Open-Domain Information Access with Talking Robots

Kristiina Jokinen and Graham Wilcock
University of Tartu, Estonia and University of Helsinki, Finland
`kjokinen@ut.ee, graham.wilcock@helsinki.fi`

Abstract

The demo shows Wikipedia-based open-domain information access dialogues with a talking humanoid robot. The robot uses face-tracking, nodding and gesturing to support interaction management and the presentation of information to the partner.

1 Introduction

The demo shows open-domain information access dialogues with the WikiTalk system on a Nao humanoid robot (Jokinen and Wilcock, 2012b). An annotated video of the demo can be seen at `https://docs.google.com/file/d/0B-D1kVqPMlKdOEcyS25nMWpjUG8`.

The WikiTalk system can be viewed from two complementary perspectives: as a spoken dialogue system or as a question-answering (QA) system.

Viewed as a spoken dialogue system, WikiTalk supports constructive interaction for talking about interesting topics (Jokinen and Wilcock, 2012a). However, using Wikipedia as its knowledge source instead of a finite database means that WikiTalk is completely open-domain. This is a significant breakthrough compared with traditional closed-domain spoken dialogue systems.

Viewed as a QA system, WikiTalk provides Wikipedia-based open-domain knowledge access (Wilcock, 2012). However, by using sentences and paragraphs from Wikipedia, the system is able to talk about the topic in a conversational manner, thus differing from a traditional QA system.

The Nao robot prototype version of WikiTalk was implemented by Csapo et al. (2012) during eNTERFACE 2012, the 8th International Summer Workshop on Multimodal Interfaces at Supélec in Metz (Figure 1). The humanoid robot uses face-tracking, nodding and gesturing to support interaction management and the presentation of new information to the partner (Han et al., 2012; Meena et al., 2012).

Figure 1: Working with the Nao humanoid robot.

2 Outline of the system

At the heart of the system (Figure 2) is a conversation manager based on a finite state machine. However, the states are not based on the domain-specific tasks and utterances for a fixed domain. In WikiTalk, the states function at a more abstract dialogue management level dealing for example with topic initiation, topic continuation, and topic switching. Further details of this approach are given by Wilcock (2012).

The finite state machine also has extensions that store various parameters of past interactions and influence the functionality of the state machine. The conversation manager communicates with a Wikipedia manager to obtain information from Wikipedia, and a Nao manager to map its states onto the actions of the robot.

To enable the robot to react to various events while getting information from Wikipedia, the Nao manager registers events and alerts the appropriate components of the system when anything of interest occurs either on the inside or the outside of the system. Figure 2 shows three examples of
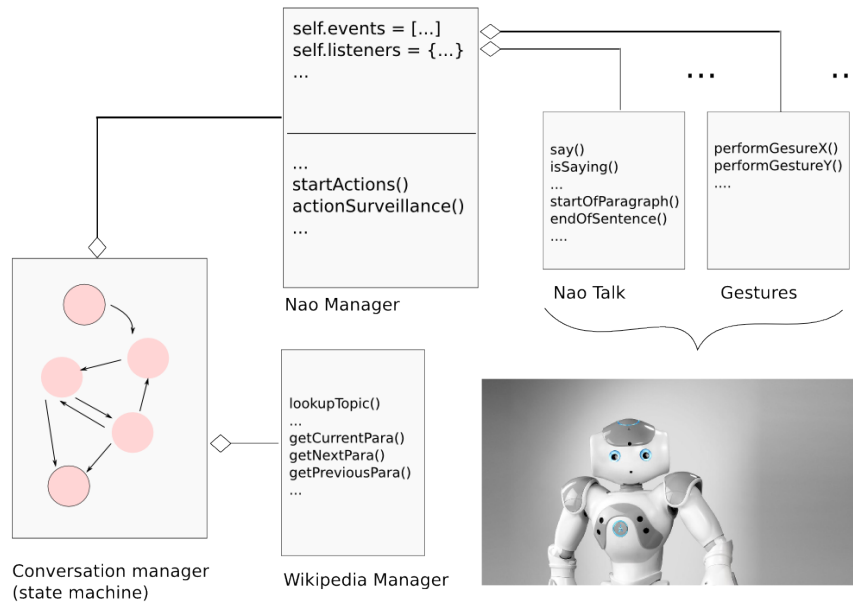
Figure 2: The system architecture, from (Csapo et al., 2012).

3 Outline of the demo

event handling within the Nao Talk module which drives the robot's speech functionality. The functions `isSaying()`, `startOfParagraph()`, and `endOfSentence()` are called periodically by the Nao manager, and return True whenever the robot is talking, reaches the start of a paragraph, or finishes a sentence, respectively. Whenever such events occur, the Nao manager can trigger appropriate reactions, for example, through the Gestures module which drives the robot's nodding and gesturing functionalities.

The history of the user's interactions is stored in a statistics dictionary in the conversation manager. Using a set of simple heuristics, it is possible to create more interesting dialogues by ensuring that the robot does not give the same instructions to the user in the same way over and over again, and by varying the level of sophistication in terms of the functionalities that are introduced to the user by the robot. For example, at first the robot gives simple instructions, allowing the user to practice and understand the basic functionalities of the system. For more advanced users, the system suggests new kinds of use cases which may not have previously been known to the user.

A corpus of videos of user trials with the system (Figure 3) was collected at the eNTERFACE 2012 workshop. The user trials and user questionnaires were used for system evaluation, which is reported by Anastasiou et al. (2013).

The demo is deliberately live, unscripted, and improvised. However, it typically starts with the robot in a sitting position. The robot stands up and greets the user, then asks what topic the user wants to hear about. The robot suggests some of its own favourite topics.

When the user selects a topic, the system gets information about the topic from Wikipedia and divides it into chunks suitable for spoken dialogue contributions. The system then manages the spoken presentation of the chunks according to the user's reactions. If the user asks for more, or otherwise shows interest in the topic, the system continues with the next chunk.

Crucially, the system makes smooth topic shifts by following the hyperlinks in Wikipedia whenever the user repeats the name of one of the links. For example, if the system is talking about Shakespeare and says "Shakespeare was born in Stratford-upon-Avon", the user can say "Stratford-upon-Avon?" and the system smoothly switches topics and starts talking about Stratford-upon-Avon using the Wikipedia information about this new topic.

The user can ask for any chunk to be repeated, or go back to the previous chunk. The user can also interrupt the current chunk and ask to skip to another chunk on the same topic.
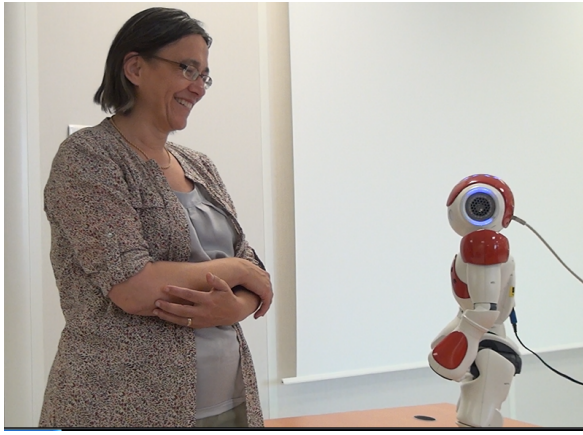
Figure 3: Testing spoken interaction with Nao.

The user can interrupt the robot at any time by touching the top of the robot's head. The robot stops talking and explicitly acknowledges the interruption by saying "Oh sorry!" and waiting for the user's input. The user can then tell it to continue, to go back, to skip to another chunk, or to switch to a new topic.

The dialogue is open-domain and typically wanders freely from topic to topic by smooth topic shifts following the links in Wikipedia. However, if the user wants to jump to an entirely unrelated topic, an awkward topic shift can be made by saying the command "Alphabet!" and spelling the first few letters of the new topic using a spelling alphabet (Alpha, Bravo, Charlie, etc.).

As well as talking about topics selected by the user, the robot can take the initiative by suggesting potentially interesting new topics. One way to do this is by using the "Did you know ...?" sections from Wikipedia that are new every day.

The demo ends when the user tells the robot to stop. The robot thanks the user and sits down.

4   Previous demos

The system was first demonstrated in July 2012 at the 8th International Summer Workshop on Multimodal Interfaces (eNTERFACE 2012) in Metz.

An annotated video of this demo can be seen at `https://docs.google.com/file/d/0B-D1kVqPMlKdOEcyS25nMWpjUG8`.

The system was also demonstrated at the 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012).

References

Dimitra Anastasiou, Kristiina Jokinen, and Graham Wilcock. 2013. Evaluation of WikiTalk - user studies of human-robot interaction. In *Proceedings of 15th International Conference on Human-Computer Interaction (HCII 2013)*, Las Vegas, USA.

Adam Csapo, Emer Gilmartin, Jonathan Grizou, Jing-Guang Han, Raveesh Meena, Dimitra Anastasiou, Kristiina Jokinen, and Graham Wilcock. 2012. Multimodal conversational interaction with a humanoid robot. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, pages 667–672, Kosice.

JingGuang Han, Nick Campbell, Kristiina Jokinen, and Graham Wilcock. 2012. Investigating the use of non-verbal cues in human-robot interaction with a Nao robot. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, pages 679–683, Kosice.

Kristiina Jokinen and Graham Wilcock. 2012a. Constructive interaction for talking about interesting topics. In *Proceedings of Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul.

Kristiina Jokinen and Graham Wilcock. 2012b. Multimodal open-domain conversations with the Nao robot. In *Fourth International Workshop on Spoken Dialogue Systems (IWSDS 2012)*, Paris.

Raveesh Meena, Kristiina Jokinen, and Graham Wilcock. 2012. Integration of gestures and speech in human-robot interaction. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, pages 673–678, Kosice.

Graham Wilcock. 2012. WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. In *Proceedings of the COLING 2012 Workshop on Question Answering for Complex Domains*, pages 57–69, Mumbai.