2015 IIAI 4th International Congress on Advanced Applied Informatics

A Climbing Plan Sharing System with a Document Converter for Machine-Readable Climbing Plans

Akihiro Nohara, Shun Shiramatsu, Tadachika Ozono, Toramatsu Shintani
Department of Computer Science and Engineering
Graduate School of Engineering, Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan
e-mail: {nakihiro, siramatu, ozono, tora}@toralab.org

***Abstract*—In this paper, we developed a climbing plan sharing system. While climbing plan documents created in Microsoft Word or PDF formats contain necessary climbing information, a computer cannot understand such information because the information is written in a natural language. Such climbing plan documents are tedious to write, but are typically required before setting out to climb. Therefore we developed a system that converts climbing plan documents into a machine-readable format, enabling much reuse and sharing of information. We implemented functionality for sharing climbing plans among many users by displaying such plans on a Web page. Our system can therefore help climbers share machine-readable climbing plan documents, progressing the accumulation of climbing information that is difficult to effectively write and share in a natural language.**

I. Introduction

In recent years, increasing number of people are enjoy climbing as a leisure sport [7]. To ensure safety, climbers create climbing plan documents before climbing. Climbing plans are important documents that typically must be submitted to the police; however, developing a climbing plan is time-consuming, because a variety of information is needed to draw up proper climbing plans. Therefore, our goal is to develop a system for sharing machine-readable climbing plans and other relevant information to assist climbers in creating new climbing plans. In this paper, we propose a climbing plan sharing system that uses machine-readable climbing plans.

As noted above, a climbing plan is an important document for all climbers that helps ensure their safety; climbing has a substantial risk of loss of life. Proper climbing plans reduce climbing accidents, because climbers with proper climbing plans can avoid the risk of deviating from their climbing routes. Rescue teams use climbing plans to estimate the location of victims of climbing accidents. When climbers submit their climbing plans to police, rescue teams can find lost climbing more easily. Climbers face a severe situation when they meet with an accident [1]. For climbing, it is important to understand the planned whereabouts 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 [6]. Finally, climbing plan documents include rich climbing knowledge, including dangers; therefore, sharing climbing plans and other relevant information is invaluable for all climbers.

Developing a computer system to assist in creating climbing plan documents is necessary because of the following three reasons. First, since it is time-consuming to create proper climbing plans, climbers do not tend to create detailed plans, endangering their safety. Second, with an automated system, climbers can efficiently collect necessary information to create a climbing plan based on past climbing plans by sharing their climbing plan documents. Third, information necessary for creating climbing plans tends to be widely distributed, and climbers must collect this information; however, current climbing plans are not machine-readable. Therefore such data collection is tedious and difficult.

Given the above problems, our goal is to implement a climbing plan sharing system that (1) automatically converts climbing plan documents written in a natural language into a machine-readable format and (2) provides a means for sharing and accumulating such machine-readable climbing plans. Unfortunately, existing climbing plan documents are largely unstructured. We therefore needed to design a system that can convert existing unstructured climbing plan documents written in a natural language into machine-readable climbing plans. Next, we needed to expand this system to accumulate and share the current machine-readable climbing information.

In addition to this introduction, this paper is organized as follows. In Section II, we succinctly define what a climbing plan document is and how it is used. Next, we discuss and analyze the existing climbing plans in Section III. Section IV provides details about how to translate existing climbing plans into a machine-readable format. In Section V, we explain our system architecture and then describe an evaluation in Section VI. We discuss problems involving information extraction from an unstructured plan document in Section VII. Finally, we conclude our paper in Section VIII.

II. Climbing Plan Documents

The reuse of existing climbing plans is a beneficial solution to help climbers; however, existing climbing plan documents are largely unstructured. Climbing plan documents differ from individual climbers and climbing organizations that have their own climbing plan formats. Templates for the documents of climbing plans are provided by the Japan Mountaineering Association, the police, and others, but the actual content remains unstructured.

978-1-4799-9958-3/15 $31.00 © 2015 IEEE
DOI 10.1109/IIAI-AAI.2015.236

Unstructured climbing plans have the following key problems. Computers cannot process or understand unstructured climbing plans. Therefore, they cannot perform any automated data collection or other such operations using unstructured climbing plans. Further, no information is provided regarding the relationships between the pieces of information within data. Therefore climbers face difficulty in finding related climbing information as they develop new climbing plans. All of this results in a decrease in safety, because climbers are strongly recommended to submit their climbing plans to the police in Japan. The purpose of this is to help expedite rescue operations in case of mountain accidents.

To create climbing plan documents, climbers obtain information from past climbing plan documents, specialized books, maps, personal blogs, and information shared by Web services and local governments. Climbers create climbing plan documents by obtaining and comparing necessary climbing information. When a climber creates a climbing plan written in a natural language, it is often difficult for a computer system to understand the relationships inherent in the available climbing information because each climbing plan has its own structure. Further, it is often difficult to grasp the relationships between climbing plans and related content such as maps. Due to the disorganized and inaccessible information, climbers are typically unable to fully scrutinize climbing information.

To solve these problems, our system performs machine-readable processing. Moreover, our system accumulates machine-readable climbing plan documents and effectively displays the information. Furthermore, our system obtains information from open data and Web APIs. To convert these data to machine-readable climbing plan documents, we need to extract appropriate information from climbing plan documents written in a natural language.

III. Analysis of Existing Climbing Plans

In this section, we analyze the styles of existing climbing plans and discuss relevant data structures of the machine-readable climbing plan document. Our goal is to realize a system that is able to automatically convert climbing plans into machine-readable climbing plans. We also discuss the relationships between items that constitute climbing plan documents.

A. Collecting Climbing Plan Documents

We collected climbing plan templates from police and local government Web sites. In general, climbers download climbing plan documents from these sites to plan by filling out these documents. Collected templates are often fixed in terms of their styles. The amount of climbing information and how such information is entered differ from template to template, because the styles of each climbing plan document are often different, even in each jurisdiction that police have over the various mountains.

A climbing plan document is divided into categories of climbing information. According to the police, it is desirable for a climbing plan document to have the following five items: (1) member information, (2) route of the planned climb, (3) climbing date, period, and schedule, (4) emergency response plan and escape route, and (5) equipment and food. Most police states that the format and description methods for climbing

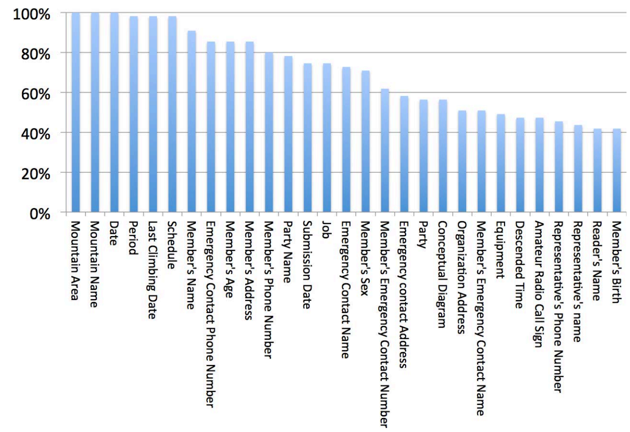

Fig. 1. Percentage of Items Described in Climbing Plans

plan documents are particularly not specified if a climbing plan document has these items. Description methods of items that are not defined are left to climbers to determine the format used. It is therefore difficult for a computer system to extract information from climbing plan documents, and items that are difficult to automatically extract need manual correction.

We analyze the frequency of items described in plans and description methods. Specifically, we analyze items using climbing plan documents available on police Web sites in Japan. Fig.1 shows the results of our analysis wherein the vertical axis shows the proportion of description items of the analyzed climbing plan and the horizontal axis shows the top 29 item names. As an example, 90% of 55 climbing plan documents include the item "Member's Phone Number." From the figure, we observe that items specifying basic climbing information and personal information rank at the top.

B. Items included in Climbing Plans

Items included in climbing plans are roughly divided into the following three groups: (1) overview climbing items, (2) supplementary information items, and (3) descent contact items. Overview climbing items are used to obtain an overview of climbing information for climbers and rescue teams. Examples of overview climbing items include "Mountain Name" or "Member's Name" (see Fig.1). Overview climbing items consist of the most basic items necessary to create climbing plans.

Items of supplementary information represent detailed information regarding local mountains, including such information as geographic data and facilities. These items are used for climbers to help ensure a safe and fun climbing experience. Examples of these items include scheduled time, sightseeing information, timetables of transportation, and hazard point lists of mountain routes.

Descent contact items are used to report that climbers have safely climbed. As an example, a questionnaire item regarding the climb is included here.

As somewhat evident in Fig.1, Overview climbing items were observed to be most important in terms of sharing climbing plans.

IV. Converting Existing Unstructured Climbing Plans into Machine-readable Climbing Plans

In this section, we explain the information extraction steps for converting existing unstructured climbing plan documents into machine-readable information. We consider items to be either easy or difficult for automated extraction and describe a machine-readable method for geographic data that are often difficult to extract.

A. Information Extraction

Items in climbing plan documents are classified into two types—Type 1 and Type 2. Automatic information extraction of Type 1 and Type 2 items is easy and difficult, respectively. In general, Type 2 items can only be semi-automatically extracted.

As an example, Type 1 items include "Member's Name," "Member's Age," and so on (see Fig.1). Notation used for climbing information fortunately does not differ too much when climbers fill in information for Type 1 items. Information extraction by computers can be almost automated since data described in these items are not complicated.

Conversely, Type 2 items are more difficult to extract from climbing plan documents. For example, in many cases, the input field of the "Schedule" item is a freeform text area; thus, the description method is left to climbers. Here, the description method is to write only a generic name of a mountain route or enumerate point names on a mountain route. Thus, the automatic extraction of Type 2 items must be manually assisted by humans, because these items are difficult to extract automatically. To support this, our system has a function to modify extraction content on the system. The system first performs the machine-readable processing of climbing plan documents to combine extraction content described in climbing plan documents with climbing data accumulated in the database. This function then performs information extraction to convert a climbing plan document into a machine-readable format.

Much of the climbing information described in climbing plan documents is written in a natural language. Climbers obtain climbing information by reading what is described in climbing plan documents; however, in a natural language form, climbers cannot query climbing data stored in a database. Therefore, they face difficulties in obtaining and understanding what various plan contents mean.

To solve this problem, a mechanical process that relates data accumulated in the database to the natural language information is needed. To convert a climbing plan into a machine-readable format, a climbing plan document needs to be subdivided into various items that constitute a climbing plan document. Because the semantics of climbing information differ for each item, we must consider the meaning of items to prevent the wrong meaning being extracted from climbing information.

In [3], information extraction processes were evaluated using various techniques, including natural language processing techniques, specifically part-of-speech-tagging, pattern recognition, and annotation. In this study, we used a pattern-matching technique to extract items from climbing plan documents written in a natural language.

Further, in this study, we used climbing plan documents created by members of the Nagoya Institute of Technology's Wonder Vogel club. Item names described in climbing plan documents are registered in the system initially. This system performs the information extraction of climbing information for each of the items after subdividing each item, as noted above.

B. Extracting Climbing Routes

Extracting climbing routes from unstructured climbing plan documents is arguably the most difficult extraction process to automate. A climbing 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 climbing. The system described in [2] adds the extracted information to an ontology. In this study, we add the extracted information to a climbing plan schema that we defined.

Geographic information used in our research was the point-of-interest (POI) information accumulated in the database. Much of the information was obtained from open-data and Web APIs. As noted above, existing climbing information described in climbing plans are written in a natural language; therefore, the information does not have details such as coordinates and links to related Web sites. Moreover, the information has fluctuating levels of quality, depending on the climbers reporting such information.

One of the difficulties in extracting geographic information from existing unstructured climbing plans is disambiguation. Place names often cannot be uniquely determined, because sometimes the same place name identifies different geographic. One such example is "the fifth station on the mountain."

When a climbing route of existing unstructured climbing plans does not include geographic locations, we need to complement the geographic locations at each point. We describe a method to relate climbing information extracted from a climbing plan document written in a natural language to climbing data accumulated in the database. Climbing information extracted from a climbing plan document is insufficient, because it does not contain detailed information such as POIs. Our system has a function that compensates for such missing climbing information.

This system can extract point names from a "Schedule" item, but it cannot find the geographic locations of the points. We aim to develop a point-matching method to match extracted point names with a set of point cases Q to find the location of the point. A point case q is a known point that consists of a mountain area identifier, a place name, and a geographic location (i.e., latitude and longitude). More specifically,

$$Q = \{q_1, q_2, \ldots, q_n\}$$
$$q_i = \langle a_i, w_i, p_i \rangle$$

Here a_i, w_i, and p_i represent a mountain area identifier, a place name, and a geographic location of a point case q_i, respectively.

```
1:  Input :
2:    An area a
3:    A point name list W = (w_1, w_2, ..., w_m)
4:    A point case set Q = {q_1, q_2, ..., q_i, ..., q_n}
5:  Q_a = {}
6:  for i = 1 to |Q| do
7:    if a_i is a then
8:      Q_a ← append(Q_a, q_i)
9:    end if
10: end for
11: Q_a = {⟨a_i, w_i, p_i⟩ | ⟨a_i, w_i, p_i⟩ ∈ Q ∧ a_i = a}
12: O = ()
13: for i = 1 to |Q_a| do
14:   for j = 1 to |W| do
15:     if CompleteMatching(w_i, w_j) then
16:       O ← append(O, ⟨w_j, p_i⟩)
17:     else
18:       O ← append(O, ⟨w_j, NAN⟩)
19:     end if
20:   end for
21: end for
22: return O
```

Fig. 2. Matching algorithm to automatically determine the locations of points (i.e., latitude and longitude) along a climbing path route

Input V of the point-matching method is defined as

$$V = \langle a, \left(w_1, w_2, \ldots, w_j, \ldots, w_m\right)\rangle$$

Here a is the mountain area identifier of a climbing plan, array $\left(w_1, w_2, \ldots, w_j, \ldots, w_m\right)$ is an array of names extracted from a climbing plan, and output O of the method is defined as

$$O = (\langle w_1, p_1\rangle, \langle w_2, p_2\rangle, \ldots, \langle w_m, p_m\rangle)$$

This method attempts to find the location of the point with name w_j. The matching algorithm is shown in Fig.2 and consists of two steps.

The step 1 is described in the lines five to nine in Fig.2. The step 1 is a process to get a POI data set that is selected by the mountain area identifier a. If the mountain area identifier a_i of qi matches to the a, the qi is appended to the Q_a in line seven.

The step 2 is described in the lines eleven to nineteen in Fig.2. The step 2 is a process to create a list of pairs of a point name and geographic data. If a complete matching of the place name w_j extracted from a climbing plan and place name w_i of a geographic data of each q_{a_i} have been successful, the system appends the pair of the place name w_j and the coordinates p_i of the geographic data q_i to output O. If matching fails, the system plug in NAN for coordinates with a place name w_j as pair to output O.

V. Climbing Plan Sharing System

We describe a climbing plan document sharing system. First, we describe the implementation of this system. Second, we describe an outline of the system based on system architecture. Finally, we describe the functions of the system.

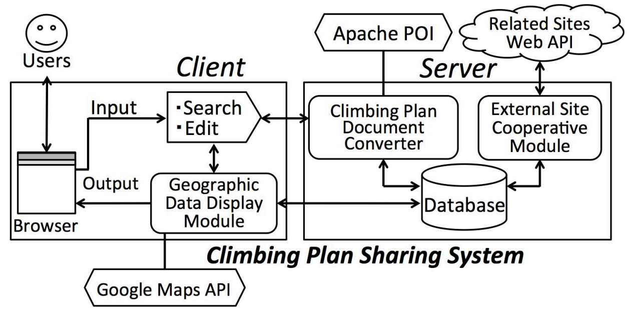

Fig. 3. System Architecture of a Climbing Plan Document Sharing System

A. System Architecture

We have implemented the system as a Web application. This system is a client-server system. We implemented the server-side system on Node.js5 version 0.10.31 and the client-side system on HTML5 and JavaScript. In addition, we implemented using the Apache POI to convert a climbing plan document into a text format and used Mongo DB in the database.

Fig.3 shows the architecture of a climbing plan document sharing system. This system consists of two parts: a server- and client-side system. This system has three modules—a climbing plan document converter, a geographic data display module, and an external site cooperative module.

A client-side system has a module of geographic data display that outputs appropriate climbing information obtained from the database to a browser depending on the input from a user. The module effectively displays climbing information by using Google Maps API.

A server-side system has two modules of a climbing plan document converter and an external site cooperative module. A climbing plan document converter performs processing to convert a climbing plan written in a natural language into a machine-readable climbing plan document. A machine-readable processing is explained in Section IV. A climbing plan that has undergone machine-readable processing by this module is accumulated in the database. An external site cooperative module is a module to get information from external sites, Web APIs and open-data. The module is displayed in a combination of information from different reference sources [5].

B. Climbing Plan Browser

The system can browse climbing information that is obtained by retrieving from the database on a Web browser. The system provides three functions—a searching function, a browsing function, and an editing function. We explained the editing function in Section IV and explain the search and browsing function.

Users can search climbing plan documents by selecting categories. We have implemented this function by reference to [4]. A user narrow down search results by selecting the search conditions of climbing plan documents. The function has three search categories—mountain names, years, and seasons to

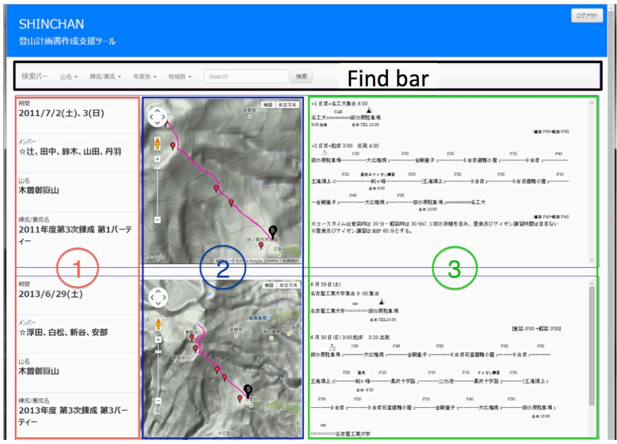

Fig. 4. Search interface for our Climbing Plan Document Sharing System

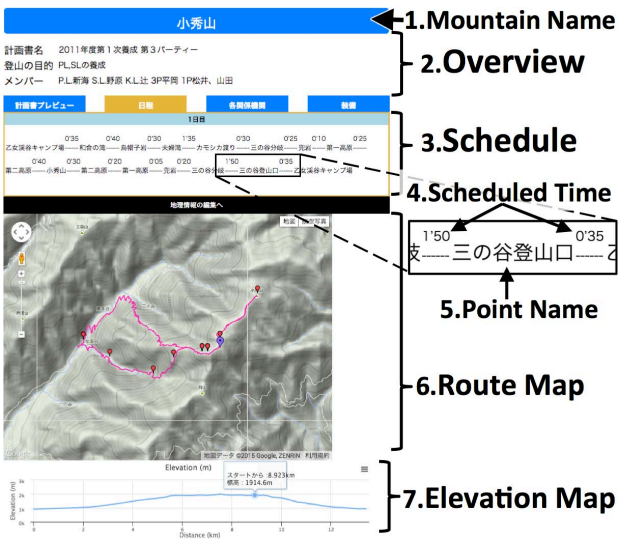

Fig. 5. Browsing interface for our Climbing Plan Document Sharing System

climbing. The search method is effective for users that have been determined climbing plans to an extent.

Fig.4 shows the search results of climbing plans. This system displays a list of search results of climbing plans; thus, it becomes easy to compare the contents of plans. A user can grasp the key points of the climbing plans at the first sight. Each element of the list of search results is composed of three parts—an overview of the climbing plan, a map and a climbing route, and the schedule of the climbing plan. The overview of the climbing plan (①)describes the following points: climbing period, climbers, mountain names, and the purpose to climbing. ② shows climbing routes planned on the map. ③ shows the itinerary of climbing.

Users can access the browsing function of a climbing plan by selecting one of the elements of the list. The browsing function provides two features: a browsing the detail climbing plan and a showing geographic data on a map.

Fig.5 shows an example of the Browsing interface. The system first displays climbing information extracted from the corresponding climbing plan document; such information includes the mountain name, an overview, the schedule, etc. (i.e., the components numbered 1-5 in the figure). Much of the information displayed in the figure for the upper portions numbered 1-3 are initially extracted from a climbing plan document written in natural language; therefore, the displayed information must be realized by converting the climbing plan into a machine-readable format. Further, users can browse the contents of climbing plan documents while checking the relationship between points and place names described in the climbing plan document. This functionality is realized by cooperating with the feature of showing geographic data on a map (described below). In the second feature, information about a climbing route is marked as number six in Fig.5. To browse the climbing plan, it is important to display the schedule on a map. Climbers use various techniques to detect their location in a mountain. While planning for climbing, it is important for climbers to detect the correct locations of each point on climbing routes.

This feature is realized by relating point names to geographic information such as coordinates, altitude, remarks, and a link of external site. Users can browse geographic information, which is described in a pop-up window of each marker by selecting the marker.

Planning the climbing, it is important to correctly grasp each point of schedule on climbing routes. There are two reasons. First, each of the points on the map are marked as landmarks so that the route is not lost. Second, the topographical map on which the climbing route is plotted gives climbers detailed information regarding the climb.

The number seven in Fig.5 shows an elevation map. The user can easily grasp the correspondence between the location information on a map and elevation information because the elevation map is related to a climbing route plotted on a map.

VI. Evaluation

In this section, we describe experiments that we performed on our item extraction system using a set of climbing plan documents. To perform information extraction on the selected climbing plan documents, the system first needs to extract each component of each climbing plan document. Part of our evaluation focuses on the necessity to extract climbing information with the proper meaning, because the meaning of words may differ for each item.

A. Experimental Protocol

We evaluated the extraction performance of our system using approximately 20 items of high importance out of the total number of items described in climbing plan documents. These items were selected by members of the Nagoya Institute of Technology's Wonder Vogel club, as were the climbing plans registered in our system. Further, we evaluated the target from 10 randomly selected climbing plan documents.

We examined whether our system extracted the correct items from a climbing plan document, evaluating our system by calculating *precision* and *recall* measures. *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

TABLE I. Distribution of Extraction Performance Across All Items

Precision	Percentage of Items	*Recall*	Percentage of Items
1.0	70%	1.0	60%
-0.8	5%	-0.8	25%
-0.6	5%	-0.6	5%
-0.4	15%	-0.4	5%
-0.2	5%	-0.2	5%
-0.0	0%	-0.0	0%

extracted, R represents the number of successful items that have been extracted by our system, and N represents the total number of items extracted by our system. Correct data used for comparison in our experiments were created by a member of The Nagoya Institute of Technology's Wonder Vogel club.

B. Experimental Procedure

TABLE.I summarizes the results of our experiments, showing the distribution of extraction performance across all items. Our experiments consisted of two key steps. First, to evaluate our item extraction system, we created correct answer data of 10 climbing plan documents by hand. Second, we calculated *precision* and *recall* for each climbing plan document covering 20 items. We then evaluated by hand whether the correct items where extracted, calculating *precision* and *recall* for each item.

C. Experimental Results

The overall *precision* and the *recall* of our extractor were 82.3% and 87.9%, respectively. Moreover, experimental results regarding the distribution of items with various *precision* and *recall* values are shown in TABLE.I. As an example, the table shows that 70% of the 20 items had *precision* 1.0 and 60% of the 20 items had *recall* 1.0. According to the table, we found a variation in the extraction performance for each item.

Our experiments yielded two key results. First, we revealed various description methods of the fluctuation of description and the items of which names are the same. We therefore need to describe our extraction rules more finely. In addition, we can improve the extraction performance of our system by further defining and enriching a climbing word ontology. By preventing extractions with the wrong meaning and fluctuations in description, it is possible to improve the performance of the entire extractor and therefore the utility of the system.

Second, our item extraction system is effective for climbing plan documents of the Nagoya Institute of Technology's Wonder Vogel club. In the future, we plan to accumulate more information by performing machine-readable processing for various styles of climbing plan documents.

VII. Discussion

Our system successfully extracted climbing information from climbing plan documents written in a natural language. In other words, we performed information extraction for each item, which is a necessary part of our system. Moreover, it is necessary to extract the proper meaning of the extracted climbing information, because the meaning of words may be different for each item.

In the above sections, we described a procedure to relate the climbing data of the database to the climbing information obtained by the information extraction method described in Section IV. There was a problem correctly relating the climbing information to climbing data in the database.

To solve this problem, we described our matching algorithm in Section IV. We solved the problem by matching geographic information in the database to extracted place names written in a natural language. Through this process, computers were able to effectively manipulate the climbing information.

According to the experiments summarized in Section VI, the *precision* and *recall* of our extractor were 82.3% and 87.9%, respectively. We found that there was a variation in the extraction performance for each item. In general, the cause of low extraction performance was that items have had fluctuations in their description and some items had multiple same names.

VIII. Conclusion

The reuse of existing climbing plans would greatly benefit climbers; however, computers cannot easily understand the existing unstructured climbing plans. Therefore, we need a system that can handle machine-readable climbing plans to implement a more effective support system for planning a climb. Our system could provide users too much information relevant to climbing plans in the database. Because our system converted climbing plan documents into a machine-readable format, users can easily obtain information about climbing plans from the database. We analyzed items of 55 climbing plan documents collected from almost all police units in Japan. We then performed information extraction processing to convert these climbing plan documents into a machine-readable format. In addition, we developed a climbing plan document sharing system. The construction and use of machine-readable climbing plans enable users to browse visualized climbing route maps and produce climbing plan documents that will help ensure their safety.

References

[1] A. Nohara, S. Shiramatsu, O. Tadachika, T. Shintani, *A Climbing Planning Support System Considering Machine-readable Climbing Plans*. IEICE technical report, 2014, 114.339: pp. 25-30. (in Japanese).

[2] R. Anantharangachar, S. Ramani, and S. Rajagopalan, *Ontology Guided Information Extraction from Unstructured Text*. International Journal of Web and Semantic Technology (IJWesT), 2013, Vol.4, No.1.

[3] P. Nesi, G. Pantaleo, and M. Tenti, *Ge (o) Lo (cator): Geographic Information Extraction from Unstructured Text Data and Web Documents*. In: 9th International Workshop on Semantic and Social Media Adaptation and Personalization. IEEE, 2014. pp. 60-65.

[4] B. Ribeiro-Neto, and R. Baeza-Yates, *Modern information retrieval*. New York: ACM press, 1999.

[5] 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.

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

[7] S. H. M. Taher, and S. A. Jamal, *Determinants of mountaineers' decision to climb: An innovative marketing for mountaineering tourism*. In: Innovation Management and Technology Research (ICIMTR), 2012 International Conference on. IEEE, 2012. pp. 646-651.