Summarizing Textual Information about Locations

Congxing Cai, Eduard Hovy
{ ccai, hovy } @isi.edu
Information Science Institute
University of Southern California
Marina del Rey, California, USA 90292-6695

ABSTRACT

This paper describes the summarization of textual material about locations in the context of a geo-spatial information display system. Both structured data and unstructured web pages are linked to maps. When the amount of associated textual data is large, it is organized and summarized before display. A hierarchical summarization framework, conditioned on the small space available for display, has been fully implemented. A case study is conducted on a fully implemented system to show how users interact with text in an integration context. Eye-tracking results are presented to illustrate and verify the observations in case study.

1. INTRODUCTION

Geospatial display systems are increasingly gaining attention, given the large amounts of geospatial data and services available online that provide images and maps of various regions. In the process of conflation (aligning different geospatial datasets, such as street maps and other geospatial imagery), the imagery may require some alteration from the original satellite photo to ensure it has the geometric properties of a map [5]. Still, although geospatial imagery and maps show geometric relations among entities, they cannot be used to present other kinds of knowledge about the temporal, topic, and other conceptual relations and entities. Given an entity on a map, a description of what happened there, in what order in time, when, and why, requires additional types of information, typically contained in text (both in structured form, such as tables and lists, and in unstructured free form), in order to support varied search and decision tasks.

In this paper, we describe the integration of maps and geospatial images with potentially large amounts of textual data, and the ways one can organize this material for optimal display. Since much textual data on the web is geographically specific or geo-referenced (containing place names, addresses, real objects with geographical features, etc.), such resources can be treated as a large geospatial database, ready for integration with both static geospatial imagery and dynamically growing information. Effective solutions to multi-modal integration and display desiderata will benefit a variety of applications, including Intelligent Geographic Search (IGS) and Complex Decision Support (CDS), for example:

IGS: Geospatial search systems are quite popular on the web nowadays, and include Google Map, Google Earth, etc. Users can issue a location name or an exact address to locate the place on the map. Efforts are also spent to link the place to the web by jumping to a general web search page\(^2\). However, with the help of integrated text data, more intelligent queries like “MJ’s memorial ceremony place in LA” can be supported, even when it requires additional knowledge from web resources.

CDS: Geospatial systems are often used as decision support tools for business and other analysis tasks. Business decisions are made based on the collection and integration of multiple kinds of data about some area, including market reports, salary distribution, local activities and big events, etc. An integrated display offering different types of data can speed up decision processes.

1.1 Challenges

Of the many different types of text-oriented resources available online, some are structured (e.g., yellow pages or business catalogs extracted from the web), and others unstructured (e.g., news stories, blogs, etc). Some of them can be geocoded accurately, such as business catalogs; others cannot, although they may be geographically specific in some aspects. For example, a news story may refer to an event at some location, or introduce a place with its history and important people. Appropriately linking the specific aspect of the textual information to the exactly relevant pixel of geospatial image is a challenging task.

A second challenge stems from the huge amounts and diversity of web material related to some geographical objects. For example, one may find millions of pages for a famous place or event at a specific map location. To textually elaborate the imagery and maps effectively, instead of simply importing masses of data, one has to organize and summarize the textual information concisely and relevantly. And given the common limitations of display space in most geospatial display systems, one must also design the interface to support dynamic browsing and search.

1.2 Contributions

In this paper, we introduce a hierarchical summarization system to organize and present large amounts of textual data in GeoXray\(^3\) an integrated geo-spatial information system. A case study is conducted to learn how users interact with the text (summaries at different levels) in an integration context. Observations are further illustrated in a preliminary eye-tracking experiment. The work de-

---

\(^2\)Google Map, http://maps.google.com

\(^3\)A product developed by GeoSemble Inc.
scribed in this paper makes the following contributions:

1. a hierarchical summarization framework to reduce displayed text and fully utilize the small display space available for textual information,

2. summarization of news pages to present both information about the location and the relation between the news and the location,

3. demonstration of a real integrated geospatial/text display system deployed on the business domain,

4. a case study to show how users interact with text in an integration context

5. eye-tracking experiments illustrate and verify the findings in case study

The remainder of the paper is organized as follows. Section 2 describes the related work. Section 3 proposes a hierarchical summarization framework to enable organization and presentation of the linked web page content. Section 4 demonstrates our design of the integrated geospatial system to support search and browsing. Section 5 presents various evaluations of the system, and analyzes the evaluation results. Section 6 concludes with future plans.

2. RELATED WORK

2.1 Problems to be addressed

Most existing multimodal geospatial/text display systems [19, 6] present news based on geographical region by identifying the location mentioned in the news, aggregating the news found, and displaying the results over a map interface. [19] identifies the locations mentioned in text using Natural Language Processing (NLP) techniques and then employs a gazetteer list to identify the location focus of each location. However, this method usually applies only at the area level, such as city or suburb presented in [19]; it cannot be used to annotate specific features (buildings, etc.) in satellite images. The problem of smaller granularity requires more-focused georeference.

A second problem relates to scale. Associating each news page individually to its location(s) may overwhelm the amount of information displayable at any point and thereby limit the scalability of the system. Since frequently the associated news contents overlap at least in part, a natural solution is to aggregate the content somehow to remove duplication. A fairly common solution, exemplified in [19] and GeoTracker [6], organizes material (at the area level) by time instead of somehow aggregating over larger numbers of related content.

The system described in this paper integrates geospatial imagery with both structured web data and unstructured news pages right down to the level of individual buildings. It includes technology for text clustering and summarization. It supports keyword search over the text collections. Taken overall, the system provides a more integrated view of the imagery, with associated complex search and decision functionality over the text as well. The geospatial management and display component of our system is developed by GeoSemble Inc.

2.2 Clustering and Summarizing Text

Within geospatial display systems, the space for textual information is often quite limited. We therefore need to summarize the most important and relevant information about each location, drawing from all the web pages linked to it. When several different events have occurred at a location, more than one distinct summary may be needed. It is therefore important to deploy topic recognition [23, 1, 4] and/or topic clustering [18] to identify and group relevant pieces of each text into single-topic “chunks”.

However, given the complexity of many geographic objects, there may be numerous quite different types of associated structured and unstructured information. For example, text pages may be linked to City Hall for a number of reasons. Some pages may provide the history of City Hall as a building; others may describe events occurring there; others may describe its operation as a civic organization. Users typically want to understand the reason a page is linked to an object. Hence it is necessary to provide more than just (simple summaries of) the page contents. It is also necessary to summarize the linking itself, supporting queries like “MJ’s memorial places”, “fun places in Culver City”, etc. Little research has been devoted to this problem to date, although some related work can be found in generating link snippets for web search results [20, 11]. The common strategy (adopted by web search engines, for example) of generating a snippet for each search result by extracting a document fragment that contains the keywords along with the surrounding texts as a “summary” does not work, given the limited space in our system. We need to generate smaller snippets and be more descriptive about the linking.

In our system, all topics and events are detected from the textual information associated with each location. All relevant pages are then clustered based on the topic. Then ‘topic signatures’ [13] can be used as top-level summary for each event associated with the location, and displayed at various levels of elaboration. We discuss our approach in more detail below.

3. TEXT SUMMARIZATION

3.1 Content Extraction and Segmentation

Multi-webpage summarization is different from traditional multi-document summarization. First, most web pages are much more complex than pure text documents. Additional information, including layout and multimedia sources, and more noise, such as advertisements and additional information, all add unnecessary distraction. Since the web contains a combination of types of information—static text, image, videos, dynamic layout, etc.—even a single page can be treated as multiple documents. Current search functions are based on keywords, making the relevant content of each relevant web page only a limited block within the page. Second, our task is oriented to locations, and hence differs from general content summarization. Hence, we need to identify and extract the essential part(s) of the webpage linked to the geospatial imagery for summarization and display. In our work, we utilize two important features, layout and semantics, to identify and extract the relevant content.

By rendering each web page into a DOM tree, it is relatively easy to segment the page into large blocks based on its layout, including header, footer, left bar, right bar, advertisements, main block, etc. We implemented a rule-based extractor to extract the most relevant block from the web page. However, the layout features can only be used to separate major blocks without semantic segmentation. Getting the smaller blocks, such as figure headings, right is very hard, and we know of no satisfactory solution.

3.2 Clustering

Given a list of text blocks relevant to a local point of interest, one can employ traditional text summarization techniques [15, 14, 13]
to produce a short summary for each one. This solution may not be helpful, however, since a long list of pages associated with each point of interest would be very hard for users to browse. Especially when the space allocated to text display by the geospatial system is also limited, a high compression ratio is typically required for the summarization system.

The solution we adopt is to deploy cluster-based multi-document summarization [8, 13]. Clustering must observe two criteria: first, the location of interest, and second, the text topic. (Clustering texts first by topic and then displaying for each topic all its relevant locations over an area is also of course possible, but requires a radically different display and interaction strategy.) Different clustering methods can be employed. To delimit topics, a simple heuristic is to introduce as additional criterion the event/article date: when the difference in document dates within a topical cluster is (far) larger than the actual duration of the topic event, we are probably dealing with multiple separate events at the same location. Better performance is obtained by using a topic and event detection module first, and then clustering documents based on the topics identified [18, 13].

Unfortunately, documents usually contain multiple locations and multiple topics. A single document may pertain to several different events (it is not unusual, for example, for a news article about a robbery to refer to previous robberies at the same location). The problem of ‘topic drift’ can cause confusion in a short summary. One approach is to segment each document into one or more ‘mini-documents’, each one devoted to a single topic, and then to perform location- and topic-based clustering over the (now larger) set of mini-documents. Several NLP techniques are available to perform such segmentation, including Latent Dirichlet Allocation [3], which, given a set of documents, simultaneously identifies the principal clusters of topics (word collections) and computes for each document the (set of) cluster(s) it ‘belongs to’.

For the system, we deployed existing text segmentation techniques [10, 21] to convert the text block(s) extracted for each page into a list of paragraphs, each devoted to a single topic. The paragraphs are then selected based on their relevance to the cluster topic and local interest in question and are treated as independent mini-documents for summarization.

### 3.3 Hierarchical Summary Generation

To generate a summary, there are primarily two types of models: extraction and abstraction. In the extraction model, each sentence in the text(s) is assigned a score for each of several features. The $N$ sentence(s) with the highest merged scores are extracted, arranged, compressed, and returned as summary. These features typically include the presence of certain cue phrases, the absence of certain penalty phrases, the position of the sentence in the text, the presence of headline words in the sentence, and so on [15, 14, 13]. [2] considers generating query relevant summaries by combining two types of scores: relevance to the query and fidelity to represent the original documents. In contrast, the abstraction model produces the summary by “translating” the text into a more concise formulation. The summary is computed as the product of the likelihood of (i) the terms selected for the summary, (ii) the length of the resulting summary, and (iii) the most likely sequencing of the terms in the content set. The extraction model cannot generate short summaries like headlines, but the abstraction model requires large amounts of training data.

Whatever the clustering approach, the result is a potentially rather large set of individual topics associated with each location. Since screen space for the summaries may be very limited next to the maps / imagery, they have to be formatted and presented for maximal interpretability and browsability. To address this problem, we adopt a hierarchical structure to display incrementally longer summaries for each location of interest. At present we have found three levels of incrementally longer summaries to be most useful. This hierarchical structure can help users find and select the materials of interest quickly, even within very limited display space.

**Thumbnail**: a very short ‘summary’ that characterizes the (clusters of) documents or segments associated with each location. We present essentially one or two single keywords — the most informative words for each cluster. We implemented a new version of our topic signature technology [13], one that uses tf-idf instead of the entropy ratio $\lambda$, as scoring measure to rank each cluster’s words.

**Title**: a headline-length phrase or short sentence (or two). The original titles of the web pages are often noisy or even unrelated to the current topic cluster (mostly when the cluster is not the principal topic of the document). Sometimes, the title may be meaningless (it might for example contain the website’s name “Pr Newswire”), or two different web pages may share the same title as shown in Figure 1. We implemented a topic-related headline generator based on our previous work [22] by incorporating a topic-based selector [2].

**Snippet**: a paragraph-length excerpt characterizing the cluster. To produce paragraph-length summaries, we implemented an extraction-based text summarizer. We built a new version of previously investigated technology [14], implementing several sentence scoring techniques and a score combination function. In addition, since high-rated sentences may be extracted from the middle of a document, we investigated the need in this domain for including so-called ‘introducer segments’ that may be rated lower but that are required to establish the context (by for example providing the names of people, organizations, and places).

![Figure 1: Original Titles of Web Pages](image)

Unfortunately directly applying these summary techniques to web pages often produces unreadable summaries. Unlike coherent text, web pages are often more like a loose combination of different segments. Also, each web page may be linked to the location for different reasons. Hence, identifying the location-related segments and eliminating unnecessary noise is essential to the success of the system. This is a topic of ongoing research.

### 3.4 Summarizing the Linking

A web page can be linked to a location of interest for various reasons. For example, an online food review may be linked to the restaurant it reviews, to another location which is the author’s contact address, and even to a third that provides other restaurants of the same chain or the same chef. Different user interests and tasks will require different (types of) linkage. We must therefore summarize the relation between the web page and the interest in the imagery. Hence, neither the most representative summary of the text nor the most relevant summary for the interest is enough. In order to capture the reason why this web page is linked to the interest, we need to capture the discourse structure between the main topic of the news (which necessitated mentioning the location) and the current specific topic of interest to the user.
Inferring from the text the semantic relationship between the text and each location it is linked to is in general beyond the scope of NLP, and certainly beyond the scope of this paper. But we can get partway there by providing:

- the principal initial topic of the text,
- the user’s interest / topic,
- the location and its role in the text.

From these three facts, the user may be able to infer the reason him- or herself.

To achieve this, we first use the extraction model to find both the sentence most representative of the current text and the sentence most relevant to the query. Assuming the original text itself is drafted in a coherent way, the text segment between these two sentences should be informative enough to express the relation(s) between them, and hence between the topic of the text and the current topic of interest. (Ideally, a discourse-level abstract would be used to provide the relation, as illustrated in [17].) In our work, we employ the introducer segments [15] to generate a discourse-level abstract for the linking as shown in Figure 2.

![Figure 2: The Summary of Linkings](image)

This type of linkage summary can be also useful in supporting intelligent geographic search, such as “MJ’s memorial places”, “great restaurants in Culver City”, etc. These queries cannot be handled directly by searching through the business knowledge bases. But they can be issued to the text collection using IR techniques. The results are linked back into the geospacial imagery, and these points of interest on the imagery identify all relevant result pages, including perhaps ones not initially found by the IR query. Such summaries provide a quick description of these locations and help users find their targets quickly.

4. SYSTEM DEMONSTRATION

In this section, we describe our design of the integrated system and demonstrate it.

4.1 Geospatial Interaction

The user can enter an address in the top search box, or search by business name. The system then centers the imagery at that address or business. The user can zoom in, zoom out, or pan on the imagery.

Clicking on “Get Features” invokes the web services to retrieve all features about the displayed image from the GeoKB and displays the features in the “AREA: Features Found” list, and also draws them as points on the maps. The features have been classified into different types, such as Company, Entertainment, Financial Institutions, Parks, Educational Institutions, Gas Stations, etc., and are displayed using a different icon for each type. The user can also select or un-select these categories of features using the “Feature Filter”.

The user can explore the map using the navigation controller. All identified points of interest in the imagery are marked using the feature types. On clicking the marker of an identified building, an information window pops up containing the associated structured web information (building name, business type, website, online images, and so on), as shown in Figure 4.

Clicking on “Get News” retrieves all news related to the displayed features; features with associated news show a small newspaper icon (see next to “Sony Pictures Entertainment” in Figure 5). Clicking on the icon displays the news that was linked with the feature, sorted by date.

To show clustered news, the user can click on the “Cluster News” link. The results are displayed in a tree, showing the title of the cluster (thumbnail or title), under which appears a small summary of the cluster, under which appear links to all the news articles belonging to that cluster.

4.2 Summarization Example

We provide an example of our text summarization system performance in Figure 5. In this example, we have selected the location of Sony Film Studios in Culver City by clicking on the map. Figure 5(a) shows the titles and dates of some of the 126 news articles that contain the words “Sony Pictures Entertainment”. As described above, these documents are clustered based on topics.
Using our current parameter settings, 20 multi-result clusters are formed, leaving 34 results unclustered. (The size of clusters, or the number of clusters desired, can be varied by the user.) As mentioned above, each cluster is presented to the users by a minimal-length thumbnail summary consisting of a few characteristic keywords; a partial list of these is shown in Figure 5(b). Figure 5(c) shows the result of selecting the cluster labeled “solar electrical system” (second from the bottom in Figure 5(b)), which contains two results. The summary contains the 5 top-ranked sentences from the two documents, presented in document order. In addition, the summary includes two hyperlinks to the two full texts for further inspection.

The summary illustrates some of the strengths but also the shortcomings of the current system. It is clearly about a solar energy system installed in 2007 on top of the Jimmy Stewart Building by EI Solutions. This is enough detail for a user to determine whether or not to read the texts any further. However, two of the extracted sentences are not satisfactory: sentence 2 is broken off and sentence 3 should not be part of the news text at all. Premature sentence breaks result from inadequate punctuation and line break processing, which is still a research problem exacerbated by the complexity of web pages.

It is not our focus to evaluate the precision of our summary techniques here; that is a topic of ongoing work. By showing the summary results, we merely demonstrate the improvement on browsability of the search system. We are relatively satisfied with the results. While the summaries are not always very good, they are uniformly understandable and completely adequate to prove that one can combine geospatial information access and text summarization in a usable and coherent manner. Techniques to improve summarization quality are among our current research topics.

5. EVALUATION

5.1 Case Study

The system is quite useful in tasks that require analysis across different media, such as searching for a restaurant that has good online reviews, that is near a theatre with free parking, etc. We next wished to explore how text interacts with other media in the users’ search process and how different levels of text (headline, summary, etc) function in the human information-seeking procedure. Hence, we conducted a case study focusing on the use of integrated data in the search for information.
3. Find the name of the Indian art event at Western Project

Ten graduate students with similar background performed the case study. We compared their behaviors on three different systems: one pure text search system with summarization (TSSM), one integrated geospatial search system without summarization (GS), and one integrated geospatial search system with summarization (GSSM).

Figure 6: Feature-based Search Workflow

5.1.1 Geographical Search Workflow

Given the two kinds of experiment, two distinct workflows were followed.

Feature-based search is looking for a location based on certain requirements. Figure 6 shows the workflow of search supported by our system. Using a pure text search system with summarization, the user can only search the news pages by forming keywords from the requirements. Retrieved pages must be examined one by one to identify the desired location. The short snippet summary of each news item can be used as a hint to filter the news pages for examination. With the help of an integrated geospatial system, the user can find the desired location by scanning all possible locations in the map. For each location, the user can make a judgment based on the associated news pages. A system with summarization will help the user locate the required information much faster.

Location-based search is looking for information about a single location. Figure 7 shows the workflow of location-based search supported by our system. A pure text search system with summarization provides little help. The user can only search for related news pages by forming search terms for the location’s name and address; related pages must be examined one by one to locate actually related information, which is time-consuming. Using an integrated geospatial system, the user can directly find all news pages for the location in question. With the help of a cluster-based summary, the user can find the required information more quickly.

5.1.2 Case Study Findings

Following [12], we selected two straightforward measures as the evaluation metric: time to finish and number of clicks. It is obvious that the less time required to complete the task, the more effective the system. Also, the greater the number of clicks implies

Figure 7: Location-based Search Workflow

the more external information that had to be checked to complete the task. Hence, the more clicks, the less effective the system. We used average search time and average number of clicks to measure user performance in all conditions.

As shown in Figure 8, the integrated system with summarization (GSSM) required fewer clicks for external information than the integrated system without summarization (GS) and pure text search system with summaries (TSSM). Summarization plus integrated media provided sufficient information for users to complete their tasks rapidly; users appeared more willing to explore the system for information. Although they were quicker to locate the target using text search for “Search for Location” queries, this may result in less information than the user really needed. Simple search fits very well in a general search environment, but not in the context of complex analysis and decision support, which requires deeper examination of information.

Also, we see that GSSM performs the best in helping users find information based on the locations. Since the data is linked with summary, it is very easy to help users filter information and obtain a better understanding. For searches of locations, users cannot do deeper exploration and analysis in the pure text search environment, and users spent less time on text search. They felt more satisfied with the integrated systems.

Detailed results for each task are shown in Figure 9. For simple information seeking tasks like SI1, SI2, SI3, the integrated multimedia system helped users find the information more quickly and
required less external information. For tasks requiring deeper analysis, users spent more time using our system but still issued few outsourcing clicks. As discussed, users did not really analyze and explore as deeply in a pure text context as in the integrated environment.

By monitoring user interaction with systems in the case study, we observed several rules to use in future as design guidelines:

Observation 1: Text search helps users locate the information, but text summaries help them understand the information. Users often relied on the search function to locate the texts they were looking for, but they preferred to read the summaries in order to understand the texts. Hence, informative keywords were important for users to locate the right texts, but they were not enough to help users understand the texts. A more readable summary is required for users to really understand the information. While existing research on summarization focuses on extracting the most informative keywords under the required compression rate, it does not guarantee good readability in real applications. This motivates our future research work on improving the readability of summaries in real integrated applications.

Observation 2: People are more willing to read texts in an integration context. Even though many people claim that images present more information than texts, we find users were more willing to read the text in the integrated context in order to understand the different media and their interconnections. Summarization played an important role in guiding user to explore different types of data sources.

Observation 3: The integration of more media requires more interaction in the interface design. Users appeared to be more active in re-forming their queries and varying their interests when working with the integrated multimedia information. In the case study, most of the users frequently switched between the different media and changed their search queries in order to obtain better results. Hence an even more dynamic summarization system is one of our goals for future research.

5.2 Eye-tracking Experiment

Finally, we conducted a preliminary eye-tracking experiment to verify our observations of the case study. Eye-tracking studies are increasingly used as a supplement to traditional user studies in different applications\(^6\) [16, 7, 9]. Researchers [9] use eye-tracking studies in the product development process to understand what users are thinking and what information they are processing to help improve the user experience. Because there are significant differences between when people try to evaluate a system and when users use a system, it is important for designers to recognize what a real user would see on the page. Eye-tracking data tells one what people

\(^6\)Eyetrack III Project: http://www.poynterextra.org/eyetrack2004

focus on the page and what they do not look at. Moreover, it tells one what triggers an action in a specified task, which can be used in design to improve design of action buttons, placement of advertisements, etc.

Eye-trackers very accurately detect and map eye traces. An eye-tracker can capture and report what people are looking at to within a centimeter\(^7\). Considering the human eye’s ability to capture information only small distances around the fixation point, the eye-tracking results are quite accurate in showing exactly what people are viewing. However, one single trace of an individual user is not sufficient to draw any conclusion. Hence, we employ heatmaps to analyze averaged user behavior over a group of users. A heatmap is an aggregate image that groups several individuals’ eye traces on a given page. The pixels are displayed in colors reflecting the percentage of people’s attention: warm colors indicate a greater overlap of attention.

In this experiment, we employed a human video-based eye-tracking system\(^8\) to gather users’ eye movements while they used the integrated system for search and browsing. Each user was asked to use the system for a set of randomly selected tasks, including both navigational and informational ones. The heat map is shown in Figure 10.

\(^7\)http://www.poynterextra.org/eyetrack2004/accuracy.htm
\(^8\)http://ilab.usc.edu/itrack/

**Figure 9: Results for Each Task**

**Figure 10: Heatmap of Eye-tracking Experiment**
6. CONCLUSIONS AND FUTURE WORK

In this paper, we describe our work on integrating geospatial (map and satellite) imagery, structured text (tables, etc.), and unstructured text (news articles from the web) into a single multimedia display system designed to support browsing and search. To enable wide-ranging coverage of large amounts of text despite very limited display space, we developed and deployed text clustering and three levels of text summarization technology. We evaluated the effectiveness of the system to support search, conducted a case study with several query types, and used eye-tracking to obtain a heat map that shows the averaged behavior and preferences of people performing geospatially oriented search within a multimedia display system.

We conclude that multi-level summarization is very helpful in such tasks; that users prefer to read shorter summaries rather than longer ones, and prefer to issue additional textual queries that generate new clusters and summaries rather than explore the geospatial context of the locations of interest. This work suggests that research focus on the length, placement, interactivity, content, and placement of different types of summary within a geospatial display system.

7. ACKNOWLEDGMENTS

Thanks to GeoSemble Inc. for providing the geographic integration support to the system. This work was performed with partial support from DARPA grant with contract No. W31P4Q-09-C-0313.

8. REFERENCES


