Conversational Natural Language interaction for Place-related Knowledge Acquisition

Srinivasan Janarthanam[1], Oliver Lemon[1], Xingkun Liu[1], Phil Bartie[2], William Mackaness[2], Tiphaine Dalmas[3] and Jana Goetze[4]

[1]Interaction Lab, Heriot-Watt University, Edinburgh
[2] School of GeoSciences, University of Edinburgh
[3]School of Informatics, University of Edinburgh
[4]KTH Royal Institute of Technology, Stockholm, Sweden
sc445,o.lemon,x.liu@hw.ac.uk, philbartie@gmail.com,
william.mackaness@ed.ac.uk, tiphaine.dalmas@aethys.com, jagoetze@kth.se

Abstract. We focus on the problems of using Natural Language interaction to support pedestrians in their place-related knowledge acquisition. Our case study for this discussion is a smartphone-based Natural Language interface that allows users to acquire spatial and cultural knowledge of a city. The framework consists of a spoken dialogue-based information system and a smartphone client. The system is novel in combining geographic information system (GIS) modules such as a visibility engine with a question-answering (QA) system. Users can use the smartphone client to engage in a variety of interleaved conversations such as navigating from A to B, using the QA functionality to learn more about points of interest (PoI) nearby, and searching for amenities and tourist attractions. This system explores a variety of research questions involving Natural Language interaction for acquisition of knowledge about space and place.

1 Motivation

Although spatial information is available to users through graphical/visual smartphone apps, there are still important problems such as how such information is delivered safely and proactively, and without cognitively overloading the user. [1] suggested that cognitive load of information presented in textual and speech-based interfaces is medium and low respectively when compared to more complicated visual interfaces. An interesting research objective, therefore, is to build a hands-free and eyes-free system that engages pedestrian users by presenting all information and receiving user requests through Natural Language.

In addition, and in contrast to other mobile applications, we focus on *conversational* interaction – meaning that the system accumulates information over time, and that the interface plans its utterances to achieve long-term communicative goals. Note that recent speech interfaces such as *Siri* use Natural Language input but they are not truly conversational because they mostly handle single user inputs one at a time, without maintaining and updating models of the dialogue context.

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 33-38

Our Natural Language interface integrates with a city model and a visibility engine [2] to identify points of interest and visible landmarks for presentation, a pedestrian tracker to improve the GPS positioning of the user, and a question-answering (QA) system to enable users to explore information about the city more freely than with a graphical interface.

Table 1 presents an example dialogue interaction with the interface showing the use of visibility information and Question-Answering.

User: I want to go to a museum.
System: How about the National Museum of Scotland? It is a five star rated attraction.
User: Okay.
System: Okay. Directions to the National Museum of Scotland.
System: Turn right on to South Bridge and walk towards the tower in front of you.
...
User: I am looking for an Indian restaurant.
System: Okay. There is an Indian restaurant named Kushis Diner on West Nicholson Street.
User: Okay. Take me there.
...
System: Near you is the famous statue of David Hume.
User: Tell me more about David Hume.
System: David Hume is a Scottish philosopher who ....

Table 1. An example interaction with the interface

2 Related work

There are several mobile apps such as *Triposo, Tripwolf*, and *Guidepal* that provide point of interest information, and apps such as *Google Navigation* that provide navigation instructions to users. However, they demand the user's visual attention because they predominantly present information on a small screen of a mobile device. In contrast, we are developing a speech-only interface in order to keep the user's cognitive load low and avoid users from being distracted (perhaps dangerously so) from their primary task.

Previously, generating navigation instructions in the real world for pedestrians has been an interesting research problem in both computational linguistics and geo-informatics [3, 4]. For example, *CORAL* is an NLG system that generates navigation instructions incrementally by keeping track of the user's location, but the user has to ask for the next instruction when he reaches a junction [3]. *DeepMap* is a system that interacts with the user to improve positioning [5]. It asks users whether they can see certain landmarks, and based on their answers improves the user's GPS position estimate. However, in many such current systems, interactions happen through the use of GUI elements such as drop-down lists and buttons, and not by using speech interaction. The *Edinburgh Augmented*

Reality System (EARS) was a prototype system that presented point of interest information to users based on visibility [6].

In contrast to these earlier systems our objective is to present navigational, point-of-interest and amenity information in an integrated way using Natural Language dialogue, with users interacting eyes-free and hands-free through a headset connected to a smartphone.

3 Architecture

The architecture of the current system is shown in figure 1. The server side consists of a dialogue interface (parser, interaction manager, and generator), a City Model, a Visibility Engine, a QA server and a Pedestrian tracker.

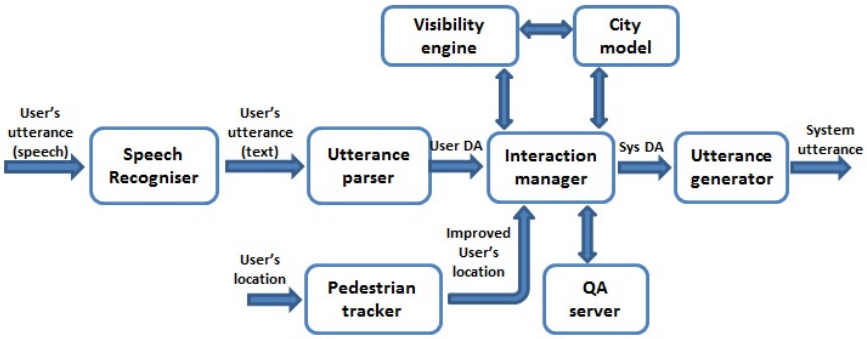

Fig. 1. System Architecture

3.1 Dialogue interface

The dialogue interface consists of a speech recogniser, an utterance parser, an Interaction Manager and an utterance generator. The speech recognition module recognises the user's utterance from the user's speech input. The utterance parser translates user utterances in to meaning representations called *dialogue acts*. The Interaction Manager is the central component of this architecture, which provides the user navigational instructions and interesting PoI information. It receives the user's input in the form of a dialogue act and the user's location in the form of latitude and longitude information. Based on these inputs and the dialogue context, it responds with system output dialogue act (DA), based on a dialogue policy. The utterance generator is a natural language generation module that translates the system DA into surface text, using the Open CCG toolkit [7].

3.2 Pedestrian tracker

Global Navigation Satellite Systems (GNSS) (e.g. GPS, GLONASS) provide a useful positioning solution with minimal user side setup costs, for location aware applications. However urban environments can be challenging with limited sky views, and hence limited line of sight to the satellites, in deep urban corridors. There is therefore significant uncertainty about the user's true location reported by GNSS sensors on smartphones [8]. This module improves on the reported user position by combining smartphone sensor data (e.g. accelerometer) with map matching techniques, to determine the most likely location of the pedestrian [2].

The output includes a robust street centreline location, and a candidate space showing the probability of the user's more exact position (e.g. pavement location). This module ensures that any GNSS-reported location placing the user at a rooftop location would be corrected to the most likely ground level location, taking into consideration user trajectory history and map matching techniques. User orientation is inferred from their trajectory.

3.3 City Model

The City Model is a spatial database containing information about thousands of entities in the city of Edinburgh. These data have been collected from a variety of existing resources such as Ordnance Survey, OpenStreetMap and the Gazetteer for Scotland. It includes the location, use class, name, street address, and where relevant other properties such as build date. The model also includes a pedestrian network (streets, pavements, tracks, steps, open spaces) which can be used to calculate minimal cost routes, such as the shortest path.

3.4 Visibility Engine

This module identifies the entities that are in the user's *vista space* [9]. To do this it accesses a *digital surface model*, sourced from LiDAR, which is a 2.5D representation of the city including buildings, vegetation, and land surface elevation. The visibility engine uses this dataset to offer a number of services, such as determining the line of sight from the observer to nominated points (e.g. which junctions are visible), and determining which entities within the city model are visible. A range of visual metrics are available to describe the visibility of entities, such as the field of view occupied, vertical extent visible, and the facade area in view. These metrics can be then used by the interaction manager to generate effective Natural Language navigation instructions. E.g. "Walk towards the castle", "Can you see the tower in front of you?", "Turn left after the large building on your left after the junction" and so on.

3.5 Question-Answering server

The QA server currently answers a range of Natural Language *definition* questions. E.g., "Tell me more about the Scottish Parliament", "Who was David

Hume?", etc. QA identifies the entity focused on in the question using machine-learning techniques [10], and then proceeds to a textual search on texts from the Gazetteer of Scotland and Wikipedia, and definitions from WordNet glosses. Candidates are reranked using a trained confidence score with the top candidate used as the final answer. These are usually long, descriptive answers and are provided in spoken output as a flow of sentence chunks that the user can interrupt. This information can also be offered by the system when a salient entity appears in the user's viewshed.

4 User interface

Users can interact with the system using a smartphone client that communicates with the system via the 3G network. The client is an Android app running on the user's mobile phone. It consists of two parts: the user's position tracker and the interaction module. The position tracker module senses user's position (latitude and longitude) and accelerometer readings. This information is sent to the system. The interaction module captures the user's speech input and relays it to the system. It also receives the system's utterances, which then is converted in to speech using the Android text-to-speech service.

We also built a web-based user interface to support the development of the system modules. It allows web-users to interact with our system from their desktops. It uses Google Street View to allow users to simulate pedestrian walking. An interaction panel lets the user interact with the system using Natural Language text or speech input. The system's utterances are synthesized using the Cereproc text-to-speech engine and presented to the user. For a detailed description of this component, please refer to [11]. A demonstration of this system will be presented at [12].

5 Future work

There are many remaining challenges in this research area for discussion, for instance:

- interleaving question-answering and navigation dialogue in a coherent manner;
- optimising the action selection of the dialogue interface (i.e. what to say next in the conversation), using machine learning techniques similar to [13–15];
- robustly handling the uncertainty generated by GPS sensors, speech recognition, and ambiguity of Natural Language interaction itself;
- generating useful referring expressions (e.g. the church on your left with the spire) which combine spatial and visual information;
- evaluating this system with real pedestrian users (this phase of the project is imminent).

Acknowledgments

The research has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270019 (SPACEBOOK project http://www.spacebook-project.eu/).

References

1. Kray, C., Laakso, K., Elting, C., Coors, V.: Presenting route instructions on mobile devices. In: Proceedings of IUI 03, Florida. (2003)
2. Bartie, P., Mackaness, W.: D3.4 pedestrian position tracker. Technical report, The SPACEBOOK Project (FP7/2011-2014 grant agreement no. 270019) (2012)
3. Dale, R., Geldof, S., Prost, J.: CORAL : Using Natural Language Generation for Navigational Assistance. In: Proceedings of ACSC2003, South Australia. (2003)
4. Richter, K., Duckham, M.: Simplest instructions: Finding easy-to-describe routes for navigation. In: Proceedings of the 5th Intl. Conference on Geographic Information Science. (2008)
5. Malaka, R., Zipf, A.: Deep Map - challenging IT research in the framework of a tourist information system. In: Information and Communication Technologies in Tourism 2000, Springer (2000) 15–27
6. Bartie, P., Mackaness, W.: Development of a speech-based augmented reality system to support exploration of cityscape. Transactions in GIS **10** (2006) 63–86
7. White, M., Rajkumar, R., Martin, S.: Towards Broad Coverage Surface Realization with CCG. In: Proc. of the UCNLG+MT workshop. (2007)
8. Zandbergen, P.A., Barbeau, S.J.: Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones. Journal of Navigation **64(3)** (2011) 381–399
9. Montello, D.: Scale and multiple psychologies of space. In Frank, A.U., Campari, I., eds.: Spatial information theory: A theoretical basis for GIS. (1993)
10. Mikhailian, A., Dalmas, T., Pinchuk, R.: Learning foci for question answering over topic maps. In: Proceedings of ACL 2009. (2009)
11. Janarthanam, S., Lemon, O., Liu, X.: A web-based evaluation framework for spatial instruction-giving systems. In: Proc. of ACL 2012, South Korea. (2012)
12. Janarthanam, S., Lemon, O., Liu, X., Bartie, P., Mackaness, W., Dalmas, T., Goetze, J.: Integrating location, visibility, and question-answering in a spoken dialogue system for pedestrian city exploration. In: Proc. of SIGDIAL. (2012)
13. Janarthanam, S., Lemon, O.: Learning Adaptive Referring Expression Generation Policies for Spoken Dialogue Systems. In: Empirical Methods in Natural Language Generation. Springer (2010)
14. Janarthanam, S., Lemon, O.: Learning to adapt to unknown users: referring expression generation in spoken dialogue systems. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics (2010) 69–78
15. Rieser, V., Lemon, O.: Reinforcement Learning for Adaptive Dialogue Systems: a Data-driven Methodology for Dialogue Management and Natural Language Generation. Theory and Applications of Natural Language Processing. Springer (2011)