

The Senior Companion: a Semantic Web Dialogue System

Debora Field*
d.field@shef.ac.uk

Alexiei Dingli
alexiei@dingli.org

Roberta Catizone
roberta@dcs.shef.ac.uk

Simon Worgan
s.worgan@dcs.shef.ac.uk

Yorick Wilks
yorick@dcs.shef.ac.uk

WeiWei Cheng
w.cheng@dcs.shef.ac.uk

Lei Ye
l.ye@dcs.shef.ac.uk

1. APPLICATION DOMAIN

The Senior Companion (SC) is a fully implemented Windows application intended for intermittent use by one user only (a senior citizen) over potentially many years. The thinking behind the SC is to make a device that will give its owner comfort, company, entertainment, and some practical functions. The SC will typically be installed at home, either as an application on a personal computer, or on a dedicated device (like a Chumby¹) or an intelligent coffee table (like Microsoft's *Surface*). By means of multimodal input and output, and a graphical interface, the SC provides its 'owner' with different functionalities, which currently include:

- conversing with the user about his personal photos
- learning about the user, user's family, and life history
- telling the user jokes
- reading the news (via RSS feed from the internet)

Chatting about photos is currently the SC's main activity. This initial direction was chosen on the assumption that senior citizens enjoy browsing photos and being reminded of events and people from their lives. The SC will currently also tell its owner jokes and read the news, if requested.

2. AGENT TECHNIQUES

Goals

The goals of the SC are more vague than those of most dialogue systems, which are often focused on task fulfilment. The overarching goal of the SC is to be a friendly, entertaining, and useful companion for its owner. One subgoal of this overarching goal is to encourage the user to talk about his life, while using personal photos as a prompt. Another subgoal is for the SC to learn the details of significant events in the user's life, so as to be able to construct a story or timeline of the user's life, and to be a knowledge source for other family members (grandchildren, for example). Another subgoal is for the SC to learn about the user's personal photos

*Affiliation of all authors is *Department of Computer Science, University of Sheffield, S1 4DP, UK*.

¹<http://www.chumby.com>

Cite as: The Senior Companion: a Semantic Web Dialogue System, Debora Field, Roberta Catizone, WeiWei Cheng, Alexiei Dingli, Simon Worgan, Lei Ye, Yorick Wilks, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi(eds.), May, 10–15, 2009, Budapest, Hungary, pp. 1383–1384
Copyright © 2009, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org), All rights reserved.

and what they depict, so that the SC can retrieve photos on the basis of what they depict, to enhance conversation.

Dialogue Manager

The agency of the SC is embodied in its dialogue manager (DM). The DM uses the stack architecture (after [2] and COMIC²) to manage dialogue, employing a set of hand-crafted augmented transition networks we call 'Dialogue Action Forms' (DAFs). DAFs are individuated by conversational topic—when the conversation moves to a new topic, a new DAF is pushed onto the stack. Although the DAFs are hand-crafted, their design is informed by data from a spoken dialogue corpus,³ and we are making preparations to add decision theory to the DAFs to enable the DM to make probabilistic decisions based on current context.

At login the SC speaks to the user, welcoming him, and begins a conversation. The SC asks the user questions, remembers some content from the user's answers, and makes statements. Some statements are motivated by inferences the SC has made, and some come from a chatbot. User utterances are processed first as input to an Automatic Speech Recogniser (ASR). The DM sends the output of the ASR to the Natural Language Understannder (NLU). We are using an Information Extraction approach to NLU exploiting GATE[1] plug-ins. Named Entity Recognition is a main feature of the NLU. In many cases, the DM sends the NLU a specification of the type of information it is expecting to be contained in the user response. For example, the DM might tell the NLU that it is looking for a family relationship in the user's utterance content. The information types the NLU can recognise are as follows:

- person names
- person relationships (family and other)
- location names
- prepositional phrases that describe locations
- dates
- time phrases not containing an explicit date
- occasions (e.g., weddings, funerals, birthdays)

If the DM does not receive from the NLU the information it is expecting at that point, it will typically apologise and repeat the question to the user up to three times, or it will invoke the chatbot.

²<http://www.hcrc.ed.ac.uk/comic>

³Collected under *Companions* [3].

The system can recognise when the user's utterance shows he has changed his mind about something he has just said (typically by using phrases like "Oh no" and "that's not right"). The system will then use a clarification routine to ensure that it has understood the correct information.

Knowledge and reasoning

Information that the DM learns from a user is stored as triples (binary predicates) in an RDF triplestore. The knowledge base is essentially monotonic, with the exception of the clarification routine just mentioned above, in which the system replaces a fact it discovers is false with one it now thinks is true.

The knowledge base also contains a small set of inference rules describing family relationships. At the end of each user utterance the DM calls a reasoner to infer new information. The reasoner is called forwards to infer everything it can about everything it knows. The newly inferred information is then exploited by the DM in its subsequent utterances at appropriate points.

With regard to ascertaining the date when a particular photo was taken, if the user uses a time phrase that is not an explicit date (for example, "It was taken six years ago"), the system invokes procedures for working out the approximate date when the photo was taken.

Object recognition

As the conversation progresses, the system not only learns about events and people in the user's life, it also makes an inventory of what is depicted in each photo, and it links the photos and the things depicted in them to the facts being stored in its knowledge base. In order to recognise that distinct objects are depicted in a photo, the system has to be able to see the photo, at least in some limited sense. Currently the system is able to detect front-facing human faces (using OpenCV⁴), but it cannot distinguish between faces, and so cannot recognise people. We are, however, about to replace this with a face recognition system developed by Polar Rose.⁵ We also intend to add the ability to recognise other object types, such as monuments and landmarks.

3. INNOVATIONS

The key innovation of the Senior Companion is the use of Semantic Web technologies to build a dialogue system. This approach was chosen in order to provide a seamless join between a dialogue system and the internet, and therefore to maximise the potential for exploiting open access knowledge when planning the content of a system's conversational utterances. Currently, the ways in which we exploit the internet to enhance conversational abilities are:

- to enable easy access to one's own or other people's personal digital photographs via Facebook.
- to look for tourist attractions near a place mentioned by the user, so as to chat to the user about them.
- to invoke an online chatbot at appropriate points.
- to provide live, up-to-date news.
- to supply particular kinds of jokes.

The above are just the first steps towards an internet-driven dialogue system.

Other innovations under development include:

⁴<http://sourceforge.net/projects/opencvlibrary/>

⁵<http://www.polarrose.com>

- using machine learning (ML) to develop a theory of how to monitor and respond to user emotions during conversation, and building that into the DM.
- using ML to derive dialogue structure from a corpus.
- using reasoning to guide the conversation towards the system's topics of greatest ignorance about the user (which we call 'grounding in the user').

4. LIVE AND INTERACTIVE ASPECTS

Hardware

The SC is a Windows application that runs on a personal computer. To interact with the SC, a microphone is needed for user input, and speakers for system output.

Personal photos

Before a user interacts with the SC some of his personal photos must have been uploaded. They can either be stored statically on the hard disk, or they can be uploaded from a Facebook album via the internet.

Input modalities

The principal input modality is speech. The first time only that the user interacts with the SC, the SC leads him through a ten-minute voice training session with the ASR.⁶

The user may also type his utterances into a text box, which is provided mainly for error correction. For example, if the ASR repeatedly misinterprets part of a user utterance, and this leads to the DM not receiving anything useful from the NLU, the user can type his utterance. (The user can see from the interface when the ASR makes mistakes.)

A third input modality not fully exploited yet is touch. The user can point on the touch-sensitive screen with an electronic pen. He might typically do this when saying, for example, "This is my sister", while touching the image of his sister on the screen. The DM knows the co-ordinates of where the screen has been touched, and knows which areas on a particular image depict faces. By aligning these, the system will be able to exploit the user's touch input in its conversational utterances and inferences.

5. ACKNOWLEDGMENTS

This work was funded by *Companions*[3], European Commission Sixth Framework Programme Information Society Technologies Integrated Project IST-34434.⁷

6. REFERENCES

- [1] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications, 2002. In *Proc. 40th Anniversary Meeting of the Association for Computational Linguistics (ACL)*.
- [2] O. Lemon, A. Bracey, A. Gruenstein, and S. Peters. The WITAS multimodal dialogue system, 2003. In *Proc. Eurospeech 2003*.
- [3] Y. Wilks. Artificial companions, 2005. *Interdisciplinary Science Reviews*, June, Vol. 30, pp. 145–152.

⁶The current ASR is *Dragon Naturally Speaking*.

⁷<http://www.companions-project.org/>