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

Yueguo Chen[§], Su Chen[§], Yu Gu[*], Mei Hui[§], Feng Li[†], Chen Liu[§]
Liangxu Liu[‡], Beng Chin Ooi[§], Xiaoyan Yang[§], Dongxiang Zhang[§], Yuan Zhou[§]

[§]*School of Computing, National University of Singapore, Singapore*
[†]*Department of Computer Science and Engineering, Fudan University, China*
[*]*College of Information Science and Engineering, Northeastern University, China*
[‡]*School of Computer Science and Technology, Donghua University, China*
Contact:{chenyueg, ooibc}@comp.nus.edu.sg

ABSTRACT

The tagging technique has been widely applied in existing Web 2.0 systems, where users label resources with tags for effective classification and efficient retrieval of resources. 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 to the effective organization of community resources using tags. Issues such as guaranteeing the quality of tags and supporting various tag-based queries emerge. In this demo, we present MarcoPolo, a Web 2.0 community system that allows users to define the hierarchical textual geo-tags and mark resources to a map using geo-tags. Statistical and feedback mechanisms are applied to guarantee the quality of tags (including geo-tags). The MarcoPolo system provides two effective interfaces for users to browse and search 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 content to a 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 published content [3, 4, 5, 15]. Users of these systems are usually required to label their published content (e.g., blogs, articles, images, videos, web links) with tags. The systems organize published content based on users' tags to support efficient browsing and search of the content. Such collaborative tagging systems [11] provide convenience for users to describe resources. The collaborative way of organizing community resources with user-created metadata is known as folksonomy [16]. 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 location-aware information. There have been applications such as Flickr [3], which allow users to upload and mark their photos to a map. Such a way of integrating information from multiple sources into one system is called mashup [7]. Examples are Amazon [4], Google [8] and Flickr [3]. However, existing map-based mashup systems such as Flickr [3, 10] mark resources to the map using real geo-tags which are geographical coordinates. As a result, to mark resources, users have to drop them to some specific positions on the map. It will be inconvenient if users mark resources in other systems (e.g., blog systems) rather than a map system itself. In this paper, we propose to use textual geo-tags so that users can easily mark resources to maps in any community systems.

We build a system called MarcoPolo which integrates travel related information from various community systems such as blogs, wiki pages, and maps. The main contribution of the MarcoPolo system is to propose a convenient mechanism for communities to publish, purify and search location-sensitive information on digital maps. Our existing solutions on keyword search [12, 16] allow us to build efficient indexes of shared resources based on the large volumes of user-created tags. Moreover, leveraging Google Maps [6] mashup API, we allow users to mark resources to maps simply by using hierarchical textual geo-tags. The system provides two interfaces for users to browse and search resources. The keyword-based interface supports efficient tag-based information retrieval while the map-based interface provides users with a compact and convenient interface to browse location-aware travel information.

2. THE MARCOPOLO SYSTEM

2.1 System Architecture

The MarcoPolo system consists of three layers: tagging, indexing and the search engine. The architecture of the Mar-

EDBT 2009, March 24–26, 2009, Saint Petersburg, Russia.

coPolo system is shown in Figure 1. The tagging layer provides interfaces for users to link any resources to the system by tagging resources with any tags and geo-tags. The tables of *Resources*, *Geotags* and *Markers* (correlation between *Resources* and *Geotags*) are updated when new resources are linked to the system. Besides resources, the tagging layer also supports marking text type geo-tags to a map. This is because many geo-tags are simple names of places (scenes) when they are first created by users. The positions of places represented by these geo-tags have to be marked in the map to support location-sensitive information retrieval.

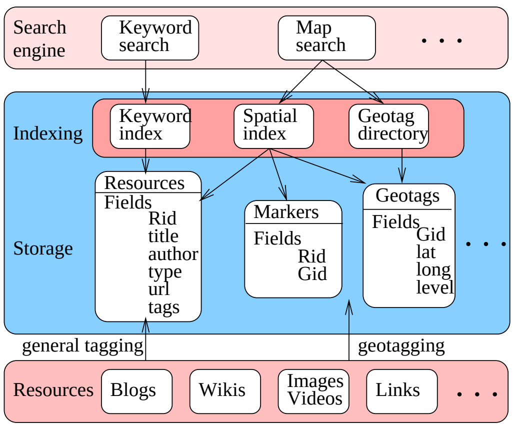

Figure 1: The architecture of the MarcoPolo system.

With the accumulation of users' contributions, the system comes to contain a large volume of resources and corresponding tags and geo-tags. Indexes have to be 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 keyword index of resources which is proposed in [12] to support efficient keyword queries on tags. The other two are the directory and the 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 where users present a query as a combination of keywords (tags). Resources labelled with those tags are retrieved and ranked based on relativity. The other is map search where users present a query as a map range or a specified geo-tag. Resources related to geo-tag(s) are retrieved and ranked based on their popularity.

2.2 Keyword Search

The general tag-based keyword search can be aptly handled by existing keyword search solutions on relational database [13]. However, as identified in [12], the lineage of resources is ignored in existing keyword search solutions. For example, a user may write a series of posts after a travel. When a user presents a query containing two tags which appear in two different posts of this series respectively, these two posts may not be well ranked with traditional keyword search. Therefore, the lineage information between the two correlative posts is ignored. To address this problem, we use the EASE model proposed in [12] to build a graph index over resources, based on the hyperlinks between resources. Efficient and effective keyword search over heterogenous data is supported by EASE.

Like [10], we also conduct tag aggregation on maps so that effective tag clouds [9] can be proposed to users when they browse to any arbitrary region. In Flickr [3], photos have to be clustered based on their locations before representative tags can be extracted from local resources [10]. However, in our system, the resources have been already well organized using the hierarchical geo-tags. Therefore, clustering on resources is unnecessary, and the accuracy of aggregating local resources is well guaranteed. An effective tag cloud can be easily created on any geo-tag by aggregating the tags attached to the resources under the geo-tag.

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 a 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 user-defined hierarchical text-based geo-tags provide the advantage of flexibility – users may label resources with any geo-tags which are names of places or sights. Once a text-type geo-tag is marked on maps, those resources labelled with the geo-tag are then link to the map.

Geo-tags are organized hierarchically (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 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*.

Note that when users mark resources using some common geo-tags, related hierarchical geo-tags can be proposed to users based on users' input, so that resources can be linked to the map more accurately.

2.4 Data Enhancement

As a Web 2.0 system, the content of MarcoPolo grows as users' contributions accumulate. The system maintains not only resources such as posts and articles, but also high-level knowledge nodes such as tags, geo-tags, links between

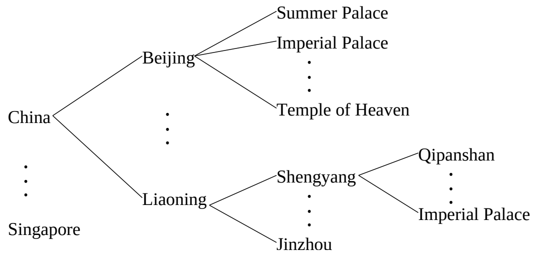

Figure 2: The hierarchical organization of geo-tags in the MarcoPolo system.

resources and tags. Large and accurate indexes on resources, tags and geo-tags need to be built so that resources can be accurately marked to the map. The central problem to achieve this goal is to guarantee the quality of geo-tags and links between geo-tags and 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. Moreover, it is possible that different users may use different names for the same sight. Since we do not want a sight to be marked with a number of messy geo-tags, one representative geo-tag will be chosen by users from those similar geo-tags (close with each other) 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 provides feedback mechanisms for users to modify wrongly marked geo-tags.

We propose some techniques to improve data quality in the MarcoPolo system as follows:

2.4.1 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. Feedbacks on resources are used to measure their quality. Resources are ranked based on their popularity so that good resources are presented first when the number of related resources is too large.

Feedbacks on knowledge nodes are 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 is automatically transformed into 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 of the node can modify, finalize or transform it back to the *reviewed* state. A knowledge node with good reviews will be directly moved to the *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.

2.4.2 Users' reputation

Good Web 2.0 systems should provide good mechanisms to attract users. MarcoPolo features a credit system that evaluates users' contributions and classifies 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. Users providing bad resources and knowledge nodes are penalized following the review process.

In the credit system, credits of users' contribution are linked with corresponding geo-tags. 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 one another 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 sends 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, marking geo-tags to the map, keyword search, browsing and subscribing location-aware travel information through the map interface, and managing feedbacks and user reputation. An example of the map interface is shown in Figure 3. A real URL will be used to demonstrate the system if the conference site is internet-enabled.

We first show how resources can be labelled and linked to the MarcoPolo system. Users can label a resource with multiple common tags and geo-tags. When users input the name of a geo-tag for labeling, all geo-tags with the same name but different geo-paths are proposed to the users for selection. Users can mark a geo-tag and create a parent-child link of two geo-tags at any position of the map. An example of managing geo-tags is shown in Figure 4.

Figure 4: Create and manage geo-tags (in Chinese)

We will show how effective keyword search and map-based keyword search are achieved in our system by using some query examples. We will also show that location-aware travel information such as posts and photos can be dynamically

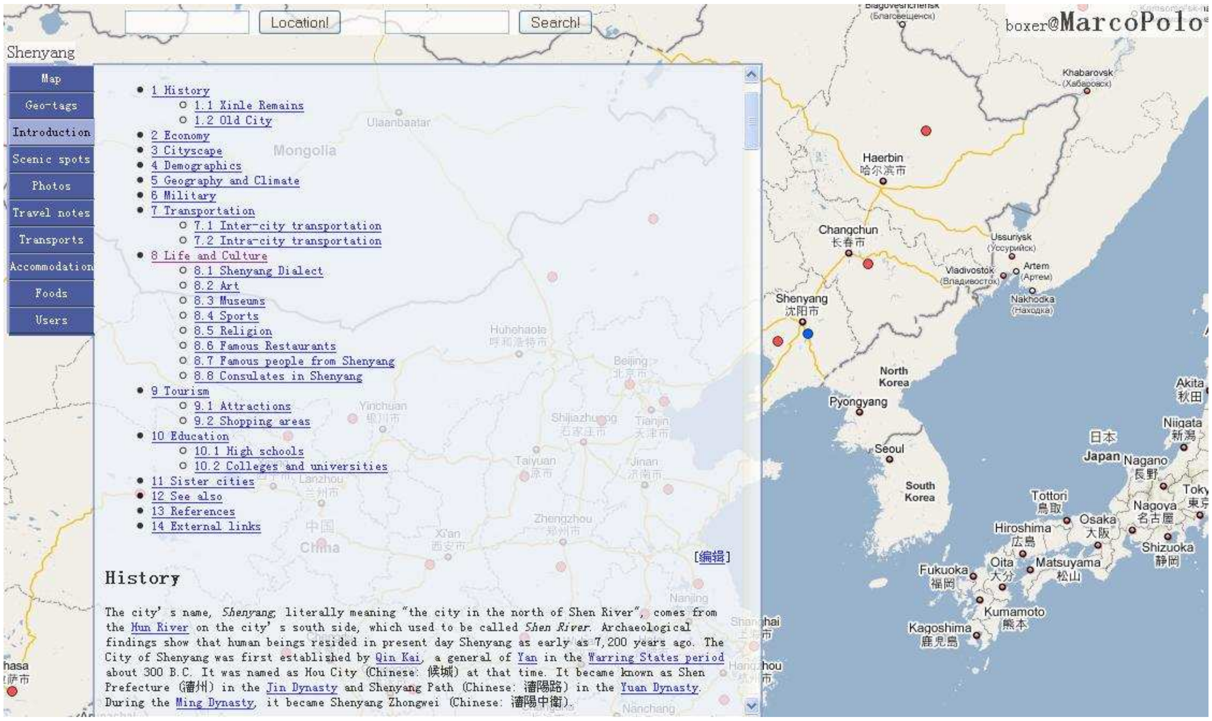

Figure 3: An example of Wiki page integrated in the MarcoPolo system

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 their popularity and feedbacks. 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 conduct discussion 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. 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 feedbacks. 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] http://http://en.wikipedia.org/wiki/Tag_cloud.

[10] S. Ahern, M. Naaman, R. Nair, and J. Yang. World explorer: Visualizing aggregate data from unstructured text in geo-referenced collections. In *In Proc. JCDL 2007*.

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

[12] 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.

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

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