The Companions paradigm as a method of eliciting and organising life data

Yorick Wilks, Roberta Catizone and Oli Mival.

COMPANIONS is a major EU project that aims to change the way we think about the relationships of people to computers and the Internet by developing virtual conversational 'Companions', agents or 'presences' that stay with the user for long periods of time, developing a relationship and 'knowing' their owner's preferences and wishes. They will communicate with the user primarily by using and understanding speech, but also using other technologies such as touch screens and sensors. The project is coordinated at University of Sheffield and consists of a consortium of 14 partners from across Europe and the US. The relationship of this project to the Memories for Life concept is that its two first prototypes (the Senior Companion and the Health and Fitness Companion) both aim to extract and organize substantial quantities of life information from a user over a long period.

The Health and Fitness Advisor is a conversational presence that can appear on a screen, a mobile phone or as a plastic rabbit with mobile ears sitting on a table: it will keep track of your diet, encourage and monitor your exercise, and help set appropriate fitness goals, communicating by conversational dialogue, touch screens and sensors. Its current state can be seen on http://www.youtube.com/watch?v=KQSiigSEYhU

The Senior Companion is another artificial agent that communicates with its user on a long-term basis, adapting to their voice, needs and interests: a companion that would entertain, inform and react to emergencies. It aims to provide access to information and services as well as company for the elderly by chatting, remembering past conversations, and organising (and making sense of) the owner's photographic and image memories or accessing resources on the Internet.

A user can discuss their photos with an on-screen avatar, their PhotoPal:
the following captures the functionality we are seeking to install in the Senior Companion's Photopal aspect:

U: This is a picture of me with my friend Bill, at the TT races
earlier this year.
SC: You told me there was a Bill in THIS photo from 2003. Is it the
same guy? (Second photo displayed beside first)
U: Yes.
S: Where did you meet him?
U: We were at college together, on the same course.
S: Okay.
.
(1 week later)
.
S: Who's in this picture?
U: I'm in the pub, with Bill, and his wife Mary.
S: Is that your friend Bill from college?
U: Yes, that's right.

Both these Companions can be seen are conversions to knowledge bases of human life data from conversation with the user; at a certain levelof abstraction it can be seen as a form of Information Extraction (Wilks and Catizone, 1999), a technology now pursued almost entirely by machine learning methods, as we also intend to do. This requires a means for acquiring the data, updating the data and retrieving bits of the data through the course of conversation with the user. Both application domains represent the world knowledge to be gained about their respective domain, as well as user-specific knowledge that varies with each user and with changes over time. In the Senior Companion, an example of the system's fixed

knowledge is an ontology of personal relations, including family relations. This information is needed so that the system can infer further relationships between people: if the system knows that a user Kate has a mother Sue and a daughter Sarah, the system can infer, through the ontology of family relations that Sue is the grandmother of Sarah and may refer to this fact in the dialogue. The ability to make inferences demonstrates basic intelligence and encourages more interesting dialogue.

In the Health & Fitness Companion, the system's knowledge is about food, drink and exercise. There is an ontology of food and drink with information about food categories such as carbohydrates, proteins and fats and includes calorie information. Knowledge about exercise includes information about burning calories with respect to both means-of-transportation, such as walking, biking and driving as well as to popular and traditional means of exercise such as jogging, swimming, rowing etc. With this knowledge, the system is able to build a model of a person's daily health regime and can advise on how to improve it.

The Senior Companion discusses information in the user's photos; initially the focus of the information is two-fold :

1) basic information about the photo itself; when and where the photo was taken and for what reason

2) basic information about the people in the photograph ; name, age and their relationship to the user. The information is stored in an ontology in RDF format; the more data we have about the photos, the more interesting the dialogue with the user can/will be.

This task is different from most goal-directed dialogue systems where there is a fixed set of information slots to fill from the user, and without it the system could not be viewed as having perform successfully, as in, for example, train or plane reservations. In the SC, we aim to collect information that will allow us to form a picture of the user, but not in a typical slot-filling way. The more information we have from/about the user, the more interesting the conversation with the user will be, but we do not have a rigid fixed set of slots to fill in order to carry on the conversation. Instead, we have a set of possible discussion points related to photographs such as: where and when the photo was taken and if there are people in the photos, then the name, age and user's relation to each person in the photo. These discussion points are meant to provide a means for motivating the user to engage in a rich discussion about their photographs, the links between them, and to reveal facts about their lives unprompted.

The SC is "mixed-initiative" which means the user can change topic at any time in the dialogue. We can elicit and extract information for any topic we are prepared for, using a technology that grabs content from the speech and language string, called Information Extraction (Wilks and Catizone, 1999).

Populating the photo-database with rich metadata

One of the most powerful features of conversational interactions is the ability to pull specific metadata from a dialogue and add this to different photos, the more information about information there is, the more relevant the power law. For example, identifying an event, "this was Sarah's birthday..." means the photo can be tagged with two levels of metadata "Birthday" and "Sarah's Birthday". Building a database of metadata rich photos allows for tag related, searching, organizing and viewing: for example, "Let me see all my photos from birthdays". Things become even more powerful when this user-generated metadata is combined with automatically and semi –automatically generated metadata to populate the photo database. An example of automatically generated metadata comes every time a photograph is taken on a modern digital camera, certain information is added to the digital file, known as exif (exchangeable image file format) data. This metadata includes things such as date and time (and shortly GPS location), as well as image specific information such as aperture, shutter speed and f-stop. Interesting inferences can be made when combining

automatic and user generated tags. To continue the example above regarding “Sarah’s Birthday”, The SC can match subsequent photos taken within a specific timeframe and/or location and append the tag “Sarah’s Birthday” without explicit confirmation (although of course it can do so if required). Additionally, a probability can be applied to the accuracy of the tag depending on how well it fits the time/location parameters that are set. For example, photos taken seconds (or even a few minutes) apart are likely to be at the same event, but not necessarily of the same people. Thus, if event metadata, such as “Sarah’s Birthday”, has been established through conversation, it can be automatically applied to other photos matching the time/location parameters set by the system.

In parallel is the idea of semi-automated tagging, where user input is combined with automatic technologies, for example recognizing people in photos. i.e. when someone has been confirmed as "Sarah Burton" 3 or 4 times, the SC can apply a background image search combined with face recognition technology to add the "Sarah Burton" tag to other photos in which she appears, but which have yet to be discussed. This allows for a far smarter and more realistic conversation, in particular with images that the user has yet to say anything about “Aha, I see Sarah’s in this photo, but who is that she is with?”

A simple example session demonstrating making inferences shows the first user photograph with the user’s friend John and his girlfriend Sue. The fourth photograph contains a person named John and a person named Mike. When the system leans that there is a person named John in the fourth photo, it naturally (due to the defined strategy) asks the user if it is the same John that was in the first photo and displays the first photo for the user to see. The user can then confirm or decline this claim.

There are of course privacy problems with the use for a large set of images which we have overcome by using an image set from the researchers’ own lives. We have paid little attention to the physical form of the Companion: the H&F companion can currently appear, as the same “person, as plastic rabbit on a table, a mobile phone rabbit or a PC screen rabbit. The SC is currently a female avatar on a screen with a pleasant voice, but again, the precise form is irrelevant. The important and defining aspect of a Companion is the ability to establish something approaching a relationship with a user over a long period, and to elicit and organise life details that may not be known to relatives or survivors. It is an assumption of this work, possibly a wrong one, that a Companion will store and organise all the memories it can, and forget nothing; this seems to be a feature crucially different from people who seem to need to forget a great deal of what they experience (see Marshall, 2008).

Acknowledgement: This work was funded by the Companions project (www.companions-project.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434.

References:

Catizone, R., Dalli, A. and Wilks, Y. Evaluating Automatically Generated Timelines from the Web, LREC 2006, Genoa Italy

Catizone, R., A. Dingli and H. Pinto, Y. Wilks, "Information Extraction tools and methods for Understanding Dialogue in a Companion", to appear in Proceedings of the Language Resources and Evaluation Conference (LREC) 2008, Morocco

Marshall, J., Unforgettable, New Scientist, 16th Feb, 2008.

Pinto, H., Catizone, R., Wilks, Y., "The Senior Companion Multiagent Dialogue System", to appear in Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems, Portugal, 2008

Wilks, Y. and Catizone, R. (1999) Can We Make Information Extraction More Adaptive? In Pazienza (ed.) In Proc. of SCIE'99 Workshop. Springer-Verlag, Berlin. Rome.