USHER: An Intelligent Tour Companion

Shubham Toshniwal
IBM Research
shtoshni@in.ibm.com

Parikshit Sharma
IBM Research
parshar2@in.ibm.com

Saurabh Srivastava
IBM Research
saurabhsrivastava@in.ibm.com

Richa Sehgal
IBM Software Group
risehgal@in.ibm.com

ABSTRACT

Audio Guides have been the prevalent mode of information delivery in public spaces such as Museums and Art Galleries. These devices are programmed to render static information to their users about the collections and artworks present and require human input to operate. The inability to automatically deliver contextual messages and the lack of interactivity are major hurdles to ensuring a rich and seamless user experience. Ubiquitous smartphones can be leveraged to create pervasive audio guides that provide rich and personalized user experience. In this paper, we present the design and implementation of "Usher", an intelligent tour companion. Usher provides three distinct advantages over traditional audio guides. First, Usher uses smartphone sensors to infer user context such as his physical location, locomotive state and orientation to deliver relevant information to the user. Second, Usher also provides interface to a cognitive Question Answer(QA) service for the inquisitive users and answers contextual queries. Finally, Usher notifies users if any of their social media friends are present in the vicinity. The ability to seamlessly track user context to provide rich semantic information and the cognitive capability to answer contextual queries means that Usher can enhance the user experience in a museum by multitudes.

Author Keywords

Context Based Services; Museum Tour Companion; Natural Language and Speech Processing; Social Media.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Audio and multimedia guides are the most prevalent form of information delivery in public spaces such as Museums, Historical sites and Art galleries. These are generally handheld devices that provide a spoken recorded commentary about the collections and art works present in the space. Such 'unintelligent' devices are used for self guided tours where in the user can ask for information about the thing being viewed by providing its unique identifier. The static content in the form of recorded commentary is a major drawback for visitors as they have to satisfy their curiosity with the limited content available. Moreover, constant human input for providing the identifier of every new art work is painstaking and a major bottleneck in ensuring seamless user experience. The goal of this paper is to design a pervasive audio companion that satisfies the curiosity of users about the collections present while ensuring seamless and personalized experience.

Context based services have gained widespread popularity in recent years. Emergence of smartphones with their plethora of sensors such as Wifi, Accelerometer, Gyroscope etc has enabled accurate inference of user context such as his position in an indoor space, his orientation and his locomotive state. Accurate inference of such context opens up a wide range of possibilities from targeted advertisements in shopping malls to dealing with emergencies in hospitals. Recent efforts have been targeted towards using users location information to deliver contextual information to him about the things nearby in museums [1, 2]. These systems however, have been limited in the sense that they have exploited this technology for mainly streaming of relevant audio content based on users location. No efforts have been made in the direction of dynamic interaction and answering contextual queries. Moreover, these systems have completely ignored the power of social media groups in the form of finding friends nearby, knowing their preferences etc.

In this paper, we present "Usher" that tries to overcome the above limitations. Usher is powered by a cognitive QA service which supports contextual querying in the most natural form. For example, if user is standing in front of Mona Lisa and he asks *"Who painted this ?"*, Usher would be able to deduce the actual query *"Who painted Mona Lisa ?"* using user's context i.e. his location and his orientation and answer accordingly. Usher allows the query itself to be asked in two modes, namely speech and text. Usher uses user's orientation and direction of movement to create dynamic audio messages, such as *"On your left is the magnificent Mona Lisa"*, on the fly which precede the actual description. The delivery of such information depends on the locomotive state of a

IUI'15 Companion, Mar 29–Apr 01, 2015, Atlanta, GA, USA
ACM 978-1-4503-3308-5/15/03.
http://dx.doi.org/10.1145/2732158.2732187

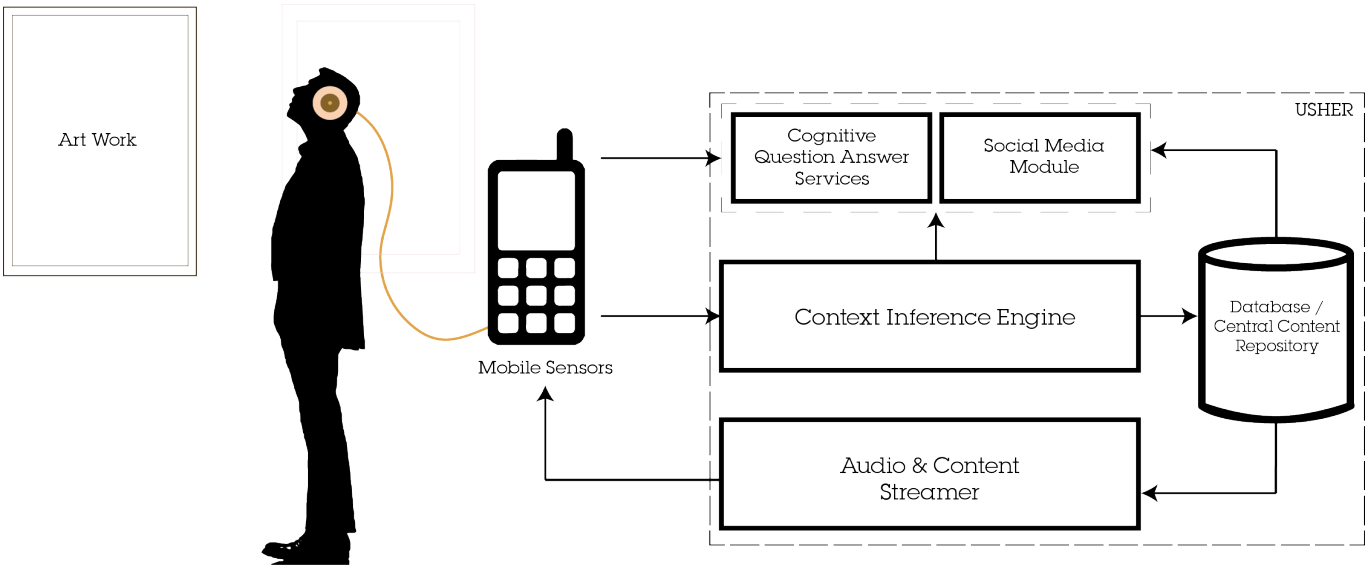

Figure 1: System Architecture of USHER showing interaction among various components

user to ensure that he is notified of relevant content only when he is gazing at an art work or looking around as opposed to when he is rushing to a particular section. Usher also notifies user if it finds a friend from user's social media sites such as Facebook, LinkedIn etc. By relaying the real time location of user's friends, it allows user to either tour the museum along with them or track their movements. Usher is thus a personalized tour companion which seamlessly tracks user context to deliver relevant information on the fly and answers contextual queries to satisfy the curiosity of a user.

The rest of the paper is organized as follows: First, we present an overview of the relevant research in this field. Then we describe our system which is followed by implementation details. Finally, we conclude our paper with discussion on implications of our work and future directions.

RELATED WORK

Audio and multimedia guides feeding static information to visitors about objects inside the museum have been prevalent since long. Recent advances in ubiquitous computing devices has seen increasing interests from museums in providing dynamic context rich information to their visitors [1]. Analysis of perception of adaptive audio guides [3] and support of computer based devices [4] indicate that users would not be willing to spend too much time on understanding how a technology works. However, they do find the information provided by museums (e.g. the short textual labels) rather limiting and acknowledge the need for dynamic help and guidance.

Advances in indoor positioning technology for handheld devices [5] has led to the use of mobile technology to provide personalized multimedia tours to enhance visitor's museum experience [2, 6]. The Hippie [7] system was one of the first to provide context based information by locating users using IR beacons installed on the entrance of each section and emitters installed on the artwork. Aixplorer [1] is a multimedia tour guide being used in the town hall of Aachen. It uses wifi hotspots to localize a person to a room level accuracy. The GUIDE project [8] supports visitors in outdoor settings using WLANs. Advances in proximity sensors such as NFC and Bluetooth Low Energy(BLE) can now enable even fine grained localization. [9] proposes object recognition based tour guide to support large scale museum guidance. Augmented reality based museum guides [10] have also been proposed.

However, prior work at large, seems to ignore two major components of an enriching museum experience: (i) Dynamic interaction (ii) Social Media Presence. Dynamic interaction itself consists of two themes: (a) Streaming dynamic content (b) Allowing interaction rather than just passive consumption. The prior work doesn't leverage the digital presence or social media presence of a user to enhance his/her experience. Usher tries to innovate on these two fronts. It allows for interaction with cognitive QA service along with, a limited dynamism in it's content. It also tries to bridge the gap between user's social medial self and real self by allowing him to find friends who are present in the same location thus, enhancing the experience of user's visit.

SYSTEM DESIGN

Figure 1 shows a graphical overview of the architecture of Usher. The context inference engine and the cognitive QA service are two core components that power Usher. Sensor data collected from user's smartphone are relayed to the context inference engine to infer attributes such as "'User is looking at Mona Lisa"'. This information is provided to the cognitive QA service to help it answer contextual queries such as *"'Who painted this?"'*. In the next few paragraphs we describe each of these components in detail.

User Context

User context is of paramount importance in our system design. Current smartphones are equipped with rich sensors

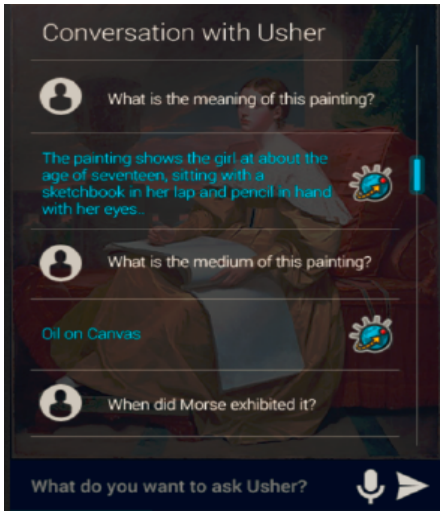

Figure 2: Question Answering with Usher.

Figure 3: Tracking friends present in Museum using Usher

which allow inference of multiple user context such as location inside a building, locomotive state and user orientation. Such context information is used to derive more semantically meaningful information. Questions such as *"Is a user looking at a painting"* can be answered based on the aisle he is in, his orientation(is he facing towards the painting?) and his locomotive state(is he stationary?). The derivation of user context and the associated semantics help the system to dynamically generate and render information that can make the tour more enriching for the user.

Audio Streaming

The context inference engine not only provides the current location of a user but also his direction of movement and his orientation. This information can be used in some innovative ways. For example, whenever a user reaches the vicinity of an artifact, an audio message such as *"On your left is Mona Lisa"* can be generated on the fly based on the orientation of the user relative to the artifact. This contextual message often leaves a deep impact on a user. Such messages can be followed by streaming the audio content related to the artifact.

Question Answering

Users can ask questions to Usher by either speaking or typing, the former requiring speech to text conversion. Usher supports interaction in the form of queries such as *"When did Morse exhibited it ?"*, as depicted in Figure 2, which do not specify the context or in this case the painting name.[1] The context here is implicitly specified by the physical location where the query is made. Such cases are to be expected because it is the natural way we would have interacted with a tour guide. So along with the user query we also send the context information to the QA service which is used for making the appropriate interpretation.

Notification For Nearby Friends

While using Usher, users can connect the application with their social media accounts like Facebook, LinkedIn etc. Usher toasts the user if a friend from their social networking is present in the museum and allows their location to be tracked (if their location is shared with them to address privacy concerns), as shown in Figure 3. This enables users to connect to their social groups present in and around the museum space. Such a service can also be used to locate and navigate to friends or family members if you get separated for some reason.

[1]Reference to entities in such a way is called "coreference" in the NLP literature and the process to correctly interpret such text is called "coreference resolution".

IMPLEMENTATION

Usher relies on three core technologies - the ability to accurately infer user context, the cognitive abilities to answer contextual queries and the ability to integrate multiple social media platforms. In this section we explain the technologies that enable Usher to provide the above functionality.

Context Inference

To determine the indoor location of a user, USHER uses an adaptive version of the widely used Bayesian inference technique called Horus [5]. Wifi based localization can achieve coarse grained information such as current aisle of the user. More fine grained information can be achieved by installing proximity sensors such as BLE based iBeacons to detect if the user is near a particular painting or object. All this information can be derived based on Wifi or bluetooth sensors present in current smartphones. In addition to location, we capture two other context attributes, specifically, user's locomotive state and his orientation. Accelerometer readings are used for step detection and the number of steps detected in used to classify locomotion into three distinct states : Stationary, Loitering and Walking. We adapted a step counting algorithm [11] and used a Kalman Filter to smooth the noise in the readings. Orientation detection is done using sensor fusion of accelerometer, gyroscope and compass.

TextToSpeech for Customized Audio Messages

Our ability to produce customized audio messages such as *"On your left is XYZ"* is built on the native

android API `android.speech.tts.TextToSpeech` which can synthesize speech from text for immediate playback.

SpeechToText for querying

We allow users to query Usher using speech. We use the native android API `android.speech.RecognizerIntent` for this with the option of `LANGUAGE_MODEL_FREE_FORM` for using a language model based on free-form speech recognition. The API gives back the top text interpretations, which are presented to the user in form of a list, . The user chooses the most appropriate result which is then selected as the user query and passed to the QA service.

Watson for QA

We use Watson [12], an artificially intelligent computer system capable of answering questions posed in natural language, developed in IBM, to answer user queries. The instance of Watson used is trained for open domain i.e. meant to answer general questions and is exposed as a REST service. The service takes the query and it's context variables, such as the artifact name, as parameters and returns a list of answers along with their confidences. We choose the top answer and display it as the response in our UI.

Mobile ID

MobileID is an IBM internal service which allows aggregating user's social media presence under one umbrella thus, reducing the number of logins required to sync user's social contacts to one. After a user logs in, the central database creates an entry corresponding to the ID for storing the location related details in a table. It then checks for the presence of user's social media contacts by searching for their corresponding IDs in the table. If a match is found then, a toast message is sent to both the users. Usher also has a separate interface to track the real time location of user's friends relayed from the central database.

CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrate Usher which provides a seamless personalized context aware experience to visitors. Usher innovates over the current state-of-the art on two major fronts: (i) Context tracking and (ii) Dynamic interactions with contextual queries. Usher allows streaming of content relevant to the user's location which are often intermixed with dynamic messages. Usher also allows the user to query via 2 modes: speech and text. The user can frame queries which might be under specified i.e. without the complete context just as if they are talking to an actual guide. Usher also brings in user's social media life into real world experiences by finding their social media friends who are also visitors thus, allowing users to enjoy their visit even more. Moreover, Usher is not limited to museums and the architecture by itself is extensible to multiple indoor domains.

The ever enriching list of mobile sensors means that Usher can always leverage these to offer even richer experience. Experience of past visitors can be used to enhance the experience for future ones. This could be achieved in multiple ways one of them being making the experience more personalized by matching user's personality type with that of past visitors and assigning user the closest match.

We believe that Usher has the potential to greatly enhance the visitor's experience and is another leap towards providing cognitive experience to users.

ACKNOWLEDGMENTS

We are really thankful to people concerned with developing and exposing the APIs of BlueZen, Watson and MobileID. These APIs allowed us to focus on innovation in integrating these services.

REFERENCES

1. Gero herkenrath and jan borchers. aixplorer - mobile audio and multimedia tour guide. url http://www.aixplorer.de/.
2. Ivo Roes, Natalia Stash, Yiwen Wang, and Lora Aroyo. A personalized walk through the museum: The chip interactive tour guide. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, 2009.
3. Daniela Petrelli, Antonella De Angeli, and Gregorio Convertino. *A user-centered approach to user modeling*. Springer, 1999.
4. Carmine Ciavarella and Fabio Paternò. The design of a handheld, location-aware guide for indoor environments. *Personal and ubiquitous computing*, 2004.
5. Moustafa Youssef and Ashok Agrawala. The horus wlan location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, 2005.
6. Paul Anderson and Adam Blackwood. Mobile and pda technologies and their future use in education. *JISC Technology and Standards Watch*, 2004.
7. Reinhard Oppermann and Marcus Specht. A context-sensitive nomadic exhibition guide. In *Handheld and Ubiquitous Computing*, 2000.
8. Keith Cheverst, Nigel Davies, Keith Mitchell, Adrian Friday, and Christos Efstratiou. Developing a context-aware electronic tourist guide: some issues and experiences. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2000.
9. Erich Bruns, Benjamin Brombach, Thomas Zeidler, and Oliver Bimber. Enabling mobile phones to support large-scale museum guidance. *IEEE multimedia*, 2007.
10. Miyashita et al. An augmented reality museum guide. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2008.
11. Najme Zehra Naqvi and Kumar et al. Step counting using smartphone-based accelerometer. *International Journal on Computer Science and Engineering (IJCSE),(May 2012)*, 2012.
12. David A. Ferrucci. Introduction to "This is Watson". *IBM Journal of Research and Development*, 56(3):1, 2012.