Integrating Location, Visibility, and Question-Answering in a Spoken Dialogue System for Pedestrian City Exploration

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 demonstrate a spoken dialogue-based information system for pedestrians. The system is novel in combining geographic information system (GIS) modules such as a visibility engine with a question-answering (QA) system, integrated within a dialogue system architecture. Users of the demonstration system can use a web-based version (simulating pedestrian movement using StreetView) 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 the integration of multiple information sources within conversational interaction.

1 Motivation

Although navigation and local information are available to users through smartphone apps, there are still important problems such as how such information is delivered safely and proactively, and without cognitively overloading the user. (Kray et al., 2003) 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. Our objective, therefore, is to build a hands-free and eyes-free system that engages the pedestrian user by presenting all information and receiving user requests through speech only.

In addition, and in contrast to other mobile applications, this system is conversational – meaning that it accumulates information over time, and plans its utterances to achieve long-term goals. It integrates with a city model and a visibility engine (Bartie and Mackaness, 2012) to identify points of interests and visibile 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 system showing the use of visibility information and Question-Answering.

User: Take me to Princes Street.
System: Turn left on to South Bridge and
walk towards the tower in front of you.
...
System: Near you is the famous statue of David Hume.
User: Tell me more about David Hume.
System: David Hume is a Scottish philosopher....

Table 1: An example interaction with the system

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 mobile screen. In contrast, ours is a speech only interface in order to keep the user's cognitive load low and avoid users from being distracted (perhaps danger-

Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pages 134–136,
Seoul, South Korea, 5-6 July 2012. ©2012 Association for Computational Linguistics

ously so) from their primary task.

Generating navigation instructions in the real world for pedestrians is an interesting research problem in both computational linguistics and geoinformatics (Dale et al., 2003; Richter and Duckham, 2008). *CORAL* is an NLG system that generates navigation instructions incrementally upon user requests based on the user's location (Dale et al., 2003). *DeepMap* is a system that interacts with the user to improve positioning using GUI controls (Malaka and Zipf, 2000). *SmartKom* is a dialogue system that presents navigation information multimodally (Reithinger et al., 2003). There are also several mobile apps developed to help low-vision users with navigation instructions (see (Stent et al., 2010) for example). In contrast to these earlier systems we present navigational, point-of-interest and amenity information in an integrated way 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. On the user's side is a web-based client that consists of the simulated real-world and the interaction panel.

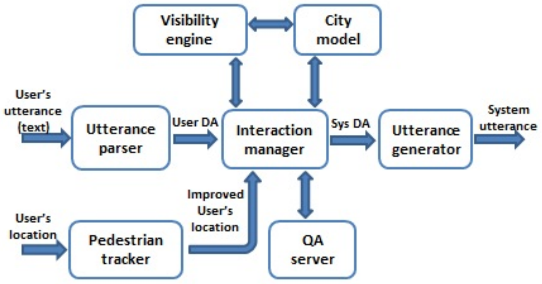

Figure 1: System Architecture

3.1 Dialogue interface

The dialogue interface consists of an utterance parser, an interaction manager and an utterance generator. 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 (White et al., 2007).

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 (Zandbergen and Barbeau, 2011). 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 (Bartie and Mackaness, 2012).

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* (Montello, 1993). 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. These metrics can be then used by the interaction manager to generate effective 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 *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 (Mikhailian et al., 2009), 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. This answer is provided as a flow of sentence chunks that the user can interrupt. This information can also be pushed by the system when a salient entity appears in the user’s viewshed.

4 Web-based User interface

For the purposes of this (necessarily non-mobile) demonstration, we present a web-based interface that simulates users walking in a 3D city environment. Users will be able to provide speech or text input (if the demonstration environment is too noisy for usable speech recognition as is often the case at conference demonstration sessions).

The web-based client is a JavaScript/HTML program running on the user’s web browser. For a detailed description of this component, please refer to (Janarthanam et al., 2012). It consists of two parts: the Streetview panel and the Interaction panel. The Streetview panel presents a simulated real world visually to the user. A Google Streetview client (Google Maps API) is created with an initial user coordinate which then allows the web user to get a panoramic view of the streets around the user’s virtual location. The user can walk around using the arrow keys on his keyboard or the mouse. The system’s utterances are synthesized using Cereproc text-to-speech engine and presented to the user.

Acknowledgments

The research has received funding from the European Community’s 7th Framework Programme (FP7/2007-2013) under grant agreement no. 270019 (SPACEBOOK project http://www.spacebook-project.eu/).

References

P. Bartie and W. Mackaness. 2012. D3.4 Pedestrian Position Tracker. Technical report, The SPACEBOOK Project (FP7/2011-2014 grant agreement no. 270019).

R. Dale, S. Geldof, and J. Prost. 2003. CORAL : Using Natural Language Generation for Navigational Assistance. In *Proceedings of ACSC2003, South Australia*.

S. Janarthanam, O. Lemon, and X. Liu. 2012. A web-based evaluation framework for spatial instruction-giving systems. In *Proc. of ACL 2012, South Korea*.

C. Kray, K. Laakso, C. Elting, and V. Coors. 2003. Presenting route instructions on mobile devices. In *Proceedings of IUI 03, Florida*.

R. Malaka and A. Zipf. 2000. Deep Map - challenging IT research in the framework of a tourist information system. In *Information and Communication Technologies in Tourism 2000*, pages 15–27. Springer.

A. Mikhailian, T. Dalmas, and R. Pinchuk. 2009. Learning foci for question answering over topic maps. In *Proceedings of ACL 2009*.

D. Montello. 1993. Scale and multiple psychologies of space. In A. U. Frank and I. Campari, editors, *Spatial information theory: A theoretical basis for GIS*.

N. Reithinger, J. Alexandersson, T. Becker, A. Blocher, R. Engel, M. Lckelt, J. Mller, N. Pfleger, P. Poller, M. Streit, and V. Tschernomas. 2003. SmartKom - Adaptive and Flexible Multimodal Access to Multiple Applications. In *Proceedings of ICMI 2003, Vancouver, B.C.*

K. Richter and M. Duckham. 2008. Simplest instructions: Finding easy-to-describe routes for navigation. In *Proceedings of the 5th Intl. Conference on Geographic Information Science*.

A. J. Stent, S. Azenkot, and B. Stern. 2010. Iwalk: a lightweight navigation system for low-vision users. In *Proc. of the ASSETS 2010*.

M. White, R. Rajkumar, and S. Martin. 2007. Towards Broad Coverage Surface Realization with CCG. In *Proc. of the UCNLG+MT workshop*.

P. A. Zandbergen and S. J. Barbeau. 2011. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. *Journal of Navigation*, 64(3):381–399.