International Journal of Geographical Information Science
Vol. 00, No. 00, Month 2009, 1–21

RESEARCH ARTICLE

Semantic Analysis of Research Trends in OpenStreetMap

Sukhjit Singh Sehra[a*], Jaiteg Singh[b], Hardeep Singh Rai[c], Sumeet Kaur Sehra[d]

[a,d] *Computer Science & Engineering*
[b] *School of Computer Science*
[c] *Department of Civil Engineering*;
[a,c,d] *Guru Nanak Dev Engineering College, Ludhiana, Punjab, India*
[b] *Chitkara Institute of Engineering, Rajpura, Punjab, India*

(*June 6, 2016*)

OpenStreetMap (OSM) has emerged as one of the key streams in mapping world. Given this interest in OSM and the rapidly increasing volume of OSM literature, it is important to understand and discern the core themes and trends emerging in OSM research, and its implications for broader research. There is previously reported studies to identify the core of OSM research. The manual studies would certainly prone to bias and interpretation of a limited number of reviewed papers. There would a lack of comprehensive, quantified and systematic classification of the OSM literature. This research brings some clarity by synthesizing and labeling a large corpus of OSM research studies published from 2007 through 2016. Latent Semantic Analysis (LSA), a natural language processing technique was applied to the abstracts of 485 academic papers. This objective analysis reveals twelve principal research areas. Various specific research themes associated with each principal area have been identified. These principal research areas and research themes indicate the patterns and trends in OSM research.

Keywords: OpenStreetMap, Crowdsourcing, Research Trends, Latent Semantic Analysis

*Corresponding author. Email: sukhjitsehra@gmail.com

ISSN: 1365-8816 print/ISSN 1362-3087 online

DOI: 10.1080/1365881YYxxxxxxxx
http://www.informaworld.com

1. Introduction

Web 2.0 encouraged greater collaboration among internet users and other users, content providers and enterprises (Hudson-Smith *et al.* 2009). This movement provided new methods of sharing and computing data (Hudson-Smith *et al.* 2009, Rana and Joliveau 2009, Turner 2006, Walsh 2008, Goodchild 2009, Haklay *et al.* 2008) by crowdsourcing movement similar to Wikipedia (Howe 2006). In regard to the geographical data the crowdsourced movement is known as VGI (Volunteered Geographic Information) and others call it collaborative mapping (Fischer 2008), so it is a special case of this web phenomenon and has been applied in many popular websites such as Wikimapia, OpenStreetMap, Google Map, Flickr (Kounadi 2009). The OpenStreetMap (OSM) project a prime example within the field of Volunteered Geographic data (VGI), providing free editable map of world, available under CC-BY-SA license.

OSM has emerged as one of the key streams in navigation and other areas and received a considerable amount of attention by researchers within the last decade, with a rapid increase in the number of related publications. The rapid increase in the volume of OSM literature poses a critical challenge in terms of identifying the research direction and trends, because despite the increasing volume of the literature, the core research and related themes remain unclear and poorly understood. Even there has been no such efforts to review the trends in academic OSM literature are typically qualitative, subjective and based on manual review.

Most of the research studies are dedicated to quality issues of OSM similar to (Neis *et al.* 2011). These studies put light onto issues related to OSM and provide an elaborate guidance, categorization and framework for implementation, adoption to the researchers.

The manual review, while insightful, is prone to be biased and limiting in terms of the number of articles that can be reviewed. In such manual reviews there is a potential tendency to read the more cited and influential papers, which while being a valid approach for review, might not be a comprehensive representation of the patterns in the literature. That is, it is important to note at the outset that we distinguish between critical review and patterns in literature. Instead of conducting a qualitative and subjective evaluation as to which themes are more important, we are interested in quantitative assessment based on textual analysis on how often the different themes have emerged in the the literature. The frequency and occurrence of themes can be assessed through semantic grouping of the keywords using techniques such as Latent Semantic Analysis (LSA), and such methods are well established as valid and useful methods to understand the patterns in literature.

Latent semantic analysis (LSA) is a text mining technique in natural language processing, in particular distributional semantics, of analysing relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. LSA assumes that words that are close in meaning will occur in similar pieces of text. The new low-dimensional space typically can be used to Compare the documents in the low-dimensional space (data clustering, document classification), Find similar documents across languages, after analyzing a base set of translated documents (cross language retrieval), Find relations between terms (synonymy and polysemy), Given a query of terms, translate it into the low-dimensional space, and find matching documents (information retrieval) and Find the best similarity between small groups of terms, in a semantic way.

As OSM data is being collected by the contributors, researchers should focus on devising methods for the analysis of such a huge dataset. In last eight years, a large number of articles have published concerning OSM. So this research brings a structured representation of the trends in research using OSM data by systematically applying LSA techniques to a large corpus of academic articles published from 2008 through 2016. This systematic analysis of 915 paper abstracts reveals principal research areas. In addition, ninety specific research themes associated with the twelve principal areas have been identified and clustered. These principal research areas and research themes comprise the comprehensive patterns of OSM research, which indicate the directions and trends in OSM research. The outcomes of analysis presented in this paper also provide opportunities for the researchers/professional to position their future research and/or implementation strategies.

The paper is organized into four section main sections. The first section discusses about OSM and second section describes the research data acquisition and the research methodology. This includes a detailed description of the data collection process, the LSA technique and its application to the collected data. The second section describes the results and findings from the analysis. Results are presented at various levels of details, ranging from twelve key principal research areas to ninety specific research themes. A cross analysis across the twelve principal research areas and ninety specific research themes is presented to identify their links. The third and final section concludes the paper with a brief discussion on the usefulness of the research findings, the limitations of the reported research, and the implications for future research.

2. About OpenStreetMap

OSM started in 2004, is a editable and freely available database of mapping information which can be used for many different purposes with very few restrictions (Neis and Zipf 2012). OSM has three main components (Haklay and Weber ????), which are Node, Way and Relation. Relation is most important element of OSM data structure, consisting of other members i.e. node and ways. Relation is a multi-purpose data structure that describes the logical relationship between elements. Any member of relation can have an optional role to describe the part it plays in a relation. All types of data elements (nodes, ways and relations) can have tags. Tags describe the meaning of the particular element to which they are attached.

OpenStreetMap, a revolutionary movement has encouraged nearly 1,742,729 as registered users, 247,0855,510 uploaded nodes, 246,912,924 ways and 2,738,864 relations. The contribution of new data to the OSM data can be accomplished using different approaches by users of varying backgrounds of mapping experiences and using different devices.The most common approach is recording map data using a Global Positioning System (GPS) receiver and edit the recorded information using one of the various freely available editors such as JOSM (Steiniger and Hunter ????). The users provide additional information about the collected data by adding attributes and store the final results in the OSM database. Users do not require any specialized GPS receiver for mapping, as it has been tested that smartphones can be considered as device for mapping (Sukhjit Singh Sehra *et al.* 2013, Golicher ????). The accuracy of GPS has been checked and found satisfactory, but the accuracy depends upon the quality of GPS chip. In addition if the accuracy of any smartphone is not good as compared to the professional GPS receiver, the mapper can map to good accuracy as Microsoft Bing supports the project (Neis and Zipf 2012) by providing various aerial images as background layer. It allows the OSM users to digitize data such as streets from the traces very easily, but for completeness of

Table 1. Criteria for paper search and collection

Sr. No.	Steps	Number of Papers
1	Online/Bibliographic Databases Search	2889
2	Filtered for particular Search phrases/Within databases/	877
3	Elimination of Duplications	165
4	Elimination of non-focussed and unwanted	227

the attributes local knowledge is still required. The only problem using satellite imagery is that it can be outdated. Other web based tools are Polatch 2 and iD editor, but only registered users can upload the changes to OSM.

3. Research Data Acquisition

The research data comprised of academic journals and conference papers collected through multiple academic databases and sources. Conference papers are in general less rigorously reviewed compared to the journal articles. However, they are included in the corpus, because instead of comparing or evaluating the strength or quality of studies, this paper aims to reflect the interest of research community on OSM.

There are three steps applied to collect and finalize the data. Initially the articles were identified using the search phrase OpenStreetMap (OSM), "Voluteered Geographic Information (VGI)", "Crowdsourced Map" and "Callaborative Mapping" in their titles and/or abstracts and/or keywords. Spatial Data related textbooks, trade magazines, brochures and product/software related white papers were not considered, because the analysis focused on research publications. Since most research reports and theses result in academic articles, they were also not considered to prevent duplication in the corpus. Initially 2887 paper abstracts were collected in the corpus. Open-Source tool Jabref is used for collection, screening, selection and corpus preparation.

> "JabRef is an open source bibliography reference manager. The native file format used by JabRef is BibTeX, the standard LaTeX bibliography format."

JabRef provides functionality to download the various bibliographic databases using Application Program Interface (API). Bibliographic database searched are IEEE, ScienceDirect, DBLP, Arxiv, DOAJ, ACM, Citeseerx. In addition to it Taylor and Francis, Wiley Bibliographic Database, MPDI journals database, Zotero repository and Medlay Repository are also searched. In this step, the queried papers were manually screened to decide whether it is relevant to include in the corpus. If the searched keyword exists in an example, citation or acknowledgement, the related paper was not added to the corpus. Table 1 summarizes the document collection process.

Large number of abstracts (documents) collected through different sources were reviewed to eliminate duplication in the final step. After the final processing, 485 documents were saved in the research corpus that includes 19 papers which are converted from German, Spanish and Italian language to English. **??** and **??** summarize the top sources of the included documents.

As the data used by the algorithm is .csv, which includes titles and abstracts, an export filter for JabRef was exclusively designed and available at `https://sukhjitsehra@`

Table 2. Top 20 Journals

Journal Name	Number of Related Publication
ISPRS International Journal of Geo-Information	32
Transactions in GIS	24
International Journal of Geographical Information Science	22
The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences	14
GeoJournal	12
Computers, Environment and Urban Systems	10
Future Internet	7
International Journal of Digital Earth	7
Crowdsourcing Geographic Knowledge	5
OpenStreetMap in GIScience	5
Progress in Location-Based Services	5
Advances in Spatial Data Handling and Analysis	4
Applied Geography	4
Applied Geomatics	4
Remote Sensing	4
Sensors	4
The Canadian Geographer	4
Geoinformatics FCE CTU	3
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	3
International Journal of Applied Earth Observation and Geoinformation	3

`bitbucket.org/sukhjitsehra/topic-modelling.git`.

4. Methodology

4.1. *Data Analysis*

The OSM literature dataset (corpus) is subjected to Latent Semantic Analysis (LSA), a data/text mining technique used to facilitate retrieving and querying large corpus of data [23], [24] and [25]. LSA has been taken as it outperform other approachs for topic modelling such as pLSA, LDA etc. As a mathematical and statistical method, LSA is used to identify the latent concepts within the textual data at the semantic level [26]. LSA employs a set of algorithms to convert unstructured text into structured data objects, and analyse these data objects to identify patterns for the discovery of knowledge [27]. The main idea behind LSA is to collect all the contexts belonging to the words in the corpus, and derive associated factors that represent related concepts. In this paper, a factor can be defined as a latent class representing multiple observed entities which have similar patterns that associated with the latent class. For example, in a 12 factor analysis, all the relevant entities that were found in the corpus relevant to the analysis are classified or represented through 12 latent classes, such that each of the entities in corpus is associated with or shows similar patterns to one of the 12 latent classes. Thus, each factor reserves and represents a certain amount of the overall observed entities, and the factors are organized in the order of how many entities they explain. A sample representation of factors and their associated entities are presented in Table A1 for information science discipline in [25]. A similar example is also presented in 3 for a

Table 3. Five Topic Based Term loading Tokens

Topics	Topic Label	Loading Terms
T5.1	Contribution and Analysis	activ contribut pattern commun edit region volunt social event flow dataset evolut behavior effort assess
T5.3	Assessment Models	semant similar heterogen qualiti tag trust assess attribut classif comparison measur shape polygon traffic pattren
T5.3	Mobile Applications	mobil queri interfac visual haptic assist interact simul user travel devic ontologi googl lexic technologi
T5.4	Routing and Navigation	traffic simul network road algorithm vehicl sensor swarm citi urban model light scenario sumo transport
T5.5	Indoor Routing	indoor land transport build public rout plan floor footprint germani outdoor evacu smartphon geodata

five-factor solution to articulate the relationship between factors and corresponding high-loading terms. In any text, multiple words may share the same meaning and one word may have many synonyms in different contexts. LSA loads the words that share the same meaning to their associated concept and also loads one word to various latent semantics other than its main associated concept LSA differs from traditional factor analysis by applying Singular Value Decomposition (SVD) to reduce the dimension of the original data. This helps to present the collected 912 papers under the categories aggregated variously. Using SVD, LSA generates two sets of loadings in matrix format, one for the terms and one for the documents. Each factor in the matrices is associated with a high-loading term or a high-loading document, where each factor represents a research theme in the corpus. Higher term/document loading values for each factor indicate a greater possibility that the related term/document discloses certain theme. The researcher can alter the level of detail for identifying the research themes by changing the number of factors in the analysis. Lower level factor aggregation represents common research areas, and higher level factor aggregation represents principal research themes [25].

The steps in applying LSA to the corpus is similar to previously reported studies [24], [25] and [28]. These steps include the following.

4.2. *Preprocessing*

Preprocessing phase involve the identification and then elimination of the non-textual and irrelevant content i.e. removal of noisy vague parts like tags, slang words, URLs etc. Most of the words contained in the text do not effect its general orientation. Keeping such words only contributes in making the problem multi-dimensional and thus classification task becomes more hard as every word occurring in the text is represented along one dimension. The following steps have been incorporated during preprocessing:

The following series of text-mining procedures is applied to the corpus using gensim, a python library for scalable semantics:

(1) Abstracts are tokenized with non-letter separators such that each abstract is aggregated to singular words that make the abstract an individual word-bag;
(2) all the letters in the word bags are transformed to lowercase;
(3) erasing punctuation characters of terms, include the period, exclamation point, comma, apostrophe, question mark, quotation mark and hyphen;
(4) Filtering the numbers so as to contain only textual terms;
(5) applied N char filtering, to filter all those terms that consist of words with less than the pre-specified number of N characters. Here N is taken as 3, so all the terms having only one or two characters will be filtered out;
(6) Stop words of English such as and, the, and so, are pre-defined in nltk python package (In addition, the keywords (OpenStreetMap, Geographic, Voluteered, information, Crowdsource, maps and OSM) which were used during document collection; and such complementary words are custom defined to remove from the word-bags.);
(7) The tokens of the words that are less than two letters are removed, because these tokens are grammatical necessity but do not add any meaning to the words;
(8) term stemming techniques are applied to the word-bags, whereby the variations of the words are removed, evaluated with their grammatical roots and considered as a token (For example the words communicate, communication, communicative are transformed to a single token communica-.);
(9) The last step of text mining is applying n-gram model. N-gram is a contiguous sequence of text or speech, which creates word groups that are considered as an individual token in the word-bag.

4.3. *Term-Document Matrix Preparation*

One more step is added to text processing, corpus was further refined to remove terms that exist only once in a document. These terms are local to a particular document. Initially dataset had 87348 tokens. But after removal of undesired words, n char words, numbers, encoded characters, English and custom stop-words, 4978 unique tokens were generated, after applying “SnowballC Stemmer” algorithm, only 2510 terms were left. A 2510 485 term-document matrix was generated.

For converting term-documents matrix to intergar format only, a document representation called bag-of-words is used. Here each document is represented by one vector where each vector element represents a question-answer pair, in the style of:

“How many times does the word system appear in the document? Once.”

It is advantageous to represent the questions only by their (integer) ids. The mapping between the questions and ids is called a dictionary. This dictionary is used to convert the documents to a vector form, which consists of counts the number of occurrences of each distinct word, converts the word to its integer word id and returns the result as a sparse vector. The sparse vector for each document of the corpus e.g. [(0, 1), (3, 2)], represents that term (id0) apears once and term(id3) appeared teo times in that document.

4.4. *Term frequency-Inverse document frequency (TF-IDF)*

The 2510 485 raw td matrix was transformed with weighting and normalization methods. Term frequency (tf) and inverse document frequency (idf) values of each entity were calculated consecutively [24], [30] and [31]. The tfidf value of each entity represents how important that entity is to a document collection or corpus. The tfidf values increase with the increase in the occurrence of the entity. However, as a counterbalance to this increase, the frequency of the word in the corpus helps to account for the fact that some entities are more common across the documents than the others [32]. The tfidf, which is transformed from the raw td matrix, was then subjected to LSI for SVD is heartt, to decompose it to term-to-concept matrix, document-to-concept matrix and square roots of eigenvalues (singular values), as concept strengths appearing in descending order. Multiplication of term-to-concept and singular value matrixes gives the factor loadings of terms for each research pattern, and the multiplication of document-to-concept matrix and singular value matrix gives the factor loadings of documents for each research pattern. The maximum number of factors generated in this way is equal to the number of documents (in this study, 485). To assess the principal research areas and the broader research themes, the number of factors is adjusted by reducing the dimension of factor loading matrix. In this study, the principal research areas are identified with the 3,5,10 factor solutions, while detailed research themes are explored with the 50 factor solution.

4.5. *Rotation of factor loadings and factor threshold values for research areas/themes*

In factor analysis, the rotation of the matrix displays the loadings of variable on factors and simplifies the factor associations. Entities with high loading value in each factor associate with a theme. Through rotation of factor axes, the loading of an entity on one factor is maximized while minimizing its loading on other factors. Thereafter, interpretation of the factor structure becomes easier. In other words, associated high-loading terms and documents of each research pattern are clustered and matched. Varimax rotation [33] was applied to the term and document loadings. To interpret both the term and document factors in the same semantic space, the same rotation matrix that was applied to term loadings was applied to the document loadings as well.

Aggregation level varies for each factor solution. While the number of entities/documents loaded in the two or three-factor solution is high, this number decreases for the ninety-factor solutions. To parse the significant and non-significant entities/documents, an appropriate threshold value should be selected for each factor solution. In traditional factor analysis, fixed threshold values are used to determine the significant variances. These values can be 0.3, 0.4 or 0.5. In LSA, fixed threshold values are not appropriate, because in LSA calculation is performed based on the covariance matrix instead of correlation matrix [25]. To determine the threshold value, a heuristic assumption based on empirical probability distribution was applied. This assumption states: “each article should relate on average, to one topic, therefore, each document should load, on average, on one factor” [28]. For example to define the threshold values of documents for different factor solutions, the loading values of documents in each factor are transformed to a vector from its matrix form; and sorted in descending order. The tail distribution of this vector is generated to define the threshold. Although the distribution graph is quite dense to interpret, for n factor solution, the threshold is defined by retaining the 1/n of the loadings. According to this assumption, threshold values for the 3, 5, 10 and 50 factor solutions are calculated as 0.0936, 0.0655, 0.0874 and 0.183 respectively for document loadings. Documents with factor loading below the related threshold value were eliminated.

4.6. *Factor evaluation and research theme labeling*

In each factor solution, high-loading entities and documents within the calculated threshold value are grouped with the absolute values of the loadings in descending order. High-loading entities and documents are reviewed together and labeled by the authors individually for convergence for all the factors (2 through 8, 12 and 90 factors). To analyse the evolution of the research areas and themes, the factor solutions and high-loading documents are grouped by beginning and ending of years 2007-2012 and 2013-2016.

5. Results and findings

5.1. *Summary of factor solutions*

Table 4 summarizes the principal research areas at various levels of detail, with the two-factor solution providing clustering of the research areas at the broadest level. The principal research areas are identified with 3, 5, and 10 factors. The number of papers loaded for each factor, grouped according to years, is listed alongside. Factors are numbered and represented as Fi.j, which refers to the jth factor of i-factor solution. For example F5.4 refers to 4th factor of the 5-factor solution.

After an iterative process, the 5-factor solution was considered optimal to represent the principal research areas. The number of articles associated with each factor shows the importance of the corresponding research area within the corresponding factor. New research areas were derived during the interpretation of factor solutions, and same research areas can be seen in multiple factors. For example "assessment and analyis" appears across (F3.1) to (F10.1), but the number of high-loading articles reduces accordingly. The popularity of each research area is relative to the factor solution.

Mapping the factor solutions aggregated at different levels shows connections across the different OSM research areas. To detail this mapping across broader and specific research themes, the findings from the 50-factor solution are presented in Table 5.

Since the 5-factor solution was deemed optimal for principal research areas, and 50-factor solution gives more specific research themes, a cross-loading analysis was performed to establish the interrelation between the 5-factor and 50-factor solutions. Cross-loading indicates how strongly an item in a low-aggregated factor solution, loads on two or more factors on another highly-aggregated factor solution. [34] refer 0.32 as the minimum loading value (threshold) for two or more factors to be considered in cross-loading.

5.2. *The principal OSM research areas*

As shown in Table 4, at the broadest level, the principal osm research areas are clustered around Quality Assessment and Analysis (T3.1), Routing and Navigation (T3.2) and Haptic and Diaster Management (T3.3) (Figure 1). These articles emphasise on development of methods for quality assessment of crowdsourced data, its applications and issues pertaining to crowdsourced data.

Similarly, the principal research areas to emerge from the results of 5-topic solution

Table 4. Core Research Areas for OpenStreetMap

Topic No.	Topic Label	2007-2014	2007-2011	2012-2016
T3.1	Quality Assessment and Analysis	365	62	303
T3.2	Routing and Navigation	21	5	16
T3.3	Haptic and Diaster Management	4	2	2
T5.1	Quality Assessment and Analysis	329	57	272
T5.2	Quality, Completeness and Assessment of Contributors Behaviour	4	1	3
T5.3	Applications to Disaster, Navigation and others	12	7	5
T5.4	Traffic Simulation and Mobility	16	0	16
T5.5	Indoor Navigation Models	8	0	8
T10.1	Quality Assessment and Analysis	400	78	322
T10.2	Quality, Completeness and Assessment of Contributors Behaviour	9	1	8
T10.3	Algorithmic Model Development	23	0	23
T10.4	Model for Haptics in navigation assistance	1	1	0
T10.5	Semantic Similarity	18	3	15
T10.6	Contributors and Trust in Data	3	1	2
T10.7	Indoor Navigation Models	15	0	15
T10.8	Street Network Models	7	1	6
T10.9	Disaster Management	5	2	3
T10.10	Routing and Navigation	4	3	1

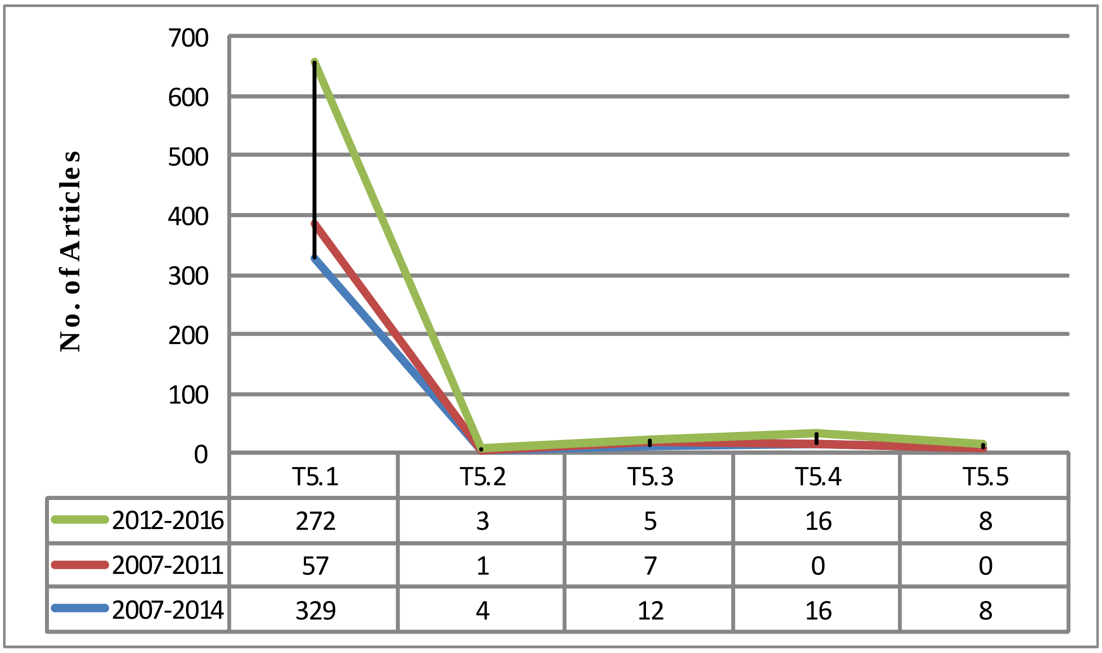

Figure 1. Paper Distribution for 5-topic solution

the Quality Assessment and Analysis (T5.1) remains the strong area of research, Other areas emerged here are Quality, Completeness and Assessment of Contributors Behaviour (T5.2), Applications to Disaster, Navigation and others (T5.3), Traffic Simulation and Management(T5.4) and Indoor Navigation Models (T5.5). Selected top five high-loading articles for each factor of the 12-factor solution, are presented in Table 7.

The 10-factor solution further expands the principal research areas. Two trending areas emerged during 10-topic solutions are "semantic similarity" (T10.5) and "Indoor Navigation models" (T10.7). It has concluded that the "Quality Assessment and Analysis" (T10.1) is clearly dominated area in all the other topic solutions also, 2012-2016 are

Table 5. Research Themes for OpenStreetMap

Topic No.	Topic Label	2007-2014	2007-2011	2012-2016
T50.1	Quality Assessment and Analysis	185	39	146
T50.2	Mobile based Routing	7	2	5
T50.3	Mobile based Serives	1	1	0
T50.4	Traffic Simulation and Management	8	0	8
T50.5	Indoor Planning and Simulation	2	0	2
T50.6	Land-use Patterns	13	0	13
T50.7	Road Matching Algorithm	4	0	4
T50.8	Smart Cities and Mobility	6	0	6
T50.9	Haptic for Navigation	5	3	2
T50.10	Street Networks	8	2	6
T50.11	OSM for Routing	6	1	5
T50.12	Wireless Sensor Networks	4	0	4
T50.13	Disaster Management	5	2	3
T50.14	Reducing Travel Times	1	0	1
T50.15	Assesment of Building Models	2	0	2
T50.16	Shortest Path Computation	6	0	6
T50.17	Automated Comparison	1	0	1
T50.18	Location Based Services	3	1	2
T50.19	NLP on OSM	1	0	1
T50.20	Data Extraction from OSM	1	0	0
T50.21	Semantic Annotations	2	0	2
T50.22	Diverisity of OSM Road Network	1	0	1
T50.23	Trust In OSM Data	3	0	3
T50.24	Digital Maps	1	0	1
T50.25	Estimating Buliding Types	1	0	1
T50.26	Contributors Patterns	1	0	1
T50.27	Generative models and Maps for Blinds	2	0	2
T50.28	OSM for Pervasive applications	4	2	2
T50.29	Conflation of Maps	1	0	1
T50.30	Managing OSM Tags	4	1	3
T50.31	Web Mapping Interface	5	0	5
T50.32	Humanitarian Efforts	2	0	2
T50.33	Mining Approaches for OSM Data	4	0	4
T50.34	Evacuation Modeling	2	0	2
T50.35	Scale of Features	1	0	1
T50.36	Matching POIs	2	0	2
T50.37	Integration and Tranformation	1	0	1
T50.38	Crowd Sourced Map	2	1	1
T50.39	Autonomous robot navigation	3	1	2
T50.40	Real Time Tracking	3	2	1
T50.41	Flood Modelling	3	0	3
T50.42	Computer Vision Based	4	1	3
T50.43	Geocoding Models	2	0	2
T50.44	Traffic Generator	1	0	1
T50.45	Event Related Mining	3	0	3
T50.46	Mapping APIs	1	0	1
T50.47	Approach for improving situational awareness	2	0	2
T50.48	Automatic Recommendation OSM Categories	1	0	1
T50.49	Turn Restrcitions	1	0	1
T50.50	Polygon Inspection in XML Data	1	1	0

Table 6. My caption

Topic No.	Labels	High-Loading Papers	Loading
T5.1	Quality Assessment and Analysis	(Ciepuch *et al.* 2011)	0.359507539
		(Al-Bakri and Fairbairn 2012)	0.330378263
		(Sachdeva 2015)	0.314754342
		(Brovelli *et al.* 2016)	0.311835763
		(Neis *et al.* 2013)	0.300522386
T5.2	Quality, Completeness and Assessment of Contributors Behaviour	(Bgin *et al.* 2013)	0.321112335
		(Keler and de Groot 2013b)	0.286861888
		(Keler and de Groot 2013a)	0.280018661
		(Arsanjani *et al.* 2013)	0.25761061
		(Yang *et al.* 2016)	0.195301071
T5.3	Applications to Disaster, Navigation and others	(Jacob *et al.* 2011)	0.25913736
		(Jacob *et al.* 2009)	0.239955536
		(Jacob *et al.* 2010)	0.181300088
		(Zook *et al.* 2010)	0.16549283
		(Kaklanis *et al.* 2013)	0.142287804
T5.4	Traffic Simulation and Mobility	(Dallmeyer *et al.* 2014)	0.287994729
		(Stolfi and Alba 2013)	0.282757478
		(Stolfi and Alba 2015a)	0.2748784
		(Su *et al.* 2014)	0.273285008
		(Backfrieder *et al.* 2013)	0.251780649
T5,5	Indoor Navigation Models	(Goetz and Zipf 2013)	0.319311775
		(Czogalla and Naumann 2015)	0.313969706
		(Czogalla 2015)	0.304831225
		(Goetz and Zipf 2012)	0.221645035
		(Nikander *et al.* 2013)	0.22064829

the years, when it is focussed by the researchers. While these research areas increased their popularity through 20142015, the assessing the behaviour of contributors "Quality, Completeness and Assessment of Contributors Behaviour" (T10.2) also become trending area of reearch.

As It has been discussed that OSM is outcome of crowdsourced mapping, and data is available for researchers to analyse and devising methods to correct and assess the data. The topic "Quality Assessment and Analysis" (T5.1) has been most widely worked as shown in fig 1. [WRITE A PARAGRAPH ABOUT ANALYSIS]

5.3. *Detailed OSM research themes*

The 50-topic analysis, revealed detailed OSM research themes as seen in Table 5 with number of papers loaded for particular topic solution. The distribution of high-loading articles in Table 5 indicates that "Quality Assessment and Analysis" (T50.1) remains the most popular research theme in the 50-topic analysis, consistent with the findings in the 5, 10 topic analysis.

The five most emerging areas of research are "Land-use Patterns" (T50.6),"Traffic Simulation and Management" (T50.4), "Street Networks" (T50.10), "Smart Cities and Mobility" (T50.8), "Shortest Path Computation" (T50.16), "Mobile based Routing" (T50.2). Topic "Land-use Patterns" (T50.6) has for the years 2007-2011 ha no contribution, but see jump during 2012-2016, same as for "Traffic Simulation and Management" (T50.4). As shown in Table 5. The researchers have contributed and used OSM data at large during 2012-2015. "Mobile based Routing" (T50.2) and "OSM for Routing" (T50.11) are hot areas where OSM has been extensively used but only after necessary preprocessing. There are many applications, Tools and APIs that use OSM data to provide online and offline route planning. The few highly loaded paper for these topics are addressing Indoor-Outdoor navigation using mobiles (Czogalla and Naumann 2015, Nikander *et al.* 2013), Issues (Graser *et al.* 2015) and execution (Hultman *et al.* 2016)

Another significant work is in the field of "Disaster Management" (T50.13), where OSM is a part of the critical infrastructure behind disaster management efforts. OSM is heavily used by disaster response organisations, ranging from the World Bank to university research centers and NGOs. This topic is also trending and uses infrastructure data mapped in OSM to run preparedness analysis and impact modelling. Highly Loaded articles focussed on OSM for disaster management (Zook *et al.* 2010, Eckle and de Albuquerque 2015, Rahman *et al.* 2012)

The topic "Haptic for Navigation" (T50.9) uncovers the use of Haptic tools for navigations, the highly loaded papers concentrates on such tools and exploration of street network (Jacob *et al.* 2009, 2011, 2010, 2012, Kaklanis *et al.* 2013). With increasing OSM data usage and the expansion of contributor base capabilities, "Contributor Pattern" (T50.26) remains another popular theme. Related high-loading paper considered after excluding cross-loading focuses on Co-editing Patterns OSM Contributors (Mooney and Corcoran 2014).

The 50-topic solution reveals research themes, but the papers loaded for each topic is single approximately, this is because that research oriented analysis of OSM is very recent. The important research topic uncovered using 50-topic solution are focusing on "Location Based Services"(T50.18), "Autonomous robot navigation" (T50.39), "Conflation of Maps" (T50.29) and "Mobile based Serives" (T50.3).

6. Mapping of principal research areas and research themes

In this section, a semantic link between the papers, the 12 principal research areas and the 90 research themes is established via crossloading of high-loading papers. Links are presented with the selected high-loading papers in Table 7

6.1. *Quality Assessment and Analysis(T5.1)*

"Quality Assessment and Analysis" (T5.1) topic is most trending and areas, which the OSM data is being used for analysis purposes, model creation etc. As in Table 7, it is visible that 38 topics in 50-topic solutions, emerged from this highly important T5.1 topic, includes "Land-use Patterns" (T50.6) (Arsanjani and Vaz 2015, Estima and Painho 2013), "Road Matching Algorithm" (T50.7) (Will 2014), "Street Networks"(T50.10) (Jilani *et al.* 2016), "Wireless Sensor Networks" (T50.12) (Mehdi *et al.* 2014), "Assesment of Building Models" (T50.15) (Yuan and Cheriyadat 2016, Fan *et al.* 2014), "Shortest Path Computation" (T50.16) (Aridhi *et al.* 2015, 2014), "Managing OSM Tags" (T50.30) (Codescu *et al.* 2014, Jilani *et al.* 2014) are areas emerged in recent years.

In conclusion quality assessment and analysis of OSM data is prominent area of research. The topics are related to the assessment and analysis of OSM in many aspects.

6.2. *Quality, Completeness and Assessment of Contributors Behaviour*

Crowdsourced data heterogeneity is a direct consequence of the collaborative processes used to produce these maps. But users are of different backgrounds and have varying level of mapping experience. This makes generated map data more vulnerable to errors and incomplete. OSM uses loose coordination and no top-down quality assurance processes.

Table 7.

Topic No.	5-Topic Labels	50-Topics No.	50-Topic Labels
T5.1	Quality Assessment and Analysis	T50.1	Quality Assessment and Analysis
		T50.6	Land-use Patterns
		T50.7	Road Matching Algorithm
		T50.10	Street Networks
		T50.11	OSM for Routing
		T50.12	Wireless Sensor Networks
		T50.15	Assesment of Building Models
		T50.16	Shortest Path Computation
		T50.17	Automated Comparison
		T50.18	Location Based Services
		T50.19	NLP on OSM
		T50.20	Data Extraction from OSM
		T50.21	Semantic Annotations
		T50.22	Diverisity of OSM Road Network
		T50.24	Digital Maps
		T50.25	Estimating Buliding Types
		T50.27	Generative models and Maps for Blinds
		T50.28	OSM for Pervasive applications
		T50.29	Conflation of Maps
		T50.30	Managing OSM Tags
		T50.31	Web Mapping Interface
		T50.33	Mining Approaches for OSM Data
		T50.35	Scale of Features
		T50.36	Matching POIs
		T50.37	Integration and Tranformation
		T50.38	Crowd Sourced Map
		T50.39	Autonomous robot navigation
		T50.40	Real Time Tracking
		T50.41	Flood Modelling
		T50.42	Computer Vision Based
		T50.43	Geocoding Models
		T50.44	Traffic Generator
		T50.45	Event Related Mining
		T50.46	Mapping APIs
		T50.47	Approach for improving situational awareness
		T50.48	Automatic Recommendation OSM Categories
		T50.49	Turn Restrcitions
		T50.50	Polygon Inspection in XML Data
T5.2	Quality, Completeness and Assessment of Contributors Behaviour	T50.23	Trust In OSM Data
		T50.26	Contributors Patterns
T5.3	Applications to Disaster, Navigation and others	T50.3	Mobile based Serives
		T50.9	Haptic for Navigation
		T50.13	Disaster Management
		T50.34	Evacuation Modeling
		T50.32	Humanitarian Efforts
T5.4	Traffic Simulation and Mobility	T50.4	Traffic Simulation and Management
		T50.8	Smart Cities and Mobility
		T50.14	Reducing Travel Times
T5.5	Indoor Navigation Models	T50.2	Mobile based Routing
		T50.5	Indoor Planning and Simulation

Only Two research themes matched up with “Quality, Completeness and Assessment of Contributors Behaviour” (T5.2) are “Trust in OSM Data” (T50.23) (Groot 2012, Keler and de Groot 2013b) and “Contributors Patterns” (T50.26) (Mooney and Corcoran 2014). These topics identifies the research trends focusses on identifying the contributor experience/category and trust in map data generated by contributors.

6.3. *Applications to Disaster, Navigation and others*

The benefit of crowdsourced map data, if it can used for society at large. Many team around the globe working on development of IT models for navigations like “Mobile based Services” (T50.3) and “Haptic for Navigation” (T50.9) (Jacob *et al.* 2009, 2012, 2010, 2011, Kaklanis *et al.* 2013), web-servers, for using map data in crisis situation like “Disaster Management” (T50.13) (Zook *et al.* 2010, Ridwan *et al.* 2012, Eckle and de Albuquerque 2015), “Evacuation Modeling” (T50.34) (Kunwar *et al.* 2014, Kunwar and Johansson 2015) and "Humanitarian Efforts" (T50.32) (Palen *et al.* 2015).

6.4. *Traffic Simulation and Mobility*

Micro-simulations of traffic systems are becoming more important as highly disaggregated data, such as mobility diaries or GPS traces, become available. For accurate results, a high-quality model of the road network is required. Recently, OSM has proven to be a valuable data source. This topic (T5.4) uncovers four such themes, where OSM is used like "Traffic Simulation and Management" (T50.4) (Dallmeyer *et al.* 2014, Stolfi and Alba 2015a, Backfrieder *et al.* 2013), "Smart Cities and Mobility" (T50.8) (Stolfi and Alba 2013, 2014, 2015b), "Reducing Travel Times" (T50.14).

6.5. *Indoor Navigation Models*

The availability of mobile devices for all has boosted the demand for indoor location-based services such as indoor routing or indoor navigation. The topic (T5.5) reveals the recent areas for indoor navigation and covers two research themes "Mobile based Routing" (T50.2) (Nikander *et al.* 2013, Czogalla 2015, Link *et al.* 2011, Czogalla and Naumann 2015) and "Indoor Planning and Simulation" (T50.5), where researchers started working and speeding devising methods for the same, even augmented reality area is also being explored for OpenstreetMap.

7. Limitations of the study

The revealed research areas and themes provide the patterns in OSM research. Nonetheless, the following limitations of this research need to be considered while reviewing the findings. First, we used only OpenStreetMap and OSM phrases in titles, abstracts and keywords of the publications through data collection. There may be some other publications which are about OSM but do not include the specified phrases in their abstracts, titles or keywords. To overcome this limitation, we run the second round through document collection by searching and OpenStreetMap and OSM phrases together in full text and 43 more abstracts were added. However, the decision to add the related publication into the corpus is based on researcher's interpretation. Similarly, there may be some other set of keywords that could have been used to form the corpus such as Crowdsourced Map, Volunteer Geographic Information.

As the result of this analysis, some of the high-loading papers for core research subjects are selected and presented. However, the loading values of selected papers neither reflect an academic value nor represent the number of citations of them in OSM literature. The methodology applies the number of occurrences of each term in each abstract through the entire corpus and calculates the loading values. Therefore, the high-loading papers in this research should not be considered as the most popular ones associated with the research area/theme.

Other limitations are related to the methodology of this study. The high-loading terms and articles are determined with respect to a threshold value. We calculated variances of term and document loadings to specify the threshold and select the terms and documents which have a bigger loading value than the threshold. This calculation estimates the statistically optimum threshold value for term and factor loadings. However, it does not mean that the terms and documents which have a loading value below the threshold are not related with the associated factor. In other words a lower threshold value increases the number of selected terms and documents. In addition, we used abstracts of the publications instead of the full text. This can be considered as a limitation due to ignoring the rest of the information. However, the derived entities (word/word groups) specifically from the literature review and methodology sections of publications increase the noise data and make the computer supported analysis more difficult. Besides, the factor labeling was done by researchers which may cause the discussions about subjectivity as a limitation of this study. However, to address this limitation, the factor labeling was independently done by the authors of this paper and then put together.

8. Future research and conclusions

This research sought to investigate the patterns and trends in OSM research by systematically analyzing the abstracts of 485 articles to empirically identify the key research areas and themes. The main contribution of this paper is the analytical decomposition and interpretation of the body of OSM research into five principal research areas, while revealing the expansive variety of the discipline with fifty substantial research themes. Through a cross-factor analysis, this research also presents the links between the principal research areas and the corresponding themes, drawing a map of OSM research structure.

This analysis establishes an objective and empirical foundation for the future discussion about the structure of BIM research. The identified research areas and themes indicate trends and future opportunities, which can be utilized by the industry as well as academia. More practically, the results of this study facilitates the existing OSM researchers to discover the position of their research studies within a broader and detailed frame; and the new researchers to understand the nature of OSM research and assess their areas of interest for their potential research with the corresponding research trend.

We revealed five main research areas with associated detailed research themes in this study. For potential further studies, researchers can pick one or more research areas and make another analysis with the same or another methodology. Researchers used LSA for this study. However, there are other statistical factor analysis methods which can be applicable to this research. For future studies, the researcher can apply another method to a similar collection to see the affinity and diversity of research subjects and themes with associated articles. To increase the application areas of this research, a dynamic query system can be developed with the same corpus by applying similar methodologies

References

Al-Bakri, M. and Fairbairn, D., 2012. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. 26 (8), 1437–1456.

Aridhi, S., *et al.*, 2014. SHORTEST PATH RESOLUTION USING HADOOP. *In*: *MOSIM 2014, 10me Confrence Francophone de Modlisation, Optimisation et Simulation*, 06 23. HAL CCSD.

Aridhi, S., *et al.*, 2015. A MapReduce-based approach for shortest path problem in large-scale networks. 41, 151 – 165.

Arsanjani, J.J., *et al.*, 2013. Assessing the Quality of OpenStreetMap Contributors together with their Contributions. Bibtex: arsanjani_assessing_2013.

Arsanjani, J.J. and Vaz, E., 2015. An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. 35, Part B, 329 – 337.

Backfrieder, C., Mecklenbruker, C.F., and Ostermayer, G., 2013. TraffSim – A Traffic Simulator for Investigating Benefits Ensuing from Intelligent Traffic Management. *In*: *Modelling Symposium (EMS), 2013 European*, 11., 451–456.

Bgin, D., Devillers, R., and Roche, S., 2013. ASSESSING VOLUNTEERED GEOGRAPHIC INFORMATION (VGI) QUALITY BASED ON CONTRIBUTORS' MAPPING BEHAVIOURS. XL-2/W1, 149–154.

Brovelli, M.A., *et al.*, 2016. Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets. n/a–n/a.

Ciepuch, B., *et al.*, 2011. Assessing the Quality of Open Spatial Data for Mobile Location-based Services Research and Applications. 105–116.

Codescu, M., *et al.*, 2014. OSMonto – An Ontology of OpenStreetMap Tags. [online] [????].

Czogalla, O. and Naumann, S., 2015. Pedestrian Guidance for Public Transport Users in Indoor Stations Using Smartphones. *In*: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Sept., 2539–2544.

Czogalla, O., 2015. Smart phone based indoor navigation for guidance in public transport facilities. 48 (10), 233 – 239 2nd {IFAC} Conference on Embedded Systems, Computer Intelligence and Telematics {CESCIT} 2015Maribor, Slovenia, 22-24 June 2015.

Dallmeyer, J., Lattner, A.D., and Timm, I.J., 2014. GIS-Based Traffic Simulation Using OSM. *Data Mining for Geoinformatics*. Springer.

Eckle, M. and de Albuquerque, J.P., 2015. Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes. *Retrieved September*, 24, 5–1.

Estima, J. and Painho, M., 2013. Exploratory analysis of OpenStreetMap for land use classification. *In*: *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, GEOCROWD 2013, Orlando, FL, USA, November 5, 2013*, 39–46.

Fan, H., *et al.*, 2014. Quality assessment for building footprints data on OpenStreetMap. 28 (4), 700–719.

Fischer, F., 2008. Collaborative mapping–How wikinomics is manifest in the geo-information economy. *Geoinformtics*, 11 (2), 28–31.

Goetz, M. and Zipf, A., 2012. Using Crowdsourced Geodata for Agent-Based Indoor Evacuation Simulations. 1 (2), 186.

Goetz, M. and Zipf, A., 2013. Indoor Route Planning with Volunteered Geographic Information on a (Mobile) Web-Based Platform. .

Golicher, D., ????. Accuracy of an Android cell phone GPS in the UK, Accessed on 13th August 2014. [online] [????].

Goodchild, M., 2009. NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, 3 (2), 82–96.

Graser, A., Straub, M., and Dragaschnig, M., 2015. Is OSM Good Enough for Vehicle Routing A Study Comparing Street Networks in Vienna. .

Groot, R.T.A.d., 2012. Evaluation of a volunteered geographical information trust measure in the case of OpenStreetMap. [online] [????].

Haklay, M. and Weber, P., ????. Openstreetmap: User-generated street maps. 7 (4), 12–18.

Haklay, M., Singleton, A., and Parker, C., 2008. Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass*, 2 (6), 2011–2039.

Howe, J., 2006. The rise of crowdsourcing. *Wired magazine*, 14 (6), 1–4.

Hudson-Smith, A., *et al.*, 2009. NeoGeography and Web 2.0: concepts, tools and applications. *Journal of Location Based Services*, 3 (2), 118–145.

Hultman, T., Boudjadar, A., and Asplund, M., 2016. Connectivity-optimal shortest paths using crowdsourced data. *In*: *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, March., 1–6.

Jacob, R., *et al.*, 2010. Haptic-GIS: Exploring the Possibilities. (3), 13–18.

Jacob, R., *et al.*, 2011. Integrating Haptic Feedback to Pedestrian Navigation Applications. *In*: *GISRUK 2011: GIS Research UK 19th Annual Conference*, 04.

Jacob, R., Mooney, P., and Winstanley, A., 2012. *In*: *What's up that Street Exploring Streets Using a Haptic GeoWand*. Springer.

Jacob, R., *et al.*, 2009. Wiimote as a Navigation tool for Pedestrians. *In*: *China-Ireland information and communications technologies conference.*, 08. National University of Ireland Maynooth, Dept. of Computer Science, 23–24.

Jilani, M., Corcoran, P., and Bertolotto, M., 2014. Automated highway tag assessment of OpenStreetMap road networks. *In*: *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Dallas/Fort Worth, TX, USA, November 4-7, 2014*, 449–452.

Jilani, M., Corcoran, P., and Bertolotto, M., 2016. Probabilistic Graphical Modelling for Semantic Labelling of Crowdsourced Map Data. .

Kaklanis, N., Votis, K., and Tzovaras, D., 2013. Open Touch/Sound Maps: A system to convey street data through haptic and auditory feedback. 57, 59 – 67.

Keler, C. and de Groot, R.T.A., 2013a. Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap. *In*: *Geographic Information Science at the Heart of Europe.*, 21–37 Springer International Publishing.

Keler, C. and de Groot, R.T.A., 2013b. Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap. *In*: *Geographic Information Science at the Heart of Europe - International AGILE'2013 Conference, Leuven, Belgium, 14-17 May 2013* SAGE Publications, 21–37.

Kounadi, O., 2009. Assessing the quality of OpenStreetMap data. Thesis (PhD). University College of London, Department of Civil, Environmental And Geomatic Engineering.

Kunwar, B. and Johansson, A., 2015. Measuring Disaster Preparedness of UK Cities from Open Spatial Databases. .

Kunwar, B., Simini, F., and Johansson, A., 2014. Large Scale Pedestrian Evacuation Modeling Framework Using Volunteered Geographical Information. 2, 813 – 818 The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands.

Link, J..B., *et al.*, 2011. FootPath: Accurate map-based indoor navigation using smartphones. *In*: *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, Sept., 1–8.

Mehdi, K., *et al.*, 2014. CupCarbon: A Multi-Agent and Discrete Event Wireless Sensor Network Design and Simulation Tool. [online] [????].

Mooney, P. and Corcoran, P., 2014. Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. 18 (5), 633–659.

Neis, P., Zielstra, D., and Zipf, A., 2011. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, 4 (1), 1–21.

Neis, P., Zielstra, D., and Zipf, A., 2013. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. 5 (2), 282.

Neis, P. and Zipf, A., 2012. Analyzing the contributor activity of a volunteered geographic information project–The case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1 (2), 146–165.

Nikander, J., *et al.*, 2013. Indoor and Outdoor Mobile Navigation by Using a Combination of Floor Plans and Street Maps. .

Palen, L., *et al.*, 2015. Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events. *In*: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Republic of Korea, April 18-23, 2015*, 4113–4122.

Rahman, K.M., Alam, T., and Chowdhury, M., 2012. Location based early disaster warning and evacuation system on mobile phones using OpenStreetMap. *In*: *Open Systems (ICOS), 2012 IEEE Conference on*, Oct., 1–6.

Rana, S. and Joliveau, T., 2009. NeoGeography: an extension of mainstream geography for everyone made by everyone?. *Journal of Location Based Services*, 3 (2), 75–81.

Ridwan, S.B., Ferdous, H.S., and Ahmed, S.I., 2012. Vol. 7546 of *Lecture Notes in Computer Science*, Human-Computer Interaction, Tourism and Cultural Heritage. *In*: *The State of OpenStreetMap in Bangladesh.*, 133–143 Springer.

Sachdeva, A., 2015. Collective Enrichment of OpenStreetMap Spatial Data Through Vehicles Equipped with Driver Assistance Systems. [online] [????].

Steiniger, S. and Hunter, A.J.S., ????. OpenJUMP HoRAE–A free GIS and toolbox for home-range analysis. 36 (3), 600–608.

Stolfi, D.H. and Alba, E., 2013. Red Swarm: smart mobility in cities with EAS. *In*: *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, 1373–1380.

Stolfi, D.H. and Alba, E., 2014. Eco-friendly reduction of travel times in european smart cities. *In*: C. Igel, D.V. Arnold, C. Gagne, E. Popovici, A. Auger, J. Bacardit, D. Brockhoff, S. Cagnoni, K. Deb, B. Doerr, J. Foster, T. Glasmachers, E. Hart, M.I. Heywood, H. Iba, C. Jacob, T. Jansen, Y. Jin, M. Kessentini, J.D. Knowles, W.B. Langdon, P. Larranaga, S. Luke, G. Luque, J.A.W. McCall, M.A. Montes de Oca, A. Motsinger-Reif, Y.S. Ong, M. Palmer, K.E. Parsopoulos, G. Raidl, S. Risi, G. Ruhe, T. Schaul, T. Schmickl, B. Sendhoff, K.O. Stanley, T. Stuetzle, D. Thierens, J. Togelius, C. Witt and C. Zarges, eds. *GECCO '14: Proceedings of the 2014 conference on Genetic and evolutionary computation*, Vancouver, BC, Canada, 12-16 07. ACM, 1207–1214.

Stolfi, D.H. and Alba, E., 2015a. An Evolutionary Algorithm to Generate Real Urban Traffic Flows. .

Stolfi, D.H. and Alba, E., 2015b. Smart Mobility Policies with Evolutionary Algorithms: The Adapting Info Panel Case. *In*: S. Silva, A.I. Esparcia-Alcazar, M. Lopez-Ibanez, S. Mostaghim, J. Timmis, C. Zarges, L. Correia, T. Soule, M. Giacobini, R. Urbanowicz, Y. Akimoto, T. Glasmachers, F. Fernandez de Vega, A. Hoover, P. Larranaga, M. Soto, C. Cotta, F.B. Pereira, J. Handl, J. Koutnik, A. Gaspar-Cunha, H. Trautmann, J.B. Mouret, S. Risi, E. Costa, O. Schuetze, K. Krawiec, A. Moraglio, J.F. Miller, P. Widera, S. Cagnoni, J. Merelo, E. Hart, L. Trujillo, M. Keswsentini, G. Ochoa, F. Chicano and C. Doerr, eds. *GECCO '15: Proceedings of the 2015 on*

Genetic and Evolutionary Computation Conference, Madrid, Spain, 11-15 07. ACM, 1287–1294.

Su, Y., Cai, H., and Shi, J., 2014. An improved realistic mobility model and mechanism for VANET based on SUMO and NS3 collaborative simulations. *In*: *2014 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, Dec., 900–905.

Sukhjit Singh Sehra, Jaiteg Singh, and Hardeep Singh Rai, 2013. Assessment of OpenStreetMap Data - A Review. *International Journal of Computer Applications*, 76 (16), 17–20.

Turner, A., 2006. Introduction to Neogeography. *O'Reilly Media, MA, USA*.

Walsh, J., 2008. The beginning and end of neogeography. *GEOconnexion International Magzine*, (7), 28–30.

Will, J., 2014. Development of an automated matching algorithm to assess the quality of the OpenStreetMap road network. [online] [????].

Yang, A., Fan, H., and Jing, N., 2016. Amateur or Professional: Assessing the Expertise of Major Contributors in OpenStreetMap Based on Contributing Behaviors. 5 (2), 21.

Yuan, J. and Cheriyadat, A.M., 2016. Automatic Generation of Building Models Using 2D Maps and Street View Images. [online] [????].

Zook, M., *et al.*, 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*, 2 (2), 7–33.

8.1. *Via e-mail*

This Guide for Authors, the Class file and the associated open-source LaTeX packages are also available by e-mail. Requests should be addressed to `latex.helpdesk@tandf.co.uk` clearly stating for which journal you require the Guide for Authors and/or Class file.