High-level geospatial information discovery and fusion for geocoded multimedia

Andrew Ennis, Liming Chen and Chris D. Nugent
University of Ulster, Belfast, UK, and
George Ioannidis and Alexandru Stan
IN2 Search Interfaces Development Ltd, Bremen, Germany

Abstract
Purpose – Improvements and portability of technologies and smart devices have enabled a rapid growth in the amount of user-generated media such as photographs and videos. Whilst various media generation and management systems exist, it still remains a challenge to discover the right information, for the right purpose. This paper aims to propose an approach to reverse geocoding by cross-referencing multiple geospatial data sources to enable the enrichment of media and therefore enable better organisation and searching of the media to create an overall picture about places.

Design/methodology/approach – The paper presents a system architecture that incorporates the proposed approach to aggregate several geospatial databases to enrich geo-tagged media with human readable information, which will further enable the goal of creating an overall picture about places.

Findings – Implementation of the proposed approach shows that a single geospatial data source does not contain enough information to accurately describe the high-level geospatial information for geocoded multimedia. However, fusing several geospatial data sources together enables richer, more accurate high-level geospatial information to be tagged to the geocoded multimedia.

Originality/value – The contribution in this paper shows that high-level geospatial information can be retrieved from many data sources and fused together to enrich geocoded multimedia which can facilitate better searching and retrieval of the multimedia.

Keywords GPS, Multimedia, Geospatial, Geotag, Reverse geocoding, Semantic enrichment

Paper type Research paper

1. Introduction
With the prevalence of the web, smart devices and the availability of communication infrastructures, it has become effortless to access data and information. An abundance of technologies are available for creating and distributing data with an ever growing diversity of formats and modalities. More and more information sources and tools are being made accessible to the public, e.g. information kiosks, emerging social networking applications and various business specific networks. As a result we are suffering from information deluge, yet it has become increasingly time-consuming and painstaking to discover the “right” information from the data. In many occasions we are starving for actionable information upon which an informed decision can be made. This is particularly the case for multimedia content, such as photographs and videos.

This work was funded by the Department of Employment and Learning, Northern Ireland and IN2 Search Interfaces Development Ltd. The authors would like to thank IN2 Search Interfaces Development Ltd for their support and funding in this research.
Multimedia content has recently undergone an explosive growth with the increasing availability and access of smart gadgets such as smart phones, tablets and smart digital cameras. Ofcom in the UK released results, in August 2013, showing that there are 82.7 million mobile phone subscriptions in the UK (Ofcom, 2013b). They also said that people in Northern Ireland are increasingly accessing the internet from their mobile device, with an increase of smartphone ownership up to 45 percent from last year. Gartner’s also predicts this trend, by saying that in 2013 mobile phones will outweigh PCs as the common device to access the web (Pettey, 2012). Gartner’s also predicts that over 80 percent of the handsets sold worldwide by 2015 will be smartphones (Pettey, 2012). The Office for National Statistics found that the main users of the internet in the UK are 16-24 year olds and the main activity that they carried out online was social networking, which is a key player in the generation and spread of user generated multimedia (Office for National Statistics, 2011). Ofcoms report also shows that 53 percent of people in Northern Ireland access and use social media, further showing the growing trend of user generated content (Ofcom, 2013a). These figures are a clear indication that more and more people have the technology available to them to generate their own content and share it with the rest of the world.

To feed this need for user generated content, there are a number of media generation and asset management systems currently available. These media asset management systems, such as Flickr (2013), YouTube (YouTube LLC, 2013) and Instagram (Facebook, 2013) allow users to upload their photographs and videos and share them with the rest of the world. The majority of these systems also provide an application programming interface (API) to allow third party systems to integrate with them. A growing trend in mobile devices is to have direct integration with the most popular media sharing systems such as Twitter being directly integrated into both iOS and Windows Phone 8. This allows the device user to upload their media directly from their mobile while on the move.

Nevertheless, this growing trend creates vast amounts of data and is overloading users with information. The biggest problem with information overload is determining what is relevant to the user and what is not. To determine what is relevant and what is not, we need to have detailed and useful metadata, to enable the categorization and searching. Unfortunately during media capture there is very limited information that can be attached to the media (Nandipati, 2011).

Generally when searching photographs, it is performed against the metadata that is attached to the photographs or the metadata that can be generated from analysing the photographs. Basic metadata such as the owners name, copyright and contact information, etc. and more recently, due to the increase of smart devices, Global positioning system (GPS) coordinate metadata can be used to search against (Zheng et al., 2010). It is also possible to derive metadata by analysing the image data to determine dominant colours, if the photo is indoors or outdoors, or to determine the image location based on the photometric effects of the sun casting shadows in the scene (Zheng et al., 2010). This metadata does not, however, provide semantic human understanding to the media. Therefore, enabling semantic searching and categorization of the media is not possible. To manually assign human understandable information to each and every piece of media would be a laborious and time consuming task and would lead to many errors.

In the last several years GPS embedded within media has become common place, mainly due to smart phones, tablets and smart digital cameras enabling more useful
metadata to be collected and attached to the multimedia at the point of media capture. This metadata is lacking in high-level semantics that are human understandable, thus providing fewer benefits than it could.

To address the aforementioned problem we propose to enhance reverse geocoding to extract high-level metadata, in a human readable format such as street name or place name, using the GPS coordinates and compass heading information embedded in the multimedia. This high-level metadata will then be used to facilitate better searching and categorization of the media. To achieve this we develop an approach to retrieve this high-level metadata from several data sources and attach it to the media. The approach uses cross-referencing of datasets to improve the accuracy and granularity of the metadata and to identify the subject in the multimedia. We use three public geospatial data sources as a reference to develop our approach. However, the approach could be implemented against any number of geospatial databases.

The remainder of this paper is presented as follows: Section 2 provides an overview of related work in the areas of extracting GPS coordinates from media, reverse geocoding to determine human readable information and current reverse geocoding platforms. Section 3 describes the proposed system architecture to support the proposed cross-referencing technique. Section 4 describes the developed technique for querying the geospatial databases and ranking the results in a meaningful order ready for the proposed cross-referencing technique. Section 5 describes the proposed technique of cross-referencing several geo-datasets to determine accurate human readable results that can be used to enrich media. Section 6 demonstrates a case study scenario to demonstrate the usage of our proposed cross-referencing technique. Section 6.2 discusses some foreseeable issues that the proposed technique might encounter. Section 7 discusses the conclusions and presents the potential areas of future work.

2. Related work
Geo-tagging (geocoding) commonly refers to attaching GPS coordinates to media. The GPS coordinates can be attached in different ways depending on the type of media (Valli and Hannay, 2010). Photographs are one of the most common types of media to be geo-tagged. Currently this is performed automatically by smart phones, tablets and smart digital cameras which are location aware (Valli and Hannay, 2010). The GPS coordinates are typically stored within images using the exchangeable image file format (EXIF) industry specification (Valli and Hannay, 2010). The EXIF can also store information about the device that captured the image and there are tools that can be used to display and use this information.

Reverse geocoding is referred to as the process of reversing GPS coordinates to a readable address or place name. In order to reverse geocode media the raw geospatial data such as the GPS coordinates are required. In addition, a reference geospatial database is required to facilitate the mapping of GPS coordinates to addresses or place names and the extraction of various high-level metadata. Given that most modern gadgets are location aware, the acquisition of GPS coordinates is not a major problem. With regards to geospatial databases, there are several widely used datasets available which are described as follows.

GeoNames is a geographical database that contains over 100 million geographic names, consisting of over 8 million unique features and is considered the most extensive gazetteer available (GeoNames, 2013; Gelernter and Mushegian, 2011; Luo et al., 2010;
DBpedia is a database of structured information that has been extracted from Wikipedia (Parundekar et al., 2010). The dataset contains a wide variety of information ranging from geospatial data, to people and music albums, etc. OpenStreetMap is an open, public map dataset that is generated through user contributions. This dataset contains large amounts of information relating to road ways, which we use in our cross-referencing technique (OpenStreetMaps, 2013).

Both the GeoNames and DBpedia databases have an ontology that represents the relationships between each of the objects, such as places or people, in the dataset. There has been previous work carried out to align and cross-reference these ontologies in order to infer information across the datasets. This ontology alignment is achieved by creating relationship links between the data, which are often expressed using “owl:sameAs” URIs (Parundekar et al., 2010). Parundekar et al. (2010) discusses how they use an algorithm that uses data analysis and statistical techniques to match two ontologies. Their algorithm uses sets to determine the instances that belong to each ontology. They use the class properties of the GeoNames and DBPedia ontologies, such as the featureCode and featureClass of the GeoNames database, to match against similar classes in the DBPedia database (Parundekar et al., 2010).

Reverse geocoding has been studied in previous research to enable enrichment of places. Alves et al. (2009) uses reverse geocoding to enrich a point on a map that only has a name and coordinates, by searching Yahoo and mining more human meaningful information from the resulting web sites. Their approach works well with a point of interest (POI) that has a specific name, such as a restaurant name, due to the granularity of the reverse geocoding only being to the city level. This is because enough of a geospatial area is covered to allow many search results to be returned and mined. A disadvantage of the method is that the granularity of the reverse geocoding is not enough to accurately mine relevant information for POIs that do not contain a precise name or any name at all, such as a simple geo-tagged photograph.

Lacerda et al. proposed an approach to improve the quality of geotags by propagating the geotags across similar photographs (Lacerda, 2013). Their research was primarily focused on personal photograph collections where some of the photographs have been tagged and others that are taken in the same region have not. They segment the images into temporal clusters based on the time difference between when the photographs were taken (Lacerda, 2013). When they get a photograph that is not geo-tagged they search for a photograph that is temporally close within the cluster and is geo-tagged (Lacerda, 2013). Their system then automatically propagates the geo-tag to the non-geo-tagged photograph which the user can verify. Once the system has propagated the geo-tag it then recursively repeats the procedure, so that the geo-tag can be propagated to other photographs. One of the issues they discovered is that if a photograph is taken on an airplane, train, or any vehicle moving at speed, the photographic process will have occurred over a considerable spatial distance, from the point the camera starts to take the photograph to when it is actually captured (Lacerda, 2013). So if the photograph is being clustered by GPS location it could be located in two areas at once, which is not feasible. Lacerda et al. proposed to capture the speed information using the GPS chip on a camera (Lacerda, 2013). They then used this information to determine if there was a geo-reference inconsistency between two photographs by calculating if the user reaches the location of the second photograph in the time it takes to capture the photograph (Lacerda, 2013). They validated their results by comparing the geo-tags that have been
propagated with the real location of the non-geo-referenced photographs. Based on this they were able to produce metrics for the precision of their system. Their results indicated that their system achieved good results with a precision of 97.08 percent and recall of 73.97 percent (Lacerda, 2013). Their future work proposed to apply new photograph clustering and machine learning techniques to improve their system further by enabling them to estimate better input parameters for their methods (Lacerda, 2013).

The “FollowThePlace” platform developed by IN2 Search Interfaces Development Ltd (2013) allows a user to upload photographs about places and add annotations that facilitate with searching and browsing the photographs of these places. “FollowThePlace” currently reverse geocodes the geo-tagged photographs, however, they only go to the city level granularity and not the precise location. This makes it difficult to create strong relationships between the media, to allow reasonable inferences to be made.

In summary, the challenges in this area are to retrieve and annotate multimedia with appropriate geospatial information in order to enrich and facilitate searching and retrieval of the multimedia. The general trends are that user generated multimedia content is on the rise and the need to categorise and organise this multimedia is a major area of focus. The work in this paper focuses on these issues to work towards an approach to enrich multimedia and therefore enabling better searching and retrieval of multimedia.

3. Proposed system architecture

We propose an overall system architecture for media harvesting, geospatial information retrieval and geospatially enriched multimedia search and display. The system architecture enables the extraction of media and its associated metadata, from many different sources, such as smart devices, social networks, the Web and more. As shown in Figure 1 the multimedia is extracted by the multimedia extractor.

![Figure 1. High-level system architecture of the proposed approach, showing the main components involved and how they interact](image-url)
The system extracts the metadata from the media and uses this to query several public geospatial databases using the GPS coordinates, to determine the high-level semantic location of the media, shown as the high-level metadata component in Figure 1. The system then takes the result from each geospatial database and cross-references them to disambiguate and improve the accuracy of the results. We model the metadata to facilitate querying and searching of the media based on the newly enriched metadata. The multiple view classifier categorization component in Figure 1 allows for the newly enriched media to be queried, searched and visually presented in various ways.

The details of this paper focus on the multimedia extractor, external database handler and high-level metadata components, shown in Figure 1. The other components of this architecture are not being considered at this moment because they will be the focus of future studies. These studies will focus on modelling and presenting the enriched multimedia.

4. Geospatial data characterisation and retrieval method

Each geospatial database contains different types of data relating to each place. For example, when searching GeoNames using the coordinates of 54.596992, −5.931239 (this is the north Western corner of Belfast City Hall grounds), we get a number of results. One result being a Boutique Hotel in Belfast. The attributes returned with this result are as follows:

- place name;
- latitude;
- longitude;
- feature type and feature description;
- country;
- administrative level;
- population;
- elevation; and
- time zone.

When DBpedia is searched we obtain some similar attributes such as name, latitude and longitude, however, we also obtain some different attributes returned. Some of these different attributes in the case of Belfast City Hall are:

- the architect who designed the building;
- the cost to build the building;
- who the current tenant of the building is; and
- the architectural style of the building.

The Google geocoding API provides the nearest address to the GPS coordinates along with the administrative levels of the address.

It is not, however, possible to obtain all of this information from just a single dataset, given that no single dataset contains all of this information. It is also impossible for a single dataset to manage and maintain all this information on its own, therefore it makes sense that various datasets focus on a specific area of information.
These varying datasets still need to be able to interlink so that the various types of information can be used. The information from the different datasets is also at different levels of abstraction, meaning that for example the architects name describes a very precise person, whereas the country describes a very large area.

If it is possible, all of this additional information should be annotated onto the media, as it can be useful for enabling better searching and retrieval. How useful each piece of information will be for enabling better searching will depend on what the user is searching for and what they are wanting to find out. For example: a user could search “Show me buildings in Belfast designed by Alfred Brumwell Thomas?” In this case the most important information would be the location information, so we can identify Belfast and also the architect that designed the building.

We query the databases by extracting the GPS coordinates from the media, for example the EXIF header in a photograph, and find all the POI points closest to the photograph. We also extract the compass heading from the media if it exists, which we use in the fusion process to weight and bias the POIs in the direction the photograph was taken.

Given a GPS coordinate, the first step is to query the database for all the POIs within a 500 meters radius of the GPS coordinate. This radius is incrementally increased if there are no POIs within the radius, until the radius becomes so large and it is clear that there are no related POIs in the vicinity. We assume that anything outside a radius of 1 mile can be considered not relevant and therefore we do not increment the search radius beyond 1 mile. A radius of 1 mile allows for places such as large parks or open areas to be considered. We implement the Spherical Law of Cosines formula for searching finding the nearest POIs from each database. We use Spherical Law of Cosines over the Haversine formula due to the faster computation speed, while maintaining accuracy.

The next step is to bias the results towards the POIs that lie closest along the line of sight, which is determined by the compass heading. This is undertaken by calculating the azimuth between the coordinates of the POIs and the coordinates of the media, to determine how closely they lie to the photographs line of sight. Figure 2 shows this in a graphical form, showing the results returned from GeoNames. The letters on the map pins show the order that each POI has been ranked based on the algorithm described above.

This algorithm will bias towards POIs that maybe far away from the media GPS coordinates, such as Pin B in our example in Figure 2. This is because they are closest to the line of sight of the photograph, compared to POIs that maybe very close to the photograph, however, are further form the photographs line of sight, or even behind the direction the photograph was taken. Therefore, in our example in Figure 2 pin B has a higher preference than pin D. Our reasoning for doing this is that if a photograph is taken in an open area, then the POI of focus in the photograph may in fact be further in the distance and is most likely to be the POI in the photograph. This may not, however, be the case in a built up area. In Section 5 high-level geospatial information fusion, we determine the context of each POI and the photograph.

We carry out this query for each geospatial database and we end up with a set of results from each, ordered by distance and azimuth from the multimedia location.

5. High-level geospatial information fusion through cross-referencing
Now that we have results from the various geospatial databases ordered by distance and degree from line of sight, we build a fusion model of the POIs so that we can
further refine the context of the multimedia and present a representation of the relationships between the POIs and the multimedia. This model will also represent the abstraction hierarchy of the POIs and their attributes.

The first step is to determine if any of the POIs, retrieved from the various databases, which lie close to each other are in fact the same POI. This is done by checking several of the POIs attributes.

Initially we check if either dataset has a reference to the other, such as a “sameAs” attribute. This is in the form of a “sameAs” attribute with a URI to the other POI. If it does have an associated “sameAs” attribute that links to the second POI, then we trust this and classify them as the same in the model. Otherwise we continue and check the other attributes to determine similarity. The procedure of checking the other attributes is explained below.

The process of checking the POIs is that we only compare a POI against its nearest neighbors. This improves performance as we are not comparing every POI, to every other POI, as there is no need to compare POIs that are distant from each other.

The first thing we check between two POIs is to see how close they are to each other. The distance threshold between the two POIs will vary depending on what feature type they have. If a place has a feature type of “statue” then it will have a relatively small spatial footprint and so any other POI that may possibly be the same would have to be very close. Whereas a POI that has a feature type of “building” then the footprint will be relatively large, so a POI from another dataset can be several meters away, yet still possibly be considered as the same POI.

We also compare the feature types of the POIs to ascertain if the POIs have the same feature type, such as both have a type of “building”. We also classify feature types such
as “building” and “architecturalStructure” the same, since an architectural structure is a building. We base our classifications on the DBpedia feature types ontology to determine what is classified as the same feature, using the hierarchy of feature types, therefore giving the relationships so we can classify an “architecturalStructure” the same as a “building”. This matching of feature types is achieved by mapping the feature types of each dataset to the feature types of the others.

The second thing we do is carry out a simple similarity match on the POI names to determine if they match. For this we implement the Levenshtein distance algorithm. We look for a relatively high percentage of similarity because they are names of places, so if they truly are the same POI, then they should have a very similar name. We aim for over an 80 percent similarity match on the place names.

Lastly, we determine if the two POIs are separated by a road or not. We do this by searching the OpenStreetMap dataset using the GPS coordinates of the two POIs. We then determine if there is a street polyline intersecting between the two POIs. In the case where a road does intersect between the two POIs, then we classify them as a different POI. In the case where no road is located between them, then the confidence percentage weighting that they are the same POI is increased. Our reasoning behind this is if a place is split between two buildings on opposite sides of a road, they are still two distinct buildings and so are two different POIs. They are, however, related to each other and maybe owned by the same company but physically they are different POIs.

We define two POIs as the same if the following conditions are met:

- the two POIs have matching feature types;
- POI name similarity is above 80 percent;
- the two POIs are in close proximity to each other; and
- the two POIs are not separated by a road.

Each of the checks on the POIs’ attributes add to the confidence percentage of how likely they are to be the same POI. In the end we have a confidence percentage of how likely the two POIs are to be the same.

Once we have fused the POIs into the fusion model and fused duplicate POIs from each dataset together, we can now use the feature attributes of the POIs to determine a semantic understanding of the POIs and therefore enable POIs to be eliminated that are not likely to be the subject in the media. An example would be if the photograph was taken in a built-up area, then it is likely that the subject in the photograph will be within a close range, as opposed to a photograph taken in an open area where the subject could be several hundreds of meters away, or further.

For each POI we derive some extra information that is used to determine the relationship between the POI and the photograph and between each POI.

GeoNames and DBpedia have feature attributes that describe what type the POI is. In the case of DBpedia this is stored in “rdf:type” and can contain elements such as “dbpedia-owl:Building” and “dbpedia-owl:ArchitecturalStructure”. GeoNames has feature classes and feature codes that together describe what type the POIs are. We can use these attributes to determine if the POI is a building and then combine this knowledge with the line of sight, we can determine that POIs located behind the building cannot be seen. Based on this derived knowledge we can eliminate those POIs located
behind as not being likely candidates for the photograph subject and so we increase the confidence weighting on the POI in the foreground.

We also use the GeoNames populated place feature code, if available, to determine if the area where the media was taken is a built up area. If this is the case then we can say that the photograph subject is likely to be close to where the media was taken from. Taking this into consideration we place more bias on the POIs closer to where the media was taken from.

To help improve performance we process the POIs in the order they were organized into from the previous retrieval algorithm. That is in order of distance and angle from the line of sight. By processing them in this order we can very quickly eliminate POIs blocked by other POIs that are situated in front, without having to process those POIs that are going to be eliminated.

After the initially retrieved list of POIs has been processed through our high-level geospatial information fusion and cross-referencing process, we end up with a reduced list that is ordered by a confidence percentage of how likely each POI is to be the subject in the media. This refined list has fewer POI’s than the initial list, due to the elimination process.

6. Case study
To demonstrate the usage of our approach, below we outline a case study using an example. For this case study we use a photograph taken of Belfast City Hall, as shown in Plate 1.

As previously mentioned our technique does not focus on the accuracy of the GPS coordinates or compass bearing. We assume that the GPS coordinates and compass bearing are accurate and precise.

6.1 Implementation
The first stage of the process is to extract the GPS coordinates and compass heading from the photograph. These are stored in the EXIF header and are easily extracted and

Plate 1.
Photograph of Belfast City Hall that was used for case study scenario
stored for use throughout the process. The GPS coordinates and compass heading retrieved from our example photograph are as follows: 54.596992, −5.931239, 125°.

The next step is to query the various geospatial databases for the nearest POIs to the photographs coordinates. In the case study we choose GeoNames and DBpedia as our example datasets. We also use the OpenStreetMaps dataset at a later stage in the process.

The GeoNames dataset was downloaded and populated into a MYSQL database to enable easy and timely querying of the POIs. A SQL query was constructed that used the Spherical Law of Cosines algorithm to determine the distances from the photograph to each POI in the GeoNames database. We found in our example, for the GeoNames database, above the 20th result the POIs are too far away to be considered relevant. So we limited the results for this scenario to the 20 closest, however, this will vary from place to place. The results are ordered by angle from the line of sight of the photograph and distance from the photograph. These results are presented in Table I.

The second dataset we use is DBpedia which we query through their SPARQL endpoint. Again we query for the closest POIs to the photographs coordinates. Due to the SPARQL query language we specify a limiting distance of 1 mile to initially restrict the POIs returned. There is also a limit on the endpoint to retrieving up to a maximum of 500 results at a time. The results are then sorted and ordered in the same manner that we do for the GeoNames POIs. That is ordering the results by angle from the line of sight of the photograph and distance from the photograph location, as presented in Table II.

The next stage is to determine if any of the POIs from the two datasets are the same. The first thing we look for is a “sameAs” attribute and if this attribute links to the POI in the other dataset. In the case where a “sameAs” attribute is found and it links to the

<table>
<thead>
<tr>
<th>GeoName id</th>
<th>Name</th>
<th>Distance</th>
<th>Nearest bearing</th>
</tr>
</thead>
<tbody>
<tr>
<td>7280023</td>
<td>Belfast City Centre</td>
<td>0.06629523</td>
<td>8.622299368</td>
</tr>
<tr>
<td>6691943</td>
<td>Belfast Central Railway Station</td>
<td>0.57550321</td>
<td>22.7322708</td>
</tr>
<tr>
<td>6466176</td>
<td>Hilton Belfast</td>
<td>0.49464916</td>
<td>27.04061263</td>
</tr>
<tr>
<td>6530001</td>
<td>Ten Square Boutique Hotel Belfast</td>
<td>0.08750617</td>
<td>35.24310177</td>
</tr>
<tr>
<td>6466849</td>
<td>Holiday Inn Belfast</td>
<td>0.29779547</td>
<td>47.36039854</td>
</tr>
<tr>
<td>7304564</td>
<td>Victoria Square</td>
<td>0.25781638</td>
<td>52.91972241</td>
</tr>
<tr>
<td>6478377</td>
<td>Europa Hotel</td>
<td>0.345907</td>
<td>59.90364881</td>
</tr>
<tr>
<td>6952104</td>
<td>Botanic Railway Station</td>
<td>0.50690098</td>
<td>61.00963389</td>
</tr>
<tr>
<td>6690632</td>
<td>Banjax Ltd</td>
<td>0.63424593</td>
<td>62.47968555</td>
</tr>
<tr>
<td>6467473</td>
<td>Crescent Townhouse</td>
<td>0.63713368</td>
<td>64.62532205</td>
</tr>
<tr>
<td>7117252</td>
<td>Malmaison</td>
<td>0.35603665</td>
<td>68.95407675</td>
</tr>
<tr>
<td>6945194</td>
<td>Hastings Europa Hotel</td>
<td>0.44704348</td>
<td>71.8713133</td>
</tr>
<tr>
<td>6465868</td>
<td>Jurys Inn Belfast</td>
<td>0.448461</td>
<td>72.45865258</td>
</tr>
<tr>
<td>6943929</td>
<td>Days Hotel Belfast</td>
<td>0.27642765</td>
<td>74.90435838</td>
</tr>
<tr>
<td>6944135</td>
<td>Travelodge Belfast Central</td>
<td>0.1411363</td>
<td>89.57381478</td>
</tr>
<tr>
<td>6952688</td>
<td>Great Victoria Street Railway Station</td>
<td>0.27780659</td>
<td>100.6279987</td>
</tr>
<tr>
<td>7117041</td>
<td>Fitzwilliam</td>
<td>0.18827845</td>
<td>107.6854066</td>
</tr>
<tr>
<td>6527879</td>
<td>Park Avenue Hotel</td>
<td>0.22135762</td>
<td>121.2211784</td>
</tr>
<tr>
<td>2653983</td>
<td>Belfast County Borough</td>
<td>0.22404093</td>
<td>213.066205</td>
</tr>
<tr>
<td>2638139</td>
<td>Shankill</td>
<td>0.22404093</td>
<td>213.066205</td>
</tr>
</tbody>
</table>

Table I. Top 20 results from GeoNames ordered by nearest bearing and distance from 54.596992, −5.931239, 125°.
other POI, then we trust this and assign the two POIs as the same. If a “sameAs” attribute is found but it does not link to the other POI then this new information is included into the model for future use. In the case where a “sameAs” attribute is not found then we analyse and compare several attributes from the two POIs and try to determine if they are the same.

First, check to ascertain if the two POIs are within a reasonable distance from each other to be considered the same place. The feature type is also looked at to help determine what a reasonable distance is. If the feature type is a building, then two POIs that are 20 meters apart, can be considered reasonably close and possibly the same POI. We do this by using an implementation of the Spherical Law of Cosines algorithm to calculate the distance between the two POIs. In the GeoNames results, most of the POIs are spread out, however, in the DBpedia results there are several POIs concentrated in the area around Belfast City Hall. One of which is labelled Belfast City Centre and represents the city centre area. This POI exists in both GeoNames and DBpedia and they are both located very close to each other and are marked as a populated area.

If this distance criterion is satisfied, then we check to ascertain if both POIs have the same feature type. In this case POI point “A, Belfast City Hall” and “D, Ten Square Boutique Hotel Belfast” shown in Figure 3, have a feature type of building, hence this increases the probability that they are the same place. In the case where the two POIs did not have the same feature type, where one POI is a building and the other is a statue,
then they would not be classified as the same place and further processing would not continue. The feature type similarity is done using the feature type ontologies of GeoNames and DBpedia and the alignment of them, so that an architectural structure and a building are classified as the same. We simply then just check to see if the feature types of each POI are classified as the same.

If the above criteria are satisfied, then check if there is a road intersecting between the two POIs. This is done by searching the OpenStreetMaps dataset using the GPS coordinates of both POIs to find the nearest street to them. This then gives the polyline coordinates of the street, which we check if it intersects between the two POIs. An example of this is POI “A, Belfast City Hall” and “D, Ten Square Boutique Hotel Belfast” shown in Figure 3. These places meet the criteria of being spatially close so we proceeded to see if the rest of the criteria were met. As shown in Figure 3 there is a road between the two, so a human can already determine that they are not the same place. For our algorithm to determine this, we have to query the OpenStreetMaps dataset to gauge what the closest road to the two POIs is and calculate if the road intersects between the two. The closest road in this case is “Donegall Square South”, which does in fact intersect between the two POI points.

We do not solely rely on the fact they are divided by a road, so the final check is the name similarity. So for our two potential POIs, POI “A, Belfast City Hall” and “D, Ten Square Boutique Hotel Belfast” shown in Figure 3, we use an implementation of the Levenshtein distance algorithm, which in this case gives a 24 percent match. This is considerably lower than our 80 percent limit, hence we can now be sure they are not the same place, based on the fact it has not matched the two sets of criteria mentioned above.

Figure 3. Shows the photograph location and two potential POIs that are identified as different locations due to the road intersecting between them.
In this case study example, our approach determines that POI A and POI B shown in Figure 3 are not the same place.

After the POIs from the geospatial datasets have been fused together, we then have a list of potential POIs, ranked in order of probability of being the subject in the media. The POI that is chosen for the media, is the POI with the highest probability. In our case study example, it is POI “A, Belfast City Hall”. This is correct and was manually verified.

From our initial case study example we have proven that our technique for improving reverse geocoding by cross-referencing multiple geospatial data sources works and provides a greater degree of accuracy than using a single data source.

6.2 Discussion
One of the biggest issues with reverse geocoding is not having enough or any geospatial information relating to the area. In our case study example, we found that in the area immediately surrounding Belfast City Hall there is still a large number of places that are not listed in either GeoNames or DBpedia. In an ideal situation every building and object would be listed. Our approach attempts to overcome this problem by enabling any number of geospatial databases to be cross-referenced, so that as much information and POIs are gathered as possible. The majority of geospatial information for all the buildings and objects in the example area around Belfast City Hall does exist. Nevertheless, a lot of it is either not in a machine readable format or is not accessible for public use. This limitation is improving through the various linked data initiatives, and the publishing of government data into the public domain.

Another issue is a picture taken from a rooftop which focuses on an object in the distance, but our approach would determine an associated POI that is close by. Nevertheless, in reality the photograph overlooks the POI.

This issue cannot be solved easily without further media information, for example altitude could be used to help determine this. The GPS system provides devices with altitude measurements, however, they are wildly inaccurate and could not provide the level of detail required. Also for this very reason many of the digital cameras do not include the altitude measurement in the media headers.

A major issue for our approach is inaccurate GPS coordinates. As previously stated, it is not our intension to improve GPS coordinates or develop methods to assist with the inaccuracy. Our approach assumes that the GPS coordinates are relatively accurate to within several meters. Our reasoning for assuming this is because the current GPS system is improving and accuracy is being greatly improved with the use of Assisted GPS (A-GPS) technologies, such as using cellular towers for triangulation and Wi-Fi hotspots. So our approach focuses beyond this on semantic enrichment of the media.

Another major issue is the lack of compass bearing data. Our approach can handle this situation by simply not ranking the POIs by the degree from the line of sight. They can, however, still be ranked by distance, which may be enough to determine the correct POI. Particularly in an area that is not densely populated with POIs.

7. Conclusion
This paper has proposed a generic approach to reverse geocoding by cross-referencing several geospatial data sources with the intension to facilitate semantic enrichment of multimedia.
We have developed a method to extract relevant POIs from several geospatial databases for a given piece of multimedia using the embedded GPS coordinates from the multimedia.

We have also developed a method to cross-reference and fuse the retrieved POIs from the geospatial databases. For this we proposed a fusion model to facilitate the cross-referencing process.

Our approach finds the most relevant POIs for the multimedia and from these POIs our goal is to extract the information from them and annotate it to the multimedia.

A case study scenario was discussed and from this study we found that we can indeed find the relevant POIs from several geospatial databases and fuse the results to determine the context of the multimedia. The ultimate goal is to extract and use the information contained within the POIs from the fused geospatial databases to semantically enrich the multimedia. From our initial research and findings we have found that we can use our approach as outlined in this paper to retrieve and extract semantic information which can be used to enrich the multimedia by annotating the multimedia.

The research is still in the early phases and the next stage of the future work is to implement the method proposed into a working prototype. Also with the collaboration with IN2 Search Interfaces Development Ltd we would like to integrate our system with their “FollowThePlace” platform. Our scenario set out how the approach works for a geo-tagged photograph, and our developed implementation would be able to handle other media types, such as videos.

Future areas that could be considered are the use of image analysis techniques to determine if a photograph was taken indoors or outdoors. This would help us to determine what kind of place the photograph was taken, for example: inside a restaurant, maybe of the food, or externally, in which case the photograph has a different meaning, it is of the restaurant, not within it. Knowing this information would also allow us to narrow down the number of potential POIs and know if distant POIs need to be considered or not.

References


**Corresponding author**

Andrew Ennis can be contacted at: ennis-a1@email.ulster.ac.uk

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com

Or visit our web site for further details: www.emeraldinsight.com/reprints