Analyzing Existing Unstructured Mountaineering Plans for Machine Readability

Akihiro Nohara* , Shun Shiramatsu* ,
Tadachika Ozono* , Toramatsu Shintani*

Abstract

A mountaineering plan is a diagram or list of steps with timing and resources for climbing, and mountaineering plans are worth accumulating and sharing; however, existing mountaineering plan documents are almost unstructured. The documents are important because they are used in case of mountaineering accidents. In this study, we developed a data model of a machine-readable mountaineering plan to reuse existing mountaineering plans. A machine-readable mountaineering plan helps us to adapt an existing plan to a new one. In this paper, we analyzed existing unstructured mountaineering plan forms to find a common item set for the machine-readable mountaineering plan. We collected various existing unstructured mountaineering plan forms from different sources for the analysis. As a result, we found frequent items in the existing mountaineering plan forms and classified the items into four common parts. Moreover, we developed a mountaineering plan editing support system that can automatically import existing unstructured mountaineering plans as machine-readable ones.

Keywords: mountaineering plan, machine readable, web service, geographical data

1 Introduction

In recent years, an increasing number of people have been enjoying mountaineering as a leisure sport [1]. To ensure safety, climbers make a mountaineering plan (MtP) before mountaineering. A MtP is important for all climbers because it helps ensure their safety; the mountaineering has a substantial risk of loss of life. Usually, a climber needs to make a MtP document, a written MtP, by filling out a MtP form because the document is required to submit for her/his safety. However, making a MtP is time-consuming because MtP information is needed to draw up proper MtPs.

In this study, developing a computer system to assist making MtPs is necessary for the following three reasons. First, since making proper MtPs is time-consuming, climbers do not tend to make detailed plans, endangering themselves. Second, with an automated system, climbers can efficiently collect necessary information from shared MtPs. Third, information necessary for making MtPs tends to be widely distributed, and climbers must

---

* Nagoya Institute of Technology, Aichi, Japan

collect this information; however, currently, many MtPs are not machine-readable. There-
fore, such information collection is tedious and difficult.

We proposed a MtP document converter to import unstructured MtPs with converting
them to structured MtPs [2]. We described the processes for extracting MtP items in a
previous paper [2]. In this paper, we analyze the frequency of items described in MtP doc-
uments and their description methods for generating machine-readable MtPs. Furthermore,
we describe a system for machine-readable MtPs in detail.

We have two goals in this study. First, we examined MtP items to develop machine-
readable MtPs based on the analysis of MtP forms. We therefore need to design MtP
forms for properly processing existing unstructured MtP by computers. Second, we need
to implement an editing support system that (1) automatically converts unstructured MtPs
into machine-readable MtPs and (2) provides a functionality for sharing and reusing such
machine-readable MtPs. Furthermore, we need to expand this system to make a new MtP
by editing current machine-readable MtPs.

This paper is organized as follows. In Section 2, we succinctly define a mountaineering
plan, and we analyze and discuss MtP forms. In Section 3, we explain our system architec-
ture and then describe an evaluation in Section 4. We discuss for important items in MtPs
found by the analysis in Section 5.

2 Analysis of Mountaineering Plan Forms

In this section, we first describe MtPs and second analyze the forms of MtPs.

2.1 Mountaineering Plans

A mountaineering plan (MtP) is a diagram or list of steps with timing and resources (mem-
bers and equipment) used to achieve a mountaineering objective or rescue climbers in case
of mountaineering accidents. A MtP is described in a document called a MtP document.

A climber needs to submit a MtP document to the police. Proper MtPs reduce the po-
tential of mountaineering accidents because climbers with proper MtPs can avoid the risk
of deviating from their mountaineering routes. Furthermore, MtPs have valuable moun-
taineering knowledge including dangers; therefore, sharing this knowledge and other rele-
vant information is valuable for all climbers. Although know-how related to adventure has
been accumulated [3], the information is not sufficiently accumulated as machine-readable
formats. Therefore, we need to accumulate the information. Moreover, it is possible to
widely share knowledge by combining the stored data and the heterogeneous data [4].

Rescue teams use MtPs to estimate the location of victims of accidents. Moreover,
when climbers submit MtPs to the police, rescue teams can find lost climbers more easily.
Climbers face a severe situation if they have an accident. For mountaineering, it is important
to understand the planned route of climbers in the mountains, enabling rescue teams to help
when problems arise. Further, it is difficult to estimate and track the location of the rescue
team and climbers in the mountains [5].

In general, climbers download MtP forms from the police and local government Web
sites and fill in the forms, or climbers use existing climbing support services, e.g. Yamareco
[1]. To make MtPs, climbers obtain information from past MtPs, books, maps, personal blogs,
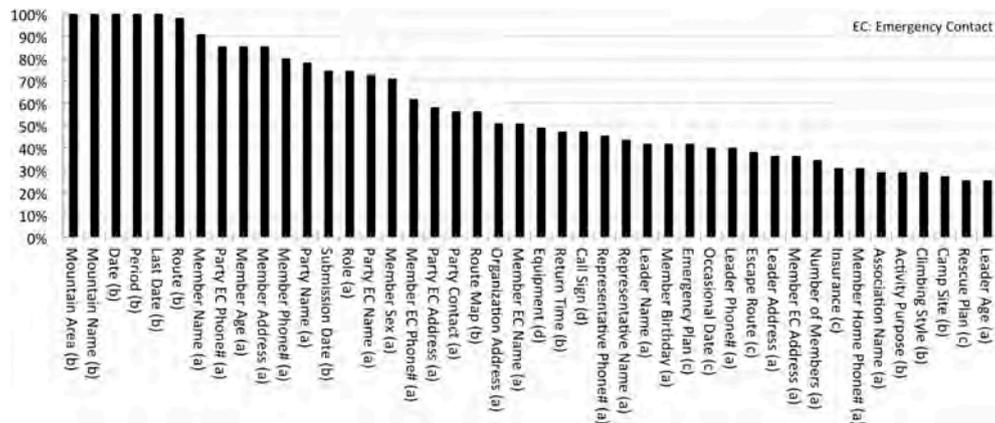
---

[1]http://www.yamareco.com/

Figure 1: Document frequency of items in existing unstructured mountaineering plans form.

Web services, and local governments. Climbers make MtPs by obtaining and comparing necessary information.

The reuse of existing MtPs is beneficial for climbers; however, existing MtPs are largely unstructured. MtPs differ between individual climbers and mountaineering organizations that have their own forms. For example, MtP forms are provided by the Japan Mountaineering Association [2], the police and others, but the actual content remains unstructured. Unstructured MtPs have the following serious problems.

Computers cannot process unstructured MtPs in accordance with the meaning of the terms in the plans. With an unstructured MtP, it is often difficult for a computer to understand the relationships inherent in the available information because each MtP has its own structure. Therefore, they cannot perform any automated data collection or other such operations when using unstructured MtPs.

Furthermore, no information is provided regarding the relationships between the pieces of information within the data. It is often difficult to grasp the relationships between MtPs and related content such as maps. Due to the disorganized and inaccessible information, climbers are typically unable to fully scrutinize information. Therefore, climbers face difficulty in finding related information as they make or compose new MtPs.

To solve these problems, our system performs machine-readable document processing. *Machine readable* is not synonymous with *digitally accessible* information. A computer may not be able to *understand digitally accessible* information; however, a computer is able to understand *machine-readable* data [6]. We need to extract appropriate information from unstructured MtPs to convert them to machine-readable MtPs. Moreover, our system accumulates these plans and effectively displays the information.

2.2 Analysis of Existing Mountaineering Plan Forms

This section explains how we analyzed existing MtP forms for realizing a machine-readable MtP. We analyzed MtP forms on the Web because we need to know frequent items in the existing MtP forms widely used by climbers, it makes possible to find common items for developing a generic data model by using the MtP forms.

---

[2]http://www.jma-sangaku.or.jp/cominfo/

First, we describe how we collected existing MtP forms for the analysis. We manually collected existing MtP forms on the Web sites of the police and mountaineering associations in Japan because a climber basically submits a MtP document using a MtP form provided by the regional police of the target mountain. We checked the police Web sites of all the 47 prefectures in Japan and the Web sites of mountaineering associations found by search engines, e.g. Japan Mountaineering Association.

Second, we explain the collected MtP forms. We manually collected 55 MtP forms from the Web site. Almost the collected MtP forms have different styles. However, the items of the collected MtP forms are similar in meaning. In other words, we found a lot of synonyms on the item names from the MtP forms. Even though the item names are same, requested information of the placeholder are different.

Last, we illustrate common items in the collected MtP forms as the result of the analysis. We analyzed the frequency of items described in the MtP forms and their description methods. We enumerated all items in all the forms and calculated the document frequency of all the items. The document frequency is the number of forms containing an item in the collection. We manually aligned synonyms of the items for disambiguation. Subsequently, we categorized the items into four parts.

Figure 1 shows the rate of the document frequencies of items that is 25% or more. The vertical and horizontal axes show the document frequency and item names, respectively. For instance, the document frequency of "Mountain Name" is 100% as shown in Figure 1; therefore, all the MtP forms include this item.

According to the analysis, a MtP form consists of the following four parts: (a) Party, (b) Schedule, (c) Risk assessment, and (d) Equipment. We classified these items into the four parts manually. The labels $(a)\cdots(d)$ after the item names in Figure 1 indicate the parts.

Part (a), *Party*, contains items of the team and its members, e.g. "Party Name" and "Member Name". Part (b), *Schedule*, has concrete action plans, items of mountaineering dates, camp site and mountaineering route, e.g. "Date", "Camp Site" and "Route Map". Part (c), *Risk assessment*, consists of items about alternative plans in the event of accidents, e.g. "Emergency plan" and "Rescue". Part (d), *Equipment*, has items of mountaineering equipment, e.g. "Equipment" and "Call Sign[3]".

We found the ambiguity of vocabulary about MtP items by this analysis. We disambiguated the items in each MtP form by classifying the item to the labels. Furthermore, by using the synonym, we can effectively extract these items from the unstructured MtPs.

3   Mountaineering Plan Editing Support System

In this section, we explain the functions of our system. First, we describe the system architecture and implementation. Second, we explain our MtP importer. Finally, we describe the user interfaces of our system.

3.1   System Architecture

We implemented the system as a Web application. We implemented the server-side system on the Node.js version 0.10.31 and the client-side system on HTML5 and JavaScript. In addition, we used the Apache POI to translate a MtP document into a text format. Moreover, we used the Mongo DB as a database. In the paper [7], the Geospatial data management

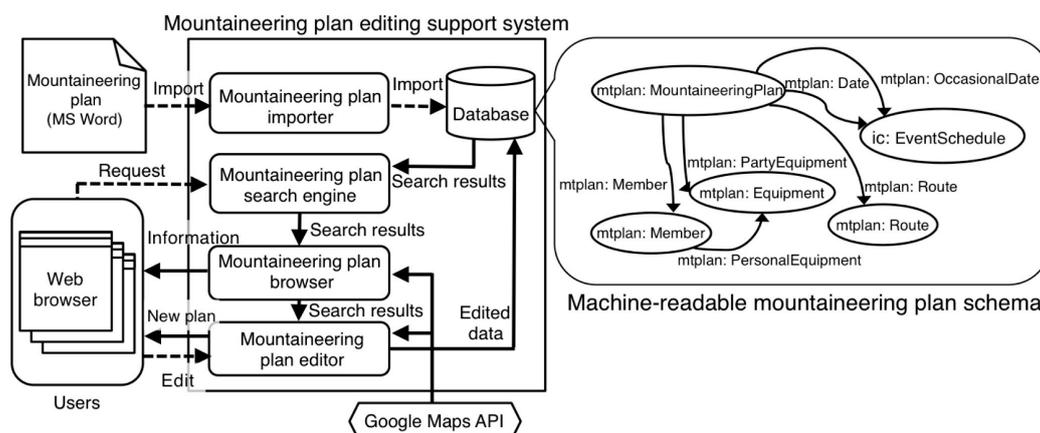---

[3]a unique designation for a transmitting station

Figure 2: System architecture of mountaineering plan editing support system.

and integration of geospatial web services with some APIs are proposed. In our system, we used Web API, e.g. Yamareco Web API[4] and Google Maps API[5].

The left part of Figure 2 shows the system architecture of our system. In our study, we developed the following four modules; a MtP importer, a MtP search engine, a MtP browser, and a MtP editor. In the left part of Figure 2, the solid and the dashed arrows indicate the flow of data and processes, respectively. The MtP importer converts a MtP document in natural languages to a machine-readable MtP and stores the converted MtP into the database. The machine-readable MtP is represented as the data model shown in the right part of Figure 2. The MtP search engine searches the database for the MtPs according to a user's request. The user's request consists of a mountain name, year of a mountaineering, and a difficulty level of a mountaineering. The search engine returns the list of MtPs matched to the queries. The MtP browser displays a MtP in detail, for example, the mountaineering route and each points on the route are plotted on a map with detailed information. The browser also displays with a combination of information from different reference sources [8]. The MtP editor gives a function to add feedback information such as comments of dangerous points on the route after the mountaineering. Furthermore, the editor can also register a new place name to the system by selecting the point on the map.

Based on the analysis of the MtP form described in Section 2, we developed a MtP data model. The right part of Figure 2 shows a part of the MtP data model. The data items defined in the infrastructure for multi-layer interoperability (IMI) [6] are shown as prefix "ic:" as data types. The prefix of data types we defined is "mtplan:." In Figure 2, the node "mtplan: MountaineeringPlan" is a root node and is linked to other nodes with curved solid arrows. Each node belongs to each group described in Section 2. For example, the "mtplan: Member" is included in the "Party" part, and the "mtplan: Route" is included in the "Schedule" part. By using the data model, the system can efficiently process the machine-readable MtPs.

---

[4]https://sites.google.com/site/apiforyamareco/

[5]https://developers.google.com/maps/

[6]http://goikiban.ipa.go.jp/node756

Figure 3: Search interface for our mountaineering plan editing support system.

3.2 Importing Unstructured Mountaineering Plan Documents

In this section, we describe the MtP importer to import machine-readable MtPs.

Extracting mountaineering routes from unstructured MtP documents is arguably the most difficult extraction process to automate. Since a mountaineering route of an existing unstructured MtP document does not include geographic locations, we need to complement the geographic locations at each point. We proposed the MtP document converter to import unstructured MtP documents and converting them to structured MtPs [2]. In this paper, a mountaineering route is an array of place points, where a place point is a tuple that includes the place name, its geographic location (i.e., latitude and longitude), and the scheduled time for mountaineering. The system described by Anantharangachar et al. adds the extracted information to ontology [9]. The development of geo-ontology spatial features and geographic names is the challenge addressed to consolidate aspects of spatial metadata [10]. In this study, we added the extracted information to the MtP data model.

Information extraction processes were evaluated using various techniques including natural language processing, specific part-of-speech tagging, pattern recognition, and annotation [11]. In this study, we used a pattern-matching technique to extract items from MtPs written in a natural language.

Furthermore, we used MtPs made by members of the Nagoya Institute of Technology's Wonder Vogel Club (NITWV). Items described in the MtPs were registered in the system initially. Geographic information used in our study was the point-of-interest (POI) information accumulated in the database. Lacasta et al. crawled the web in search of available geospatial services and enriched their descriptions with concepts from common knowledge organization models [12]. Sedbrook et al. developed the system so that it could semantically enhance knowledge search and retrieval by combining the open source WordNet lexical database of English and the OpenCalais semantic metadata web service [13]. Much of the information we used was obtained from open-data and Web APIs.

3.3 User Interface

In this section, we explain the functions of our system. In the MtP search engine, a user narrows down search results by selecting search items. We implemented this module by reference to Ribeiro-Neto et al. [14]. Figure 3 shows the search results of MtPs. This
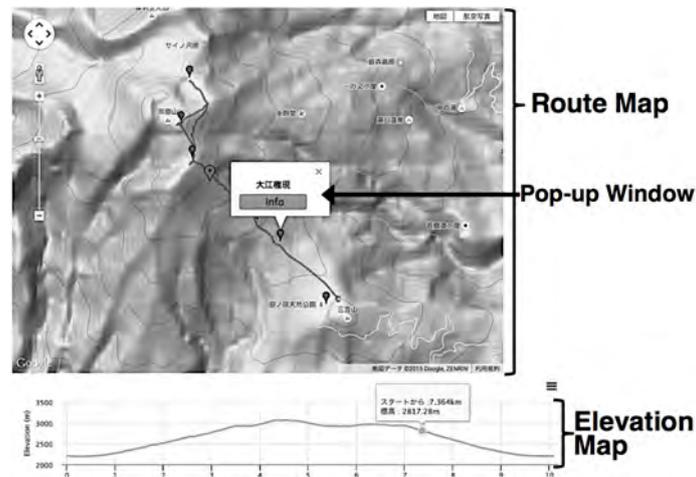
Figure 4: Part of browsing interface for our mountaineering plan editing support system.

system displays a list of search results of MtPs based on a user query; thus, it helps users to compare MtPs. Therefore, a user can grasp the key points of the MtPs at first sight. Each element of the list is composed of three parts: an overview, mountaineering routes on the map, and a mountaineering schedule. The overview of the MtP (①) shows the following points: a mountaineering period, climbers, mountain names, and purposes of mountaineering. The map (②) shows the mountaineering route. Users can easily grasp the mountaineering schedule by visualizing the route on the map. Part ③ shows the schedule of the mountaineering.

Users can access the MtP browser by selecting one of the elements of the list. In this paper, we especially explain the browser of mountaineering route information. The mountaineering route information is shown in the top part of Figure 4. This function provides geographic information such as coordinates, altitude, remarks, and an external site link. Users can browse geographic information, which is presented in a pop-up window of each marker by selecting the marker.

To grasp each point on a map correctly is important in planning mountaineering because of two reasons. First, each of the points on the map is marked as landmarks so as not to get lost. Second, the topographical map gives climbers detailed information regarding the mountaineering. The graph shown at the bottom of Figure 4 is an elevation map. The user can easily grasp the correspondence between the location information on the map and the elevation information because the elevation map is related to the route plotted on the map.

The MtP editor gives users the chance to edit, for example, a user can give feedback information about the mountaineering. In the editing screen, when user selects a point on a map, a pop-up window is displayed on the map. A user enters a title and feedback content in the pop-up window and saves them. This feedback information is displayed in the MtP browser. Users can determine whether the MtP is appropriate by reading the feedback information. Furthermore, the system achieves the sharing of MtPs by using the editor.

4   Evaluation

In this section, we describe an experiment that we performed on our importer using a set of MtP documents. To extract MtPs with correct meaning, the system first needs to extract each component of each MtP. Our evaluation focuses on the necessity of extracting MtP items with the proper meaning because the meaning of words may differ for each item.

We evaluated the extraction performance of our importer using approximately 20 items of high importance out of the total number of items described in MtPs. These items were selected by members of the NITWV as were the MtPs registered in our system. Further, we randomly selected 10 MtP documents for the experiment. These documents were made for the actual mountaineering and accumulated by the NITWV.

It is difficult to widely collect MtP documents having personal information. However, in order to evaluate our importer, the importer needs to adequately extract the information from MtP documents. The MtP documents made by the NITWV have personal information and schedules of mountaineering in detail. Therefore, for the evaluation, we used the actual MtP documents.

We examined whether our system extracted the correct items from a MtP document, evaluating our system by calculating the *precision* and *recall* values. *Precision* is expressed as R/N while *recall* is expressed as R/C. Here, C represents the total number of successful items to be originally extracted, R represents the number of successful items extracted by our system, and N represents the total number of items extracted by our system. The correct data used for comparison in our experiments were created by a member of the NITWV.
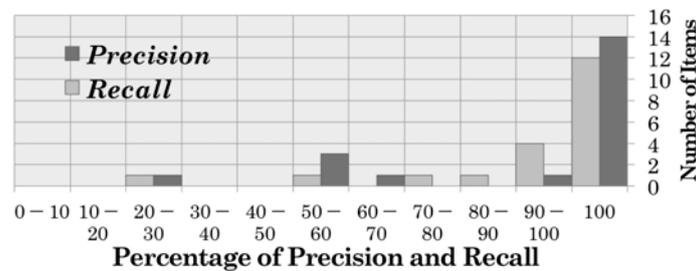

Figure 5: Extraction performance across all items.

Figure 5 summarizes the results of our experiments, showing the distribution of extraction performance across all items. Our experiments consisted of two steps. First, we created correct answer data of the 10 MtP documents manually. Second, we calculated *precision* and *recall* for each MtP item extracted from MtP documents. We then evaluated manually whether the correct items were properly extracted.

The overall *precision* and *recall* of our importer were 82.3% and 87.9%, respectively. Moreover, the experimental results regarding the distribution of items with various *precision* and *recall* values are shown in Figure 5. As shown in this graph, many items can be extracted with 100% performance by the importer

By extracting the items from MtP documents in accordance with the data set described in Section 2, the MtPs used in this experiment are suitable for importing into the system.

5   Discussion

One of our contribution is the analysis described in Section 2. We analyzed MtP forms collected from the police and local government Web sites in Japan. By the analysis, we found the common important items in MtP forms and the various description methods of these items. In addition, we disambiguated MtP items. Furthermore, we enriched the vocabulary of MtPs on the basis of the common important items. The analysis described in Section 2 is beneficial to the development of the MtP data model. The police recommend submitting a MtP document to the police. Therefore, MtP forms we collected are likely to contain items needed to develop the MtP data model. According to the analysis, a MtP consists of the following four parts: (a) Party, (b) Schedule, (c) Risk assessment, and (d) Equipment. By classifying and layering each label, we found the meaning of the MtP items that were described in unstructured MtPs. Furthermore, this classification was used to develop the MtP data model. We developed the data model of MtPs based on the analysis. We described the part of the data model in Section 3. By using the data model, we defined the relationship between information. Furthermore, computers can process the information with understanding the meaning of the words by linking the items according to the data model.

The other contribution is that we implemented a MtP editing support system. This system has the four modules: the MtP importer, the MtP search engine, the MtP browser, and the MtP editor. We developed the importer for automatically converting unstructured MtPs to machine-readable MtPs. The importer successfully extracted information from unstructured MtP documents. In other words, we can perform information extraction for each item. It is necessary to extract MtP items with proper meaning because the meaning of words may be different for each item. We demonstrated that existing unstructured MtP documents can be automatically imported, and the overall *precision* and *recall* were 82.3% and 87.9%, respectively. This experiment shows the effectiveness of our importer. Therefore, by utilizing this data model, we can import MtP documents into the system. The search engine and browser were implemented by using information extracted from MtPs in natural language. Furthermore, the editor can provide feedback information to MtPs. A user can evaluate whether the MtP was successful by using feedback information. Furthermore, a user can register a new place name to the system with the editor. This function enriches the vocabulary of Mused to extract information from unstructured MtP documents.

6   Conclusion

We developed a mountaineering plan editing support system that helps climbers to make a MtP document effectively. Our main contribution is to find common items for achieving machine-readable MtPs by analyzing 55 existing unstructured MtP forms widely collected in Japan. Our analysis provides significant insight into finding the four important common parts of MtPs. Furthermore, we demonstrated that existing unstructured MtP documents can be automatically imported, and the overall *precision* and *recall* were 82.3% and 87.9%, respectively.

Acknowledgments

References

[1] S. H. M. Taher, and S. A. Jamal, "Determinants of mountaineers' decision to climb: An innovative marketing for mountaineering tourism," ICIMTR2012, 2012, pp. 646-651.

[2] A. Nohara, S. Shiramatsu, O. Tadachika, T. Shintani, "A Climbing Plan Sharing System With a Document Converter for Machine-Readable Climbing Plans," ESKM 2015, 2015, pp.97-102.

[3] Shane Winser, "The Royal Geographical Society's Expedition Handbook," Royal Geographical Society (with the Institute of British Geographers), 2004.

[4] Erkimbaev, A. O., Zitserman, V. Y., Kobzev, G. A., Serebrjakov, V. A., Teymurazov, K. B, "Publishing Scientific Data as Linked Open Data," Scientific and Technical Information Processing, 2013, Vol.40, No.4, pp. 253-263.

[5] P. Georgopoulos, B. McCarthy, and C. Edwards, "Location Awareness Rescue System: Support for Mountain Rescue Teams," In: NCA. 2010, pp. 243-246.

[6] "A Primer on Machine Readability for Online Documents and Data," Data.gov. 2012-09-24.   www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data

[7] Cepicky, J., Gnip, P., Kafka, S., Koskova, I., Charvat, K., Nagatsuka, T., Ninomiya, S, "Geospatial data management and integration of geospatial web services," IAALD AFITA WCCA Tokyo, 2008, pp. 175-190 .

[8] P. Vohnout, O. Cerba, S. Kafka, J. Fryml, Z. Krivanek, and S. Holy, "SmartTouristData approach for connecting local and global tourist information systems," IST-Africa Conference Proceedings 2014. IEEE, 2014. pp. 1-6.

[9] R. Anantharangachar, S. Ramani, and S. Rajagopalan, "Ontology Guided Information Extraction from Unstructured Text," International Journal of Web and Semantic Technology, 2013, Vol.4, No.1, pp. 19-36.

[10] Gimenez, P. J., Tanaka, A. K., Baio, F. A, "A geo-ontology to support the semantic integration of geoinformation from the National Spatial Data Infrastructure," XIV GEOINFO, 2013, November, pp. 24-27.

[11] P. Nesi, G. Pantaleo, and M. Tenti, "Ge (o) Lo (cator): Geographic Information Extraction from Unstructured Text Data and Web Documents," SMAP14, 2014. pp. 60-65.

[12] Lacasta, J., Lopez-Pellicer, F. J., Renteria-Agualimpia, W., Nogueras-Iso, J, "Improving the visibility of geospatial data on the Web," In Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries,IEEE Press ,2014, pp. 155-164.

[13] Sedbrook, T., Lightfoot, J. M., "DEAR: A New Technique for Information Extraction and Context-Dependent Text Mining," Communications of the IIMA, 2010, Vol.10, Issue. 3, pp.32-48.

[14] B. Ribeiro-Neto, and R. Baeza-Yates, "Modern information retrieval," ACM Press, 1999.