A spoken dialogue interface for pedestrian city exploration: integrating navigation, visibility, and Question-Answering

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 navigation and QA conversations.

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 the cognitive load of information presented in textual and speech-based interfaces is medium and low respectively when compared to more complicated visual interfaces. Another important challenge is to bring different sources of data together and present information appropriately based on the dialogue context. 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 inputs through speech only. The system integrates a City Model and a Visibility Engine to identify visible landmarks, 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 (see example in table 1).

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 was a Scottish philosopher....

Table 1: An example interaction with the system

2 Architecture

The architecture of the current system is shown in figure 1.

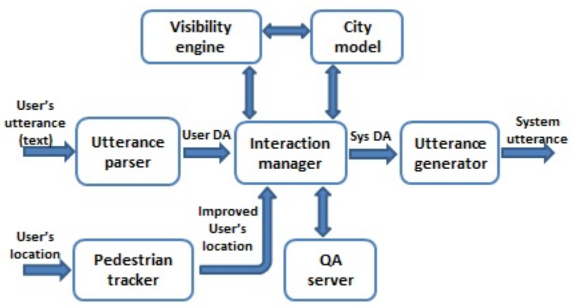

Figure 1: System Architecture

2.1 Dialogue interface

The dialogue interface consists of an utterance parser, an Interaction Manager and an utterance generator. The Interaction Manager (IM) is the central component of the system, which provides the user with timely navigational instructions and interesting PoI information. It receives the user's input in the form of a dialogue act (DA) and the user's location in the form of latitude and longitude. Based on these inputs and the dialogue context, it estimates the user's orientation, and it responds with system output dialogue act, based on a dialogue policy. The utterance generator is a natural language generation module that translates the system's DAs into surface text, using the Open CCG toolkit (White et al., 2007).

2.2 Pedestrian Tracker

Using Global Navigation Satellite Systems (GNSS) (e.g. GPS, GLONASS) this module provides user positioning information. Since urban environments can be challenging with limited sky views, and hence limited line of sight to satellites, 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).

2.3 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, and street address of many entities. The model also includes a pedestrian network (streets, pavements, etc) which can be used to calculate routes for the user.

2.4 Visibility Engine

This module identifies the entities that are visible to the user using a 2.5D representation of the city. This information is used by the IM 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.

2.5 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.

3 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. 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). A simulated real world is presented to the user visually using a Google Streetview client (Google Maps API). It allows the user to simulate walking around in real streets using arrow keys. The user can interact with the dialogue system using speech or text, which is sent to the system along with the user's location. The system's utterances are synthesized using the Cereproc text-to-speech engine.

Acknowledgments

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

References

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

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*.

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

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