MarcoPolo: A Community System for Sharing and Integrating Travel Information in Maps

Yueguo Chen[§], Feng Li[†], Su Chen[§], Mei Hui[§], Chen Liu[§], Liangxu Liu[‡], Yuan Zhou[§]

[§]National University of Singapore
[†]Fudan University
[‡]Donghua University

{chenyueg, lifeng, chensu, hui_mei, liuchen, liuliang, yuanzhou}@comp.nus.edu.sg

ABSTRACT

The tagging technique has been widely applied in the existing Web 2.0 systems, where users label their resources with tags for resources to be efficiently retrieved through keywords search. Location-aware geographical tags (geo-tags) are required if users want to mark location sensitive resources to digital maps. Large volumes of different kinds of user-created tags pose challenges for effectively organizing community resources based on tags, guaranteeing the quality of tags, and supporting various tag-based queries efficiently. In this demo, we present MarcoPolo, a Web 2.0 community system that allows users to label various kinds of resources with tags, and mark them to the map simply by labelling them with user-defined hierarchical geo-tags. The MarcoPolo system provides two effective interfaces for users to browse and search the resources: one is the keyword-based interface and the other is the map-based interface.

1. INTRODUCTION

Recent years have witnessed the success of many Web 2.0 systems such as blogs [1], Wikipedia [2], and Flickr [3], where users contribute contents to the system. People like Web 2.0 systems because the systems allow users to display their creativity, earn respect from others, and obtain shared information from the community. A Web 2.0 system serves as a platform for users of a community to interact and collaborate with each other. Web 2.0 systems have been successfully applied in an extensive range of communities because of their effectiveness in collecting and organizing the wisdom of crowds [14].

Tagging has been widely used in Web 2.0 systems to facilitate the retrieval of the published contents [3, 4, 5, 15]. Users of these systems are usually required to label their published contents (e.g., blogs, articles, images, videos, web links) with tags. The systems organize the published contents based on users' tags to support efficient browsing and searching of the contents. Such collaborative tagging systems [9] provide convenience for users to describe resources. The collaborative way of organizing community resources with user-created metadata is coined as folksonomy [13]. However, the lack of controlled vocabulary and systematic taxonomy in such systems makes the classification of resources imprecise and imperfect. Therefore, guaranteeing the quality of tags is quite important in such systems.

Digital maps such as Google Maps [6] provide convenient interfaces for users to browse and share the location-aware information. There have been applications such as Flickr [3] allowing users to upload and mark their photos to the map. Such a way of integrating information from multiple sources into one system is called mashup [7], which has been supported by many Web 2.0 systems such as Amazon [4], Google [8], Flickr [3]. The mashup method motivates us to build a system integrating travel related information from various community systems. Our existing solutions on keyword search [11] and universal table storage [13] allow us to build efficient indexes of the shared resources based on the large volumes of user-created tags. Moreover, taking the advantage of Google Maps [6] mashup API, we allow users to mark resources to the maps simply using user-defined geo-tags. Therefore, we develop a system called MarcoPolo that provides two interfaces for users to browse and query resources in the system. The keyword-based interface supports efficient tag-based information retrieval. While, the map-based interface provides users a compact and convenient interface to browse location-aware travel information, mark and evaluate resources and related tags.

2. THE MARCOPOLO SYSTEM

2.1 System Architecture

The MarcoPolo system consists of three layers: tagging, indexing and search engine. The architecture of the MarcoPolo system is shown in Figure 1. The tagging layer provides interfaces for users to link any resources to the system by tagging the resources with any tags and geo-tags. The tables of *Resources*, *Geotags* and *Geotagging* (correlation between *Resources* and *Geotags*) will be updated when new resources linked to the system. Besides resources, the tagging layer also supports marking geo-tags to the map. This is because many geo-tags are simple names of places (scenes) when they are firstly created by users. The positions of places represented by these geo-tags have to be marked in

VLDB '08, August 24-30, 2008, Auckland, New Zealand

the map to support location sensitive information retrieval.

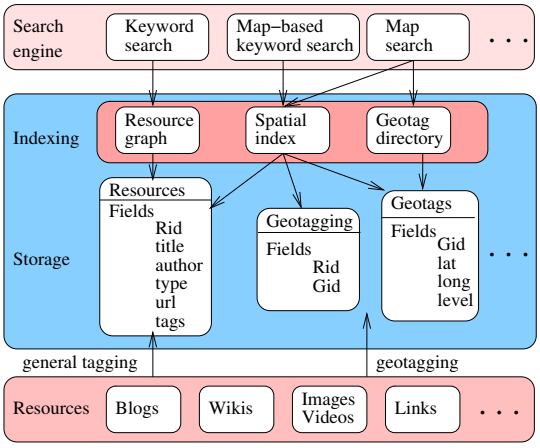

Figure 1: The architecture of the MarcoPolo system.

With the accumulation of users' contribution, the system will contain a large volume of resources and corresponding tags and geo-tags. Indexes have to been maintained to support efficient tag-based queries. In the MarcoPolo system, three indexes for tags are maintained within the indexing layer (shown in Figure 1). One is the graph index of resources which is proposed in [11] to support efficient keyword queries on tags. The other two are directory and spatial index of geo-tags. The geo-tag directory maintains the hierarchical structure of geo-tags. The spatial index maintains the positions of geo-tags in the map.

The search engine of MarcoPolo supports two kinds of tag-based queries. One is keyword search in which users present a query as a combination of keywords (tags). Resources labelled with those tags are retrieved and ranked. The other is map search in which users present a query as a map range or a specified geo-tag. Resources related to the geo-tag(s) are retrieved and ranked.

2.2 Keyword Search

The general tag-based keyword search can be well handled by exiting keyword search solutions on relational database [12]. However, as identified in [11], the lineage of resources is ignored in existing keyword search solutions. For example, a user wrote a series of posts after travelling to some places in Beijing. He linked those posts to MarcoPolo using a tag *Summer Palace* for one post and another tag *Forbidden City* for another post. Suppose some other user wants to retrieve posts about both *Summer Palace* and *Forbidden City* by presenting a query containing these two tags. However, the previous separated posts will be not well ranked using traditional keyword search because each post only contains one tag. The correlative information between the two posts is ignored. To address this problem, we use EASE model proposed in [11] to build graph index over resources, based on the hyperlinks between resources. Efficient and effective keyword search over heterogenous data is supported by EASE.

There is another kind of keyword search, called map-based keyword search, in which geographical information of resources has to be considered. For example, a user may want to find a place within a map region, containing keywords of *beach* and *casino*. The given two query tags are not required to be contained by the same resources or linked to the same geo-tag. As long as they are linked to some geo-tags that are not far away in the map region, the querying place can then be retrieved. In such map-based keyword search, spatial index of common tags has to be built based on the links of common tags to the marked geo-tags.

To effectively organize resources based on the attached tags, we also apply the tag-based clustering approach proposed in [13], to categorize resources of similar topics into clusters.

2.3 Map Search

An important feature of the MarcoPolo system is the compact and convenient map interface. With the support of location-aware geo-tags, related resources are presented when users browse the map. Given a map view, a number of popularized geo-tags whose level is equal to or sightly larger than the map view are shown within the map. Links or summaries of high-ranking resources related to geo-tags within the map view are presented to users. When a user points to a special geo-tag, the presented resources will focus on the specified geo-tag.

Users in MarcoPolo can also subscribe to location-aware travel information with regards to geo-tags or map regions. Newly published resources related to subscribed geo-tags or map regions are automatically delivered to users.

2.3.1 Hierarchical Geo-tags

Geo-tags are quite important in the MarcoPolo system as the logics of map search runs on them. Although some existing digital maps maintain many geo-tags of cities and places, the self-defined geo-tags in MarcoPolo provide the advantage of flexibility – users may label resources with any geo-tags. A geo-tag is basically the name of a place labelled with Lat/Long. They are indexed by spatial index techniques such as [10].

Geo-tags are organized in a hierarchical way (see Figure 2), where each geo-tag has a parent specifying that the geo-tag is a sub-sight of the parent geo-tag. Such a way of organization provides better precision and recall in location-based information retrieval. First, it extends the scope of geo-tags, and therefore, improves the recall of keyword search. For example, when a user wants to retrieve posts related to *Beijing*, those posts tagged with *Summer Palace* (without *Beijing*) can also be retrieved because *Summer Palace* is a child geo-tag of *Beijing*. Second, the hierarchical organization helps distinguish different sights with the same names, and therefore, improve the search precision. For example, there are two geo-tags with the same name of *Imperial Palace* in Figure 2. Within the hierarchical structure, they can be easily distinguished as one is under *Beijing* and the other is under *Shenyang*.

2.4 Data Enhancement

As a Web 2.0 system, the contents of MarcoPolo grow as users contributions accumulate. The system maintains not only basic resources such as posts and articles, but also high-level knowledge nodes such as tags, geo-tags and links between resources and tags. One goal of the MarcoPolo system is to build large and accurate indexes on resources, tags, and geo-tags so that resources can be accurately marked to the map. The central problem in achieving this goal is to

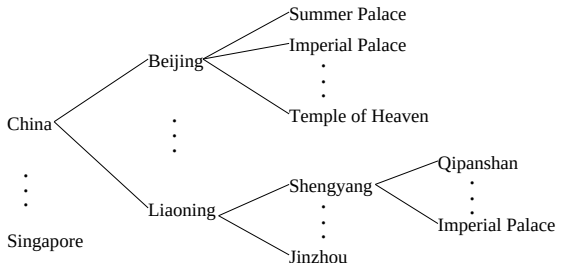

Figure 2: The hierarchical organization of geo-tags in the MarcoPolo system.

guarantee the quality of geo-tags and links between geo-tags and related resources.

We now consider some issues in tagging resources and marking geo-tags to maps. First, users may label resources with wrong or inappropriate geo-tags. As many geo-tags do not exist at the early stage of the system, users can create any name as a geo-pretag to label a resource. A geo-pretag only has the attributes of ID and name, and is not marked to a map yet. A geo-pretag can be later marked to maps by users. However, it is possible that different users may use different names (although they are likely to be similar) for the same sight. Since we do not want a sight to be marked with a number of messy geo-tags, we should keep only one geo-tag, and remove or modify the other geo-pretags or geo-tags representing the same sight. Second, a geo-tag can be wrongly marked in a map in terms of its position, map level and parent geo-tag. To guarantee the quality of geo-tags, the system has to provide feedback mechanisms for users to modify wrongly marked geo-tags. Moreover, as mentioned before, there may be many geo-tags with the same name representing different sights. They have to be identified by the name path, which requires the correct match of parent geo-tags. We propose some techniques to improve data quality in the MarcoPolo system as follows:

2.4.1 Location-aware geo-pretag proposal

Geo-tags and geo-pretags exist in the system. The relative number of geo-pretags will be large in the early stage of the system. It will be beneficial if we can propose some appropriate geo-pretags for users to mark when they are browsing a map. Although users can give any marks (with any name) to a map, it is better to guide users to mark using existing geo-pretags so that existing resources can be well linked to maps.

Since the location attributes in geo-pretags are vacant, we can only guess their possible locations by using existing geo-tags. This is achieved through the concurrency of tags in the same posts or articles. If a geo-pretag (e.g., *Summer Palace*) appears concurrently with a geo-tag (e.g., *Beijing*) in several posts or articles, there is a good chance that the location of the geo-pretag is close to that of the geo-tag. Therefore, we may propose the geo-pretag for marking when users browse to the region of the geo-tag. In implementation, we maintain a bipartite graph measuring the concurrency of geo-pretags and geo-tags so that context sensitive geo-pretags can be quickly identified and proposed when users browse the map.

2.4.2 User feedback

All user-created resources and knowledge nodes (tags, geo-tags and links) in the MarcoPolo system provide feedback mechanisms so that the quality of resources and knowledge nodes can be guaranteed through peer review. Feedback on resources are used to measure their quality. The resources are ranked based on their quality (e.g., measured by click rates) so that good resources are presented first when the number of related resources is too large.

Feedback on knowledge nodes is used to guarantee the precision of knowledge in the system. Each knowledge node has three states: *reviewed*, *protected* and *finalized*. A knowledge node is automatically under the *reviewed* state when it is created. In this state, users can provide feedback and also modify the knowledge node (e.g., adjust the position or change the name). A knowledge node automatically transforms to the *protected* state if it is not modified for some duration. In this state, normal users cannot modify the node any more. Only power users (to be explained later) of the node can modify, finalize or transform it back to *reviewed* state. A knowledge node with good reviews will be directly moved to *finalized* state if it is not processed by power users during the *protected* state for some duration. Once a knowledge node enters the *finalized* state, only system administrators can change its state.

2.5 Users' Reputation

Good Web 2.0 systems should provide good mechanisms to attract users. MarcoPolo features a credit system that evaluates users' contributions and classify their rights in the system. Users obtain credits when they link resources to the system, build knowledge nodes, or make good modifications to some knowledge nodes in the system. While, users providing bad resources and knowledge nodes are penalized following the review process.

In the credit system, credits of users' contribution are linked with the corresponding geo-tags. A contribution of a user to a geo-tag (or links and contents directly related to the geo-tag) is recorded with the ID of the geo-tag. In this way, we can aggregate and rank users' contribution within any map regions or on any geo-tags. For example, the top three contributors for the geo-tag *Beijing* can be ranked by the aggregated users' contributions over the geo-tag *Beijing* and all its offspring geo-tags.

Such a way of managing users' credits offers at least two benefits. First, we allow users to compete with each other at any local regions, just as empires conquer territory in computer games. Second, users can be assigned different rights at different regions based on their local credits. A user can be a power user within those regions where he contributes more while he is a normal user in other regions. The system will send decision requests of knowledge nodes to power users of the regions around the knowledge nodes, since they are most likely to be familiar with those regions.

3. DEMONSTRATION

We will demonstrate the MarcoPolo system with the following major functions: linking resources with tags and geo-tags, mark geo-tags to the map, keyword search, browse and subscribe location-aware travel information through map interface, manage feedback and user reputation. An example of the map interface is shown in Figure 3.

We first show how resources can be labelled and linked to MarcoPolo system. Users can label a resource with multiple

Figure 3: The map interface of MarcoPolo.

tags and geo-tags. When users input the name of a geo-tag for labelling, all geo-tags with the same name but different geo-paths are proposed to the users for selection. If there is no such geo-tag, a new geo-pretag is created.

Users can mark a geo-tag and create a parent-child link of two geo-tags at any position of the map. The map interface proposes some geo-pretags which are likely to be close to those geo-tags in the map view, so that they can be conveniently marked by users. The resources labelled with the marked geo-pretags will then be automatically linked to the map.

We will show how effective keyword search and map-based keyword search are achieved in our system by using some query examples. We also show that location-aware travel information such as posts and photos can be dynamically shown when users browse a map. A number of popular geo-tags within the map view will be selected and shown if there are too many geo-tags. The related resources are ranked according to the popularity and feedback. Highly ranked resources are proposed first to users when the number of related resources is large.

Users can give feedback to any knowledge nodes such as tags, geo-tags and links which are in the *reviewed* state when they browse a map. They can also see others' feedback and discuss in the feedback panel. The decision requests of knowledge nodes are proposed to those users who are power users in the region of the related geo-tags. Top k high reputation users within the map view can be dynamically aggregated, ranked, and shown.

4. CONCLUSIONS

In this demo, we have presented MarcoPolo, a community system integrating resources from multiple community systems such as blogs, wikis and Google maps. In MarcoPolo, users describe resources in a free-and-easy way by providing some tags and geo-tags. They help build hierarchical geo-tags and mark resources to a map with geo-tags. The MarcoPolo supports effective keyword search over tags and geo-tags. We have presented an effective method for browsing, creating and evaluating location-aware travel information based on the map interface. The system manages user reputation dynamically in terms of locations and feedback. Information quality can be aptly controlled by power users who are likely to be familiar with local regions.

5. REFERENCES

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

[2] http://en.wikipedia.org.

[3] http://www.flickr.com.

[4] http://www.amazon.com.

[5] http://del.icio.us.

[6] http://maps.google.com.

[7] http://en.wikipedia.org/wiki/Mashup.

[8] http://editor.googlemashups.com.

[9] S. Golder and B. A. Huberman. The structure of collaborative tagging systems. In *Technical report, Information Dynamics Lab*, 2005.

[10] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *SIGMOD Conference*, pages 47–57, 1984.

[11] G. Li, B. C. Ooi, J. Feng, J. Wang, and L. Zhou. Ease: Efficient and adaptive keyword search on unstructured, semi-structured and structured data. In *SIGMOD Conference*, 2008.

[12] Y. Luo, X. Lin, W. Wang, and X. Zhou. Spark: top-k keyword query in relational databases. In *SIGMOD Conference*, pages 115–126, 2007.

[13] B. C. Ooi, B. Yu, and G. Li. One table stores all: Enabling painless free-and-easy data publishing and sharing. In *CIDR*, pages 142–153, 2007.

[14] T. O'Reilly. *What Is Web 2.0.* http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

[15] F. Wang, C. Rabsch, P. Kling, P. Liu, and J. Pearson. Web-based collaborative information integration for scientific research. In *ICDE*, pages 1232–1241, 2007.