View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Swedish Institute of Computer Science Publications Database

An Embodied Question Answering System for Use in the Treatment of Eating Disorders

Oscar Täckström[1ǂ], Cecilia Bergh[2], Magnus Sahlgren[1], Marie Sjölinder[1], Per Södersten[2], Modjtaba Zandian[2]

Abstract. This paper presents work in progress on implementing an embodied question answering system, *Dr. Cecilia*, in the form of a virtual caregiver, for use in the treatment of eating disorders. The rationale for the system is grounded in one of the few effective treatments for anorexia and bulimia nervosa. The questions and answers database is encoded using natural language, and is easily updatable by human caregivers without any technical expertise. Matching of users' questions with database entries is performed using a weighted and normalized *n*-gram similarity function. In this paper we give a comprehensive background to and an overview of the system, with a focus on aspects pertaining to natural language processing and user interaction. The system is currently only implemented for Swedish.

1 INTRODUCTION

Eating disorders, in the form of anorexia and bulimia nervosa, pose a large challenge to society, and gravely affect the lives of many young women. About one percent of all females between the age of fourteen and nineteen develop anorexia, characterized by self starvation, while one to three percent of all females in the age of 20-23 develop bulimia, characterized by the eating of large amounts of food with subsequent vomiting [1].

Though a large body of research have been conducted over the last fifty years, eating disorders are still by many considered chronic disorders. A recent review of available treatments for anorexia and bulimia nervosa showed that most, if not all, currently employed methods fail to provide any improvement for the patients [2].

AB Mando[3] is one of the few providers of a treatment for patients with eating disorders that has proved effective. Inspired by this treatment method, we are developing an embodied question answering system, *Dr. Cecilia*, specifically targeted at young women with eating disorders. Our aim is that the system could form a part of the treatment at AB Mando in the future.

In this paper we give an overview of the *Dr. Cecilia* system, its rationale, design and implementation. The next section gives a background to and a rationale for the system. We then describe the embodied conversational agent and the user interaction in sections 3 and 4. Following this, in section 5 we describe the characteristics of the domain and the users, which lead to the choice of similarity metric used in matching users' questions, described in section 6. We then describe the structure of the questions and answers database and the interface for updating this knowledge base, in section 7. Finally, in section 8, we present some conclusions and suggestions for future work.

2 BACKGROUND & RATIONALE

Patients with eating disorders form a rather homogenous group with respect to their thoughts and questions regarding their disorder; especially in the earlier stages of treatment. Anecdotal clinical experience indicate that patients tend to ask the same kinds of questions using similar language, and patients at the onset of treatment constantly think about their disorder and need to be repeatedly ensured that they won't be fooled by the treatment, and reminded of why they should stay in the program. For example, it is common for patients to over and over ask their caregivers questions such as “Will I get fat by the treatment?” and “Why do I panic after eating?”.

The aim of *Dr. Cecilia* is to be able to answer and confirm these patients, at times when a human caregiver is not available. *Dr. Cecilia* is a question answering system, embodied as a virtual caregiver, accessible through a web interface. Questions are posed to the system using written natural language. By providing an embodiment of the question answering system, the hope is to support a more natural feeling dialogue and to promote users' trust in the answers. Furthermore, by leveraging body language and facial expression, the degree of solemnity of the issues discussed can be conveyed across several modalities. In addition it is hoped that an embodied system will be more interesting and “fun” to use, lessening the focus on the patient's problems.

The reader might raise concerns regarding the use of a question answering system in treating patients whose condition is traditionally viewed as caused by a mental disorder. The rationale for the use of *Dr. Cecilia* is that questions related to the psychological state of the patient, such as those listed above, are secondary to the physical condition of the patient and that the patient's eating behaviour is the cause of her psychological state. The primary focus in the treatment should thus be on teaching the patients how to eat again. The scientific basis for this rationale is discussed in detail in [2]. By focusing on providing answers to questions related to the disorder, based on scientific facts and expressed in a professional, yet comprehensible, encouraging, motivating and empathetic language, the hope is that the system will help in disrupting the often ruminating questions and thoughts of the patients.

There are of course potential problems with Internet based treatments. For example, Internet access might be a limiting factor [3]. However a recent survey shows that 95 per cent of the Swedish population in the age group of 15-24 year olds, have

[1] Swedish Institute of Computer Science, Box 1263, SE-164 29 Kista, Sweden.

[ǂ] Contact author: **oscar@sics.se** (other authors in alphabetical order).

[2] Karolinska Institutet, AB Mando, Novum, S-141 04 Huddinge, Sweden.

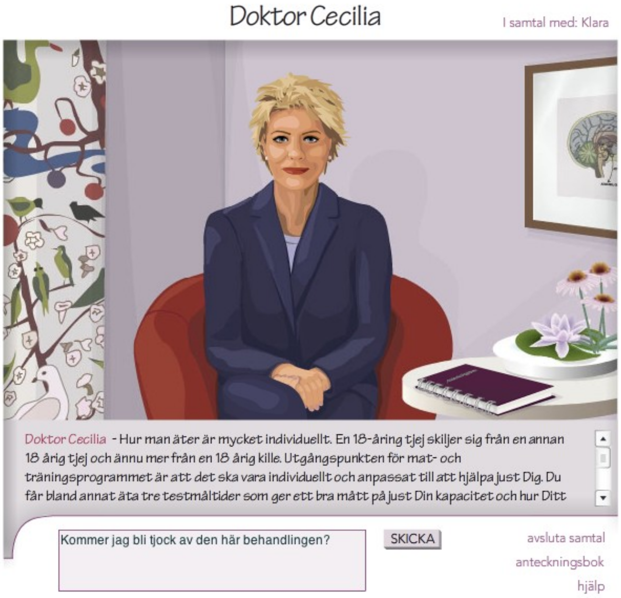
[3] www.mando.se

access to the Internet in their homes, and that 86 per cent use the Internet on a daily basis [4]. Furthermore, it should be noted that the aim of leveraging a virtual caregiver is not to replace the human caregivers. Instead one should view these as complementary parts of the treatment.

Internet based treatments have previously been used successfully for example in the treatment of agoraphobia [5] and ADHD [6]. These naturally focused on the psychological aspects of the disorders and were based on Cognitive Behavioural Therapy and working memory training, respectively. The present work is not directly inspired by this previous work, since the nature of the disorders are very different.

3 EMBODIED CONVERSATIONAL AGENT

While one should not take for granted that the use of an embodied conversational agent (ECA) will benefit all information system applications, there are some properties of ECAs that make them especially beneficial to this specific application. As discussed earlier, patients with eating disorders are not easily motivated, and need constantly to be reminded of and encouraged to as why they should stay in treatment. Numerous studies have shown that these aspects are not easily conveyed through written text alone. For example, internal states of the system can be conveyed more directly, by use of facial expressions and body gestures, than is possible with text alone [7]. Studies have also shown that by using friendly small talk, verbal and non-verbal expression of emotion and expressions of expertise, empathy and trust can be conveyed [8].[4] Furthermore, by using artificial characters a sense of confidence, which can have motivational effects, can be conveyed [9].



Doktor Cecilia - Hur man äter är mycket individuellt. En 18-åring tjej skiljer sig från en annan 18 årig tjej och ännu mer från en 18 årig kille. Utgångspunkten för mat- och träningsprogrammet är att det ska vara individuellt och anpassat till att hjälpa just Dig. Du får bland annat äta tre testmåltider som ger ett bra mått på just Din kapacitet och hur Ditt

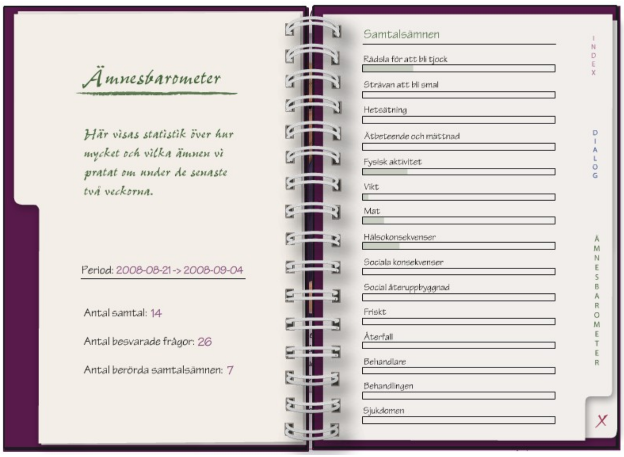
Kommer jag bli tjock av den här behandlingen?

Figure 1. Screenshot of Dr. Cecilia answering the question: "Will this treatment make me fat?"

The *Dr. Cecilia* system is embodied as a conversational agent in the form of a virtual caregiver, placed in an environment resembling the eating disorder clinic at AB Mando, see figure 1. By mimicking the environment in which patients meet their human caregivers, we hope to promote a sense of familiarity and trust. A small initial user study suggests that these goals have been met successfully. Most of the interviewed patients liked the character, and thought that it gave a professional and trustworthy impression. All of the ten patients interviewed in the study pointed out that they liked the environment, and eight of them specifically commented on the similarities with the actual clinic as something good.

4 USER INTERACTION

Although we make use of an ECA, the primary communication channel used in the system is written text. Users interact with *Dr. Cecilia* by posing questions using natural language entered using a standard text-field, see figure 1. If the posed question has a sufficiently high similarity (see section 7) with a question in the database, compared to an empirically determined threshold value, the answer is presented directly to the user by *Dr. Cecilia*. If no such question is found, up to three questions from the database are presented to the user, who can then choose to read the answer to one of the alternatives or to ask another question. An empirically determined lower threshold is used to filter out questions that are unlikely to match the user's question. Since a specific topic often has several different, but highly related, questions and answers, most of the time the reason for presenting more than one answer is that the question posed by the user is in some way underspecified with respect to the questions in the database.



Här visas statistik över hur mycket och vilka ämnen vi pratat om under de senaste två veckorna.

Period: 2008-08-21 -> 2008-09-04

Antal samtal: 14

Antal besvarade frågor: 26

Antal berörda samtalsämnen: 7

Figure 2. Screenshot of the "diary" showing a graphical representation of the user's topic history for the last two weeks.

After each session, the user is provided with a rating from a scale of 1-3 "stars", and a written (pre-formulated) assessment of treatment progress. Each category of questions is assigned a degree of severity in terms of how far the patient is considered to be from remission if she asks questions in this category. The rating of the patient's progress is calculated based on which categories of questions has been in focus during the session.

[4] The cited study used only a Wizard of Oz setting, not a complete and functional system.

The user can also see, in the form of a bar chart, which categories she has asked most questions about, see figure 2. This allows the patient to keep track of her own development towards recovery, and gives her positive and negative feedback, hopefully pushing her in the right direction. This kind of feedback is an integral part of the treatment method on which *Dr. Cecilia* is based. The user study mentioned above also showed that the patients found this kind of feedback very useful, and they especially liked the bar chart representation.

5 DOMAIN CHARACTERISTICS

A number of characteristics of the domain and the specific patient group, in combination with the rather small scale of the project, has influenced the choices made in designing the system.

First of all, the domain which the system needs to cover is too broad for the system to make use of any encoded expert or common sense knowledge. Secondly, since the system is to be employed in a real-world situation at an eating disorder clinic, caregivers with no technical training should be able to quickly expand the knowledge base using only natural language, in case the system lacks coverage of a certain sub-domain. It follows that the knowledge base in principle should consist of a set of more or less frequently asked questions and answers to these questions.

Matching users questions to this database on the surface looks like a standard information retrieval (IR) task. However, the domain and task at hand differs from those considered in traditional IR in a number of ways:

- *Small data set* – the number of different answers is quite small (currently about 500).
- *Short documents* – each answer contains quite a few number of words (on average about 70)
- *Narrow sub-domains* – several different answers are often related to different aspects of the same topic.
- *Fatigued users* – patients with eating disorders often find it hard to concentrate for a longer period of time.

Traditional IR methods, using the bag-of-words (BoW) representation or latent semantic indexing (LSI) [10], work well on large data sets, with longer documents, however they often place a high burden on the users to sift through search result sets in order to find what they are really looking for. In the case of patients with eating disorders, we want to keep search result set size down to a minimum, preferably delivering the correct answer to a question directly.

Due to the short length of the answers, the system must handle morphological variation. The system should also be able to handle spelling errors gracefully, since the users are young and often to some degree fatigued spelling errors are expected to be common. Furthermore, since the users interact with the system in Swedish, compound words must be handled appropriately.[5] These issues are not as important when documents are long and drawn from a large domain, and when there is no limit on search result size, since the probability of a certain morphological form, or spelling variation, occurring at least somewhere in at least some relevant document is high.

Given these considerations, we instead link each answer in the database to a set of different, but "synonymous", question formulations, and use an *n*-gram based similarity measure for matching users' questions to those in the database.

6 WEIGHTED *N*-GRAM SIMILARITY

As discussed, the standard BoW representation used in IR is not suited to the present task. Instead we need a representation and similarity measure that handles morphological variation, spelling errors, and compound words. String similarity measures such as Levenshtein distance is commonly used in spelling correction [11], and handles morphological variation satisfactory, however, it is not suited for use with compound words. Furthermore, it is only useful on the word level and can not make use of word order information.

Character based *n*-gram similarity measures form a family of similarity measures that handles morphological variation, spelling errors, and compound words, and can make use of word order information. Though BoW based representations, using e.g. cosine similarity, usually gives somewhat higher precision, *n*-gram similarity measures usually give a much higher recall in the presence of noise of the types discussed above. This is an important property given that the set of relevant answers to a specific question is very small in this domain. By empirical inspection, a value of n=3 was chosen. This gave the best balance between precision and recall.

In order to take the importance of different *n*-grams into account, we use a weighted version of *n*-gram similarity, in which certain *n*-grams are down-weighted. This is analogous to the inverse document frequency weighting function commonly used in IR. Unfortunately we cannot use inverse *n*-gram frequency directly, since some short but highly relevant words such as "mat" ("food") is common in the database. Instead we currently use the heuristic of feeding the *n*-gram index with words and phrases that are commonly used in questions, without carrying much content, e.g. "varför" ("why"), but that are potentially useful in disambiguating between similar questions with different answers.

Computing the *n*-gram similarity of two strings is quite fast, with time complexity in $O(mp \log p)$ using *n*-gram sorting and $O(mp)$ using suffix arrays, where m is the number of questions in the knowledge base and p is the number of unique *n*-grams in the string. By using cover trees, finding the closest k matches to a question is potentially much faster [12]. It is also possible to use inverted *n*-gram indexing to further improve performance [13]. The current system uses the implementation based on sorting, which is fast enough given the current database size.

7 QUESTIONS AND ANSWERS DATABASE

As described above, the knowledge base of the system consists in a database of questions and answers formulated in natural language. Each answer can be linked to several different question formulations, in order to capture language variations that are not handled by the *n*-gram similarity function.

Answers are assigned to topics, e.g. *Anxiety* and *Food,* with topics hierarchically related, e.g. *Depression* forms a subcategory of *Health Consequences*. The topic hierarchy is used to give the user a visual presentation of how much she has

[5] Where English has multi-word terms, Swedish instead has compound words, e.g. "medical nurse" translate to "sjuksköterska", with "sköterska" being the Swedish translation of "nurse".

focused on specific topics as described above. It is also useful to aid the human caregivers in the addition of new questions and answers to the knowledge base. Each answer is further assigned a mood, which is currently used to control the body language and facial expressions of *Dr. Cecilia* as described above.

In order to collect data and involve the users early on in the design of the system, a small scale pilot study was performed. This took the form of a "Wizard of Oz" study [14], in which patients communicated with a human caregiver through a simple text based interface. The experiment resulted in an initial database of questions and answers, which has subsequently been refined. It also gave valuable information on the characteristics of the language use of the patients, the degree of spelling errors and other aspects of language variation.

A web based "editor interface" aid in the expansion of the questions and answers database. Using this interface, a human caregiver can add, remove and update questions and answers, add alternative formulations to an answer, and browse and search questions posed to the system that were left unanswered. We are currently experimenting with using hierarchical clustering on the set of unanswered questions, which through the weighted *n*-gram similarity is endowed with a metric. This allows the editor to find sets of commonly asked questions, and to work on all questions in such a set of related questions simultaneously, to add alternative formulations until all formulations in the subset are covered by the database. The actual clustering, in the form of hierarchical single linkage clustering, is currently performed off-line, while model selection, i.e. selecting the number of clusters, is performed interactively.[6]

While the database of questions and answers is necessary for the embodied question answering system, it could also be potentially useful in providing human caregivers with a common ground, so that the patients are not given conflicting answers when asking different caregivers the same question. It could also potentially be useful in highlighting mildly frequently asked questions and topics, allowing the human caregivers to better prepare for these questions.

8 CONCLUSIONS & FUTURE WORK

In this paper we presented a complete and usable embodied question answering system for use in the treatment of eating disorders, based on one of the few available effective treatments for anorexia and bulimia nervosa. An initial interview-based user study shows that there is indeed a want for this kind of system among patients with eating disorders, and suggests that the current system fulfils the demands of these patients satisfactorily.

Currently data collected in a more extensive user test is being evaluated, the results of which will be used to determine the final steps of the implementation and possible improvements to all aspects of the system. We have high hopes on the utility of the final system, which is to be finished later this fall. Although the current system is well received by the patients, it is important to point out that there is yet no evidence that the system is an effective treatment for eating disorders. This can only be proved by conducting a randomized controlled trial in a clinical setting.

The system was developed on a small budget, using open source software and standard web technology throughout, showing that it is indeed possible to build usable systems in this domain as a small scale research project with limited funding. Currently, the system is only implemented for Swedish, but the natural language processing modules are in principle language agnostic.

Of course there are still much room for improvement, both to the dialogue and interaction aspects and to the natural language processing aspects of the system. We are happy to receive comments and suggestions for further improvements from the participants of this workshop.

ACKNOWLEDGEMENTS

The authors would like to thank the staff of AB Mando and the patients that have participated in the studies leading to the system described in this paper. We would also like to thank Jussi Karlgren for comments on this paper. This project was funded by VINNOVA (Swedish Governmental Agency for Innovation Systems), along with AB Mando and SICS.

REFERENCES

[1] Bergh C., Brodin U., Lindberg G., and Södersten P. Randomized controlled trial of a treatment for anorexia and bulimia nervosa. In: *Proceedings of the National Academy of Sciences*, USA 99:9486-91, (2002).

[2] Södersten P., Nergårdh R., Bergh C., Zandian M. and Scheurink A. Behavioral neuroendocrinology and treatment of anorexia nervosa. *Front Neuroendocrinol*. Jun 14. [Epub ahead of print] (2008).

[3] Carlbring P, Andersson G. Internet and psychological treatment. How well can they be combined? *Computers in Human Behaviour.* 22(3):545–553, (2006).

[4] Nordicom-Sveriges Internetbarometer 2007. *MedieNotiser* 2/2008. ISSN 1101-4539. (2008).

[5] Carlbring P. Panic! Its Prevalence, Diagnosis and Treatment via the Internet. Doctoral thesis, Department of Psychology, Uppsala University, (2004).

[6] Klingberg T., Forssberg H., & Westerberg H. (2002). Training of working memory in children with ADHD. *J Clin Exp Neuropsychol*. 24(6):781-91, (2002).

[7] Maes, P. Agents that reduce work and information overload, Communications of *the ACM*, 37(7):30-40 (1994).

[8] de Rosis, F., Cavalluzi, A., Mazotta, I. and Novielli, N. Can embodied conversational agents induce empathy in users? In: *AISB'05 Virtual Characters Symposium*, (2005).

[9] Berry, D., Butler, L., de Rosis, F., Laaksolahti, J., Pelachaud, C., and Steedman, M. Final evaluation report, the Magister project. January (2004).

[10] Dumais, S.T., Furnas, G.S., Landauer, T.K. and Deerwester, S.C. Using latent semantic analysis to improve information retrieval. In: *Proceedings of CHI'88: Conference on Human Factors in Computing*, 281–285, (1988).

[11] Levenshtein, V.I. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 1966(*10):707–710, (1966).

[12] Beygelzimer, A., Kakade, S., and Langford, J. Cover trees for nearest neighbor. In: *Proceedings of the 23rd international Conference on Machine Learning.* ICML '06, vol.148:97-104, (2005).

[13] Kim, M., Whang, K., Lee, J., and Lee, M. N-gram/2L: A Space and Time Efficient Two-level n-gram Inverted Index Structure. *Proceedings of the 31st international Conference on Very Large Data Bases*. Very Large Data Bases. VLDB Endowment, 325-336, (2005).

[14] Dahlbäck, N., Jönsson, A., and Ahrenberg, L. Wizard of Oz Studies - Why and How. *Knowledge-Based Systems*, 6(4):258-266, (1993).

[6] We use the freely available *PyCluster* interface to the *C Clustering* library, available at http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster.