I think I know you: a sharable memory model between agent and human

Joana Campos and **Ana Paiva** [1]

Abstract. The idea of 'human-like' memory has gained importance with artificial companion systems, which attempt to "change interactions into relationships". Following that perspective, it is necessary to have memory models that not only store events, but also allow semantical integration of those events into the agent and user's lifetime. In our work, we argue that a sharable memory framework between agent and human, both in structure and content, is essential to help the agent to perform in a social environment and as such establish meaningful relationships with the user. In this paper we propose a memory framework for mapping also the user's memory into three levels of abstraction like humans do. Such structure should enhance social aspects of memory and development of social bonds between agents and human.

1 INTRODUCTION

Memory in humans is more than just a set of mechanisms for retaining mundane facts of our lives. It serves other purposes such as to define personal identity, to guide future behavior or to help people perform in a social environment. This functional approach of memory, most precisely autobiographical memory (AM), explains why people retain memories for so long [3]. Socially, those mechanisms contribute to develop intimacy and maintain relationships over time [14], as they allow us to naturally communicate with our peers.

Accordingly, 'human-like' memory architectures for *artificial companions* are recognized as an essential aspect for sustaining long-term relationships between those companions and humans [10]. Undeniably, agents endowed with such mechanisms would be able to perform in a social setting using the same base line for thought.

We are not the first to note that a *sharable* framework for representing memories or events would be an essential feature to integrate into the agents' memory [12]. In our work, we argue that such *sharable* framework not only in structure, but also in content, would support social aspects of memory and help the agent to behave properly in a verbal communication.

However, the memory architecture proposed in this paper does not try to fully reproduce the human memory. It tries to capture computationally some of its relevant aspects for achieving a more 'human-like' acceptable behaviour. We grounded our work on the assumption that conceptual autobiographical knowledge is formed from abstractions of episodic memories coupled with beliefs and attitudes of the working self [6]. In other words, AM can be seen as a 'semantic network' of events contextualized in one's life, retaining knowledge about progress of personal goals.

Therefore, we formalized a model for a companions' memory based on Conway's perspective [6], who suggests a human memory division in three levels of abstraction ([4] describes the model in detail). A more abstract level (*Lifetime Periods*) to contextualize the self in his lifetime, a middle level that accounts for the experienced events (*General Events*) and a less abstract level contextualizing the events in time and emotionally (*Memory Line*). This hierarchical conceptualization allows to the system interpret the encoded and retrieved information in a meaningful way, acting as specific views over the memory.

The underlying assumption is that such model could map not only the agent's memory, but also the companion would be able to share memory content with a human. Thus, throughout this paper, we describe the implemented memory structure and its processes of remembering and forgetting, always focusing in the social functionality of AM.

2 RELATED RESEARCH

Autobiographic memory (AM) in humans empowers the integration of the past into the future. Likewise, it has been suggested that in agents this could help them to communicate and form social relationships.

To address the question of how to include autobiographical memory mechanisms in an agent and how its own emotions can increase the believability through an interaction with a user, Ho et al. [9] defined AMIA (Autobiographical Memory for Intelligent Agents) framework - an autobiographical knowledge base of significant events sensed by the agent [9, 8]. Such framework does not try to copy an adult AM, but rather to capture essential features from some psychological models suggested by Conway [6].

Ho et al. proposed an implementable computational model (yet, not implemented), divided in Life Periods,Themes, Episodes, Events and Action, with different models that can be linked and yet evaluated separately. Events are organized by goals and they can encapsulate all necessary knowledge for a particular object or situation. Further, these highly specific experiences are a central feature of AM allowing the agents to represent their own experiences for acting in a virtual environment. The defined model, formalizes components that in fact are useful for addressing the agents' memory content. However, the model suggested does not contemplate a shared

[1] Instituto Superior Tecnico - UTL and INESC-ID, Portugal, email: joana.campos@ist.utl.pt and ana.paiva@inesc-id.pt

memory between user and agent.

Focusing on a more social memory, Mei Yii Lim et al. also present an initial prototype for a social companion generic memory. The aim is to create mechanisms reflecting human memory's characteristics and then allow companions to identify, characterize and distinguish experiences [11]. They differentiate two components of an agent's memory, Short Term Memory(STM) and Long Term Memory (LTM). The main goal is to maintain active the information relevant for the agent's actual state and at the same time ensure that the agent adapts to the situations over the long-term [7]

3 MEMORY'S ARCHITECTURE

In humans, Autobiographical memory (AM) has knowledge at three levels of specificity, which are sensitive to cues and patterns of activation. As mentioned earlier, while lifetime periods identify thematic and temporal knowledge, general events are related to actions, happenings and situations in one's life. Event Specific Knowledge (ESK) details are contextualized within a general event that in turn is associated with one or more lifetime periods, linking self autobiographical memory as a whole (fig. 1).

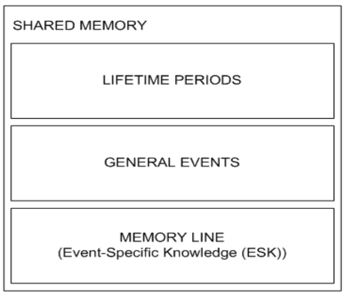

Figure 1. Knowledge base abstraction levels.

As referred by Conway [6] and other researchers[13], anything can be a cue. In our model, we will consider cues that could be represented by text due to our initial concern of developing a framework for *sharable* content between agent and human.

The underlying structure of the proposed architecture [4] is a collection of RDF (Resource Description Framework) triples consisting of $< subject, predicate, object >$. These triples are organized in graphs, one for each level of specificity, representing a simple data model for inference. The subject and object are nodes in the graph both linked by an edge. This edge represents the predicate, which establishes the relation between the two nodes.

3.1 Shared Memory

The implicit concept of shared memory (*sm*), should be clarified for a better understanding of the framework, as it is the basis for the memory's structure. Put simply, it is an experience that one had had and told to the agent. An event that occurred in the user's life. This information is defined by the tuple $< L, G, E >$, where:

- **L** – refers to one or more lifetime periods, which contextualize an event in a broad period of time.
- **G** – defines the main part of a shared memory, that is the action or event.
- **E** – specifies the details of one event.

Each one of the elements represents a level in the memory hierarchy. Besides all levels can be accessed separately, as if they were different views over the knowledge base, the 3 levels (graphs) are interconnected. This rich data integration allows us to represent knowledge in a meaningful way, while it unifies the memory as a whole.

3.1.1 Lifetime Period - L

A lifetime period (LTP) can be divided in two categories:

- **FL** – Fixed lifetime periods. Those refer to periods that are common in everyones' life. For example, the current year and the user's age.
- **SL** – When we refer to a specific time in our lives we use particular words that possibly only make sense for us. According to this idea, lifetime periods can be subjective, and cannot be defined *apriori*. Therefore, new LTPs should be created dynamically whenever it is needed.

3.1.2 General Event - G

A general event is a tuple with 6 characteristics $< A, Wo, We, Wn, Wt, Ev, Sub >$, where:

- **A** (action) – infinitive of the main verb identified in the shared memory.
- **Wo** (who) – participants that had taken part in the event
- **We** (where) – specific place where the event occurred
- **Wn** (when) – specific time when the event occurred
- **Wt** (what) – any other complement of the shared memory that not fits in the other characteristics.
- **Ev** (event) – refers to the event itself. The event is generated by linking the action to one of the inferred characteristics: $A + \{Wo, We, Wn, Wt\}$. That link is based on the underlying semantic of the verb. For example, if the verb indicates movement, such as “go” or “go out”, it links to where. Thus, the event is given by $A + We$.
- **Sub** (subevent) – link to a related G element.

3.1.3 Memory Line - E

The details of an event (G) refers to its surrounding context and the emotional details that the user may have added. This element is defined by the tuple $< T, D, Em, I, S >$, where:

- **T** (text) - sentence or set of sentences that describe the event and add personal details to it. It describes a personal view of the facts and its emotional connotation.
- **D**(date) - date object extracted from the Wn characteristic of the event. It corresponds to an instant (a specific day) or a interval with a settled begin and end.
- **Em** (emotion) - emotional state in T. The system is capable to work with this variable (and others that we might want to add), but at this stage of implementation the emotional stage is not inferred from T.

- **I** (image) - image that one can use to better describe the event.
- **S** (sound) - sound that could add some personal detail and sufficient to bring the event to one's mind.

3.2 Accessing agent's memory

Remembering is a complex process and is theoretically divided in three phases *Encoding*, *Storage*, *Retrieval.* We based our approach to accessing the knowledge base in this three stages, which are formalized below.

3.2.1 Encoding Process

The encoding process refers to the stage that information is registered [1] and is directly linked to how interesting some subject is. In this process, sequences of linked events are associated with different kinds of information, as formalized in the previous section. Each event is formalized according to the structure of a *sm*, formed by layer's rules.

Figure 2 depicts the encoding process for a written text as input. A main event (G1) is extracted from the set of sentences and sub-event (G2) is attached to it.

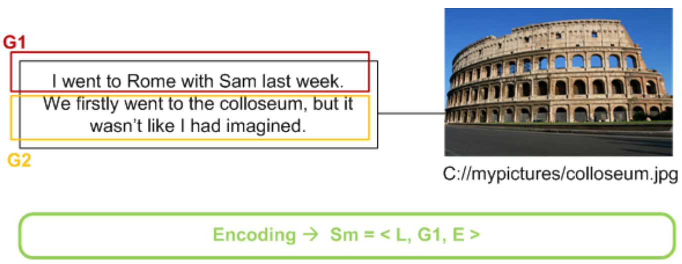

Figure 2. Encoding example.

To each one of the general events, the components of a shared memory (*sm*) are extracted and after used as cues for memory triggering.

```
L = Year 2010; Relationship with Sam

G1 = <A   = go;            G2 = <A   = go;
Wo  = I, Sam;              Wo  = I, Sam;
We  = Rome;                We  = the colosseum;
Wn  = last week;           Wn  = - ;
Wt  = - ;                  Wt  = it wasn't like I
Ev  = go to Rome;          had imagined;
Sub = G2 ; >               Ev  = go to the colosseum;
                           Sub = - >

E = < T  = I went to Rome with Sam last week.
 We went to the colosseum,
 it wasn't like I had imagined.;
      D  = 2010-03-01 , 2010-03-07
      Em = - ;    S  = - ;
      I  = c://mypictures/colloseum.jpg
```

3.2.2 Storage

The *storage* or *consolidation* is the process whereby information is maintained in memory over time [1]. In human memory only relevant and important events are retained in one's memory. Thus, as we are concerned with an agent that gathers meaningful information about the user, a filter should be applied to guarantee that only relevant events are stored in memory.

A straightforward approach for filtering relevant information, can be done weighting the action present in some event. For example, actions that may change the user's state are more important than others that do not make such change. Therefore, a *sm* is only stored in the agent's memory if the event (based on its action) adds relevant information about the user.

Continuing with the previous example, the verb 'to go' indicates movement and perhaps an important change on the user's state, so that may be a good indicator that the event associated should be retained in memory. According to that decision a shared memory object is created (see fig. 3).

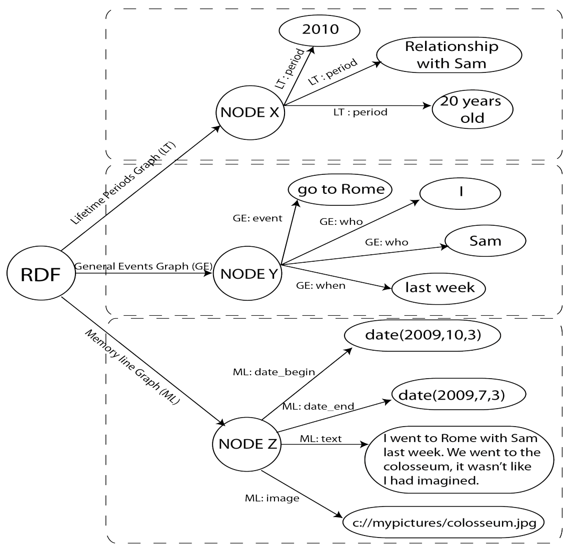

Figure 3. Graphical example of the RDF memory description.

3.2.3 Retrieval

The retrieval is a complex process for accessing information, which is possible by recognition or recall [1]. Recall is normally an intentional process but can occur when some perception - cue - triggers some experienced event. In our approach we take into account such perspective to simplify the access to the agent's memory and to take advantage of the social function of AM. Researchers have suggested that everything can be a cue for memory triggering, as long as it is linked to an event.

Therefore, any element at any level of this memory model can be retrieved, as long as, a reference to the general event is present. So to retrieve any element from memory three elements should be present $O =< G, Lv, Wr >$, where:

- **G** – reference to the specific event (SUBJECT)
- **Lv** – level in the database or graph where the search should be performed (context)
- **Wr** – what to retrieve from the performed search. Note that this element represents an edge in the graph, more precisely a relation between two nodes. (PREDICATE)
- **O** – obtained element (OBJECT)

To get any event based on information from any of the other levels, it is only necessary to perform the inverse process $G =< Wr, Lv, O >$.

4 SOCIAL MEMORY

An important function of memory in a social environment is to increase one's responsiveness, as it allows listeners to make empathetic and contextually grounded responses to what the speaker is saying [2]. To enhance this social characteristic, the described structure plays an essential role providing a very cue sensitive database, which facilitates building the social bond.

This memory model was integrated into the architecture of a companion system (MAY - my Memories are Yours) [4], which interacts with the user through dialogue. We verified that the memory content in conjunction with its structure, offers support to a more interesting interaction. Not only does this sharable framework increase the agent's responsiveness, but also its subsequent utterances are sensitive to textual cues in user's input.

4.1 Companion's overview

MAY is an agent created to assist a teenager user on self-reflection about what happens in his/her life. The communication between the agent and the user is done through dialogue by which the shared memories are collected and saved in a diary form (or timeline). The process flows as follows:

- Every sentence will be analyzed using the natural language tools, which are responsible for (1) identifying a sentence's verbal tense and to separate future from past events; (2) identifying the event (action) and its characteristics: *when* it happened, *who* participated and *where* it took place. Those components are responsible for indexing an event.

- The *shared* memory base is updated with every new relevant "sensed" event. Apart from this main module, which stores all relevant past events in user's life that the agent knows. Another one, similar in structure, accounts for events that have not happened yet.

- To produce an adequate response, the agent starts by searching its memory for anything appropriate to say. It looks for active goals, past events with some relevant information for the current situation or even go beyond the present and infer future plans. Other tools are also available for enriching the dialogue (For instance, ConceptNet[2]).

[2] http://csc.media.mit.edu/conceptnet

4.2 *Shared* structure and content

When recollecting some event from memory, humans follow a pattern that starts by establishing a broad period in their lives, which cannot always be mapped into a date ("When I was 15 I went to Brazil"). Then, they specify an event and after that they start describing its details [5]. The described memory structure tries to map this line of thought, enabling the agent to 'think' as similar as possible to the user. Allowing him/her to 'frame' the experienced events in a semantically meaningful space.

In the previously described scenario, the memory has an essential role on assisting the dialogue and again on increasing the agent's responsiveness using the previous *shared* events. This task is performed by 'sensing' if any internal stimulus lead to a pattern identified in data, either after a memory had been created or at any point of the interaction. So far, we focused on three views of data 1. *Tracking Goals*; 2. *Virtual Sensing*; and 3. *Forecast*. Also, the user is able to ask direct questions to the agent.

The former function decides, with user acquiescence, whether to store permanently or simply eliminate some event in the active goals database. The other two functions, which we call *Virtual Sensing* and *Forecast*, extract specific generalities while focusing on specific information. These cognitive functions are supported on the fact that we nearly always interpret new events based on available knowledge about the world and about our selves. In this case, the agent interpret the past or new events based on the shared events on previous interactions.

- *Virtual Sensing* – is concerned with the agents ability to sense that something is missing in the told event. When this search in the database is performed, we compute a view of the events that match the criterion of the event introduced. We set that the agent would be confident in believing that some fact is true in proportion with the number of events in memory.

- *Forecast* – normally refers to future, but in this case it refers to anything that the agent does not actually know but still can be inferred using the data in memory. The aim of this feature is particularly useful in diversifying the conversation when the system asks about yesterday or tomorrow. In contrast with the previous feature, this takes in account the day of the week to make a prediction.

4.3 Memory access through layers

The aforementioned cognitive functions use differently the layered structure of memory. In this section we describe situations wherein the companion uses the information in memory to enrich the dialogue.

4.3.1 Virtual Sensing

In this situation the agent makes a prediction about some missing element in a sentence. To clarify this situation, consider the following sentence: "I'm going to the cinema with Lyam." . In this sentence typed by the user the element 'where' is missing, so the agent combines two things that it knows: most of the times the event 'go to the cinema'(event) with 'Lyam'(who) happens every other 'Thursday' (when). This query occurs in two different levels of the data base: General Events, which match 'event' and 'who' and also the Memory Line, which gathers more specific details of each event (day of the week).

Figure 4 (left) depicts preliminary results of query time[3] for getting the required information. The time for query response increases in pair with the database size. At this point, the superior level would have an important and relevant role in decreasing the search space and to guarantee agent better performance. Figure 4 (right) shows the size of memory in terms of number of nodes.

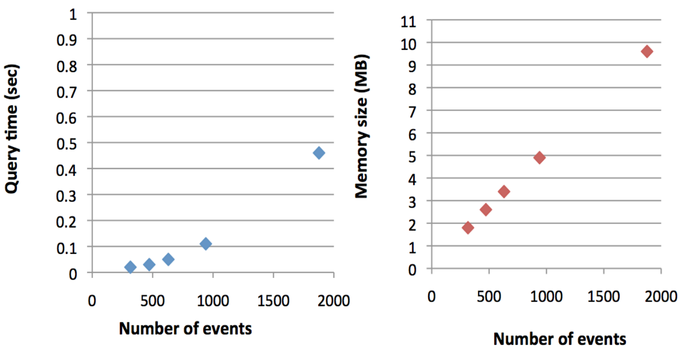

Figure 4. Storage space in function of the number of nodes

4.3.2 Forecast

When the agent tries to diversify the conversation asking about yesterday or tomorrow, a sentence can take the following structure: "Are you going jogging to the Stadium, tomorrow?" . The agent knows that several times in the past the user told it that at Saturdays he/she used to going jogging to the stadium.

5 CONCLUSION

In this paper, we described a framework which allows the representation of the agent's mental information about the user similarly to how humans do. Furthermore, this implemented model enables to share content and structure allowing the agent to behave more naturally in a social setting. In a few steps the agent is capable of recalling the exact episode in different granularities of time at any level of the RDF structure.

We consider that this structure provides an acceptable time for reaction in dialogue (< 0.5 *seconds* for large databases). However, the system would benefit if the agent encompass a lifetime period that could map the 'current period' in one's life. It would work as a 'window' over the events in memory establishing the search space constant independently of the memory size. Therefore, the memory (autobiographical memory) would be considered, as it should be, a "transitory representation " [5].

[3] The time for each query is the average of 10000 loops over the same function. We followed this procedure due to sightly variations on the Python processor and background processes during a normal interaction.

As mentioned in the opening section the social feature is only one of the characteristics of memory and this framework offers a base line for exploring the remaining functions enabling the agent to operate in situations and relations that are not present to the senses [15].

ACKNOWLEDGEMENTS. This work is partially supported by the European Community (EC) and is currently funded by the EU FP7 ICT-215554 project LIREC (LIving with Robots and IntEractive Companions), and FCT (INESC-ID multiannual funding) through the PIDDAC Program funds. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

REFERENCES

[1] A. D. Baddeley, 'The psychology of memory', *in The Handbook of Memory Disorders*, (Apr 2002).

[2] S. Bluck, 'Autobiographical memory: Exploring its funtions in everydaylife', *Memory*, **11:2**, 113–123, (2003).

[3] S. Bluck, N. Alea, T. Habermas, and D. C. Rubin, 'A tale of three functions: The self-reported uses of autobiographical memory', *SOCIAL COGNITION*, **23(1)**, 91–117, (2005).

[4] J. Campos and A. Paiva, 'May: My memories are yours', in *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, volume 6356, pp. 406–412. Springer Berlin / Heidelberg, (2010).

[5] M. A. Conway, 'Sensory-perceptual episodic memory and its context: autobiographical memory', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **356**, 1375–1384, (2001).

[6] M. A. Conway, 'Memory and the self', *Journal of Memory an Language*, **53**, 594–628, (2005).

[7] W. C. Ho, K. Dautenhahn, M. Y. Lim, P.A. Vargas, R. Aylett, and S. Enz, 'An initial memory model for virtual and robot companions supporting migration and long-term interaction', in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pp. 277 –284, (272009-oct.2 2009).

[8] W. C. Ho and S. Watson, 'Autobiographic knowledge for believable virtual characters', in *Intelligent Virtual Agents*, eds., Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin, and Patrick Olivier, volume 4133, 383–394, Springer Berlin / Heidelberg, (2006).

[9] W. C. Ho, S. Watson, and K. Dautenhahn, 'Amia: A knowledge representation model for computational autobiographic agents', in *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on*, pp. 247 –252, (2007).

[10] J. E. Laird and N. Derbinsky, 'A year of episodic memory', *In: Workshop on Grand Challenges for Reasoning from Experiences, 21st IJCAI (2009)*, (2009).

[11] M. Y. Lim, R. Aylett, W. C. Ho, S. Enz, and P. Vargas, 'A socially-aware memory for companion agents', in *IVA '09: Proceedings of the 9th International Conference on Intelligent Virtual Agents*, pp. 20–26, Berlin, Heidelberg, (2009). Springer-Verlag.

[12] N. Mavridis and M. Petychakis, 'Human-like memory systems for interactive robots: Desiderata and two case studies utilizing grounded situation models and online social networking', in *AISB - Symposium on Human Memory for Artificial Agents.*, (2010).

[13] A. R. Mayes and N. Roberts, 'Theories of episodic memory', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **356**, 1395–1408, (2001).

[14] K. Nelson, 'The psychological and social origins of autobiographical memory', *Psychological Science*, **4(1)**, 7–14, (1993).

[15] E. Tulving, 'What is episodic memory?', *Current Directions in Psychological Science*, **2**, 67–70, (1993).