Efficient remote image-based situational queries through mobile devices

Conference Paper · November 2016

DOI: 10.1145/2999508.2999524

CITATIONS
0

READS
23

5 authors, including:

Börje Karlsson
Microsoft
57 PUBLICATIONS **552** CITATIONS

SEE PROFILE

Efficient remote image-based situational queries through mobile devices

Xiaoming Leng*, Ying Yan†, Yang Chen†, Börje F. Karlsson†, Thomas Moscibroda†
†Microsoft Research Asia *Beijing University of Posts and Telecommunications

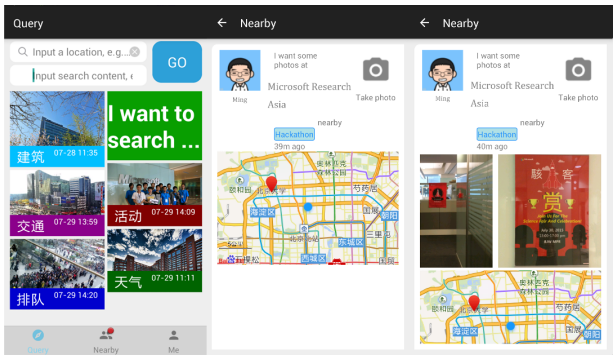

Figure 1: *ThereNow query scenarios. Left: Query categories. Center: User ties a question to a location on the map. Right: Replies start to come in for a given location and query topic.*

Abstract

This paper presents ThereNow, a LBS[Junglas and Watson 2008] mobile application designed to get close-to-real-time answers for situational queries about real-world locations. Two key issues in this scenario are: extracting information from existing data to answer user queries; and easily acquiring more data or information, if it doesn't exist yet in the system. This can be problematic due to the format and semantics of the data or to the cost (time or resources) of collecting it. ThereNow takes a unique design approach where it relies on images/photos as data and in the actual users looking at those images to 'see' if they provide enough information to answer their queries. This approach can both bypass the difficulties in information extraction from data available on the Internet and make collecting more data as easy as just taking a picture. Thus, by leveraging mobile phones being everywhere and the "an image is worth a thousand words" effect, users can easily request and quickly receive information about what is happening now at a certain location. Moreover, ThereNow makes use of an image crawler to bootstrap the system with location-tagged images and utilizes computer vision techniques to extract additional potentially useful information from images.

Keywords: Location-based Services, Crowdsourcing, Computer Vision, Web of Things.

Concepts: •**Human-centered computing** → **Ubiquitous and mobile computing systems and tools;** *Social tagging systems;* •**General and reference** → *Design;* •**Information systems** → Image search;

SA '16 Symposium on Mobile Graphics and Interactive Applications, December 05-08, 2016, Macao.
ISBN: 978-1-4503-4551-4/16/12
DOI: http://dx.doi.org/10.1145/2999508.2999524

1 Introduction

Imagine you need to know what is happening at a certain location right now. How could you achieve that? A search engine might be a good choice, but the information provided may be outdated since the last time the page was crawled. Radio and TV news can broadcast in real-time about different events or locations, but they usually focus only on hot events. Social networks could also be real-time, but how to request the precise information you want and quickly spread your question? And how do you even know who to ask and if they will answer you?

Even with the continued explosion of the number of Internet-connected mobile devices equipped with cameras, GPS modules, and a multitude of other sensors - which led to the increased popularity of crowdsourcing systems[Lane et al. 2010] - the scenario of getting time-sensitive answers for a location-specific question has not been properly addressed. Most crowdsourcing approaches require complex special apps or turn out to be a burden for users. We argue that by cleverly manipulating design constraints we can accomplish the goal of providing users with timely answers about the specific locations they are interested in.

Many location-based information needs are transient[Ye et al. 2011]. Responses would be useful now or for a few minutes, but maybe not useful after a couple of hours. For example, imagine if you need to pay a utilities bill at the bank, but you don't have much free time at work. Is it worth it to take time off your lunch and go to the bank now? What if it is crowded and the lines are huge? You might miss lunch and still not pay that bill.

In order to satisfy such needs, users need to be provided with up-to-date information about the constantly changing location. One could resort to social networks, but without friends all around the world, information is subject to friends' spatial distribution. And if you ask questions publicly, there is no guarantee that your questions will be read or answered accurately nor timely. We present ThereNow, a location-focused mobile app and backend system designed to address this complex scenario and to give people the answers they need about specific places as soon as possible, while asking for as little extra effort from other users as manageable.

ThereNow aims to enable users to always get fresh up-to-date answers to their questions about the situation in specific places. But two key issues exist in this scenario: i) how to properly extract information from existing data to answer a user question; and ii) how to easily collect more data if it is not currently available exist (i.e it does not exist in the system yet). Also, both of these issues need to be dealt with as quickly and reliably as possible. Accomplishing this can be problematic due to the format and semantics of the data (Is the information too complex? Does it answer the user question? Does it require domain-specific knowledge to properly answer a query?) or due to the cost of collecting more data (e.g. price, resource usage, or user effort).

ThereNow's design takes a unique approach where it relies on image/photo content as data and information to answer queries. Users can quickly 'see' if responses provide the required information to answer their questions and meet their interests. This approach both circumvents the problems in information extraction and makes collecting more data as easy as just taking a photo. By implementing an efficient way to spatially broadcast queries and collect images

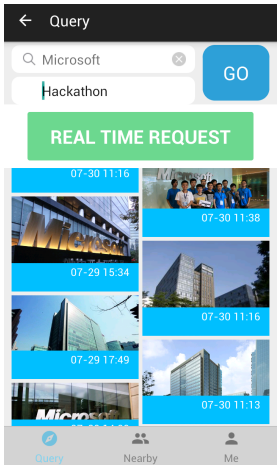

Figure 2: *Combined images sources as a single reply. Photos from online maps, previously used as responses, or new images are automatically ranked.*

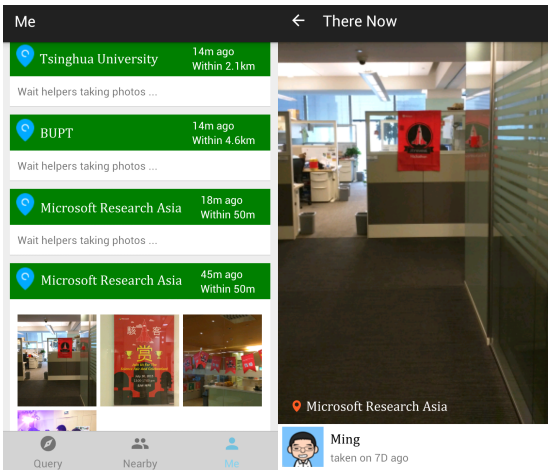

Figure 3: *Screenshots of the app. Left: List of latest queries and replies coming in. Right: Image details of one of the responses.*

we can leverage the ubiquity of mobile phones and the ”an image is worth a thousand words” effect, so users can easily and quickly get the information they need about what is happening now at a given location. Nonetheless, ThereNow can also utilize computer vision techniques to extract additional potentially useful information from images to help answer user queries.

2 Related Work

The demands of a ThereNow-like services are substantial. However, not entirely unique in a general sense. [Goodchild and Glennon 2010] and [Elwood et al. 2012], for example, discuss the use of crowdsourcing geographic information over disaster areas or as a new social practice. And according to one report from American Life Project, 74% of smartphone users get real-time location-based information with their phones [Brunsdon 2013]. ThereNow is different in that it provides a flexible way to request for additional situational information, not only to receive previously existing processed data streams.

[Yan et al. 2009] and [Kazemi and Shahabi 2012] exemplify traditional crowdsourcing efforts where spatial tasks are assigned to and performed by users. [Alt et al. 2010] describes a platform to allow users to push location-based tasks, while other users can apply for them and resolve the tasks. The system uses both implicitly and explicitly user-generated content. This approach frequent requires high user investment, complex interactions, and usually consumes too much battery power - which is always a scarce resource on mobile devices. The design behind ThereNow attempts to bypass these issues by utilizing easily available or obtainable images as its data source; and by leveraging human cognition to simplify information extraction to answer user queries.

[Kazemi and Shahabi 2012] also details the complexity of creating a general taxonomy and its data store issues. And [Franklin et al. 2011] describes an efficient database design for use in query answering scenarios; implementing not only implement traditional database operators, but also crowd operators. [Shankar et al. 2012] utilizes social network interactions to improve crowdsourcing results (but not close-to-real-time). In contrast, ThereNow covers multiple scenarios and simplifies tasks by focusing on a single image-based model, which can also be replaced by historical cached replies if no users are available at a given time. All replies, be them from other users of from the system image cache, are provided as a ranked stream to the querying user that than can quickly select an image that answers her question. Additionally, the same images on a given location can potentially be utilized to provided answers to queries from other users; even with different semantics.

3 System description

ThereNow follows a set of principles in order to let users ask timely questions about remote locations and get useful answers back.

3.1 Design principles

ThereNow is designed to provide users timely information needs, but basic functionality is not enough. The primary bottleneck in crowdsourcing systems is the high burden placed on the user who must manually collect sensor data or respond to multiple queries [Xu et al. 2015]. In general, the system must be easy to use (i.e. don’t overburden users), it must work even if few other users are available, and its functions should motivate users to keep using the app/system.

Therefore, we considered a few design principles for the prototyped system:

- Targeted questions. To avoid annoying users and maximize the chance of useful replies, queries should be broadcast only to users near the location of interest.

- Timely replies. Users asking questions should be provided with data as soon as possible. I.e. a tradeoff between reply speed and freshness of answers needs to be balanced. The system should at least provide a best effort to answer with relevant data as soon as possible, and then update the user with new fresh data as soon as it becomes available.

- Cold start ready. The system should also work from the beginning of its utilization, even if not enough users are using it yet. This means that the system needs to also utilize other streams of location-tagged images or try to reuse previous responses to answer new questions. Both as a bootstrapping mechanism, but also as a fallback strategy.

3.2 System overview

There are three roles a user may play in ThereNow: *Requester*, *Helper*, and *Reviewer*. (1) Requesters are those that ask questions. A Requester submits a question with a certain location (required), question text (recommended, but optional), and tags (optional); (2) Helpers are those who upload photos (with optional descriptions/tags) in response to questions; and (3) During feedback time (when answers are back, but before the requester chooses one), Reviewers can 'like' or 'dislike' images to provide feedback which might help the Requester or filter out bogus or inappropriate images.

Let's look at a scenario to exemplify the work flow of the system. Let's say a person wants to know if a given subway station is too crowded at this moment, she can mark the station spot on the map in the app and fire a question. People around the station are then notified, and someone takes a photo inside the station. Now the user has a sense of the situation at the station and can decide whether to go there now or not.

Fig. 4 shows an overview of the basic usage of the system. Steps 1 and 2 happen in sequence, but the delay before potential Helpers are notified is kept as short as possible based on the number of possibly available users and content already in the system (e.g. previously crawled images or response cache).

ThereNow is not just passive; users are actively asking for new information. Leveraging image content allows the system to avoid requiring the creation of accurate or complex models of the data. It is very easy for a person to analyze a picture's image content and decide if it has the information they want. But, while images are easy for people to parse and share, there are still limitations.

It's at the same time unfeasible and annoying for users to receive questions about locations too far away from them. As such, ThereNow only notifies nearby helpers of specific requests. Additionally, people using the app can tag themselves by interests. These tags can then be used as filters to prioritize who to send the request if an area has many users. However, sometimes there may be no users to answer a query, causing feedback delay. For such situations, ThereNow uses two other data sources to fulfill requests.

Firstly, the system utilizes cached image replies (that came from other previous requests). These photos have been marked with location, tags, and review counts, which serve for filtering. And by applying computer vision techniques (for example, using Project Oxford [Microsoft 2015]), even more features and categories can be extracted from images.

The last source works as a fallback and uses online photos stream providers (such as microblogs, location review sites, or online maps). These photos also usually have have accurate location information and category information. Furthermore, all image sources can be mixed, with photos sorted by a combination of criteria before being returned to users (as illustrated in 2). Sources as Twitter may contain very specific information, and can even be updated more frequently, than official sites [Power et al. 2013]. However, while of great use, extracting information from social media is also harder.

4 Implementation

ThereNow consists of a prototype Android app that allows user to make queries and interact with content (i.e. images) and a cloud-based backend that handles geo-spatial indexing of the data and users, caching of image responses to queries, and applying com-

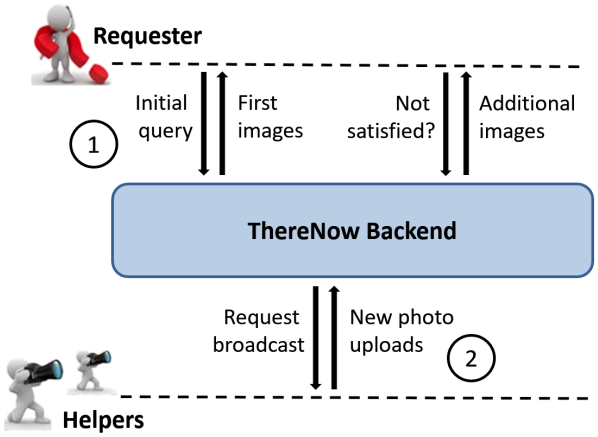

Figure 4: *System overview diagram. First, a user posts a question and quickly starts getting replies. Later, if the user still has not select an image as response, or if she actively asks for more options, requests are broadcast to nearby users.*

puter vision (CV) techniques to extract more information from photos for future use (and filter potentially bad images).

Fig. 5 shows the different tiers and components of the prototyped end-to-end system.

In order to efficiently retrieve the relevant image content to return to users and to find out which users are close to a location to notify them of relevant questions, the backend service makes use of a geohash-like structure, where basically hashes can be used to divide geographic regions into a hierarchical structure [Balkić et al. 2012]. Details of the data-structure are outside the scope of this paper as it focuses on the interactive end-to-end system. The backend is built with Node.js [Node.js 2016] server, a event-driven I/O server-side environment, and utilizes MongoDB [MongoDB 2016] as a cross-platform document-oriented database.

Fig. 1 shows some screenshots of the Android app. The user can start by choosing a query category (left screenshot) or by tying a question to a location on the map in the app - which will trigger notifications to potential helpers around the location. The screenshot on the right shows response images as they start to arrive in the form of an image stream.

ThereNow caches all responses to previous question for potential intelligent reuse in new requests. By combining this cache with the output of image analysis algorithms [Microsoft 2015], previous image content can be used to answer new questions even if they are for different categories of information. For example, if a user asks if a given restaurant is open late today, the same reply image can be used for another user asking if the restaurant is full. Or a question about if it is raining on the other part of town can potentially also be used to answer a question about the traffic situation over there (provided the CV APIs can identify these elements).

5 Future Work

This work represents our ongoing effort to investigate a novel way to combine images, crowdsourcing, and spatial data that allows users to quickly get answers about real world locations in a flexible way.

This paper presents the potential of using image content as the main data format in the case of getting answers for time-sensitive location-specific queries as images are easy to acquire and easy to

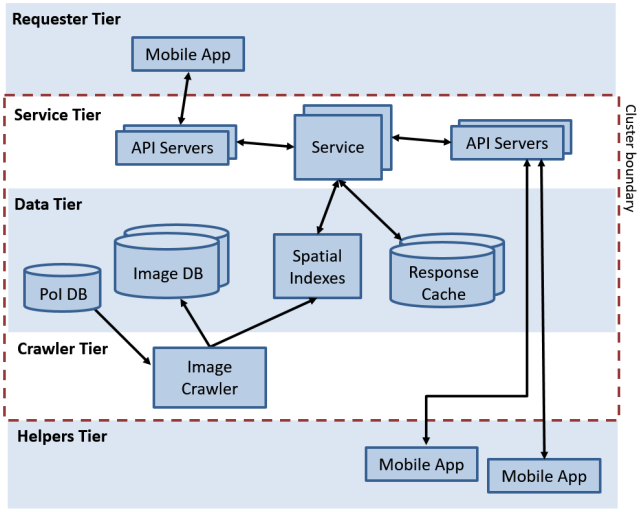

Figure 5: *System tiers. Web crawler is used to bootstrap the system from a database of Points of Interest (PoI) and an image cache is kept of previous image replies associated to categories extracted from the image via CV techniques.*

process by humans, but also amenable to reuse by leveraging time and location correlations as well as categorizations generated by computer vision techniques.

The immediate next step in our work is to analyze the usability of the system, especially on its bootstrapping cases and utilize the findings to refine and improve the implementation and the user experience. Exploring gamification strategies to promote motivation and engagement is also a target, specifically to incentivize users in less densely visited areas, or to answer questions that heave been idle for longer.

Acknowledgments

We'd like to thank all colleagues at MSRA for the valuable input and feedback on the ThereNow app idea.

References

ALT, F., SHIRAZI, A. S., SCHMIDT, A., KRAMER, U., AND NAWAZ, Z. 2010. Location-based crowdsourcing: extending crowdsourcing to the real world. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, ACM, 13–22.

BALKIĆ, Z., ŠOŠTARIĆ, D., AND HORVAT, G. 2012. Geohash and uuid identifier for multi-agent systems. In *Proceedings of the KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*, Springer, 290–298.

BRUNSDON, C. 2013. Computing with spatial trajectories, edited by yu zheng. *International Journal of Geographical Information Science 27*, 1, 208–209.

DOAN, A., RAMAKRISHNAN, R., AND HALEVY, A. Y. 2011. Crowdsourcing systems on the world-wide web. *Communications of the ACM 54*, 4, 86–96.

ELWOOD, S., GOODCHILD, M. F., AND SUI, D. Z. 2012. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the association of American geographers 102*, 3, 571–590.

FRANKLIN, M. J., KOSSMANN, D., KRASKA, T., RAMESH, S., AND XIN, R. 2011. Crowddb: answering queries with crowdsourcing. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, ACM, 61–72.

GOODCHILD, M. F., AND GLENNON, J. A. 2010. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth 3*, 3, 231–241.

ITTYCHERIAH, A., FRANZ, M., ZHU, W. J., RATNAPARKHI, A., AND MAMMONE, R. J. 2000. IBM's Statistical Question Answering System. In *Proceedings of the Ninth Text REtrieval Conference (TREC 9)*, NIST, 229–.

JUNGLAS, I. A., AND WATSON, R. T. 2008. Location-based services. *Communications of the ACM 51*, 3, 65–69.

KAZEMI, L., AND SHAHABI, C. 2012. Geocrowd: enabling query answering with spatial crowdsourcing. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, ACM, 189–198.

LANE, N. D., MILUZZO, E., LU, H., PEEBLES, D., CHOUDHURY, T., AND CAMPBELL, A. T. 2010. A survey of mobile phone sensing. *IEEE Communications magazine 48*, 9, 140–150.

MICROSOFT, 2015. Cognitive Services - Computer Vision API. https://www.microsoft.com/cognitive-services/en-us/computer-vision-api.

MONGODB, 2016. MongoDB Manual 3.2 - 2dsphere Indexes. http://docs.mongodb.org/manual/core/2dsphere/. [Online; accessed 26-Sep-2016].

NODE.JS, 2016. Node.js v6.7.0 Documentation. https://nodejs.org/api/. [Online; accessed 30-Sep-2016].

POWER, R., ROBINSON, B., AND WISE, C. 2013. Comparing web feeds and tweets for emergency management. In *WWW'13 Companion Proceedings of the 22nd International Conference on World Wide Web*, ACM, 1007–1010.

SHANKAR, P., HUANG, Y.-W., CASTRO, P., NATH, B., AND IFTODE, L. 2012. Crowds replace experts: Building better location-based services using mobile social network interactions. In *Proceedings of the International Conference on Pervasive Computing and Communications (PerCom)*, IEEE, 20–29.

XU, L., HAO, X., LANE, N. D., LIU, X., AND MOSCIBRODA, T. 2015. More with less: Lowering user burden in mobile crowdsourcing through compressive sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 659–670.

YAN, T., MARZILLI, M., HOLMES, R., GANESAN, D., AND CORNER, M. 2009. mCrowd: a platform for mobile crowdsourcing. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, ACM, 347–348.

YE, M., YIN, P., LEE, W.-C., AND LEE, D.-L. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, ACM, 325–334.