IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 2, 2019, accepted March 17, 2019, date of publication March 26, 2019, date of current version April 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907570

Visualization of Location-Referenced Web Textual Information Based on Map Mashups

HANG ZHANG[1], LIN LI[1,2,3], WEI HU[1], WENJING YAO[4], AND HAIHONG ZHU[1]

[1]School of Resource and Environment Sciences, Wuhan University, Wuhan 430079, China
[2]Geospatial Information Science Collaborative Innovation Center, Wuhan University, Wuhan 430079, China
[3]The Key Laboratory for Geographical Information Systems, Ministry of Education, Wuhan University, Wuhan 430079, China
[4]Wuhan Winsse Security Technology Co., Ltd., Wuhan 430079, China

Corresponding author: Lin Li (lilin@whu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2017YFB0503701.

ABSTRACT Space is the most fundamental organizing dimension for information that forms the basic spatial understanding around which all other temporal and semantic details are situated. Various types of web information are present in our daily lives, and spatial content can facilitate the understanding of that information. Hence, many studies have been conducted to extract the implicit spatial location from different web resources, and they have mainly investigated web textual information. However, the existing studies mainly focus on the extraction or simple presentation of the location of the information, while further exploration and visualization of the information through the comprehensive consideration of its spatial, temporal, and semantic dimensions have rarely been performed. Thus, this paper proposes a novel modeling and visualization framework for location-referenced web textual information. The framework applies an information model to structurally extract and resolve spatial content along with the temporal and semantic elements from the location-referenced information items. Based on the constructed information model, the information items are hierarchically clustered and visualized by combining map mashups with cartographic processes and methods. This framework enables users to interactively index and browses individual information items. Furthermore, the association relationships involved in the information set are explored in our work. With knowledge graphs that are automatically generated based on the information model, a high-level understanding of the information set can be obtained by users.

INDEX TERMS Geotagging, information visualization, location-referenced information, map mashups, web textual information.

I. INTRODUCTION

The mainstream digital resources used in various web applications and social media, such as the text, images and video, may implicitly contain abundant geographic location content [1]–[3], and textual resources are the most popular of these digital resources. Studies show that more than 57% of data reference locations or have geographic descriptions [1]; thus, geographic content is essential for the formation of a comprehensive spatial understanding of web content. However, most traditional web interfaces do not consider these implicit locations and instead list the digital items sequentially on a conventional linear interface; thus, spatial information is neglected. Because the map is the carrier of geographic content, if the location information is presented along with its corresponding geographic reference on a map interface, more intuitive spatial knowledge can be obtained.

Creating knowledge by combining information and services from different sources is known as a mashup. Mashups first appeared in 2004 as a result of the application of social media and Web 2.0 technology [4]–[7], and map mashups quickly became the most popular form of mashup [8]–[10]. Map mashups combine (or ''mash up'') multiple sources of data that are displayed in some geographic form [11]. With the emergence of Web Mapping 2.0, geo-browsing activities increased quickly [12], leading to the rapid development of map mashups. In these map mashups applications, web maps generally ''function as an interface or index to additional information'' [13], [14]. On this basis, the up-to-date, dynamic, and interactive presentation and dissemination of various geospatial information can be implemented effectively.

The associate editor coordinating the review of this manuscript and approving it for publication was Waleed Alsabhan.

Many related studies have been performed to achieve map-based information visualization or map mashups, and some of the studies are devoted to locating and tagging the geographic coordinates of the web content to extract the implicit locations [15], as geographic locations are usually implicitly contained in location-referenced information. These studies are known as geotagging, geo-referencing or geocoding in the field of geographic information system [16]–[18], most of which consider textual data. Amitay *et al.* [19] proposed a system named Web-a-Where to identify the geographic focus of a document using a hierarchical gazetteer and a simple scoring algorithm. Blessing *et al.* [20] used named entity recognition (NER) and a knowledgebase gazetteer to recognize and locate the geographic names for German content. Kordopatis-Zilos *et al.* [21], [22] refined language models for geotagging social media content based on text. Various geotagging tasks for different types of digital resources have been conducted in an international benchmarking initiative named MediaEval [23].

Based on these geotagging methods, map mashup applications that use geographic locations as the medium to present and index information have been explored by researchers. Ahern *et al.* [24] built an application system called World Explorer that analyzes the location-referenced textual tags associated with Flickr images and aggregated and presented these images with the map interface. Carmo *et al.* [25] completed MoViSys, a map visualization system for geo-referenced data, organized by categories with intuitive and adaptive icons for mobile devices. Teitler *et al.* [26] and Sankaranarayanan *et al.* [27] proposed automated systems to associate news articles or Twitter messages with geographic locations and displayed location-referenced information based on the map interface. Gao *et al.* [28] proposed the NewsViews system, which leverages text mining to identify key concepts and locations discussed in articles and to automatically create visually thematic maps based on an extensive repository of databases. Troudi *et al.* [29] detected events from social media by using descriptive dimensions as the topic, time, and location and developed a mashup-based framework to present detailed event information.

The use of geotagging technologies and map mashup applications offers a new perspective for browsing and perceiving location-referenced information. However, the existing studies are insufficient in terms of several factors. First, the existing studies mainly focus on the acquisition and presentation of location content, and the comprehensive resolution and modeling of the information that the temporal and semantic content should be considered has not been sufficiently explored. Second, regarding the visualization strategies of these studies, cartographic processing of map mashups is lacking, and the related cartographic principles and solutions have not been utilized. Because of the richness of location-referenced information, the quality of the visual representation and man-machine interaction of the map mashups cannot be guaranteed. In addition, because the multidimensional content of the information has not been explored, visualization is limited to roughly present the location of information; thus, the further level knowledge of the information set has not been thoroughly extracted or displayed.

This paper proposes a novel modeling and visualization framework for location-referenced web textual information. In this framework, a location-referenced information model is established to comprehensively resolve and extract basic spatial, temporal, and semantic elements such as the location, time, and participant (person, organization) from the information items. The information processed in the study mainly exists as text because, for other types of information, to enable automatic interpretation using a computer, the information content is generally processed and extracted to the form of text. Based on the constructed information model, hierarchical clustering of the information is applied, and a visualization strategy that combines map mashups with cartographic processes so that users can interactively index and browse the clustered information is developed. In addition, the association relationships between the entities (locations, persons, etc.) in the information are explored according to their co-occurrences in the information model to automatically generate the corresponding knowledge graphs, which enable the users to obtain entity-level knowledge from the information set.

This paper is organized as follows. Section II describes modeling, information processing, and the visualization framework in detail, while Section III provides the experiments and the evaluation of the result. Finally, Section IV discusses the results, and a summary of the conclusions is presented.

II. METHODOLOGY

To address the limitations mentioned previously, we propose an information processing and map mashup framework to meet the needs of the users by providing a convenient way to interactively browse and index location-referenced web textual information and obtain knowledge graphs of the information set on a map interface. The detailed architecture is presented in Figure 1.

Through this framework, we crawl and model the location-referenced information from web sources. By resolving the web textual content, the basic elements of the information, including the time, location, activity, and participant, are analyzed and extracted using natural language processing (NLP) and geotagging techniques; the location-referenced information model is constructed accordingly. Then, the information items are hierarchically clustered, and the relationships between the entities in the information are explored according to the information model. Finally, based on the map mashup strategy, the integrated information items are effectively presented on a map interface using different labeling modes. Meanwhile, the knowledge graphs of the relationships between the entities involved in the information set can also be obtained.

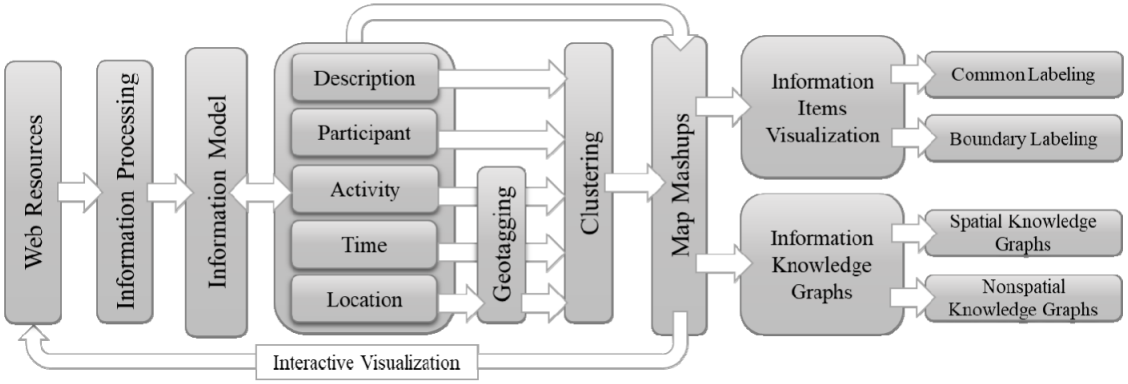

FIGURE 1. General framework for the information processing and map mashup visualization.

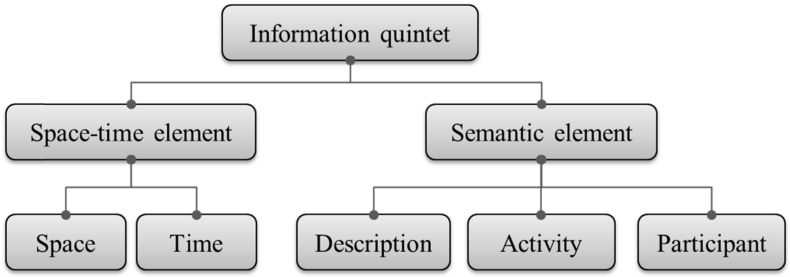

FIGURE 2. Location-referenced information model illustration.

A. LOCATION-REFERENCED INFORMATION RESOLUTION

1) LOCATION-REFERENCED INFORMATION MODEL

With the swift development of the internet, diverse, massive, and heterogeneous information has been generated and transmitted in different forms, such as news, articles, tweets, and blogs. To identify and describe the information precisely, basic multidimensional elements, including spatial, temporal, and semantic content, are necessary. Because information is usually presented in an unstructured form, the basic elements of the information should be analyzed, extracted, and modeled prior to utilization. In this paper, various types of web data are abstracted into an information model, i.e., {*Location*, *Time*, *Participant*, *Activity*, *Description*}, which is denoted with the tuple:

$$Q =< L, T, P, A, D > \tag{1}$$

where Q is the quintet for the information item; L and T stand for the specific location and time related to the information, respectively; P represents the agent (organization, person, etc.) associated with the information; A is the action or activity illustrated in the information; and D is the overall description of the information, which uses semantic primitives [30] to define and identify the specific information item. In this research, the lexical patterns of nominal and verb phrases are considered semantic primitives. The information model is illustrated in Figure 2.

2) INFORMATION SEMANTIC INTERPRETATION

Based on the abovementioned model, the information could be further extracted and modeled. The information on the internet takes different forms, such as text, video, audio, and graphics, among which, text occupies the largest proportion. For other types of information, to enable automatic recognition and processing using a computer, the information is generally mined and extracted to the form of text. Therefore, this article focuses on the interpretation of web textual information; other types of information can be processed in a similar way by applying information preprocessing based on semantic data mining techniques.

To extract information from free text, NLP is most effective [31], [32]. In this research, the Stanford CoreNLP toolkit [33] is used for text processing such as sentence splitting, NER, lexical analysis, and part of speech (POS) tagging to preprocess the information to obtain the basic semantics for the information model. The main steps of textual information preprocessing are as follows, and an example is illustrated in Table 1.

(1) The textual information is inputted and segmented into single sentences;
(2) For each sentence, NER, lemmatization, and POS tagging (see Table 2) are performed;
(3) To discard the unremarkable terms in the sentences, the well-known Term Frequency-Inverse Document Frequency (TF-IDF) [34] score for each term is

TABLE 1. Location-referenced information preprocessing example.

Processed sentence	Oxford professor Tariq Ramadan is charged with the alleged rape of two women and jailed in Paris.																
Sentence analysis result	**Lemmas**	Oxford	professor	Tariq	Ramadan	be	charge	with	the	allege	rape	of	two	woman	and	jail	in	Paris
	POS	NNP	NN	NNP	NNP	VBZ	VBN	IN	DT	JJ	NN	IN	CD	NNS	CC	VBN	IN	NNP
	NER	Organization		Person														Location
Extracted quintet	***Location**:* Paris									***Participant**:* Oxford | Tariq Ramadan							
	***Time**:* 2018-02-02									***Activity**:* charge | jail							
	***Description**:* Oxford | professor | Tariq Ramadan | charge | rape | women | jail | Paris																

TABLE 2. Summary of the PENN treebank part of speech tag sets in english.

Part of Speech Tag	Abbr	Part of Speech Tag	Abbr	Part of Speech Tag	Abbr
Adjective	JJ	Exclamation	UH	Possessive wh-pronoun	WP$
Adjective comparative	JJR	Existential	EX	Predeterminer	PDT
Adjective superlative	JJS	Foreign word	FW	Proper noun plural	NNPS
Adverb	RB	Gerund	VBG	Proper noun	NNP
Adverb comparative	RBR	List item marker	LS	Symbol	SYM
Adverb superlative	RBS	Modal verb	MD	to	TO
Article	DT	Participle past	VBN	Verb base form	VB
Cardinal number	CD	Particle	RP	Verb present tense	VBP
Common noun plural	NNS	Past tense verb	VBD	Verb 3rd person singular	VBZ
Common noun singular or mass	NN	Personal pronoun	PRP	Wh-determiner	WDT
Conjunction coordinating	CC	Possessive ending	POS	Wh-pronoun	WP
Conjunction subordinating	IN	Possessive pronoun	PRP$	Wh-adverb	WRB

calculated based on the following equation, where $n_{i,j}$ is the number of occurrences of t_i in d_j, $\sum_k n_{k,j}$ is the total number of terms in d_j, and $\left|\left\{j : t_i \in d_j\right\}\right|$ is the number of articles in D that contain t_i. Based on this score, the remarkable terms are selected;

$$TFIDF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log \frac{|D|}{\left|\left\{j : t_i \in d_j\right\}\right|} \tag{2}$$

(4) Based on the results, NER, POS, and lemma tags for the selected terms, multidimensional textual descriptions such as the location name, time, organization, and person for the information model can be generated. The location name geotagging is performed in a subsequent process stage.

3) INFORMATION GEOTAGGING

The abovementioned process extracts most of the content required by the information model, but the geographic location tuple in Equation 1 needs to be confirmed. The geotagging process described earlier unifies the textual location name and the specific geography to allow for the spatial exploration of the information. This paper uses the GeoNames gazetteer database [35] and an existing geotagging method [19], which mainly includes three stages: geographic location assignment, geographic location disambiguation, and geographic focus determination, to complete the information geotagging. And the specific processes of geotagging are as follows:

(1) Geographic location assignment: In this stage, geographic locations are assigned for each of the location names extracted from the information based on the gazetteer.

(2) Geographic location disambiguation: Some location names may have multiple geographic locations associated with them, which causes location ambiguity. Thus, location disambiguation is applied by selecting the most likely set of assignments for each location reference based on hierarchical relations and the confidence scoring algorithm [19].

(3) Geographic focus determination: For each information item, based on the occurrence counts and the corresponding confidence scores of the locations, several locations that are mostly associated with the information are extracted to represent the geographic focus of the page.

After these steps, the information item can be tagged as one or several of the most related and unambiguous geographic locations. Each tagged location is recorded in a hierarchical structure as the country, state (or province), county, city, etc., corresponding to the GeoNames gazetteer; for the example, in Table 1, the location "Paris" is resolved as "Paris/France". Then, the modeling process of the location-referenced information is accomplished, and further operations can be carried out on this basis.

4) LOCATION-REFERENCED INFORMATION CLUSTERING

Due to the amount and diversity of web information, web information is difficult for people to browse and understand. Therefore, the information items must be aggregated and clustered based on spatial, temporal, and semantic

dimensions to realize the comprehensive expression of information according to topic and content. Since web information is constantly being generated in real time; thus, traditional clustering methods for massive incremental information are infeasible because, for every new information item inputted, all information items need to be clustered again, which would lead to unacceptable time and computing costs. Hence, a modified leader-follower clustering [36] strategy is applied in this study to incrementally aggregate the information items into clusters according to their extracted location, time, and semantic elements. The specific clustering process is applied based on the active clusters (with location, time, and semantic centroid). For each new information item, the similarity between the information and the active clusters is measured. Then, the cluster is determined according to the similarity score. The detailed location, time, and semantic similarity measurement are described as follows.

During the clustering process, location is an important factor that must be considered, and information items related to the same location need to be integrated. In our study, Formula 3 is used to measure the location similarity between the information item and the active clusters:

$$L\_sim_k(c,i) = \begin{cases} 1 & If\ i\ and\ c\ have\ the\ same\ location \\ 0 & else. \end{cases} \tag{3}$$

where k stands for the hierarchical level (country, state, county, etc.) of the location. Based on this parameter, information clustering is conducted hierarchically by location level. c refers to the specific active cluster and i refers to the information to be clustered. If i and c have the same location, then the similarity score is set to 1; otherwise, the value is set to 0, which enables the information set to be clustered based on the location factor.

To account for temporal similarity, the Gaussian attenuation function is applied to measure the time difference between the new information item and the active cluster time centroid. The time similarity measurement formula is as follows:

$$T\_sim(c,i) = e^{-\frac{(t_i - t_c)^2}{2\sigma^2}} \tag{4}$$

where c and i refer to the specific active cluster and the information to be clustered, respectively; t_i and t_c are the time element of information i and cluster c, respectively; and σ is the parameter to handle the attenuate rate, which is set to 3 in the formula. Based on this formula, temporal similarity is quantified, and the result values range from 0 to 1, depending on the time difference between the information item and the cluster.

The semantic similarity between the crawled information and the cluster is also measured based on the *Description* tuple from the information model because the *Description* tuple is the overall semantic expression, which involves the participant, activity, and other content. Thus, the cosine similarity function, which calculates the term feature vector similarity between the information and the cluster, is applied:

$$D\_sim(c,i) = \frac{\vec{V}_c \cdot \vec{V}_i}{\left\|\vec{V}_c\right\| \left\|\vec{V}_i\right\|} \tag{5}$$

where the c is the specific active cluster, i is the information to be clustered, and $\vec{V}_c$ and $\vec{V}_i$ are the term feature vectors for information i and cluster c, respectively, which are generated based on the terms from the *Description* tuple and their corresponding TF-IDF scores. Based on this formula, the semantic similarity between the information item and the cluster is calculated, and the result values range from 0 to 1.

With the abovementioned location, time, and semantic similarity measurement, the comprehensive similarity between an information item and a cluster is calculated using Formula 6, and the result values range from 0 to 1. In this formula, c and i are the cluster and the information item, respectively, and k is the location hierarchical level. Based on this formula, the incremental real-time clustering process is performed for each new information item. Specifically, the similarity scores between the new information item and the existing clusters are calculated. Then, the item is added to the most similar cluster if the score is larger than the threshold value (since the value ranges of the time and semantic similarity score are both from 0 to 1, the comprehensive threshold is set as $0.5 \times 0.5 = 0.25$ in this work), and the cluster is updated accordingly; otherwise, a new cluster is created using the information item. In this way, clustering is incrementally performed for each new information item collected in real time.

$$Total\_sim_k(c,i) = L\_sim_k(c,i) \cdot T\_sim(c,i) \cdot D\_sim(c,i) \tag{6}$$

B. MAP MASHUP VISUALIZATION

Based on the abovementioned work, location-referenced information was processed, modeled, and hierarchically clustered. Then the visualization of these structured information items in the map mashup interface is handled in this stage. There are two main concerns in our visualization implementation. First, a concise and intuitive map presentation for the information items should be developed because the fundamental function of a map mashup is for the users to index and browse the location-referenced information on the map interface. Second, the relationships between the entities (locations, persons, etc.) involved in the information set are visualized using knowledge graphs, according to their co-occurrences in the information model.

1) VISUALIZATION OF INFORMATION ITEMS

The essential function of the map mashup is to index or present location-referenced information on a map interface. However, information visualization via map mashups is usually visually chaotic [37], [38] due to the great amount of information items and the lack of cartographic processing. If the presented

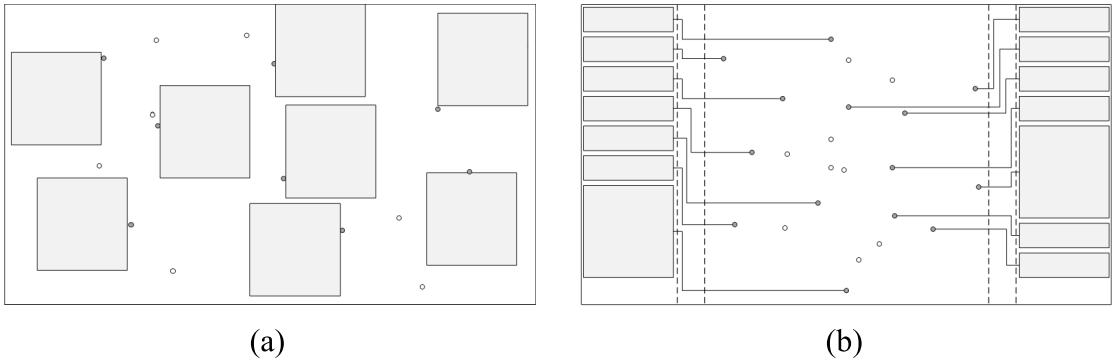

FIGURE 3. Map labeling modes for information schematic diagram. (a) Traditional labeling mode. (b) Boundary labeling mode.

information is too chaotic, the visualization would be worthless. The problem of display massive amounts of information on the map has been handled in cartography for decades, and various labeling methods have provided the most effective for interactive information visualization [39], [40]. Nevertheless, relevant methods have rarely been used to map mashup in previous studies.

Thus, we apply two types of labeling modes (i.e., the traditional labeling mode and the boundary labeling mode) [39], [41] in the interactive and real-time map mashup framework to achieve an intuitive and effective visualization of information items The traditional labeling mode, which has been most widely studied in previous cartographic studies, is often used to place labels on a map. This mode mainly considers the explicit correspondence between the placed items and their corresponding locations on a map; thus, each item is maintained close to its location while avoiding item overlap. A modified traditional labeling method [42] is applied here, and the visualization schematic diagram is shown in Figure 3(a). The boundary labeling mode places the items outside the map area, with the leader lines showing the corresponding relationship between items and locations. In our study, boundary labeling is applied to combine the heuristic search strategy with a boundary labeling method [41], [43]. The boundary labeling mode takes advantage of free space to present information; a schematic diagram is shown in Figure 3(b). These two labeling modes have their own advantages, and they offer users flexible options to conveniently index and browse the location-referenced information items under the map mashups framework.

2) INFORMATION KNOWLEDGE GRAPHS

In addition to map presentations of information items, people are also interested in the association knowledge of the information, which enables them to form a high-level understanding of the information set. With the information model, various entities are extracted and recorded for information items such as locations, organizations, and persons, based on which corresponding knowledge graphs [44], [45] can be automatically constructed. According to whether the entities involved are spatial elements (locations, countries, cities, etc.) or nonspatial elements (organizations, persons, etc.), the generated knowledge graphs can be divided into two categories: spatial knowledge graphs and nonspatial knowledge graphs.

(1) Spatial knowledge graphs: For this type of knowledge graph, the location tuple of information is mainly considered. Specifically, the spatial distribution of the referenced locations within the corresponding information items are presented as the nodes on the map interface, and their association relationships are presented by connecting links according to co-occurrences in the information set.
(2) Nonspatial knowledge graphs: For this visualization type, knowledge graphs are created for nonspatial entities such as organizations and persons related to the information. These entities are compiled as the nodes of the knowledge graph network and connected with lines based on co-occurrences in the information set to present the association relationships between them.

Knowledge graphs are constructed based on the co-occurrences of entities in the information set; the detailed process is shown in Figure 4. For convenience of illustration, we take ''Person'' entities as an example to describe this process, and the information items in the figure are assumed to be a subset that is filtered based on a user's request. For each information item (I1, I2, I3, etc.) in this subset, some ''Person'' from the Participant tuple in the information model (E1, E2, E3, etc.) may be referenced synchronously, which are referred to as co-occurrences between ''Person'' entities. By integrating the co-occurrences of each information item, the overall co-occurrence matrix for the information subset can be generated; accordingly, the co-occurrence knowledge graph between the ''Person'' entities can be automatically constructed.

With the proposed knowledge graphs, users can determine the relationships that exist for a specific information subset of interest, which can focus on a particular period of time or specific locations, organizations, or persons.

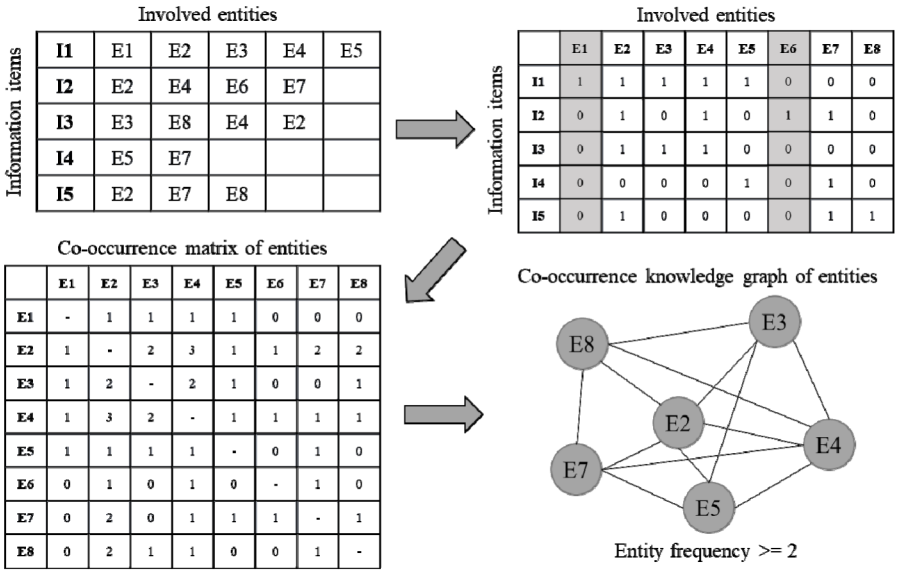

FIGURE 4. Knowledge graph based on co-occurrence analysis process schematic diagram.

III. EXPERIMENTAL VISUALIZATION

With the proposed information processing, modeling, and map mashup framework, the automated information retrieval and display application is implemented. The application is implemented in Java with JSP pages, as well as Stanford CoreNLP, Openlayers, and ECharts APIs. Through this system, we attempt to obtain location-referenced web textual data and to extract multidimensional tuples from the data to form an information model. Then, based on this model, individual information items have been visualized and corresponding knowledge graphs have been created. Practical visualization experiments are conducted in this section.

A. EXPERIMENTAL DATA COLLECTION

To demonstrate the proposed framework and corresponding application, a location-referenced data source is needed. Because online news is a typical web sources that contains massive amounts of geographic locations and complex content, it is an appropriate experimental data source for location-referenced information modeling and map mashup visualization experiments. Therefore, in this step, we use web crawlers and a data acquisition interface to obtain articles from news media sites, including China Daily and Yahoo News, as the experimental data in our application.

Data collection is carried out not only for past news but also for news items posted in real time. In our experiment, the data are collected from January 2017 until October 31, 2018; more than twenty thousand valid news articles collected. With the proposed information processing method, the news articles are handled and extracted to the proposed information model. Then, hierarchical clustering is performed based on the model. The partial clustering result is shown in Table 3, from which statistics for country-level location-referenced information is presented for several representative countries,

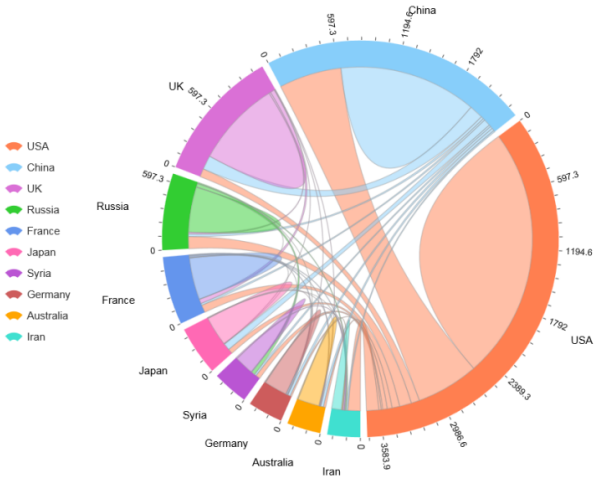

FIGURE 5. Chord diagram of the 10 countries with the most location-referenced information.

including the USA, China, the UK, etc. These countries have the largest number of information items; thus, the article count, cluster count, and the top three clusters are illustrated in the table. The USA occupies the largest proportion in the result, with 3919 articles distributed into 483 clusters, followed by China, the UK, Russia, and France.

The co-occurrences and correlations between the countries for our collected data are further explored, and a chord diagram for the top ten countries with the most location-referenced articles is presented in Figure 5. From this figure, we can see that there are a number of articles that cover more than one geographic location, and the co-occurrences of countries in these articles reflect the degree of interaction between them, i.e., articles related to the United States and China often relate to other countries

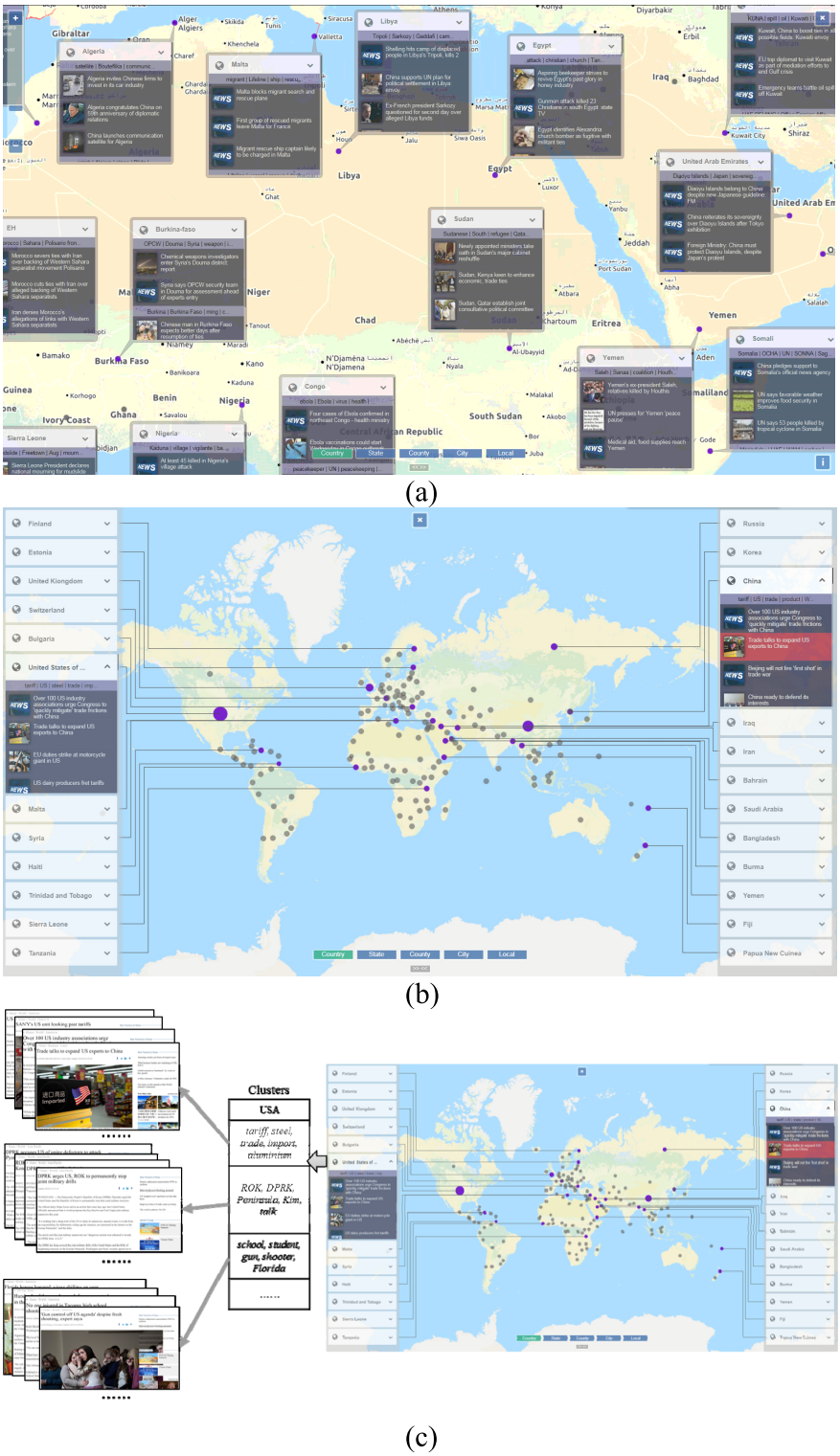

FIGURE 6. Two types of labeling modes for location-referenced information visualization. (a) Traditional labeling mode. (b) Boundary labeling mode. (c) Visual panel and clustered topic schematic diagram.

synchronously, such as the UK, Russia, France, etc. In addition, Syria and Iran are involved in the results because of the high exposure and international attention they have recently received.

Based on these collected data, a large amount of location-referenced information is covered in our experiment, and the interactive visualization for the real-time obtained data can be achieved accordingly.

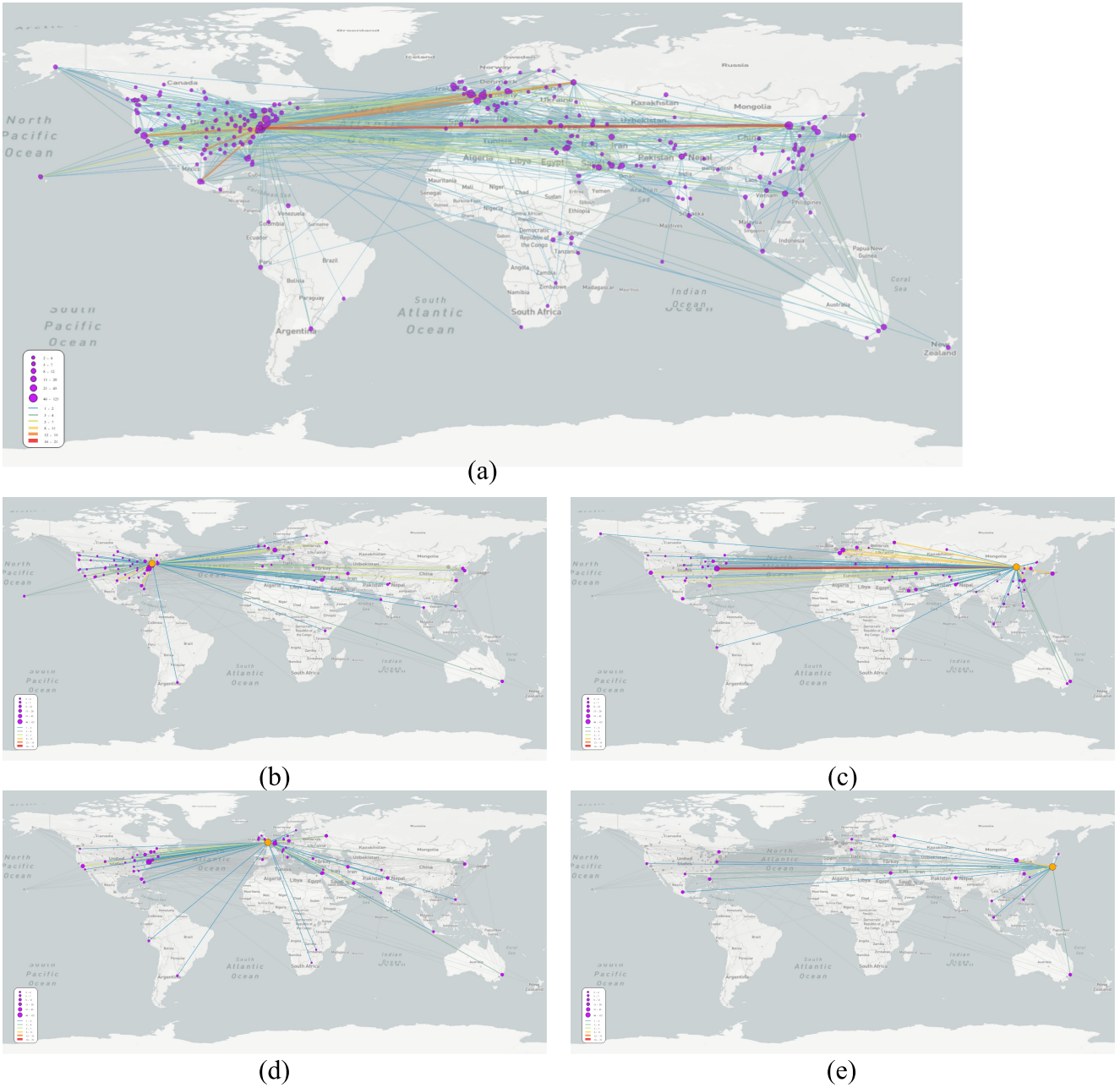

FIGURE 7. Co-occurrence knowledge graph between locations of "trade" activity. (a) Overview of the co-occurrence graph. (b) Co-occurrences for New York, USA. (c) Co-occurrences for Beijing, China. (d) Co-occurrences for London, UK. (e) Co-occurrences for Tokyo, Japan.

B. INFORMATION ITEM VISUALIZATION

The main aim of the designed visualization framework is to use the map interface as the spatial reference to index and browse location-referenced information items. The interface mainly employs a network map and corresponding visual containers to present the information items. Users can pan and zoom through the map interface to interactively browse and retrieve specific information items that have been hierarchically clustered and integrated according to their location, time, and semantic dimensions. Based on the hierarchically structured information set, users can focus on location levels of interest, such as country, state, county, etc.

Some visualizations for country-level locations are presented in Figure 6, and the results for the two types of visualization modes, traditional labeling and boundary labeling, are shown in Figure 6(a) and Figure 6(b), respectively. For a specific location on the map, a visual panel is created to provide an integrated display for the corresponding information items, which are distributed in clusters, as shown in Figure 6(c). Users can swipe up and down the panel

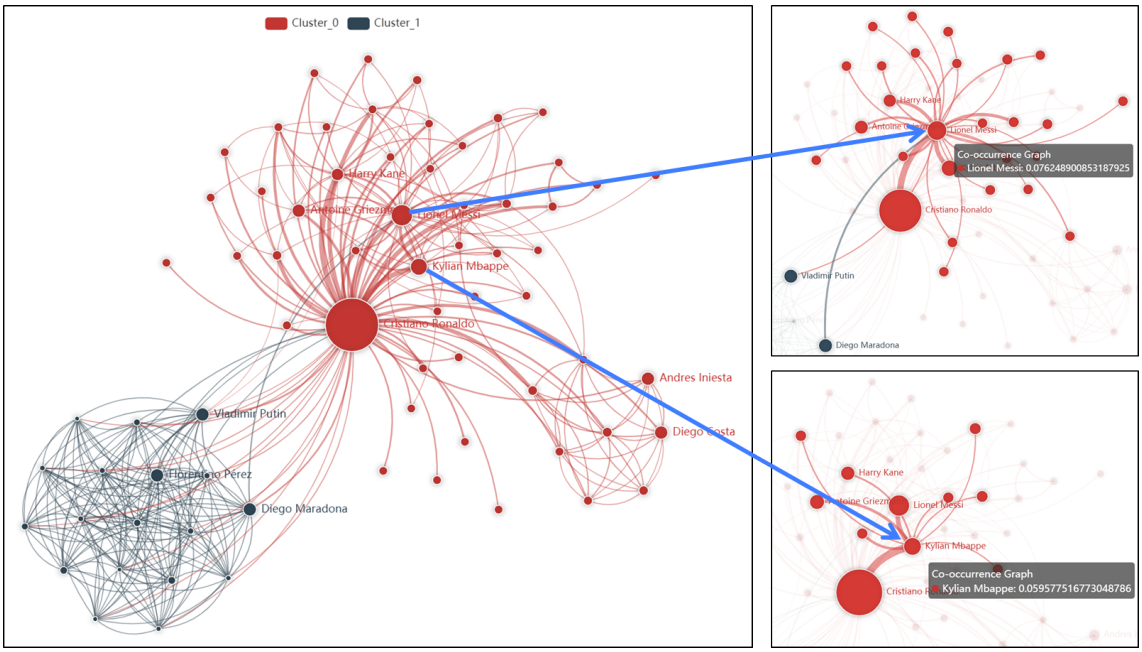

FIGURE 8. Co-occurrence knowledge graph between "cristiano ronaldo" and related persons.

TABLE 3. Articles and cluster statistics for the main countries.

Country Name	Total Article Count	Cluster Count	Main Cluster Keywords	Cluster Article Count
USA	3919	483	tariff, steel, trade, import, aluminum	283
			ROK, DPRK, Peninsula, Kim, talk	148
			school, student, gun, shooter, Florida	127
China	2487	382	tariff, trade, product, WTO, steel	191
			THAAD, Peninsula, DPRK, ROK, negotiate	136
			Africa, industry, cooperation, Tillerson, debt	98
UK	1230	213	terror, police, attack, London, London Bridge	133
			Manchester, police, bomber, concert, attack	83
			Brexit, EU, economy, growth, trade	81
Russia	708	97	agent, expel, poisoning, Britain, diplomat	109
			missile, test, Sarmat, sanction, Saratov	74
			fan, Cup, football, World, open	58
France	629	89	Brexit, EU, Britain, May, border	62
			Macron, Paris, supermarket, hostage, reform	53
			Paris, climate, accord, agreement, Trump	49

to dynamically load and browse information items for this location in an interactive way.

For the traditional labeling mode, the information panels are placed close to the locations; in this way, the corresponding relationship between the location and the information can be clearly presented. However, the information panels occupy map space and may influence the expression of the base map. For boundary labeling mode, because the information panels are placed outside the map space, there is little influence on the base map; however, the correlation between location and information in this mode is not as intuitive as that in the former mode. The two modes have their own advantages, and

both have the ability to instantly refresh the map interface according to pan and zoom operations. Thus, by applying these two labeling modes in the proposed map mashup framework, a more flexible and intuitive way to index and browse location-referenced information items is provided.

C. KNOWLEDGE GRAPH VISUALIZATION

In addition to indexing and browsing individual information items on a map interface, users are interested in the association relationship knowledge contained in the information set. With the proposed framework, the basic elements for each information item are extracted and modeled, which provides the basis to further interpret the relationships between the location, person, organization, etc., entities in the information set. With the method proposed in Section II, spatial and nonspatial knowledge graphs can be generated for various entities in which users are interested.

Knowledge graph visualization examples are shown in Figures 7 and 8 for location and person entities, respectively. The spatial knowledge graph in Figure 7 considers the ''trade'' *Activity* between various locations around the world. The graph is constructed based on the subset retrieved from the information set with the search criteria as follows: the *Time* tuple is from 2017/1/1 to 2018/10/31, and the *Activity*tuple is ''trade''. Thus, the network is constructed with the locations as the nodes (with different symbol sizes to reflect the occurrence frequencies) and the co-occurrences between locations as the edges (with different colors and widths to reflect the connection intensity between entities according to their co-occurrence frequencies), as shown in Figure 7(a). The ''trade'' condition for some locations can be interactively viewed by hovering the mouse cursor over the location, and the ''trade'' condition views for New York, Beijing, London, and Tokyo are presented in Figure 7(b)-(e), respectively.

The nonspatial knowledge graph example is presented in Figure 8. The information subset for this result is retrieved with the following search criteria: the *Time* tuple is from 2018/5/1 to 2018/10/31 and the *Participant* tuple is ''Cristiano Ronaldo''. The co-occurrence relationships between soccer player Cristiano Ronaldo and other related persons, such as Lionel Messi, Florentino, and Putin, are illustrated in the figure. From this type of knowledge graph, users can intuitively determine implicit relationships for specific persons, organizations, etc., using the visualization result. Thus, the visualization examples show that the knowledge graphs in our framework provide an intuitive and effective perspective to help users gain insight into the entity-level knowledge in the information set.

IV. CONCLUSIONS

In this research, we propose a novel information processing, modeling, and map mashup visualization framework for location-referenced web resources. The framework attempts to resolve and model real-time collected web textual information from spatial, temporal, and semantic dimensions in real time using NLP, geotagging, and hierarchical clustering methods. Based on the quintet information model, the extracted spatial, temporal, and semantic elements provide deep-level content that can intuitively and comprehensively facilitate the understanding of the information. Then, the visualization application is implemented to combine map mashups with cartographic methods to automatically generate knowledge graphs, which allows users to interactively index and browse the individual information items and to further view the overall distribution and correlations of various entities in the information set.

The proposed framework is an effective and feasible architecture for general location-referenced information. Customized extensions can be developed for specific domains to obtain more targeted and particular information models along with the corresponding visualization patterns. In addition, while the framework is currently limited to textual information, the semantic resolution of other forms of media, such as video, audio, graphs, and images, could be explored to extend the ability of the system to parse various types of information. These considerations present several directions for future work.

REFERENCES

[1] S. Hahmann and D. Burghardt, ''How much information is geospatially referenced? Networks and cognition,'' *Int. J. Geograph. Inf. Sci.*, vol. 27, no. 6, pp. 1171–1189, Jun. 2013.

[2] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, ''Analysis of geographic queries in a search engine log,'' in *Proc. 1st Int. Workshop Location Web*, Beijing, China, 2008, pp. 49–56.

[3] S. Aloteibi and M. Sanderson, ''Analyzing geographic query reformulation: An exploratory study,'' *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 1, pp. 13–24, Jan. 2014.

[4] D. Butler, ''Mashups mix data into global service,'' *Nature*, vol. 439, p. 6, Jan. 2006.

[5] D. E. Simmen, M. Altinel, V. Markl, S. Padmanabhan, and A. Singh, ''Damia: Data mashups for intranet applications,'' in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Vancouver, BC, Canada, 2008, pp. 1171–1182.

[6] B. Beemer and D. Gregg, ''Mashups: A literature review and classification framework,'' *Future Internet*, vol. 1, no. 1, pp. 59–87, Dec. 2009.

[7] F. Daniel, M. Matera, and M. Weiss, ''Next in mashup development: User-created apps on the Web,'' *IT Prof.*, vol. 13, no. 5, pp. 22–29, Sep./Oct. 2011.

[8] S. Li and J. Gong, ''Mashup: A new way of providing Web mapping/GIS services,'' in *Proc. ISPRS Congr. Commission*, Beijing, China, 2008, pp. 639–649.

[9] L. R. Johnston and K. L. Jensen, ''MapHappy: A user-centered interface to library map collections via a Google maps 'mashup','' *J. Map Geography Libraries*, vol. 5, no. 2, pp. 114–130, Jun. 2009.

[10] M. Batty, A. Hudson-Smith, R. Milton, and A. Crooks, ''Map mashups, Web 2.0 and the GIS revolution,'' *Ann. GIS*, vol. 16, no. 1, pp. 1–13, Apr. 2010.

[11] S. B. Liu and L. Palen, ''The new cartographers: Crisis map mashups and the emergence of neogeographic practice,'' *Cartography Geograph. Inf. Sci.*, vol. 37, no. 1, pp. 69–90, Jan. 2010.

[12] M. Haklay, A. Singleton, and C. Parker, ''Web mapping 2.0: The neogeography of the geoWeb,'' *Geography Compass*, vol. 2, no. 6, pp. 2011–2039, Nov. 2008.

[13] M.-J. Kraak, ''Trends in cartography,'' in *Web Cartography: Developments Prospects*, M.-J. Kraak and A. Brown, Eds. London, U.K.: Taylor & Francis, 2001, pp. 9–19.

[14] M.-J. Kraak, ''Setting and needs for web cartography,'' in *Web Cartography: Developments Prospects*, M.-J. Kraak and A. Brown, Eds. London, U.K.: Taylor & Francis, 2001, pp. 1–7.

[15] J. Luo, D. Joshi, J. Yu, and A. Gallagher, ''Geotagging in multimedia and computer vision—A survey,'' *Multimedia Tools Appl.*, vol. 51, no. 1, pp. 187–211, Jan. 2011.

[16] J. A. McElroy, P. L. Remington, A. Trentham-Dietz, S. A. Robert, and P. A. Newcomb, "Geocoding addresses from a large population-based study: Lessons learned," *Epidemiology*, vol. 14, no. 4, pp. 399–407, Jul. 2003.
[17] M. Larson *et al.*, "Automatic tagging and geotagging in video collections and communities," in *Proc. 1st ACM Int. Conf. Multimedia Retr.*, Trento, Italy, 2011, p. 51.
[18] P. Nesi, G. Pantaleo, and M. Tenti, "Ge(o)Lo(cator): Geographic information extraction from unstructured text data and Web documents," in *Proc. 9th Int. Workshop Semantic Social Media Adaptation Person.*, Corfu, Greece, Nov. 2014, pp. 60–65.
[19] E. Amitay, N. Har'El, R. Sivan, and A. Soffer, "Web-a-where: Geotagging Web content," in *Proc. 27th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Sheffield, U.K., 2004, pp. 273–280.
[20] A. Blessing, R. Kuntz, and H. Schütze, "Towards a context model driven German geo-tagging system," in *Proc. 4th ACM Workshop Geograph. Inf. Retr.*, Lisbon, Portugal, 2007, pp. 25–30.
[21] G. Kordopatis-Zilos, S. Papadopoulos, and Y. Kompatsiaris, "Geotagging social media content with a refined language modelling approach," in *Proc. Pacific–Asia Workshop Intell. Secur. Inform.*, Cham, Switzerland, 2015, pp. 21–40.
[22] G. Kordopatis-Zilos, S. Papadopoulos, and I. Kompatsiaris, "Geotagging text content with language models and feature mining," *Proc. IEEE*, vol. 105, no. 10, pp. 1971–1986, Oct. 2017.
[23] M. Larson *et al.*, "The benchmark as a research catalyst: Charting the progress of geo-prediction for social multimedia," in *Multimodal Location Estimation of Videos and Images*, J. Choi and G. Friedland, Eds. Cham, Switzerland: Springer, 2015, pp. 5–40.
[24] S. Ahern, M. Naaman, R. Nair, and J. H.-I. Yang, "World explorer: Visualizing aggregate data from unstructured text in geo-referenced collections," in *Proc. 7th ACM/IEEE-CS Joint Conf. Digit. Libraries*, Vancouver, BC, Canada, 2007, pp. 1–10.
[25] M. B. Carmo, A. P. Afonso, P. P. de Matos, and A. Vaz, "MoViSys—A visualization system for geo-referenced information on mobile devices," in *Proc. Int. Conf. Adv. Vis. Inf. Syst.*, Berlin, Germany, 2008, pp. 167–178.
[26] B. E. Teitler, M. D. Lieberman, D. Panozzo, J. Sankaranarayanan, H. Samet, and J. Sperling, "NewsStand: A new view on news," in *Proc. 16th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Irvine, CA, USA, 2008, p. 18.
[27] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, "Twitterstand: News in tweets," in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Seattle, WA, USA, 2009, pp. 42–51.
[28] T. Gao, J. R. Hullman, E. Adar, B. Hecht, and N. Diakopoulos, "NewsViews: An automated pipeline for creating custom geovisualizations for news," in *Proc. 32nd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Toronto, ON, Canada, 2014, pp. 3005–3014.
[29] A. Troudi, C. A. Zayani, S. Jamoussi, and I. A. B. Amor, "A new mashup based method for event detection from social media," *Inf. Syst. Frontiers*, vol. 20, no. 5, pp. 981–992, Oct. 2018.
[30] X. Kuai, L. Li, H. Luo, S. Hang, Z. Zhang, and Y. Liu, "Geospatial information categories mapping in a cross-lingual environment: A case study of 'surface water' categories in Chinese and American topographic maps," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 6, p. 90, Jun. 2016.
[31] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.
[32] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol, CA, USA: O'Reilly Media, 2009.
[33] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The stanford coreNLP natural language processing toolkit," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, Syst. Demonstrations*, Baltimore, MD, USA, 2014, pp. 55–60.
[34] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manage.* vol. 24, no. 5, pp. 513–523, 1988.
[35] M. Wick and B. Vatant. (2012). *The Geonames Geographical Database*. [Online]. Available: https://www.geonames.org/
[36] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: Wiley, 2000.
[37] M.-J. Kraak, "Is there a need for neo-cartography?" *Cartography Geographic Inf. Sci.*, vol. 38, no. 2, pp. 73–78, Jan. 2011.
[38] J. Korpi and P. Ahonen-Rainio, "Clutter reduction methods for point symbols in map mashups," *Cartographic J.*, vol. 50, no. 3, pp. 257–265, Aug. 2013.
[39] F. Wagner and A. Wolff, "A combinatorial framework for map labeling," in *Graph Drawing* (Lecture Notes in Computer Science), S. H. Whitesides, Ed. Berlin, Germany: Springer, 1998, pp. 316–331.
[40] M. Luboschik, H. Schumann, and H. Cords, "Particle-based labeling: Fast point-feature labeling without obscuring other visual features," *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 6, pp. 1237–1244, Nov./Dec. 2008.
[41] M. A. Bekos, M. Kaufmann, D. Papadopoulos, and A. Symvonis, "Combining traditional map labeling with boundary labeling," in *Theory and Practice of Computer Science*. Berlin, Germany, 2011, pp. 111–122.
[42] L. Li, H. Zhang, H. Zhu, X. Kuai, and W. Hu, "A labeling model based on the region of movability for point-feature label placement," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 9, p. 159, Sep. 2016.
[43] M. Benkert, H. Haverkort, M. Kroll, and M. Nöllenburg, "Algorithms for multi-criteria one-sided boundary labeling," in *Graph Drawing* (Lecture Notes in Computer Science), S.-H. Hong, T. Nishizeki, and W. Quan, Eds. Berlin, Germany: Springer, 2007, pp. 243–254.
[44] A. Singhal, *Introducing the Knowledge Graph: Things, Not Strings, Official Google Blog*. 2012, p. 5.
[45] J. Pujara, H. Miao, L. Getoor, and W. Cohen, "Knowledge graph identification," in *The Semantic Web* (Lecture Notes in Computer Science), H. Alani, Eds. Berlin, Germany: Springer, 2013, pp. 542–557.

HANG ZHANG received the B.S. degree from the Wuhan University of Technology, China, in 2012. He is currently pursuing the Ph.D. degree in cartography and geographical information engineering with the School of Resource and Environmental Science, Wuhan University. His research interests include geographic information mining, cartography, and map visualization.

LIN LI received the Ph.D. degree from Wuhan University, China, in 1997. He was a Professor with the School of Resource and Environmental Science, Wuhan University. He was with Joseph Fourier University, France, and The University of Tokyo, Japan, for over four years. His current research interests include 3D modeling and visualization, geographical ontology, 3D cadastre, the integration of ubiquitous location information, and feature extraction from point cloud data.

WEI HU received the B.S. degree in geomatics engineering from Northeastern University, China, in 2012. He is currently pursuing the Ph.D. degree in cartography and geographical information system with the School of Resource and Environmental Science, Wuhan University. His research interests include digital cartography models and technologies, web mapping, and spatial and temporal big data analysis.

WENJING YAO received the B.S. degree in geographical information system the from Wuhan University of Technology, China, in 2012. She is currently involved in research related to smart cities and big data mining.

HAIHONG ZHU received the B.S. and M.S. degrees in cartography from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1986 and 1996, respectively, and the Ph.D. degree in cartography from Wuhan University, China, in 2013, where she is currently a Professor and the Ph.D. Advisor with the School of Resource and Environmental Science. Her research interests include navigation digital map, map design, geographical ontology, and the 3D modeling and visualization of geographical information.