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 13, 2016*)

OpenStreetMap (OSM) has become the greatest example of collaborative mapping and increasing academic community interests, pushed more sensible and rigorous research in this area. In recent times, numerous researchers working with OSM, produced a large literature. It is critical to comprehend and observe the research trends rising in OSM exploration. There is no such systematic review of literature reported till date. In this study, Latent Semantic Analysis (LSA) technique is employed to uncover the research trends by analysing an extensive corpus of abstracts of 485 academic papers published from 2007 to 2016. Five core research areas and fifty research trends are identified during this study.

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 energized more prominent coordinated effort among web clients and different users, content suppliers and ventures (Hudson-Smith *et al.* 2009). This development gave new techniques for sharing and figuring information (Hudson-Smith *et al.* 2009, Rana and Joliveau 2009, Walsh 2008, Goodchild 2009, Haklay and Weber 2008) by crowdsourcing development like Wikipedia (Howe 2006). In context to geographic information, it is known as Volunteered Geographic Information (VGI) or collaborative mapping (Fischer 2008), used by many website Wikimapia, OpenStreetMap, Google Map (Kounadi 2009). The OpenStreetMap (OSM) is only crowdsourced project, which provides free editable guide of world, available under CC-BY-SA license.

Amid most recent nine years, OSM got a great deal of thought by researchers and numerous papers have been published. But, despite of the expanding volume of the literature, the research areas remain vague and ineffectively understood. Most of the research studies are dedicated to quality issues of OSM similar to conducted by Neis *et al.* (2012). These studies discuss issues related to OSM and give an involved illustration concerning quality of OSM data, tools and framework, applications to various fields etc. Due to large rise in literature, it has become important to identify the research trends for which manual systematic reviews (White and Schmidt 2005) and semi-automated topic modelling algorithms (Delen and Crossland 2008, Lee *et al.* 2010, Evangelopoulos *et al.* 2012) can be used, but first being more critical review and second more generic in finding the research trends (Yalcinkaya and Singh 2015). In this study, we are focussing on finding the research trends and patterns, which will also uncover the recent and non-emphasised trends based on textual literature dataset. For processing the large corpus Latent Semantic Analysis (LSA), a natural language processing approach, is utilised.

LSA provides methodology for automatically organising, understanding, searching, and summarizing textual dataset. It reveals the concealed concepts or research trends that pervade the collection. LSA examines the relationships between documents and terms of document and reveal concepts. LSA uses a versatile mathematical approach Singular Vector Decomposition (SVD) to create low-dimensional space for finding relationships, revealing topics, comparing documents, document similarity (Deerwester *et al.* 1990, Landauer and Dumais 1997, Landauer *et al.* 1998, Landauer 2006, Foltz 1996, Dumais 2004, Yalcinkaya and Singh 2015).

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. The motive of the study is to bring out an organised representation of the patterns in exploration of OSM literature information by efficiently applying LSA approach to substantial corpus of scholarly articles published from 2007 to 2016. During this study 485 paper abstracts are used to develop corpus. The outcome of this study is five core research areas and fifty research trends. The results of investigation would provide insight into OSM and give chances to the researchers to position their future exploration.

The paper has been divided into different sections. The next section discusses about OSM and third section describes the methodology followed to collect research literature. Fourth section elaborates methodological analysis, that includes the steps followed for application of the technique onto literature dataset. Section five discusses results obtained for different topic solution, identified from the dataset, thereafter mapping of core research areas and research trends is presented in sixth section. The paper concludes with the findings of the study and future scope.

Table 1. Paper Count during various steps of Data 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

2. About OpenStreetMap

OSM project began in 2004, is an editable and unreservedly accessible database of mapping data which can be utilised for various purposes without many limitations under CC-BY-SA license (Haklay 2010). OSM has three main elements (Sehra *et al.* 2013), which are Node, Way and Relation. Relation data structure depicts the sensible relationship between components. The elements use tags, to describe a geographic attribute i.e. physical feature on ground associated with element (node, way and relation) and changesets. A tag consists of 'Key' and a 'Value', which are free format text fields. OSM's free tagging system allows unlimited number of attributes describing each feature using tags (OpenStreetMap 2016a).

OpenStreetMap, a revolutionary movement has encouraged nearly 2,734,089 as registered users and have 3,397,863,934 uploaded nodes, 352,772,514 ways and 4,278,682 relations (OpenStreetMap 2016c). The contributors to this project are enthusiasts having varying level of experiences. The contributors utilise various devices to record GPX tracks and edit the information using online (iD3, Potlatch) or offline editors (JOSM). The contributor also add attributes to the data for making data more sensible before uploading to OSM server (OpenStreetMap 2016b).

3. Data Preparation

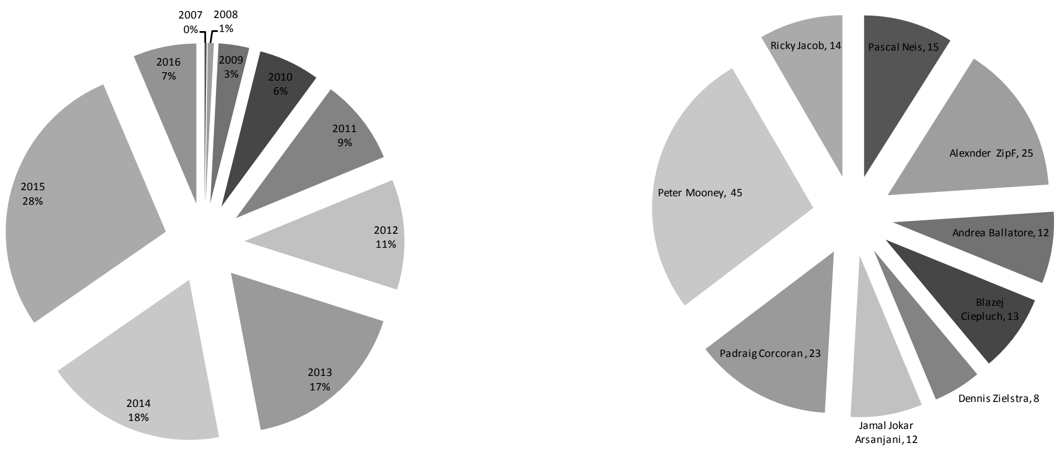
For preparing literature dataset (corpus), publications data was gathered through numerous scholastic databases and sources. Corpus preparation involved various steps, the publications were selected using "OpenStreetMap (OSM)", "Voluteered Geographic Information (VGI)", "Crowdsourced Map" and "Callaborative Mapping" as search phrase in their titles, abstracts, keywords. Spatial data related to magazines, brochures and software tool related documentation were not considered. Open-Source tool Jabref (JabRef Development Team 2016) was used for collection, screening, selection and corpus preparation and 2889 paper publication were collected.

Bibliographic databases searched for collecting literature dataset were 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 were also manually searched and added to collection. After this, the collection was manually searched to filter relevant and irrelevant papers. Table 1 summarises the document collection process.

Large number of publication literature collected from different bibliographic database

Table 2. Top 15 Journals

Journal Name	Number of Related Publication
ISPRS International Journal of Geo-Information	27
Transactions in GIS	15
International Journal of Geographical Information Science	18
The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences	14
GeoJournal	4
Computers, Environment and Urban Systems	3
Future Internet	5
International Journal of Digital Earth	4
Survey Review	3
Applied Geomatics	3
Remote Sensing	4
Geoinformatics FCE CTU	3
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	3
Annals of GIS	2
Applied Geography	2



(a) Year-wise OpenStreetMap research publications from 2007-2016

(b) Top Reseacjhers in OSM research

Figure 1. Statistical Data about OSM Research Publication

were reviewed to eliminate, non-focussed, duplication and unwanted in the final step. After the final processing, 485 documents were left in the literature dataset that includes 19 papers which are converted from German, Spanish and Italian language to English. Figure 1(a) shows year-wise distribution of literature dataset. Table 2 summarize the top fifteen journals for publishing on OpenStreetMap and figure 1(b) illustrates the top researchers having maximum publications. Once collection is prepared, whole collection is exported to .csv, which includes titles and abstracts, year of publication. The data prepared and export filter designed for JabRef are available at `https://sukhjitsehra@bitbucket.org/sukhjitsehra/topic-modelling.git`.

4. Methodological Analysis

The corpus prepared given to Latent Semantic Analysis (LSA) model, for uncovering the "latent" semantic structure (topics) (Deerwester *et al.* 1990, Sidorova *et al.* 2008, Hossain *et al.* 2011), called topic modelling (Alghamdi and Alfalqi 2015). LSA is a semi-automated unsupervised text-mining approach which give better results than manual literature reviews. LSA approach is efficiently able to conclude the required topics from large unstructured dataset (Delen and Crossland 2008). LSA has been taken as it outperform other approaches for topic modelling such as Probabilistic Latent Semantic Analysis (pLSA), Latent Dirichlet Allocation (LDA) on few aspects (Lee *et al.* 2010). The heart of this semantic analysis is powerful computational method, singular vector decomposition (SVD), capable of reducing enormous textual information to low-dimensional space (Abidin *et al.* 2010, Kuandykov *et al.* 2015). There are many implementations of SVD, but fast truncated incremental stochastic singular value decomposition algorithm (Halko *et al.* 2009) with single pass is used (Řehruřek 2011). It is implemented as python library (Řehůřek and Sojka 2010) and is efficient to handle even very large corpus. The primary thought to use this approach is to infer "latent" semantic structure (topics) hidden in corpus, input required other than the corpus is number of topics, we want to infer from the corpus. For example, as in table 3, five-topic solution is derived to express the relationship amongst five topics and associated high-loading terms along-with corresponding label. LSA finds low-rank approximation, this results in some topics combined and depends on multiple terms and documents (Sidorova *et al.* 2008). The output of approach are two matrices, term loading and document loading matrix i.e. one matrix that represent the topics and associated high loaded terms to respective topic and second matrix that represents topics and associated high-loaded documents to respective topic (Yalcinkaya and Singh 2015). Higher values of loading represents greater familiarity to a topic. The methodology followed for semantic analysis is as follows:

4.1. *Preprocessing*

The process LSA starts with creation of term-document matrix, but corpus created during data acquisition contains lots of noisy data. So preprocessing phase involves, the elimination of noisy words/characters from the dataset. The following steps have been incorporated during preprocessing to literature dataset:

(1) Sentences (Titles and Abstracts) for each publication (document) are tokenised.
(2) Tokens in each document are converted to lowercase letters.
(3) Punctuation characters of terms, include the period, exclamation point, comma, apostrophe, question mark, quotation mark and hyphen are eliminated
(4) The numbers (numerical values) filtered to contain only textual terms.
(5) Performed 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;

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

(6) Removal of English stop-words (stop-words of nltk python package) and the keywords which are very common in all the publications ("OpenStreetMap", "Geographic", "Voluteered", "information", "Crowdsource", "maps" and "OSM") are removed.
(7) SnowballC Stemmer algorithm is used to tokens in each document, It converts inflected words to base stem.

4.2. *Vector Space model and Term-Document Matrix*

In addition to previous steps, dataset is further refined to remove terms that exist only once in a document. These terms are local to a particular document. Initially dataset had 87,348 tokens. But after removal of undesired words, n char words, numbers, encoded characters, "English" and custom stop-words, 4978 unique tokens were generated and after applying "SnowballC Stemmer" algorithm, the token count reduced to 2510.

The representing documents in vector space is acknowledged as the vector space model and this important for operation like scoring documents, document classification and document clustering. For converting documents to vector format, document representation called "bag-of-words" is used Řehůřek and Sojka (2010). The document vector represents question-answer pair as below:

"How many times does the word "Assessment" appear in the document?, Once or twice or thrice."

These questions are represented using ids(integer) and mapping process is called dictionary. This dictionary is used convert the word to its id, along-with the count of occurrences of each token word in that document, and hence creates a sparse vector for each document. The vector for each document of the corpus e.g. [(0, 1), (3, 2)], represents that term (id0) apears once and term(id3) appeared two times in that document. In this study 485 sparse vectors are created containing 2510 tokens. A dataset of 485 documents converted to vector space with 2510 terms is a natural view of 2510 x 485 term-document matrix, here rows represent 2510 terms (dimensions) of the 485 columns, each of which corresponds to a document (publication).

4.3. *Term frequency-Inverse document frequency (TF-IDF) Weighting scheme*

TF-IDF (term frequency-inverse document frequency) weighting scheme is able to reflect significance of a given entity in comparison to other entities in the corpus thus helps in creating better clusters (Salton and Buckley 1988, Paik 2013). The weight given by TF-IDF is highest, when a term is observed many times within set of documents, lower when the term is observed, fewer times in a document or in large documents. It is lowest when the term observed in all the documents i.e. zero.

There are many combinations of TF-IDF weighting schemes (Lan *et al.* 2005, Wu *et al.* 2008). The approach used during this study is given in equation 1, where $tf - idf$ is weight obtained, tf is term frequency of term, n_d is number of documents in the dataset.

$$tf - idf_{i,j} = tf_{i,j} * log_2(n_d/df_i) \tag{1}$$

Using tf-idf given in equation 1, a new 2510x485 term-document weighted matrix is created.

4.4. *Latent Semantic Analysis*

LSA expects terms close by meaning occurs in similar type textual informaiton. The prepared tf-idf weighted matrix is relinquished to fast truncatedSVD singular vector decomposition (SVD) to perform rank lowering. The SVD model is $X = U\Sigma V^T$, it is basically a mathematical factorisation of matrix i.e. factorisation of matrix X into varlables: initial rotation U, scaling Σ and final rotation V (Abidin *et al.* 2010, Al-Qawasmeh *et al.* 2010, Deerwester *et al.* 1990, Halko *et al.* 2009)

$$XX^T = (U\Sigma V^T)(U\Sigma V^T)^T = (U\Sigma V^T)(V^{T^T}\Sigma^T U^T) = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T \tag{2}$$

$$X^T X = (U\Sigma V^T)^T(U\Sigma V^T) = (V^{T^T}\Sigma^T U^T)(U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T \tag{3}$$

Here XX^T and $X^T X$ provides term-loading and document-loading with respect to topics and $\Sigma\Sigma^T$ provides the weights of topics (singular values) in descending order. For extracting fewer topics, number of topics to be extracted k are provided (Evangelopoulos *et al.* 2012, Kuandykov *et al.* 2015). This will take topmost k singular values and perform k approximation. In this study, 3, 5, 10 principal research areas (topics) are extracted and while for detailed research areas 50 topics are extracted.

4.5. *Selecting threshold values for topics*

The resultant term-loading matrix and document-loading matrix obtained, consists of corresponding weights for identified topics respectively. The values in matrix are positive and negative, to interpret the results better same varimax rotation (Kaiser 1958, Abdi 2004) is performed for both matrices. This concludes in making entity loading increased for one topic relative to other topics (Sidorova *et al.* 2008, Yalcinkaya and Singh 2015). To distinguish between significant and insignificant loading a heuristic approach is used i.e. empirical tail probability distribution is applied to select the threshold values as explained by Sidorova *et al.* (2008) and Yalcinkaya and Singh (2015). During this study the terms and documents are loaded to only one topic, based on the knowledge of the concerned research area. As per the calculation performed by tail probabilty distribution, the threshold values for 3, 5, 10 and 50 topic solutions are 0.0936, 0.0655, 0.0874 and 0.183 respectively for document loadings. So any document with loading less than these values are considered insignificant for the topic.

4.6. *Topic labelling*

For each topic, high-loading terms and documents are grouped together and sensible label are given to each topic e.g. in table 3. Similarly topic labels are given to each topic of 3, 5, 10 and 50 topic solution. For each entity the loading are represented in descending order to group and labelling together.

5. Results and discussion

Table 4 illustrates the core research areas, along with count of publications during year 2007-2016 and further split into two groups. The trends during 2012-2016 shows steep increase in the interest of researcher community towards OSM. The results are more skewed towards one topic “Quality assessment and analysis” in each of 3,5,10 topic solution. This describes the most of the work on OSM is about checking its quality in different aspects and moreover each researcher assessed limited area during research work. After thoughtful analysis, 5-topic solution has been considered as describing core research areas and 50-topic solution as research trends in OSM. The number of documents loaded for particular topic is describing the publication for that topic. The document count has been taken after empirical analysis and each document is loaded to only one topic after considering suitable values for cross-loading. The topic “Quality assessment and analysis” appears across (T3.1) to (T10.1), and the number of high-loading documents is decreasing, as new topics are emerging out from the corpus.

Table 4 illustrates are core research areas and table 5 represents 50 topic solution research trends in OSM. Table 7 shows the mapping of 50-topic solution to 5-topic solution i.e. research trends to core research areas.

5.1. *Core OSM research areas*

Table 4, illustrates during 3-topic solutions, that the OSM research is focussed around “Quality Assessment and Analysis” (T3.1), “Routing and Navigation” (T3.2) and “Haptic and Diaster Management” (T3.3). These articles emphasise on development of methods for quality assessment of crowdsourced data, its applications and issues pertaining

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

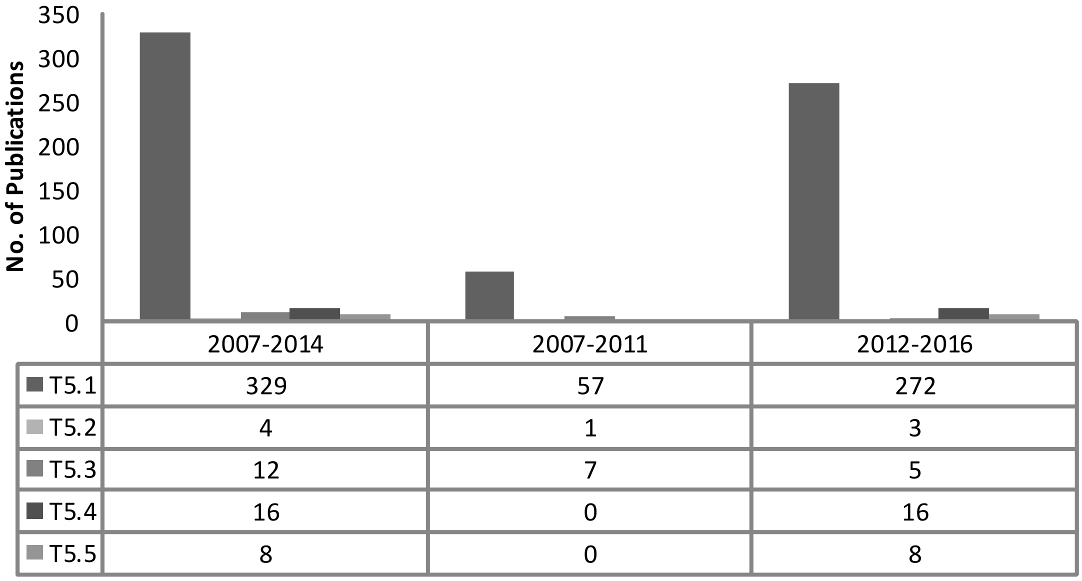


Figure 2. Paper Distribution for 5-topic solution

to crowdsourced data.

Core research areas emerged for 5-topic solution are "Quality Assessment and Analysis" (T5.1), "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)". "Quality Assessment and Analysis" (T5.1) remains the strong area of research, 5 highly-loaded document for each topic, are given in Table 3.

The 10-factor solution 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 dominated re-

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	(Ciepłuch *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	(Bégin *et al.* 2013)	0.321112335
		(Keßler and de Groot 2013)	0.286861888
		(Arsanjani *et al.* 2013)	0.25761061
		(Yang *et al.* 2016)	0.195301071
		(Neis and Zipf 2012)	0.184615268
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

search area, during years 2012-2016. The area for assessing the behaviour of contributors "Quality, Completeness and Assessment of Contributors Behaviour" (T10.2) also become trending area of reearch during 2012-2016.

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 2. [WRITE A PARAGRAPH ABOUT ANALYSIS]

5.2. *OSM Research Trends*

The 50-topic solution uncovers 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) came out as most explored in the 50-topic analysis too, in consistency with 3, 5, 10 topic solutions.

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 no contribution during years 2007-2011, but saw good interest by researchers 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 editing patterns of OSM Contributors (Mooney and Corcoran 2014).

The 50-topic solution reveals research trends, 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 core research areas and research trends

Table 7 presents mapping between the papers, 5-topic core research areas and 50 topic research trends. Connection is established on the basis of the loading via selecting high-loading papers for both solution

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.* 2014, 2015), "Managing OSM Tags" (T50.30) (Codescu *et al.* 2011, 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*

(T5.2) OSM data is consequence of crowdsourced effort, the contributors have different level of mapping interests and experience. This makes generated osm more vulnerable to errors and incomplete. OSM uses loose coordination and no top-down quality assurance processes. Only Two research themes matched up with "Quality, Completeness and Assessment of Contributors Behaviour" (T5.2) are "Trust in OSM Data" (T50.23) (Groot

Table 7. Mappig of core Research Areas and Redearch trends

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

2012, Keßler and de Groot 2013) 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*

(T5.3) 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). This topic particularly focusses on the navigation, using haptic enabled devices and preparedness of OSM for disaster/crisis management.

6.4. *Traffic Simulation and Mobility*

(T5.4) Simulations of traffic systems are become topic of research as GPS traces are available in public domian. OSM has proved as a valuable data for such researches. 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*

(T5.5) 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. Conclusion

This research is based on mathematical foundations, discovers research trends in OpenStreetMap literature. It uncovers the research trends by analysing 485 documents published by researchers. The outcome of the approach generates k-topic solution and corresponding term and documents loading. These loadings explain the proximity to a given topic, only highly loaded terms and documents above threshold are considered relevant to the topic. In this study 3, 5, 10 and 50 topics are extracted to identify the core research areas and research trends in OSM. The results of the study shown that "Quality Assessment and Analysis" topic is most explored. The quest for high quality map data for application like connected cars and navigation can be fulfilled if more researchers across the globe, explore the quality of the OSM.The outcome of the study would certainly help the researchers to identify and explore research trends in OSM.

The LSA is not only approach for semantic analysis of unstructured data to find the hidden information, but LSA has its own advantages. Other approaches like pLSA and LDA can also be applied onto same literature datasets. During this study only titles and abstract are considered for the preparation of the corpus, but full paper text can be added while corpus preparation. A web-based system can be developed in the future, where user has to provide corpus and number of topics to extracted from the corpus.

References

Abdi, H., 2004. Factor rotations in factor analyses. *In*: *The SAGE Encyclopedia of Social Science Research Methods.*, 792–795 Thousand Oaks, California: Sage.

Abidin, T.F., Yusuf, B., and Umran, M., 2010. Singular Value Decomposition for dimensionality reduction in unsupervised text learning problems. *In*: *Proc. 2nd Int. Conf.*

Education Technology and Computer, Vol. 4, 22-24 June 2010. Shanghai: IEEE, 422–426.

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

Al-Qawasmeh, A.M., Maciejewski, A.A., and Siegel, H.J., 2010. Characterizing heterogeneous computing environments using singular value decomposition. *In*: *26th IEEE International Parallel and Distributed Processing Symposium*, Shanghai, China, 21-25 May , 2012. IEEE, 1–9.

Alghamdi, R. and Alfalqi, K., 2015. A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*, 6 (1), 147–153.

Aridhi, S., *et al.*, 2014. Shortest Path Resolution using Hadoop. *In*: *MOSIM 2014, 10ème Conférence Francophone de Modélisation, Optimisation et Simulation*, November, 2014. Nancy, France: HAL CCSD.

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

Arsanjani, J.J., *et al.*, 2013. Assessing the Quality of OpenStreetMap Contributors together with their Contributions. *In*: *16th AGILE Conference on Geographic Information Science*, 14-17 May, 2013., Leuven, Belgium.

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

Backfrieder, C., Mecklenbräuker, C.F., and Ostermayer, G., 2013. TraffSim – A Traffic Simulator for Investigating Benefits Ensuing from Intelligent Traffic Management. *In*: *Modelling Symposium (EMS), 2013 European*, 20-22 Nov. 2013. Manchester, UK: IEEE, 451–456.

Bégin, D., Devillers, R., and Roche, S., 2013. Assessing Volunteered Geographic Information VGI Quality Based On Contributors' Mapping Behaviours. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-2/W1, 149–154.

Brovelli, M.A., *et al.*, 2016. Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets. *Transactions in GIS*, p. Early View.

Ciepłuch, B., *et al.*, 2011. Assessing the Quality of Open Spatial Data for Mobile Location-based Services Research and Applications. *The journal Archives of Photogrammetry, Cartography and Remote Sensing*, 22, 105–116.

Codescu, M., *et al.*, 2011. OSMonto – An Ontology of OpenStreetMap Tags. *In*: *State of the map Europe (SOTM-EU)*, 15-17 July, 2011., Vienna University of Technology, Vienna, 55–64.

Czogalla, O. and Naumann, S., 2015. Pedestrian Guidance for Public Transport Users in Indoor Stations Using Smartphones. *In*: *The 18th International IEEE Conference on Intelligent Transportation Systems*, 15 – 18 September 2015. Las Palmas de Gran Canaria, Spain: IEEE, 2539–2544.

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

Dallmeyer, J., Lattner, A.D., and Timm, I.J., 2014. GIS-Based Traffic Simulation Using OSM. *In*: *Data Mining for Geoinformatics.*, 65–82 New York: Springer.

Deerwester, S., *et al.*, 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41 (6), 391–407.

Delen, D. and Crossland, M.D., 2008. Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, 34 (3), 1707 – 1720.

Dumais, S.T., 2004. Latent semantic analysis. *Annual Review of Information Science and Technology*, 38 (1), 188–230.

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

Řehůřek, R. and Sojka, P., 2010. Software Framework for Topic Modelling with Large Corpora. *In*: *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks* Valletta, Malta: University of Malta, 46–50.

Řehruřek, R., 2011. Fast and Faster: A Comparison of Two Streamed Matrix Decomposition Algorithms. *CoRR*, abs/1102.5597, 1–7.

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*, 5 November, 2013. Orlando, Florida, USA,: ACM, 39–46.

Evangelopoulos, N., Zhang, X., and Prybutok, R.V., 2012. Latent Semantic Analysis: five methodological recommendations. *European Journal of Information Systems*, 21 (1), 70–86.

Fan, H., *et al.*, 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28 (4), 700–719.

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

Foltz, P.W., 1996. Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, & Computers*, 28 (2), 197–202.

Goetz, M. and Zipf, A., 2012. Using Crowdsourced Geodata for Agent-Based Indoor Evacuation Simulations. *ISPRS International Journal of Geo-Information*, 1 (2), 186.

Goetz, M. and Zipf, A., 2013. Indoor Route Planning with Volunteered Geographic Information on a (Mobile) Web-Based Platform. *In*: *Progress in Location-Based Services.*, 211–231 Berlin Heidelberg: Springer.

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. *In*: *Progress in Location-Based Services 2014.*, 3–17 Switzerland: Springer.

Groot, R.T.A.d., Evaluation of a volunteered geographical information trust measure in the case of OpenStreetMap. Thesis (M.Sc.), Institute of Formal Methods in computer science, 2012. .

Haklay, M. and Weber, P., 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, 7 (4), 12–18.

Haklay, M., 2010. How good is volunteered geographical information A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37 (4), 682–703.

Halko, N., Martinsson, P.G., and Tropp, J.A., 2009. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53 (2), 217–288.

Hossain, M.M., Prybutok, V., and Evangelopoulos, N., 2011. Causal Latent Semantic Analysis (cLSA): An Illustration. *International Business Research*, 4 (2), 38.

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)*, 14-18 March, 2016. Sydney, Australia: IEEE, 1–6.

JabRef Development Team, 2016. JabRef. [online] `http://www.jabref.org` [????].

Jacob, R., *et al.*, 2010. Haptic-GIS: Exploring the Possibilities. *The SIGSPATIAL Special*, 2 (3), 13–18.

Jacob, R., *et al.*, 2011. Integrating Haptic Feedback to Pedestrian Navigation Applications. *In*: *GISRUK 2011: GIS Research UK 19th Annual Conference*, 27-29 April 2011. Portsmouth: GIS Research UK, 205–210.

Jacob, R., Mooney, P., and Winstanley, A., 2012. What's up that Street Exploring Streets Using a Haptic GeoWand. *In*: *Advances in Location-Based Services.*, 91–105 Heidelberg, Berlin: Springer.

Jacob, R., *et al.*, 2009. Wiimote as a Navigation tool for Pedestrians. *In*: *CIICT 2009 : proceedings of the China-Ireland information and communications technologies conference.*, 19-21 August, 2009. Maynooth, Ireland: 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*, 4-7 November, 2014. Dallas/Fort Worth, TX, USA: ACM, 449–452.

Jilani, M., Corcoran, P., and Bertolotto, M., 2016. Probabilistic Graphical Modelling for Semantic Labelling of Crowdsourced Map Data. *In*: *Intelligent Systems Technologies and Applications.*, 213–224 Switzerland: Springer.

Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23 (3), 187–200.

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

Keßler, C. and de Groot, R.T.A., 2013. 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 Switzerland: Springer.

Kounadi, O., Assessing the quality of OpenStreetMap data. MSc Dissertation, Department of Civil, Environmental And Geomatic Engineering, University College of London, 2009. .

Kuandykov, A.A., *et al.*, 2015. Usage of singular value decomposition matrix for search latent semantic structures in natural language texts. *In*: *Proc. 54th Annual Conf Society of Instrument and Control Engineers of Japan (SICE) of the*, 28- 30 July, 2015. China: IEEE, 286–291.

Kunwar, B. and Johansson, A., 2015. Measuring Disaster Preparedness of UK Cities from Open Spatial Databases. *In*: *Traffic and Granular Flow '13.*, 265–271 Switzerland: Springer.

Kunwar, B., Simini, F., and Johansson, A., 2014. Large Scale Pedestrian Evacuation Modeling Framework Using Volunteered Geographical Information. *Transportation Research Procedia*, 2, 813 – 818.

Lan, M., *et al.*, 2005. A comprehensive comparative study on term weighting schemes for text categorization with support vector machines. *In*: *Special interest tracks and* *posters of the 14th international conference on World Wide Web*, 10 -14 May, 2005. Chiba, Japan: ACM, 1032–1033.

Landauer, T.K., 2006. Latent Semantic Analysis. *In*: *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd.

Landauer, T.K. and Dumais, S.T., 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104 (2), 211.

Landauer, T.K., Foltz, P.W., and Laham, D., 1998. An introduction to latent semantic analysis. *Discourse processes*, 25 (2-3), 259–284.

Lee, S., *et al.*, 2010. An Empirical Comparison of Four Text Mining Methods. *In*: *Proc. 43rd Hawaii Int System Sciences (HICSS) Conf*, 5-8 January, 2010. Honolulu, Hawaii: IEEE, 1–10.

Link, J..B., *et al.*, 2011. FootPath: Accurate map-based indoor navigation using smartphones. *In*: *2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN),*, 21-23 September, 2011. Guimarães, Portugal: IEEE, 1–8.

Mehdi, K., *et al.*, 2014. CupCarbon: A Multi-Agent and Discrete Event Wireless Sensor Network Design and Simulation Tool. *In*: *SIMUtools 2014: 7th International ICST Conference on Simulation Tools and Techniques*, 17-19 March, 2014. Lisbon, Portugal: The Institute for Computer Science, Social Informatics, and Telecommunications Engineering.

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

Neis, P., Zielstra, D., and Zipf, A., 2012. The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007-2011. *Future Internet*, 4 (1), 1.

Neis, P., Zielstra, D., and Zipf, A., 2013. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet*, 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 (3), 146–165.

Nikander, J., *et al.*, 2013. Indoor and Outdoor Mobile Navigation by Using a Combination of Floor Plans and Street Maps. *In*: *Progress in Location-Based Services.*, 233–249 Heidelberg, Berlin: Springer.

OpenStreetMap, 2016a. About OpenStreetMap. [online] `https://wiki.openstreetmap.org/wiki/About\_OpenStreetMap` [Accessed on June 8, 2016].

OpenStreetMap, 2016b. Comparison of editors. [online] Available from: `http://wiki.openstreetmap.org/wiki/Comparison\_of\_editors` [Accessed on June 8, 2016].

OpenStreetMap, 2016c. OpenStreetMap Statistics. [online] `http://www.openstreetmap.org/stats/data\_stats.html` [Accessed on June 8, 2016].

Paik, J.H., 2013. A Novel TF-IDF Weighting Scheme for Effective Ranking. *In*: *SIGIR '13 Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, 28 July - 01 August, 2013. Dublin, Ireland: ACM, 343–352.

Palen, L., *et al.*, 2015. Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events. *In*: *CHI 2015 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 18-23 April, 2015. Seoul, Republic of Korea: ACM, 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*: *2012 IEEE*

Conference on Open Systems (ICOS), 21-24 October, 2012. Kuala Lumpur: IEEE, 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*, The State of OpenStreetMap in Bangladesh. *In*: *Human-Computer Interaction, Tourism and Cultural Heritage.*, 133–143 Heidelberg, Berlin: Springer.

Sachdeva, A., Collective Enrichment of OpenStreetMap Spatial Data Through Vehicles Equipped with Driver Assistance Systems. Masterthesis, Faculty of Computer Science, Chemnitz University of Technology, 2015. .

Salton, G. and Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24 (5), 513 – 523.

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

Sidorova, A., *et al.*, 2008. Uncovering the Intellectual Core of the Information Systems Discipline. *MIS Quaterly*, 32 (3), 467–482.

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*, 06 - 10 July, 2013. Amsterdam, Netherlands: ACM, 1373–1380.

Stolfi, D.H. and Alba, E., 2014. Eco-friendly reduction of travel times in european smart cities. *In*: *GECCO '14: Proceedings of the 2014 conference on Genetic and evolutionary computation*, 12-16 July, 2014. Vancouver, BC, Canada: ACM, 1207–1214.

Stolfi, D.H. and Alba, E., 2015a. An Evolutionary Algorithm to Generate Real Urban Traffic Flows. *In*: *Advances in Artificial Intelligence.*, 332–343 Switzerland: Springer.

Stolfi, D.H. and Alba, E., 2015b. Smart Mobility Policies with Evolutionary Algorithms: The Adapting Info Panel Case. *In*: *GECCO '15: Proceedings of the 2015 on Genetic and Evolutionary Computation Conference*, Madrid, Spain, 11-15 July, 2015. Madrid, Spain: 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)*, 16 – 19 December, 2014. Hsinchu, Taiwan: Department of Computer Science, National Tsing Hua University, 900–905.

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

White, A. and Schmidt, K., 2005. Systematic literature reviews. *Complementary Therapies in Medicine*, 13 (1), 54 – 60.

Will, J., Development of an automated matching algorithm to assess the quality of the OpenStreetMap road network. Master Degree Thesis, Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan, Lund, Sweden, 2014. .

Wu, H.C., *et al.*, 2008. Interpreting TF-IDF Term Weights As Making Relevance Decisions. *ACM Transactions on Information Systems*, 26 (3), 13:1–13:37.

Yalcinkaya, M. and Singh, V., 2015. Patterns and trends in Building Information Modeling (BIM) research: A Latent Semantic Analysis. *Automation in Construction*, 59, 68–80.

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

Yuan, J. and Cheriyadat, A.M., 2016. Automatic Generation of Building Models Using 2D Maps and Street View Images. *CoRR*, abs/1601.07630.

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.