An Ontology-Based Question System
for a Virtual Coach Assisting
in Trauma Recollection

Myrthe Tielman[1]([✉]), Marieke van Meggelen[3], Mark A. Neerincx[1,2],
and Willem-Paul Brinkman[1]

[1] Delft University of Technology, Delft, The Netherlands
{m.l.tielman,w.p.brinkman,m.a.neerincx}@tudelft.nl
[2] TNO Human Factors, Soesterberg, The Netherlands
[3] Erasmus University Rotterdam, Rotterdam, The Netherlands
m.vanmeggelen@fsw.eur.nl

Abstract. Internet-based guided self-therapy systems provide a novel
method for Post-Traumatic Stress Disorder patients to follow therapy
at home with the assistance of a virtual coach. One of the main chal-
lenges for such a coach is assisting patients with recollecting their trau-
matic memories, a vital part of therapy. In this paper, an ontology-based
question system capable of posing appropriate and personalized ques-
tions is presented. This method was tested in an experiment with non-
patients ($n = 24$), where it was compared with a non-ontology-based
system which did not provide personalization. Results show that people
take more time answering questions with the ontology-based system and
use more words describing properties, such as adjectives. This indicates
that the ontology-based system facilitates more thoughtful and detailed
memory-recollection.

Keywords: Virtual coach · PTSD · Dialogue system · Ontology ·
Memories

1 Introduction

Post-Traumatic Stress Disorder (PTSD) is a mental disorder caused by one or
more traumatic experiences [1]. Several treatments for PTSD are available, with
the most common element being exposure, which is the process of exposing
patients to their traumatic memories [8]. One problem for PTSD treatment is
that there is often a barrier to talk about problems and a stigma on seeking help
from the mental health-care system. One new method for exposure treatment
for PTSD which addresses this issue is, a self-therapy system with a virtual
coach [7,16]. With such a system, patients follow their therapy at home behind
their computer with the assistance of a virtual coach and a human therapist is
only remotely involved. One of the main challenges for a virtual coach in such
a self-therapy system is providing the assistance PTSD patients need for expo-
sure sessions. PTSD patients often have fragmented memories of their trauma

and are very reluctant to recall them, requiring detailed questions to stimulate memory retrieval. For a virtual coach to know which questions to ask, it needs some understanding of the traumatic experience and the personal story of the patient. Aside from needing detailed and personalized questions, it is also important to get it right, because of the sensitivity of the topic. To solve this problem, we propose an ontology-based conversational system with minimal natural-language processing with which a virtual coach can pose relevant and personalized questions to assist individuals with memory recollection.

For our system we envision that our virtual coach can apply similar techniques for motivation and behavioral change as human coaches would, and can achieve the same effect, even though patients know they are interacting with a digital agent. Blanson-Henkemans et al. [6] already showed that a virtual coach with emotional facial expressions can motivate people to live a healthier life, and Bickmore et al. [4] showed the effectiveness of an application with virtual character to elicit healthier behavior in older adults. For mental health-care, virtual characters have been employed for complex user groups such as people suffering from depression [11]. For PTSD, Rizzo et al. [14] developed the SimCoach, a virtual coach guiding veterans who potentially have PTSD towards treatment. Even though virtual avatars have some limitations compared to human coaches, such as the lack of full language abilities, they also have their own advantages, like full-time availability. Moreover, the anonymous nature of a virtual character can increase self-disclosure by patients [10].

Despite a lack of full natural language capabilities, virtual agents which can communicate with humans in a meaningful way have been developed. Schulman et al. [15] developed a conversational agent using Motivational Interviewing (MI, [12]) for health-behavior change. Their method relied on multiple-choice and free text input based on which specific dialogue acts for MI were selected. Also considering MI, Friedrichs et al. [9] developed a system which repeats back utterances of the user and employs multiple-choice input to personalize the content. Both these systems have been evaluated with users, showing that even without natural language understanding a system can hold a personalized and meaningful dialogue with a user and elicit behavior change.

Ontologies are often used in dialogue systems to add additional meaning and world knowledge. For example, Bickmore et al. [5] developed an ontology-based counseling framework which described a patient's mental states and therapist's actions affecting these states. Another possible use of ontologies is to add meaning to the speech of the agent itself, something often used for chatbots [2,3]. This shows that ontologies can be used in dialogue systems to add knowledge and meaning. In this paper, we propose to use ontologies in such a way that they assist in interpreting the user input, to steer the conversation towards its goal, and giving meaning to the dialogue of the virtual agent.

2   Ontology

Several definitions of the term *ontology* have been proposed. In this paper we use the working definition as formulated by Noy & McGuiness [13], where an

ontology is *a formal explicit description of concepts in a domain of discourse (classes) properties of each concept describing various features and attributes of the concepts (slots), and restrictions on properties.* Ontologies enable a structured knowledge base of a domain. For trauma, this means that the ontology allows a system to have an understanding of traumatic events. It could understand, for instance, that whenever an abuse victim never mentions a perpetrator, something is missing from the discourse and the system should ask an appropriate question.

The first thing to consider when designing an ontology for a question system is the type of the expected answers. For this paper, we consider the possibilities of the self-therapy Multi-Modal Memory Restructuring system (3MR) [7]. This system allows users to employ different types of media, namely text, images, music, video, google maps, websites, and emotion labels. For a user, all these are available to expose them to their memory and a question system can employ these as possible answers. An example would be the question *Where did this happen?*, which might be answered in the form of text, but also through adding a map. For this reason, our ontology is based around these types of media.
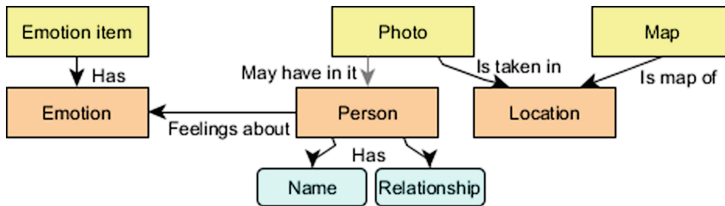

Fig. 1. Section of an ontology based on diary items in the 3MR system, showing the types Emotion, Photo and Map and its relations and properties.

Figure 1 shows an example of an ontology based on types of answers. In this way, every answer type has properties, which can again be instances of other classes again having their own properties. Which classes and properties these are, is determined by the type of memory one wants to retrieve. For a war veteran, the location of the trauma in a foreign country is very relevant, while for abuse victims this might be the type of room they were in.

3   Question System

Together with the ontology, a question dialogue based on this ontology needs to be in place. The ontology led the design of the natural language questions, whenever a specific item was entered, such as a photo, the resulting questions could be derived from the ontology. Whenever the ontology was not specific enough to decide on the order of the questions (for instance what to start with), a basic *when, where, who, what,* paradigm was used. We asked for both memories relating to a general time period and those relating to a specific moment.
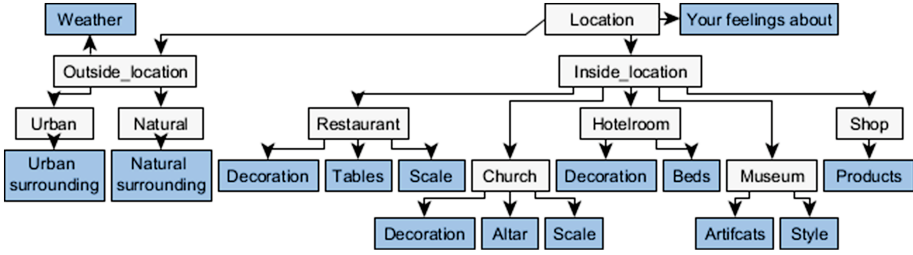
Fig. 2. Ontology of holiday moment locations. The light boxes are the classes, the darker ones their properties.

Both types are an important aspect of exposure therapy for PTSD and with the 3MR system, the therapy follows a gradual exposure paradigm where people first confront their general memories before working on the trauma itself. Two similar ontologies were in place, both asking for the when, where, who and what of the memory. The main deviation was that only the general ontology contained photo and movie questions (*Add a photo taken in that general period*) and only the specific ontology contained emotion related questions (*How were you feeling in that specific moment?*). An example of the general difference between those conversations for war veterans would be that in the general conversation the *where* question would be *Where was your mission?*, while in the specific conversation this would be *At which exact spot were you in that moment?*.

Personalization begins when the ontology is filled in and specific questions are asked. This could happen for instance, by knowing if a photo was taken in an inside or outside location. These have different properties and only those appropriate would be asked after. While a system not knowing this could still ask for properties of the location, it could only ask for those applying to all locations and give examples. To illustrate this concept, Fig. 2 shows a section of an ontology of a location of a holiday memory. The classes, or light boxes would correspond to multiple-choice questions and possible answers. The darker boxes, or properties would correspond to open questions asking after these properties. For example, once one has indicated that the location was inside in a shop, the questions would be *What kind of items were they selling in the shop?* corresponding to *products* and *How did you feel being in this place?* corresponding to *Your feelings about* (which is inherited from being a *location*). In contrast, in a naive system where a patients' answers are not taken into account, there would only be the following questions in all situations: *Can you describe your location? If you were outside, think of what your surroundings looked like and what the weather was like, and if you were inside, what type of place you were in and what it looked like?* followed by *If google streetview is available for this spot, could you find it and add it to the diary?* In this example it is clear that the naive conversation will consist of more general and longer questions. Moreover, with an ontology is that one has a clear overview of all the topics of the questions. If the ontology contains classes and properties for every concept in a certain memory, one knows

the question dialogue will also be complete. Another expected advantage is that with an ontology one can draw new conclusions. Knowing the 2 facts that 1, a shop is a type of location, and 2, that one can have feelings about a location, one can infer that one can have feelings about being in a shop. This reasoning quality is what separates the ontology from a tree-based dialogue. It can be expected, that an ontology based system can assist people better at recollecting memories than a naive system where answers are not taken into account (from this point referred to as non-ontology based). To test this hypothesis, an experiment was set up comparing an ontology-based to a non-ontology based question dialogue.

4   Experiment

A within-subject experiment with two conditions was conducted. We wished to know if an ontology-based system would allow people to recollect their memories better than a system where answers were not taken into account (non-ontology based). Our ontology-based system was compared with such a non-ontology-based question system which had the same topics and order of questions, but without any of the personalization. The effect this difference had on people's opinions and experiences with the system was tested, as well as the effect it had on the level of detail in the answers.

4.1   Participants

Giving exposure sessions to PTSD patients without providing full treatment was not considered ethically appropriate. For this reason 24 healthy participants (10 female, mean age 28.4, SD 3.1) were recruited from the University staff and student population. Eight Participants performed the experiment in their native language (English or Dutch), all others in their second language (English). Because the participants did not have traumatic experiences, the memories they had to recollect were of holidays. This topic was chosen because it was universal for this sample and could be modelled with an ontology quite well. Furthermore, in exposure treatment PTSD patients would also start with a positive memory to get familiar with the system.

4.2   Question System

The question system was based on two ontologies, one for a general holiday memory and one focusing on a specific moment within that same holiday. For the non-ontology based system, the questions followed the exact same order and topics as the ontology-based system, such as destination and travelling companions. An example of the ontology and the difference between the systems can be found in Sect. 3. Multiple-choice options were used to ensure that the ontology-based system could react to answers appropriately. The non-ontology based system did not include any multiple-choice options. Finally, in the ontology-based system it was possible to pose a constraint on the type of answers possible, in this

case the length of the answers. For the questions asking for descriptions, the answer needed to be at least six words long. Whenever this was not the case, a follow-up question would ask the participant to tell more. All open questions were answered through adding items to the diary, be it in form of media or typed text in a text item to answer the question.

4.3   Wizard of Oz and Procedure

A Wizard of Oz procedure was followed. The full dialogue was written in advance and the procedure was fully specified to avoid any influence from the wizard. The order of the questions was set and participants could signal they were finished answering through a button. The wizard was in the same room as the participants, but they could not see the wizard controlling the system. All questions appeared on the screen of the participant as typed text. The whole question would appear at once, along with multiple-choice options if applicable.

Prior to the experiment, all participants were asked to bring media (photos, video & music) from four holidays. The two memories which were used in the experiment were randomly chosen. All participants started with an introductory exercise to get to know the system. After this, the agent posing the questions gave a short introduction explaining its function and that it could communicate through text. After this, two dialogue sessions followed in which participants were asked to describe two holiday memories, one with the ontology-based and one with the non-ontology based question system. Each dialogue session consisted of 10 min of questions on the general experience and 10 min of questions on a specific moment, the order of sessions was counter-balanced. Participants had the option to take a short break between the two dialogue sessions. The experiment was approved by the University ethics committee.

4.4   Measures

Both subjective experience and the amount of detail in participant's answers were studied. The subjective experience was examined in three ways. The first was the emotion experienced when thinking back to the memory. In particular, if the subjective feeling changed when the memory was recollected. Pre and post-measures were taken with the 5-point Self-Assessment Manikin scale (SAM) scale (Lang, 1980) of Arousal and Valence for both memories recollected. The second subjective measure was how well people felt the system helped them in recollecting their memory. The third way was how people experienced the conversation with the system. These two measures were examined with 6 questions each in a questionnaire answered on a 7-point Likert scale. Examples of these questions are *The questions assisted me well in recollecting my memory* for the memory recollection and *The conversation did not run smoothly* for the conversational experience. This questionnaire was presented directly after each dialogue session. Finally, each participant answered four questions on their overall preference of one system over the other. The first was on which system helped recollect the memory best, the second which system was most pleasant to work with, the

Table 1. Annotations and definition

Annotation	Consists of
Objects	All nouns, except those referring to a person (or multiple, such as *people*), and those referring to a period of time (e.g. *day* or *moment*).
People	All nouns referring to people (e.g. *girl, tourists*) and names of people.
Descriptives	Adjectives (including terms as *very* and *three*), as well as words describing aspects of something (e.g. *cold*, in *the room is cold*). Excluding adjectives of feelings. Double adjectives were counted apart (*a very cold room*, and *a large cold room* both counting 2 descriptives)
Feelings	All words referring to feelings (e.g. *excited*), including *looking forward to* and *tired*, as well as all adjectives of feelings (*very excited* counting both words).
Time	All nouns referring to time, such as *month* or *period*.

third on which asked the best questions to trigger the memory and the fourth on which system they would use again. The objective measure considered was the amount of detail in the answers of the patients. On a general level, the number of words typed and the number of question topics posed were checked. The number of question topics could differ per participant because the dialogue sessions had a fixed time, i.e. some participants would answer only questions about location and travel, while others were quicker and would also answer questions on travel companions. To consider the amount of detail present in the texts, all answers were annotated and the number of objects, people, descriptives, feelings and time references were counted. The description of the categories in this annotation can be found in Table 1. A second annotator annotated 1235 words to ascertain reliability of the rating. Interrater reliability was assessed with Cohen's Kappa and showed a good agreement between annotators $\kappa = 0.86$, $p < 0.0001$.

4.5   Data Preparation and Analysis

Two questionnaires were designed for this experiment, one measuring how well people felt the system helped them in memory recollection, and one measuring how people experienced the conversation with the system. The validity of these was tested with Cronbach's alpha, after which one question was removed from the recollection questionnaire and two from the conversation questionnaire to improve internal validity. Internal validity after this was acceptable to good (α 0.72 to 0.81 for the recollection and α 0.54 to 0.63 for the conversation questionnaire). For objective measures we considered the answers given to the questions. For one of the participants, the answers to the questions were lost due to a technical error, so the answers of only 23 participants were taken into account. When considering the amount of detail in the answers, we only considered the comparable answers. Here *comparable*, means the answers to question

topics which were actually posed and answered in both conditions. This gives a measure where we can compare, as if per question, how detailed the actual answers were. As it is possible that more questions were posed in one condition than the other, comparing all texts could result in comparing answers of, for instance, answers to five questions to answers to three questions. This would give a distorted image of how detailed the response to each question actually was.

5   Results

5.1   Questionnaires

The first questionnaire measured if recollecting a memory changed peoples arousal and/or valence regarding the memory. A doubly multivariate repeated measures was done for both arousal and valence, with moment of measurement (pre/post) and system (ontology/non-ontology based) as within-subject factors. No significant results were found ($p >$.05). For both the recollection and the conversation questionnaire, a paired samples t-test was done to compare scores between conditions. Neither of these two questionnaires yielded any significant results between conditions (Recollection: $t(23) = $-.38, $p = $.71, Conversation: $t(23) = $-.27, $p = $.79). On the overall preference, a single-sample t-test showed no significant difference between the result of any of the 4 questions and the middle position on the scale (50), signifying that there was no significant preference for one system over the other (Recollection: $t(23) = 1.26$, $p = $.22, Pleasant: $t(23) = 1.82$, $p = $.81 Questions: $t(23) = 1.17$, $p = $.25 Use again: $t(23) = $.1.57, $p = $.13).

5.2   Answers

A paired t-test was performed on the amount of words typed in answers and the number of question topics answered in both conditions, the results of which are presented in Table 2. The table shows that there is a significantly higher number of total words in the answers with non-ontology based system compared to the answers with the ontology-based system. The result for the number of topics is similar, a significantly higher number of topics was covered with the non-ontology-based system compared to the ontology-based system.

Table 2. Comparison between the number of words and number of topics for the ontology-based and non-ontology-based system.

| | Mean(SD) | | | | | |
Measurement	Ontology	Non-Ontology	t	df	p	Cohen's d
Nr. of words	237 (114)	285 (11)	-2.37	22	0.027	0.33
Nr. of topics	8 (3)	10 (4)	-3.98	22	0.001	-0.54

Table 3. Comparison between the ontology-based and non-ontology-based system based on the total number of objects, people, descriptives, feelings and time references in the participants comparable texts.

Category	Mean(SD)		$F(1,22)$	p	η^2
	Ontology	Non-Ontology			
Objects	21 (14)	20 (13)	1.00	0.328	0.044
People	7 (7)	6 (4)	0.66	0.425	0.029
Descriptive	22 (13)	16 (9)	8.91	**0.007**	0.288
Feeling	5 (3)	6 (4)	0.34	0.567	0.015
Time	4 (3)	3 (3)	0.32	0.263	0.057

Finally, the amount of detail in comparable texts in the two conditions was considered. An omnibus test was done on the annotations of the total number of comparable texts, showing a trend of a higher number of words in the ontology-based system, but no significant result $F(5, 18) = 23.63$, $p = .084$, $\eta^2 = .87$. Table 3 shows the univariate analysis for the individual annotation categories. Here we see that there was a significantly higher number of descriptive words used in the ontology-based system even after a Bonferroni correction which sets the α level at 0.01. None of the other categories showed significant results.

6   Discussion and Conclusion

The first conclusion we can draw based on the results is that no subjective difference between the ontology-based and the non-ontology-based system was found. The second conclusion is that people answered the questions more quickly with the non-ontology based system, as shown by a higher number of topics answered with the non-ontology-based system, while both conditions lasted equally long. This result might indicate that people put more effort into answering the questions from the ontology-based system. When making statements about effort in memory recollection it is, however, also important to consider the amount of detail in the answers and not just the time taken. Concerning this, we see that there is a significantly higher number of descriptive terms for the ontology-based system. From this we can conclude that people describe memories in more detail with this system. Taken together with the result that people take more time, this suggest that people recollect their memories in more detail with an ontology-based system. This study also has some limitations, the main drawback being that the participants tested were healthy individuals, and not PTSD patients. We believe, however, that our results do provide a valid insight in memory recollection with an ontology-based system as it shows that such a system can assist in detailed memory recollection. Future work will have to study the effect of an ontology-based system on the recollection of memories which people would rather forget. One contribution of this study has been to show that aside from high-level planning [5], and adding domain knowledge [2], ontologies can also be

used to store specific knowledge of the user and steer the conversation based on this. It has also shown that the use of multiple-choice options to personalize the conversation [9,15] can be combined with such an ontology. Finally, we have shown that an ontology-based question system is effective in assisting users with detailed memory recollection, as necessary in PTSD exposure therapy [8].

References

1. Diagnostic and statistical manual of mental disorders: American Psychiatric Association, 5th edn. Washington DC (2013)
2. Al-Zubaide, H., Issa, A.A.: Ontbot : ontology based chatbot. In: Innovation in Information and Communication Technology (2011)
3. Augello, A., Pilato, G., Vassallo, G., Gaglio, S.: Chatbots as interface to ontologies. In: Gaglio, S., Lo Re, G. (eds.) Advances in Intelligent Systems and Computing, pp. 285–299. Springer, Switzerland (2014)
4. Bickmore, T.W., Nelson, R.A.S.K., Cheng, D.M., Winter, M., Henault, L., Paasche-Orlow, M.K.: A randomized controlled trial of an automated exercise coach for older adults. J. Am. Geriatr. Soc. **61**(10), 1676–1683 (2013)
5. Bickmore, T., Schulman, D., Sidner, C.: A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J. Biomed. Inform. **44**, 183–197 (2011)
6. Blanson-Henkemans, O., van der Mast, C., van der Boog, P., Neerincx, M., Lindenberg, J., Zwetsloot-Schonk, B.: An online lifestyle diary with a persuasiv computer assistant providing feedback on self-management. Tech. Health Care **17**, 253–267 (2009)
7. Brinkman, W.P., Vermetten, E., van der Steen, M., Neerincx, M.A.: Cognitive engineering of a military multi-modal memory restructuring system. J. Cyber Ther. Rehabil. **4**(1), 83–99 (2011)
8. Foa, E., Hembree, E., Rothbaum, B.O.: Prolonged Exposure Therapy for PTSD: Emotional Processing of Traumatic Experiences Therapist Guide (Treatments That Work). Oxford University Press, New York (2007)
9. Friederichs, S., Bolman, C., Oenema, A., Guyaux, J., Lechner, L.: Motivational interviewing in a web-based physical activity intervention with an avatar: randomized controlled trial. J. Med. Internet Res. **16**(2), e48 (2014)
10. Lucas, G., Gratch, J., King, A., Morency, L.P.: Its only a computer: virtual humans increase willingness to disclose. Comput. Hum. Behav. **37**, 94–100 (2014)
11. Martínez-Miranda, J., Bresó, A., García-Gómez, J.M.: Look on the bright side: a model of cognitive change in virtual agents. In: Bickmore, T., Marsella, S., Sidner, C. (eds.) IVA 2014. LNCS, vol. 8637, pp. 285–294. Springer, Heidelberg (2014)
12. Miller, W.R., Rollnick, S.: Motivational Interviewing: Preparing People to Change Addictive Behavior. Guilford Press, New York (1991)
13. Noy, N.F., McGuinness, D.L.: Ontology development 101: a guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report (2001)

14. Rizzo, A., Lange, B., Buckwalter, J.G., Forbell, E., Kim, J., Sagae, K., Williams, J., Rothbaum, B.O., Difede, J., Reger, G., Parsons, T., Kenny, P.: An intelligent virtual human system for providing healthcare information and support. Med. Meets Virtual Reality **18**, 503–509 (2011)
15. Schulman, D., Bickmore, T., Sidner, C.: An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In: AAAI Spring Symposium Series (2011)
16. Tielman, M., Brinkman, W.-P., Neerincx, M.A.: Design guidelines for a virtual coach for post-traumatic stress disorder patients. In: Bickmore, T., Marsella, S., Sidner, C. (eds.) IVA 2014. LNCS, vol. 8637, pp. 434–437. Springer, Heidelberg (2014)