la Comunicación (TIC), para poder mejorar las capacidades de recolección, almacenamiento, procesamiento y análisis de la información por parte de las entidades.

El proceso de georreferenciación masiva, que se propone en este proyecto, y su posterior metodología para la visualización de los datos del Museum de Ciencias Naturales de Barcelona, otorga a la entidad una facilitación en el proceso de digitalizar sus colecciones de muestras antiguas para lograr su preservación y, sobretodo, su difusión.

Se han propuesto mejoras en los interfaces de visualización. Los datos georreferenciados pueden ser visualizados con sistemas familiares como Google Maps, Google Earth o distintos servicios de mapas más utilizados en el momento.

Puesto que anteriormente, describir la dimensión espacial de las muestras mediante uno o varios topónimos resulta ambiguo y está destinado a la obsolescencia. Las coordenadas geográficas que se han generado para cada muestra aportan mayor precisión y facilidad de difusión, así como estabilidad a la recuperación de la información.

Este proyecto aporta una manera de llevar a cabo una georreferenciación masiva, utilizando un sistema en línea para asignar las coordenadas geográficas. Siendo un proceso simple, así la tarea de georreferenciar resulta asequible también para personal no formado en sistemas de información geográfica.

Para finalizar, los resultados obtenidos ponen de manifiesto que resulta ser un método muy válido para obtener dicha georreferenciación masiva, así como, información sobre la incertidumbre en la posición para cada registro. Siendo esto una novedad o extra en comparación con otros servicios de georreferenciación.

Referencias bibliográficas/web

- <http://www.economia3.com/2014/06/22/26988-big-data-una-tecnologia-disruptiva/>

- <http://www.rcg.cat/articles.php?id=257>

- [Leaflet](#)

- [Cartodb](#)

Anexos

-Ejemplo 1, JSON generado por OpenCage.

```
{
  "documentation" : "https://geocoder.opencagedata.com/api",
  "licenses" : [
    {
      "name" : "CC-BY-SA",
      "url" : "http://creativecommons.org/licenses/by-sa/3.0/"
    },
    {
      "name" : "ODbL",
      "url" : "http://opendatacommons.org/licenses/odbl/summary/"
    }
  ],
  "rate" : {
    "limit" : 2500,
    "remaining" : 2499,
    "reset" : 1454457600
  },
  "results" : [
    {
      "annotations" : {
        "DMS" : {
          "lat" : "36\u00b0 41' 28.57200'' N",
          "lng" : "5\u00b0 16' 11.29116'' W"
        },
        "MGRS" : "30STF9721263025",
        "Maidenhead" : "IM76iq75ov",
        "Mercator" : {
          "x" : -586631.798,
          "y" : 4370628.39
        },
        "OSM" : {
          "edit_url" :
"https://www.openstreetmap.org/edit?node=417107069#map=17/36.69127/-5.26980",
          "url" : "https://www.openstreetmap.org/?mlat=36.69127&mlon=-5.26980#map=17/36.69127/-5.26980"
        },
        "geohash" : "eys28p6whsqnbpj1tybu",
        "sun" : {
          "rise" : {
            "apparent" : 1454397720,
            "astronomical" : 1454392440,
            "civil" : 1454396100,
            "nautical" : 1454394240
          }
        }
      }
    }
  ]
}
```

```

    },
    "set" : {
      "apparent" : 1454435280,
      "astronomical" : 1454440560,
      "civil" : 1454436900,
      "nautical" : 1454438760
    }
  },
  "timezone" : {
    "name" : "Europe/Madrid",
    "now_in_dst" : 0,
    "offset_sec" : 3600,
    "offset_string" : 100,
    "short_name" : "CET"
  },
  "what3words" : {
    "words" : "annotate.positivity.noise"
  }
},
"bounds" : {
  "northeast" : {
    "lat" : 36.69132,
    "lng" : -5.2697531
  },
  "southwest" : {
    "lat" : 36.69122,
    "lng" : -5.2698531
  }
},
"components" : {
  "cave_entrance" : "Cueva de la Pileta",
  "country" : "Espa\u00fla",
  "country_code" : "es",
  "county" : "Provincia de M\u00ellaga",
  "road" : "MA-8401",
  "state" : "Andaluc\u00eda",
  "village" : "Benaoj\u00e9n"
},
"confidence" : 10,
"formatted" : "Cueva de la Pileta, MA-8401, Benaoj\u00e9n, Espa\u00fla",
"geometry" : {
  "lat" : 36.69127,
  "lng" : -5.2698031
}
},
{
  "annotations" : {
    "DMS" : {
      "lat" : "36\u00b0 43' 7.42800'' N",
      "lng" : "5\u00b0 15' 11.37600'' W"
    },
    "MGRS" : "30STF9877166037",
    "Maidenhead" : "IM76ir92ol",
    "Mercator" : {
      "x" : -584779.096,
      "y" : 4374424.772
    },
    "OSM" : {
      "url" : "https://www.openstreetmap.org/?mlat=36.71873&mlon=-5.25316#map=17/36.71873/-5.25316"
    },
    "geohash" : "eys2bkqy07rurpbumb8",
    "sun" : {
      "rise" : {
        "apparent" : 1454397720,
        "astronomical" : 1454392440,
        "civil" : 1454396100,
        "nautical" : 1454394240
      },
      "set" : {
        "apparent" : 1454435220,
        "astronomical" : 1454440560,
        "civil" : 1454436900,
        "nautical" : 1454438760
      }
    }
  },
  "timezone" : {

```

```

        "name" : "Europe/Madrid",
        "now_in_dst" : 0,
        "offset_sec" : 3600,
        "offset_string" : 100,
        "short_name" : "CET"
    },
    "what3words" : {
        "words" : "blog.thriving.crush"
    }
},
"components" : {
    "country" : "Spain",
    "county" : "Malaga",
    "local_administrative_area" : "Benaol\u00e1n",
    "point_of_interest" : "Pileta Caves",
    "state" : "Andalusia"
},
"confidence" : 10,
"formatted" : "Benaol\u00e1n, Pileta Caves, Malaga, Spain",
"geometry" : {
    "lat" : 36.71873,
    "lng" : -5.25316
}
}
],
"status" : {
    "code" : 200,
    "message" : "OK"
},
"stay_informed" : {
    "blog" : "http://blog.opencagedata.com",
    "twitter" : "https://twitter.com/opencagedata"
},
"thanks" : "For using an OpenCage Data API",
"timestamp" : {
    "created_http" : "Tue, 02 Feb 2016 11:36:23 GMT",
    "created_unix" : 1454412983
},
"total_results" : 2
}

```

-CODIGO DEL VISOR

Adjunto en el CD-ROM en la carpeta *Visor Cartográfico*