

WIKIPEDIA ENTRIES AS A SOURCE OF CAR NAVIGATION LANDMARKS

N. Binski, L. Zhang, S. Dalyot

Mapping and Geo-Information Engineering, Civil and Environmental Engineering Faculty, The Technion, Israel
(nbski, lijuan.zhang, dalyot@technion.ac.il)

Commission II, WG II/3

KEY WORDS: Navigation, Landmarks, Wikipedia, Survey Knowledge, Data Mining, VGI

ABSTRACT:

Car navigation system devices provide today with an easy and simple solution to the basic concept of reaching a destination. Although these systems usually achieve this goal, they still deliver a limited and poor sequence of instructions that do not consider the human nature of using landmarks during wayfinding. This research paper addresses the concept of enriching navigation route instructions by adding supplementary route information in the form of landmarks. We aim at using a contributed source of landmarks information, which is easy to access, available, show high update rate, and have a large scale of information. For this, Wikipedia was chosen, since it represents the world's largest free encyclopaedia that includes information about many spatial entities. A survey and classification of available landmarks is implemented, coupled with ranking algorithms based on the entries' categories and attributes. These are aimed at retrieving the most relevant landmark information required that are valuable for the enrichment of a specific navigation route. The paper will present this methodology, together with examples and results, showing the feasibility of using this concept and its potential of enriching navigation processes.

1. INTRODUCTION

Nowadays, with the increasing use of GNSS (Global Navigation Satellite System) in various applications, many people use automated navigation software to assist them with getting directions. These systems are mostly capable of dealing with giving route directions from point A (origin) to point B (destination), by giving a relatively small and simple set of instructions, which are mainly based on turn by turn navigation directions. Moreover, they usually rely on finding the shortest route that exists between points A and B according to the road network data, thus not dealing with the issue of "what is a good route direction" (Schwering et al., 2013).

To handle the topic of "what is a good route direction", we need to understand and separate between three different aspects: 1) how can one assess the quality of a good route direction: what aids the user (e.g., driver) to navigate successfully in a cognitive manner (i.e., visual aids, more informative audio instructions, turns, landmarks); 2) what criteria interest the user while using the navigating system: distance (the shortest route), time (the least time it takes), or other-interest (restaurants on the way); and, 3) how to collect and what sources to use for delivering with comprehensive set of valuable and instructive information, so the system can construct the most complete and useful navigation process.

Recent research and studies show that the human mind can be distracted while driving and listening to the verbal route instructions or concentrating on the map that appear on the system's screen (Hipp et al., 2010). Such that the navigators do not pay attention to their immediate surrounding environment. This means that the users do not construct a cognitive process during the navigation process; consecutively, they do not create a mental map of their area of interaction – a process that is immediate in case of navigating with a map and using landmarks for wayfinding. This translates to the fact that users feel lost or detached from their surroundings, which commonly leads to the outcome that they might find it hard to navigate the same route

again without using the navigation system, only based on their memory. Current research show that the wayfinding process is more efficient when involving landmarks with the instructions, i.e., relating spatially to existing features in the surroundings and thus enriching the users' survey knowledge, e.g., "you are passing the Eiffel Tower on your left", rather than giving 'dry' directions, such as "turn left" or "turn right" (Tom and Denis, 2003; Lovelace et al., 1999). In the past, when people were asked to describe how to get from one point to another, they used the most natural form of wayfinding for navigation – using and pointing to landmarks, which are recognizable features that stand out from their near surrounding. Although people today use navigation system devices for this purpose, this concept has not changed. Therefore, the use of landmarks in a navigation process should be considered when discussing issues relevant to the improvement of navigation processes that make use of navigation systems aimed at enriching the navigator's survey knowledge. Still, questions related to what is a valuable landmark that contributes to the overall objective, and to what immediate available data sources exist for this purpose need to be addressed.

It is becoming an acknowledged fact that contributed (volunteered) data can provide today with a rich source of complementary information when compared to authoritative data (Sester et al., 2014). Contributed data sources provide today large amounts of common and personal knowledge, and thus hold great potential for improving the quality of existing spatial databases, as in the case of navigation and Location Based Services (LBS). Such that these data sources can be used as alternative data sources for the retrieval of more informal information, such as vernacular places or local knowledge, which often is not represented in maps, although heavily used by people in their daily lives (Sester and Dalyot, 2015). Moreover, these data sources have the potential to provide timely and immediate information, because the update process is usually event based and continuous – instead of time based and cyclic, as in authoritative data.

This research paper focuses on the concept of enriching navigation systems route instructions by adding supplementary route information in the form of landmarks into the route direction process. This is designed to increase the cognitive awareness during navigation, enriching the navigator's survey knowledge. Here, we aim at using a contributed source of landmarks information, which is easy to access, available, have a high update rate and a large scale of information. For this, Wikipedia was chosen, since it represents the world's largest free encyclopaedia, including information about many spatial entities and vernacular places (Ponzetto and Strube, 2007). Due to the contributed nature of Wikipedia, and the fact that many spatial entities (e.g., landmarks) are geotagged (having latitude and longitude coordinates), one can refer to Wikipedia as a valuable complementary data source. This research paper will present landmark-related classification and ranking of Wikipedia entries, which are aimed at retrieving the most valuable landmarks for the enrichment task of navigation routes. The classification itself does not rely solely on the type or category of the landmark, but also uses a combined classification that is based on a list of attributes associated with the entries, which have no spatial context. The paper will present this methodology, together with examples and results, proving the effectiveness of the developed and implemented algorithms, and the potential of the proposed enrichment process.

2. RELATED WORK

2.1 Landmarks for Navigation Use

Landmarks play an important role in route direction and description. Navigation by landmark is the most natural form of navigation, and landmarks are found to be key components for constructing the representation of routes during navigation (Michon and Denis, 2001; Brenner and Elias, 2003). Instructions that are produced by navigation systems have the limitation of lacking interactions between the traveller and the environment; the main reason is that the route planning is based only on the geometric data derived from an existing route-network, thus they are communicated in a very limited form - a series of turns.

Research that deal with the issue of what is a good route direction and the role of landmarks on wayfinding concur the same conclusion: beside simple turn by turn verbal instructions, other visual aids, such as critical points along the route and landmarks can dramatically improve the chances of success in navigation, thus reducing the likelihood of getting lost (Michon and Denis, 2001; Lovelace et al., 1999; Elias, 2003; Raubal and Winter, 2002). Including landmarks in navigation instructions is investigated by many researchers. Hile et al. (2009), for example, developed landmark-based navigation method for outdoor pedestrian navigation. Goodman et al. (2005) investigated the use of landmarks in navigation for the elder. Duckham et al. (2010) suggest the generation of routing instructions that include references to landmarks using weighting model with the aim of enriching relatively simple routes.

However, if such information of landmarks are to be included in navigation systems, two challenges arise: 1) the automation of the process of defining and selecting good landmarks; and, 2) the route planning integrating landmarks to the road network. For the first, many studies try to identify the characteristics of good landmarks for navigation use. In the work of Burnett et al. (2001), the authors select useful landmarks based on the generic characteristics and significance of landmarks, including permanence, visibility, usefulness of location, uniqueness and brevity. Raubal and Winter (2002) qualify landmarks using their

visual attraction, such as facade area, colour, shape and visibility, semantic attraction that results from the cultural and historical importance and structural attraction of nodes and boundaries. Duckham et al. (2010) proposed a weighting system to calculate weights for landmark categories rather than specific instances of landmarks. Visual, semantical and structural characters are used to score the categories, deriving an overall suitability score to select good landmarks for navigation use.

Studies investigating the use of Volunteered Geographic Information (VGI) and user-generated data to infer the popularity of landmarks, such as Schlieder and Matya (2009), use the collection of photos from social web to measure the popularity of locations and landmarks of cities. Sun et al. (2013) employ geotagged photos on social media, such as Flickr, to infer users' preference on landmarks, integrating landmarks in route planning system to provide users the travel routings having most popular landmarks. Integrating landmarks in navigation process involves the spatial relations identification, as well as instruction generation. In Caduff and Timpf (2005), the landmarks-based navigation is done by selecting the referential landmarks based on distance, orientation and importance. Duckham et al. (2010) present a formal grammar notation to generate instructions for route directions using the general form "move - relative to - landmark" descriptions. Raubal and Winter (2002) also define a grammar for generating route instructions using landmarks.

2.2 Wikipedia

Wikipedia is a free online collaborative encyclopaedia existing in many languages, containing enormous textual and visual information, in which anyone can access and edit. Wikipedia is one of the most visited websites on the web, and it is possible to access and use Wikipedia pages in many applications. Wikipedia have attracted the attention of many scholars who aim to better understand the differences exist between Wikipedia entries and other sources of information, mainly authoritative. Lehmann et al. (2012), for example, have found that Wikipedia entries are stylistically indistinguishable from those found in traditional printed sources (i.e., encyclopaedias).

Due to Wikipedia's voluntarily nature, and the fact that many geographic entries have latitude and longitude coordinates (geotagged), one can also refer to Wikipedia as a VGI source, and use it to extract important information that might not exist elsewhere. Zirn et al. (2008) explain how to use Wikipedia templates and the inherent structure to extract meaningful information. Authors discuss the "Media Wiki Templates", which support sophisticated and organized template mechanism that contains different types of structured information, to create data-consistency. The main idea is that there is a template for any kind of content, such as: geographic entities (cities, rivers, mountain...), people (politicians, scientists, presidents...), organizations (companies, sports teams...) – to name a few. Wikipedia templates are also used in DBpedia (Lehmann et al., 2012). This is a community project, which extracts structured information from Wikipedia, allowing one to ask sophisticated queries from Wikipedia, to make it freely available on the web.

Consequently, it is only natural to aim and exploit this source of knowledge (Zirn et al., 2008). The main problem exists in extracting information from Wikipedia is the huge amount of data that can be retrieved, which makes it very hard to deal with; moreover, to be able and extract what is required and relevant, i.e., useful data and information. Therefore, the extraction of data from Wikipedia must include a procedure that sorts, classifies and selects the most relevant information for the process.

3. METHODOLOGY

3.1 What is Considered a Good Landmark?

To retrieve the most valuable landmarks, i.e., landmarks that will contribute best to the navigation process and help improving orientation in the environment, a study of understanding what is considered to be a good landmark from the user perspective was made. This study was based on existing research, as well as on a short survey with participants.

Burnett et al. (2000) state that to include richer instructions in the process of navigation, we must first understand what constitutes a 'good landmark', since the use of inappropriate landmarks might lead to driver confusion, and even reduce the usability of the navigation system. Authors aim to distinct between unique characteristics, and conducted an experiment with users that had to describe a route from origin to destination; for some of the users the route was familiar, and for others not. The purpose of this experiment was to see and understand which landmarks the users use to describe the route, and to pay attention to the way they describe these landmarks. Accordingly, authors took under consideration two important factors: 1) what made a participant to choose a specific landmark; and, 2) the textual concept of a landmark, meaning: how they choose to describe a specific landmark, and the differences between the descriptions for the same landmarks. These led to the establishment of landmark characteristics value, which was adopted here to correctly identify and retrieve valuable landmarks from Wikipedia entries. Figure 1 depicts the list of characteristics of the valued landmarks. The table also provides examples of different landmarks for each characteristic that typically score high or poor ranking based on its attributes.

We have decided to base our hierarchy and ranking of the different landmarks on this table, and use three characteristics: permanence, visibility, and uniqueness, which can be categorized from Wikipedia's entries, to identify the most relevant and important landmarks. Accordingly, each landmark receives a score as an attribute, from the less preferable (inferior) to the most preferable (superior). The following characterization is implemented during the selection process; a decision tree for each characterization is presented, which facilitate the subjective data classification (to some extent), a non-numeric one, in a quantitative way; at the end of every branch, the most probable rank is obtained, with values from 1 (inferior) to 5 (superior). As an example, Figure 2 depicts the decision tree for the Uniqueness character, with the different rankings retrieved for each branch.

Attributes of valued landmarks	Landmarks that score <i>highly</i> on the attribute	Landmarks that score <i>poorly</i> on the attribute
Permanence - the likelihood of the landmark being present, either in • Form (shape/size etc.) or • Label (name, logo, etc.)	Churches, Woods Monuments, Schools	Factories Shops, Petrol stations
Visibility - whether the landmark can be clearly seen in all conditions	24 hr Petrol Stations, Pedestrian crossings	Post-boxes, Street names
Usefulness of Location - whether the landmark is located close to navigational decision points	Traffic lights, Corner shops	Rivers, Railway lines
Uniqueness - the likelihood of the landmark <i>not</i> being mistaken for other objects/features, either due to: • A highly individual appearance • Being located apart from landmarks of same type	Bridges, Roundabouts Railway stations, Parks	Repairs garages Traffic lights
Brevity - the conciseness of description associated with a landmark	Traffic lights	Large white house on the left

Figure 1. List of characteristics of the valued landmarks with examples of different landmarks for each characteristic that typically score high or poor ranking (source: Burnett et al., 2000)

1. **Permanence** - indicates whether the landmark is a permanent object or a temporary one. A permanent object will receive a high score, while a temporary one will receive a low score. This characteristic is important since we must choose permanent landmarks, and tend not to use a temporal character (such as sign or form). This characteristic also contains the subject of a natural landmark, opposite an artificial landmark.

In the decision tree we examine: is it possible that the object will totally change? Meaning whether the object will change his nature, i.e. transform to another object; is the object natural or artificial? And, does the object tend to change its form, in terms of color, shape and even name?

2. **Visibility** - indicates whether a landmark is clearly visible (noticeable) in all conditions, and in relation to its surroundings. This category refers to what usually affects the visibility of an object, such as size and shape. In relation to size, tall objects can be seen more clearly, and as such will receive a high score; in relation to shape, the larger the object is (area) and complex, the higher score it will receive.

In the decision tree we examine: is it possible to notice the object from a great distance? Is the object tall? Is the object spread out on a large area? Is it possible to see the object in all possible environmental conditions (e.g., light and weather)?

3. **Uniqueness** - indicates the possibility that a certain object will be confused with other objects in its vicinity. An object will receive a high score if it has a distinct appearance or it is located apart (distant) from similar landmarks.

In the decision tree (depicted in Figure 2) we examine: does the object outstands in its environment? Is it noticeable? Is it possible to find the same or similar objects nearby? Is it easy to confuse the object with other objects?

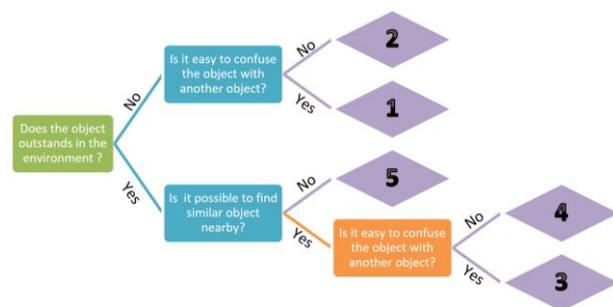


Figure 2. Decision tree for the Uniqueness character

3.2 Wikipedia Landmark Categories

The central goal of the category system in Wikipedia is to provide navigational links to all Wikipedia pages in a system of hierarchy for all categories, assisting readers to browse better and quickly in finding sets of pages on topics defined by these characteristics. Wikipedia is based on an enormous amount of categories, which are divided to sub-categories, which are also divided to sub-categories. Each entry has a list of relevant categories. This list of categories can be very long, whereas certain categories are not relevant for navigation. For example, the "Eiffel Tower" is obviously a very good landmark: the tower itself is very tall and can be seen from a great distance, and it is well-known; still, the category list of this page contains 11 different categories, where,

for example, one is “7tharrondissement of Paris”, which is not relevant to us. Therefore, landmark filtering must be conducted on the category level of all potential pages to extract only the ones that are necessary for our goal.

For this purpose, we have conducted a short survey, in which we have asked people to choose at least 5 landmarks out of 50 landmarks, i.e., serving as Wikipedia categories, that in their opinion will help them to orientate better in the surrounding environment (users could also add landmarks they thought are relevant that were not on the list). 160 people who own or drive a car have participated in this survey. From this survey, we have compiled a list of relevant categories, i.e., commonly used landmarks. Still, this list must be checked in respect to what we can actually extract from the Wikipedia database to ascertain that indeed all categories can be retrieved.

To achieve this, we have conducted a simple analysis, in which a central site was chosen to view what categories exist in its surroundings. These categories are then cross-referenced with the list from our survey. The site we have chosen is Central Park in New York, with a search radius of 1000 [m] in its vicinity; the assumption is that this site and radius should deliver us with a big enough list of categories existing in urban areas. We have retrieved more than 100 page IDs (entries) based on location only, and for each page ID we have extracted the list of all associated categories (a Wikipedia entry usually has numerous different categories). All categories extracted were compared to the list from our survey, and the combination of both, depicted in Table 1, created the list of common landmarks. This list serves for preliminary filtering in a way that only entries that contain a category from the common category list will continue to the next stage of the process.

List of common landmarks
restaurant, night club, coffee shop, pubs, roundabout, sculptures, squares, bus stations, shopping center, markets, hotels, schools, university, collage, towers, library, hospitals, parking lot, yard, court, mall, church, synagogues, museums, architecture structures, tall building, sky scrapers, airport, cemetery, highway roads, parks, railway stations, subway, natural landmarks, castles, fortress, historical sites, landmark, sea, rivers, lake

Table 1. List of common landmarks

3.3 Filtering and Ranking of Wikipedia Landmarks

Based on the information gained in the previous sections, every landmark in the “common landmark list” is given a rank based on the three characteristics (permanence, visibility and uniqueness). This rank is later combined with other parameters (as follows in the next section), and together we receive a final score, prioritized, for all possible landmarks. Table 2 depicts the rank given to each characterization according to its decision tree, summed up to a final rank based on all three characteristics. Since ranking here can be, to some extent, subjective, the ranking task was given to 10 different participants, whereas Table 2 depicts the average rankings. A normalization was then implemented on the final rank received, whereas each category is given a rank between the score of 1 and 10, depicted in Table 3.

3.4 Extraction of Landmarks and Page Properties

3.4.1 Defining of Queries: The extraction of Wikipedia entries based on location and category requires the development of query code. Queries are based on the entry’s location, information and data, and are divided into two main sections: 1) Category - the

extraction of an entry based on its category list. The specific category will point out the potential landmarks that could assist the driver in the navigation task; and, 2) Page information – the extraction of properties information of the entry that could help with the classification and ranking of landmarks to allow the retrieval of the most valuable landmark(s) for the navigation task. Many Wikipedia entries, specifically landmarks, are geo-tagged, meaning that latitude and longitude coordinates exist. Therefore, the preliminary extraction of landmarks’ entries is based on the entries’ coordinates. For the location-based query, we have used the Mediawiki tool, which is a free and very powerful scalable software, and a feature-rich wiki implementation that uses PHP to process and display data stored in the Wikipedia database. This tool uses MySQL, allowing one to query complex queries from Wikipedia, and thus the retrieval of information, such as: which pages or entries are placed nearby another entry, or what kind of information can be found for a specific entry.

Category	Permanence	Visibility	Uniqueness	Sum
restaurant	1	2	1	4
night club	1	2	1	4
coffee shop	1	2	1	4
pubs	1	2	1	4
mall	3	3	2	8
shopping center	3	3	2	8
markets	1	2	1	4
schools	3	3	1	7
university	3	3	3	9
collage	3	3	3	9
tall building	4	4	4	12
architecture structures	4	3	3	10
natural landmarks	2	2	5	9
historical sites	4	3	5	12
cemetery	4	3	5	12
museums	2	3	5	10
parks	3	3	5	11
highway roads	3	3	3	9
fortress	4	5	5	14
sculptures	2	1	4	7
roundabout	2	1	1	4
parking lot	1	3	2	6
railway stations	3	3	5	11
bus stations	1	1	1	3
airport	3	3	5	11
sea	5	5	5	15
rivers	5	5	5	15
lake	5	5	5	15
squares	3	1	1	5
church	4	2	4	10
towers	4	5	4	13
landmark	4	4	4	12
sky scrapers	4	5	5	14
synagogues	4	2	4	10
hospitals	3	3	4	10
library	1	2	3	6
subway	3	1	1	5
castles	4	5	5	14
hotels	1	4	3	8
yard	3	3	4	10
court	3	3	4	10

Table 2. Scoring of common landmarks based on the three characteristics

This process is implemented by defining a certain route from the origin point A to the destination point B. This route is then divided to segments by points in a distance of X [m] from each previous one (Figure 3). For each point in the segment, we use Wikimedia query to:

1. Find which entries are placed within a radius of Y[m] from the search point.
2. Extract the category list for a certain entry.

And then, to perform another filtering and classifying:

3. Extract page properties of a certain entry (section 3.4.4).

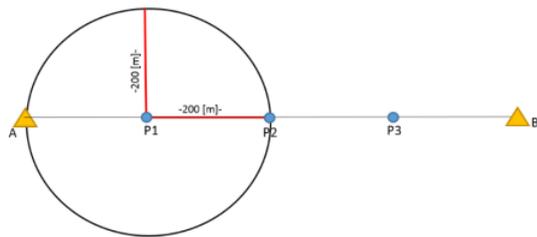


Figure 3. The route from an origine point A to a destination point B is segmented by intermediate points, and Wikipedia queries are performed on these points

Category	Final rank
bus stations	1
restaurant	2
night club	2
coffee shop	2
pubs	2
markets	2
roundabout	2
parking lot	3
squares	3
library	3
subway	3
schools	4
sculptures	4
mall	5
shopping center	5
hotels	5
university	6
collage	6
architecture structures	6
natural landmarks	6
museums	6
highway roads	6
church	6
synagogues	6
hospitals	6
yard	6
court	6
parks	7
railway stations	7
airport	7
tall building	8
historical sites	8
cemetery	8
landmark	8
fortress	9
towers	9
sky scrapers	9
castles	9
sea	10
rivers	10
lake	10

Table 3. Final score of common landmark list

3.4.2 Find which entries are placed in a radius of Y from the search points: Using the Wikipedia API one can retrieve the nearby wiki information, and a list of all pages and their location (coordinates). For example:

<https://en.wikipedia.org/w/api.php?action=query&list=geosearch&gsradius=200&gscoord=40.785119%7C-73.964973&gslimit=500&format=xml>

This query is based on a search radius of 200 m (*geosearch&gsradius=150*) around a point with geographic

coordinates with values of: *gscoord=40.785119%7C-73.964973*. The input of the coordinates is in a format of Decimal Degrees in WGS84 Datum, with negative values for South and West. This query result is a list of page IDs, which contain all the entries that are located in a radius of Y [m] from the search point. One can limit the number of the results: *gslimit=500* (here 500), and choose the format of the results: *format=xml*.

Using the query below, one can retrieve the wiki markup (content) of every page existing in Wikipedia database labelled with a page ID. The *action=query* module allows to retrieve information on a certain page and the data stored in it, such as the text of a particular page, the links and categories it relates to: *rvprop=content*. This query is based on a specific page ID, e.g., *pageids=1935441*.

<https://en.wikipedia.org/w/api.php?action=query&prop=revisions&rvprop=content&format=xml&pageids=1935441>

The linking of these two queries is simple, whereas the coordinates of a point in a route (e.g., in Figure 3, point P1) serves as input of the first query, while the output of this (first) query is a list of entry IDs that exist nearby (in the above example: radius of 200 [m]). Then, on the second query, every entry retrieved in the first query is checked according to the page ID. Accordingly, we retrieve all the text (information) exist on all entries in the vicinity, to extract the information that is relevant to us – page (entry) category and properties (see below).

3.4.3 Extract the category list for a certain entry: After retrieving the entries' text, we have implemented a Matlab code to perform parsing required to carry out the initial filtering – from a long list of page IDs, we find and choose the ones having the desired category that exists in the “common category list”. The Matlab pseudo code is as follows:

1. **Get** text from Wikipedia URL
2. **Split** text to lines by different signs
3. **For** each line: Search the word “Category”
If (word==‘Category’)
 Check all the words after the word category
If (word+1== ‘landmarks from the common category list’)
Add word+1 to cell final
Else move to the next line
4. **Repeat** until the end of the text

3.4.4 Extract page properties regarding a certain entry: Previous queries allow us to retrieve entries of desired categories in the vicinity of the route. Still, even for a fairly small search radius we might retrieve a high number of potential landmarks (categories), even when the landmarks rating system (presented in Table 3) is implemented. Communicating a high number of landmarks to the driver during navigation might lead to overload and the opposite output that we look for. Therefore, a successive filtering is implemented that is based on the entry attributes (other than the category alone). Such that the most valuable landmarks are chosen, e.g., 1 or maximum 2 landmarks for every section of the route (search point). For example:

<https://en.wikipedia.org/w/index.php?title=WLAN-FM&action=info#mw-pageinfo>

The query *action=info#mw-pageinfo* allows the retrieval of entry-related information associated with the entry, such as: article length (text-related), number of redirects (links) to this page, number of page viewers, and editing history details (date of

page creation, date of latest edit, total number of edits). It is assumed that these entry properties give a general indication to the importance and recency of a specific entry, and thus can serve as a supplementary filtering when required. The input of this query is the title of the page: *title=WFXN-FM*. A Matlab code was implemented to perform the parsing and retrieval of this information, with the pseudo code as follows:

1. **Get** text from Wikipedia URL
2. **Split** text to lines by different signs
3. **For** each line: **Search** the phrase “Number of page watchers” or “Page length” or “Date of latest edit” or “Total number of edits” or “Total number of distinct authors” or “Number of redirects to this page” or “Date of page creation”
If (phrase== one of the above)
 Add phrase +1 to cell final
 Else **move** to the next line
4. **Repeat** till the end of the text

4. EXPERIMENTAL RESULTS

In this section, we will present an examination of the proposed methodology. To do so, we choose a 2,700 [m] route in New York City, which by a quick look in Google Maps, it is clear that this area is surrounded by a variety of different landmarks covering many of the categories we are interested in. The chosen route is depicted in Figure 4.

The automatic retrieval process is implemented by defining a 150 [m] search radius, with 200 [m] route segments, resulting with 12 search points along the route. Only 3 of the 12 are presented here: P3, P5, and P12 (depicted in Figure 4). These values were chosen after an empirical analysis, which ascertained that most relevant landmarks for urban environments are retrieved, while avoiding the retrieval of distant and overlapping ones. It is assumed that for rural areas different values should be considered.



Figure 4. The route and area where the experiment is performed with P3, P5 and P12 search points (source: Google Maps)

An interesting fact is that a survey made manually on the chosen area showed that 83% of the landmarks appearing on Google Maps in the vicinity of the route have Wikipedia entries, where all were retrieved by our implementation. The rest of the 17% are mainly landmarks with categories that do not exist on our category list, mostly stores (Apple Store, for example). A total of 70 landmarks were retrieved from Wikipedia by our methodology, meaning Google Maps showed less than 20% of all available landmarks in the vicinity (12 landmarks were

identified manually). This proves the large amount of complementary landmark information that can be retrieved from Wikipedia.

Results for each search point are presented in the tables below, where main properties of retrieved entries are given, with ranks in different colors. For point P3 (depicted in Figure 4), 13 Wikipedia entries (landmarks) were retrieved after the initial filtering, depicted in Table 4. It is clear that the highest ranked landmarks are four skyscrapers (rank 9), one is a park (rank 7), and two are churches (rank 6). It is important to show that 6 out of the 13 landmarks retrieved are found to be irrelevant for our purpose, having very low ranks of 2 and 3. The four landmarks with the highest rank of 9 are too much to presented on such a short segment (200 [m]), such that the second filtering is implemented based on the entries’ attributes, depicted in Table 5.

Title	Category	Rank
Fifth Avenue / 53rd Street (IND Queens Boulevard Line)	Subway	3
Saint Thomas Church (Manhattan)	Church	6
650 Fifth Avenue	Skyscrapers	9
La Grenouille (restaurant)	Restaurants	2
Paley Park	Parks	7
21 Club	Clubs	2
Austrian Cultural Forum New York	Library	3
Olympic Tower	Skyscrapers	9
Donnell Library Center	Library	3
75 Rockefeller Plaza	Skyscrapers	9
St. Patrick's Cathedral (Manhattan)	Church	6
DuMont Building	Skyscrapers	9
Burger Heaven	Restaurants	2

Table 4. 13 Wikipedia entries (landmarks) with ranking scores for point P3 after the initial filtering

Numbers of views	Page length	Latest edit	Number of edit	Links	Creation date
Fewer than 30 watchers	7,831	26 October 2015	131	6	10 June 2006
Fewer than 30 watchers	29,781	8 December 2015	347	6	6 November 2003
Fewer than 30 watchers	3,892	31 October 2015	42	2	5 March 2009
Fewer than 30 watchers	11,658	14 November 2015	58	1	5 April 2009
Fewer than 30 watchers	4,457	17 November 2015	62	0	15 May 2006
Fewer than 30 watchers	14,550	6 November 2015	273	1	2 July 2005
Fewer than 30 watchers	5,615	31 May 2015	39	1	25 September 2008
Fewer than 30 watchers	4,570	16 November 2015	105	2	18 October 2005
Fewer than 30 watchers	7,327	16 July 2015	55	1	8 November 2007
Fewer than 30 watchers	2,024	13 May 2015	21	0	8 January 2012
68	27,755	16 November 2015	878	7	13 June 2003
Fewer than 30 watchers	5,343	12 August 2015	52	0	28 September 2008
Fewer than 30 watchers	3,012	27 July 2015	37	2	16 May 2010

Table 5. The Wikipedia page attributes of the 13 landmarks

From all landmarks ranked as 9, the "Olympic Tower" (8th row, in bold) shows the highest values for most attributes (parameters), including Creation date and Latest edit. Still, a closer inspection shows that "St. Patrick Cathedral" (with a rank of 6, 11th row) have attribute values that are higher, thus further development of the second filtering (and thus ranking) is still required.

For point P5 (depicted in Figure 4), 4 Wikipedia entries (landmarks) were retrieved after the initial filtering, depicted in Table 6. For P5 it is very easy to choose the preferable landmark, since its rank is significantly higher when compared to the other landmarks, proving the usefulness and feasibility of the proposed methodology.

Title	Category	Rank
Fred F. French Building	Skyscrapers	9
Philippine Center	Library	3
IDT Megabite Cafe	Restaurants	2
Penn Club of New York City	club	2

Table 6. 4 Wikipedia entries (landmarks) with ranking scores for point P5 after the initial filtering

For point P12 (depicted in Figure 4), 7 Wikipedia entries (landmarks) were retrieved after the initial filtering, depicted in Table 7. Again, as in P5, two landmarks are salient, with the rank

of 9 (skyscraper) and rank of 7 (park), thus it is easy to choose the preferable ones.

Title	Category	Rank
José Bonifácio de Andrada e Silva (Lima)	sculptures	4
1095 Avenue of the Americas	Skyscrapers	9
Josephine Shaw Lowell Memorial Fountain	sculptures	4
Benito Juárez (Crozco)	sculptures	4
Bryant Park	Parks	7
Johann Wolfgang von Goethe (Fischer)	sculptures	4
William Earl Dodge (Ward)	sculptures	4

Table 7.7 Wikipedia entries (landmarks) with ranking scores for point P12 after the initial filtering

In total, for the search points along the route presented here, 3 landmarks are retrieved (depicted in Figure 5), which are the most relevant ones according to the three characteristics and corresponding ranks (depicted in Table 3). Accordingly, these landmarks are the most valuable for the goal here, whereas all are easy to notice while driving and thus orientate to, serving as complementary and enriching information in the navigation process. Instead of simple two turn directions (depicted as arrows in Figure 5), the navigation is now complemented and enriched with the retrieved landmarks, which are communicated to the user, for example: “you are now passing the Olympic Tower on your left”, or “your destination point is nearby the 1095 Avenue of the Americas, which is on your left”. Also, it is clear that even for a very dense area as New York City, the preliminary classification and filtering suffice. Still, the second filtering that uses the entries’ attributes can ascertain the process to be more robust, mainly when many landmarks are retrieved and further filtering and classification is required, pointing – out the more valuable and significant landmarks.

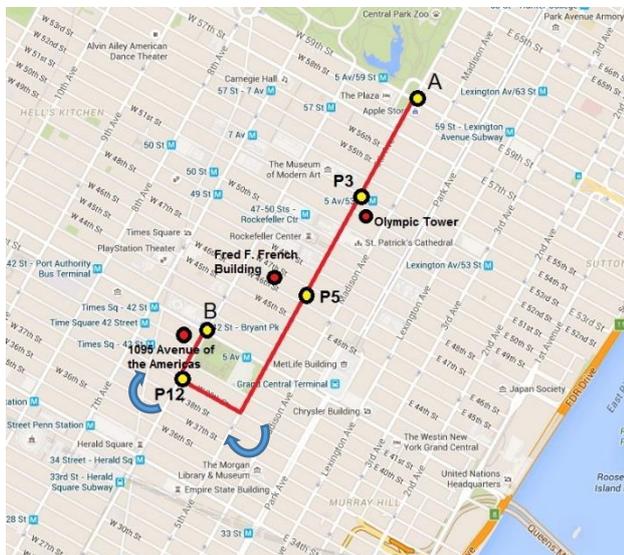


Figure 5. The route with search points P3, P5 and P12, along side the three landmarks that were retrieved for these search points (source: Google Maps)

5. CONCLUSIONS AND FUTURE WORK

A methodology for retrieving Wikipedia entries, which are assumed to represent spatial landmarks that are relevant and valuable information for navigation enrichment, was presented. The methodology exists of several classification and filtering processes, which use the category and characterization of the different landmarks, as well as their attributes. Ranking systems were devised, which proved to be robust for retrieving the landmarks that should contribute the most for the goal at hand.

An experiment was presented, which ascertained our methodology, and proved to be robust and enrichment valuable.

Future work is planned to carry out more experiments, with longer routes and on different environmental scenarios (e.g., rural areas). Future experiments will also include the testing of this system on navigation processes, where users will navigate with and without the retrieved landmarks, to assess the contribution of the landmarks on the overall navigation process, and improvement of orientation in the environment. This will also include an analysis on visibility factors, testing the importance of observable landmarks along the routes during the navigation and wayfinding process. Also, a more complex ranking system is planned for development on the entries’ attributes, which seems to be important in cases where there exist an excessive number of landmarks. Furthermore, our aim is to test our methodology on alternative volunteered sources, such as foursquare or OSM, which might add additional information that might not exist in Wikipedia. All in all we believe that this concept is important, since it serves as working ground to the integration of landmarks in navigation processes, as well as the use of volunteered and user-generated information for this process.

REFERENCES

- Brenner, C. and Elias, B., 2003. Extracting Landmarks For Car Navigation Systems Using Existing Gis Databases And Laser Scanning. ISPRS Archives, Vol. XXXIV, Part 3/W8, Munich, pp.17-19.
- Burnett, G., Smith, D., and May, A., 2001. Supporting the navigation task: Characteristics of ‘good’ landmarks. *Contemporary ergonomics*, 1, pp. 441-446.
- Caduff, D. and Timpf S., 2005. The Landmark Spider: Representing Landmark Knowledge for Wayfinding Tasks. In AAAI’05 Spring Symposium, pp. 30-35.
- Duckham, M., Winter, S. and Robinson, M., 2010. Including Landmarks in Routing Instructions. *Journal of Location Based Services*, 4(1), pp. 28-52.
- Elias, B., 2003. Extracting landmarks with data mining methods. In: Kuhn W, Worboys MF, Timpf S (eds) *Spatial Information Theory. Foundations of Geographic Information Science. Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp.375–389.
- Goodman, J., Brewster, S. A., and Gray, P., 2005. How can we best use landmarks to support older people in navigation? *Behaviour & Information Technology*, 24(1), pp. 3-20.
- Hile, H., Grzeszczuk, R., Liu, A., Vedantham, R., Kos̄ecká, J., and Borriello, G., 2009. Landmark-Based Pedestrian Navigation with Enhanced Spatial Reasoning. *Proceedings of the 7th International Conference on Pervasive Computing*, ser. *Pervasive ’09*. Springer, 2009, pp. 59–76.
- Hipp, M., Schaub, F., Kargl, F., Weber, M., 2010. Interaction weaknesses of personal navigation devices. In *Proceeding of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 129-136, Pittsburgh, Pennsylvania, 11-12 November 2010.
- Lehmann, J., Iseleg, R., Jakob, M., Jentsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Kleef, V.P., Auer, S.,

- Bizer, C., 2012. "DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia", *Semantic Web 1*, IOS Press, pp.1-5.
- Lovelace, K.L., Hegarty, M., and Montello, D.R., 1999. Elements of Good Route Directions in Familiar and Unfamiliar Environments. C. Freksa, D.M. Mark (Eds.): COSIT'99, LNCS 1661, pp. 65-82, Springer-Verlag Berlin Heidelberg.
- Michon P.E., Denis, M., 2001. When and Why Are Visual Landmarks Used in Giving Directions? D.R. Montello (Ed.): COSIT 2001, LNCS 2205, pp.292–305, Springer-Verlag Berlin Heidelberg.
- Ponzetto, P.S., and Strube, M., 2007. Deriving a Large Scale Taxonomy from Wikipedia. Proceedings of the 22nd national conference on Artificial intelligence, Volume 2, pp.1440-1445.
- Raubal, M., and Winter, S., 2002. Enriching wayfinding instructions with local landmarks. In: Egenhofer, MJ, David M (eds.) *Geographic Information Science, Lecture Notes in Computer Science*, Vol. 2478. Springer, Berlin, pp. 243-259.
- Schlieder, C. and Matyas, C., 2009. Photographing a city: an analysis of place concepts based on spatial choices. *Spatial Cognition & Computation*, 9(3), pp. 212-228.
- Schwering, A., Li, R., Anacta, V.J.A., 2013. Orientation Information In Different Forms Of Route Instructions, The 16th AGILE Conference.
- Sester, M., Arsanjani, J. J., Klammer, R., Burghardt, D., Haunert, J. H., 2014. Integrating and generalising volunteered geographic information. In: Burghardt, D., Duchene, C., Mackaness, W., (eds.) *Abstracting Geographic Information in a Data Rich World*. Springer International Publishing, pp. 119–155.
- Sester, M., and Dalyot, S., 2015. Enriching Navigation Instructions to Support the Formation of Mental Maps. In: Harvey, F.; Leung, Y. (eds.). *Advances in Spatial Data Handling and Analysis. Series: Advances in Geographic Information Science*, Springer, pp. 15-33.
- Sun, Y., Fan, H., Bakillah, M., and Zipf, A., 2013. Road-based travel recommendation using geo-tagged images. *Computers, Computers, Environment and Urban Systems*, 53, pp. 110-122.
- Tom, A., and Denis, M., 2003. Referring to Landmark or Street Information in Route Directions: What Difference Does It Make? In: W. Kuhn, M.F. Worboys, and S. Timpf (Eds.): COSIT 2003, LNCS 2825, Springer-Verlag Berlin Heidelberg 2003, pp.362–374.
- Zirn, C., Nastase, V., Strube, M., 2008. Distinguishing between Instances and Classes in the Wikipedia Taxonomy, *Lecture Notes in Computer Science*, Volume 5021, 2008, pp. 376-387.