Generation of Personalised Advisory Messages: an Ontology Based Approach

Elisabetta Erriquez
Department of Informatics
University of Bari, Italy
e.erriquez@yahoo.it

Floriana Grasso
Department of Computer Science
University of Liverpool, UK
F.Grasso@liverpool.ac.uk

Abstract

This paper presents an ontology based approach to the production of personalised motivational messages to the user of a e-Health service. The service is aimed at encouraging diabetic patients to do physical activity. The textual messages have an overall predetermined structure, and need to be personalised to the user's characteristics, and to the context in which the message needs to be generated.

1 Introduction

The context of this paper is the European research project PIPS (Personalised Information Platform for Health and Life Services)[1], which investigates the use of an eHealth platform, combined with motivational tools, for health promotion. In particular, we focus on an intervention, called "Strolling and Motivation", aimed at introducing a programme of physical activity to diabetic patients [5]. The main objective of the intervention is to act as an incentive to perform a daily walking activity, the entity and the schedule of which is agreed by the patient and the medical supervisor at enrolment time. The motivational strategy is based on personalised messages, which are automatically sent to the user's mobile phone throughout each day. A reminder is sent when it's time to start the scheduled activity, and, at the end of the day, a message is sent with feedback on the daily performance, as a way to support and encourage on how to improve, as well as summarising the performance itself. The composition of these messages is supported by the *Trans-theoretical Model of Change* ([6] as extended by [1]). A "Stage of Change", as defined by the theory, is attributed to a patient by means of a standard questionnaire. This standard characterisation of the user is however augmented with extra features, in order to take into account a range of other factors which might influence the users' decision to change their lifestyle (for more details on the strategy see [5]).

[1]http://www.pips.eu.org/ - FP6 Contract n. 507019, January 2004 - June 2008.

In a first prototype of the message composition system [5], implemented in order to carry out a pilot study, the generation of the messages was achieved by means of a composition algorithm using constraint satisfaction techniques, selecting among over 500 canned texts. A message consisted of various segments, each of which filled in by one canned message from the dataset. In order to fill in a segment, the system selected all the canned texts appropriate to the current situation of the patient, and among these, one was chosen at random. The system was well received by the sample users, however, in order to guarantee a good variety to messages, it was necessary to store a very high number of equivalent segments. This process is not scalable, obviously. The present paper reports on an improvement on this solution, based on an ontology approach.

2 Towards an Ontology Based Approach

In order to improve on the approach implemented in the prototype, and in line with the overall philosophy on knowledge representation adopted in PIPS, we set ourselves the task of collecting and cataloguing all messages, or pieces of messages, on the basis of an ontology. We started by analysing all the messages which formed the pool of choice for the first prototype. The purpose of the analysis was the identification of an overall structure, not so much, or not only, in terms of the components of the message, but also in terms of the factors that are useful for personalisation. In this section we briefly present the set of factors we took into account, and we describe the ontological analysis we performed.

2.1 Personalisation Factors

In the preliminary research towards the optimal composition of the messages, a psychologist was contacted, who provided, among other things, a set of sample messages, tailored to users with different characteristics. The material provided by the psychologist was encapsulated in the first prototype of the system, although not in a systematic fashion. Personalisation, in the first prototype, is performed on the basis of various factors. A knowledge base captures the information related to the habits and the preferences of the user, especially with respect to physical activity, collected by means of a questionnaire (the "Exercise Preference/Habits Questionnaire") that users can fill in and update at any time. The questionnaire does not lead to a classification of the users as such, but the replies are utilised by the system for several purposes, including the personalisation of the messages.

To add extra elements to the personalisation, and as a way to make the interaction with the system more appealing, the users can each day input information related to four aspects of their daily life:

1. the level of gratification from the working day;
2. the user's perception of the social relationships with friends and family on the day;
3. the weather;
4. the overall emotional orientation, or "mood", of the day (as a choice among emoticons).

A further, important input is of course how far the patient is in the programme (initial stages, advanced level, etc.), and, in case of a negative performance, the possible causes for this "failure" in the achievement of the targets, reported by the patient directly by choosing among predefined possibilities.

All the above factors are represented as flat labels in the prototype algorithm, aiding the constraint selection process. In the new system, we needed to take all this into account to be faithful to the spirit of the behavioural theory adopted.

2.2 Ontological Analysis of the Messages

We describe here the ontological analysis we performed on the messages, summarising the major characteristics we identified. We report on the various stages of the analysis, subdivided into the main topics under attention. Then, in the next sections, we present our proposed ontologies: the "Ontology of Discourse" and the "User Ontology", and we show how the two ontologies are linked and contribute to the production of the messages.

2.2.1 Message "Macro-Structure"

The starting point for the analysis of the pool of messages was the Stage of Change of the user. The first aim of the analysis was the definition of a macro structure of the message composition. This was done also in the constraint satisfaction approach, in order to define the portions of the message to select, but in our case the analysis was driven by the identification of the speaker's intention. This is a very common approach in classic template-based natural language generation systems [7].

The main components of the message emerged from our analysis are as follows:

1. A **comment**, which aims at providing a feedback on the user's performance.
2. An **argument**, which aims at supporting the final thesis that each message is set to maintain: that is that the physical activity is good for the user.
3. An **aid to introspection**, which attempts to relate the user's performance with the diary values that the user has given as input on the day.
4. A **suggestion**, aimed at providing both general and practical indication on how the users can improve their performance.
5. An **encouragement**, briefly concluding the communication.

Some of these macro-components are to be considered optional, so that the final messages can be varied also in terms of length and content. Let us analyse each of them in turn.

2.2.2 The comment component

A comment is composed of several sub-components:

1. An **assessment**: a brief, personalised message, reflecting the level of performance. This has an orientation, depending on the performance itself:
   (a) Positive for a performance meeting or surpassing the target;
   (b) Negative for a substandard performance.
2. A **report** on the actual performance.
3. A **'trend' comment** on the temporal development of the performance, to provide the user with the bigger picture of how things are going overall.

2.2.3 The argumentation component

This component aims at influencing the users, by persuading them to improve their physical activity, therefore particular attention should be paid in its modelling. In order for the message to be truly persuasive, research has shown that it should be personalised to the user's interests and characteristics [4]. The diary values are then useful to this purpose, but they are not enough, obviously. The more we know about the patients, their habits, preferences, and beliefs, either coming from the initial questionnaire or from other records, such as the medical profile, the more elements we have to strengthen the persuasive force of a message. Moreover, information about the domain (like diabetes, nutritional knowledge, etc.) gives the opportunity to produce rational arguments in favour of given choices [2]. Finally, this component is obviously influenced by the Stage of Change: for example, in case of a user well advanced in the progression through change (Behaviour in Action or Maintenance stages), the argumentative message might be superfluous, as the user should already be in a mental state acknowledging the benefits of physical activity, or might need to be refocused on another aim (for example to reinforce the behaviour and avoid relapse).

2.2.4 The aid to introspection component

The aim of this component, as devised by the psychologist, is twofold. On one hand, it aims at helping the user achieve a more holistic understanding of how all components of his life fit together. On the other hand, it aims at highlighting which specific aspects have been in fact influenced by the performance, and which have influenced the performance in turn. This component is strictly linked to the argumentative one. For instance, if a patient had a bad day at work, and also had a bad performance, this component will highlight the fact that the performance might have been influenced by the negative working day, but at the same time it will point out that more commitment to the physical activity might benefit the working performances in turn. This can be reinforced by, for instance, saying that clinical studies have shown that a good level of physical activity improves the functionality of the brain. This is a good example of how one can use information on the user to personalise the message, and how to compose the parts of the message in a more efficacious way.

2.2.5 The suggestion component

The aim of this component is to provide tips for the user to help perform better. This component is heavily related to the Stage of Change: if the user has not yet achieved full understanding of the importance of physical activity, practical suggestions are superfluous, or will risk not to be recalled at all, and a generic message might be more appropriate. A patient who is fully engaged in the changing process and working actively towards the target will, instead, welcome a practical suggestion. Such suggestion has to be, again, personalised on the basis of the habits, preferences and beliefs of the patient.

2.2.6 The encouragement component

This is just, mainly, a brief message to show support, and hopefully boost confidence, but also a way to add naturalness to the message by acting as a conclusion.

3 Ontology of Discourse

The ontology of discourse we devised contains all the information related to the messages, and all the factors that allow personalisation, but more importantly it encapsulates the structure itself of the message. Therefore, the root of the ontology is the "message" in its entirety. The structure is rendered by the macro/sub component link, reproduced by means of a classic "has-part" relationship. For instance, if the class X is related, by means of a has-part relationship, to classes X_1, X_2 and X_3, this means that message X is composed of X_1, X_2 and X_3. Each concept can have an attribute, "optional", which indicates whether it can be omitted from the message.

From the message analysis we extracted a taxonomy, which was particularly useful for the purpose of modelling the argumentative component and the suggestion. In the case of the argumentative messages, the examples provided by the psychologist were partitioned into classes, on the basis of the characteristics of the user model that were more relevant to the class. This allows to choose the argumentative messages which is more personalised to the aspects of the patient's life, the one he is likely to be more susceptible to. In the case of the suggestions, these are classified in such a way that one can select, among the various contents, the one which are "closer" to the habits and the preferences of the user, as indicated in the initial questionnaire.

The argumentative message is subdivided into subclasses, according to the taxonomy identified, for example:

- Appeal to positive consequence;
- Appeal to negative consequence;
- Message promoting the value of "Work";
- Message promoting the value of "Health";

This is not a partitioning: the classes are not necessarily non-overlapping. This allows to enter an instance of the ontology belonging to more than one class, to reflect the fact that a message can be at the same time an "appeal to positive consequences" and "promoting health". In this way we can avoid repetition and redundancy, and also move from general messages to more specific and personalised ones, according to the information we have and to the strategy we decide to adopt. Moreover, the argumentation strategies are not by any means exhaustive: we identified the ones most commonly used in the pool of messages we analysed, but we assume many more can be included. Figure 1 shows an extract of the discourse ontology.

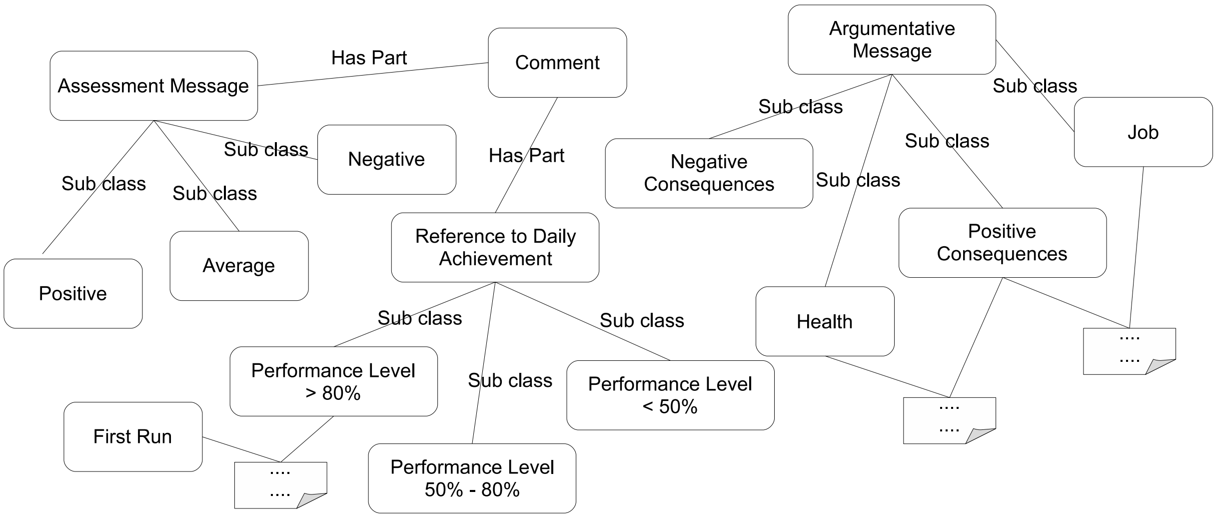

Figure 1. Discourse Ontology (extract)

4 The User Ontology

Given that the information on the user is strictly related to the message content, we need to represent the user model in a way that makes this correlation easier. Therefore, in this case also, an ontological approach has been used, in order not only to connect the user model to the discourse model, but also to give a more appropriate structural representation to the user model itself. In order to define the user model ontology, we grounded our work on a recent research effort, the General User Model Ontology (GUMO) [3]. This is an ontology aimed at a uniform interpretation of the user model when distributed in semantic web enriched environments. GUMO starts from the standpoint that a top level ontology, commonly accepted, would be of a great importance to the field of User Modelling. The greatest advantage of such ontology would be a simplification of data exchange among different systems. The problem of syntactic and structural differences among various systems would be overcome with the use of a shared ontology, specialised to user modelling tasks. The authors propose a complex ontology, including also the inference rules and a vast knowledge related to user modelling research. Therefore, the ontology presents general information (like age), but also abilities (can swim), as well as interests and preferences (reading poems, adventure games, etc.). GUMO has been designed with a Situational Statements approach. This means that the dimensions of a user model are divided in three parts, `auxiliary`, `predicate` and `range` so:

> If one wants to say *something about the user's interest in football*, one could divide this into the `auxiliary`=*hasInterest*, the `predicate`=*football* and the `range`=*low-medium-high*. If one wants to express something like *knowledge about symphonies*, one could divide this into the `auxiliary`=*hasKnowledge*, the `predicate`=*symphonies* and the `range`=*poor-average-good-excellent* [3, p. 429].

Around 1000 groups of auxiliaries, predicates and ranges have been identified in GUMO, like *hasInterest, hasBelief, hasPlan, hasProperty, hasGoal*. The list is not meant to be complete or exhaustive, but a starting point to analyse and represent a good number of facts related to the user.

We drew inspiration from GUMO in order to create our user model, as explained in what follows. The user model ontology, as in GUMO, is structured in terms of "dimensions". The main dimensions are:

- Those related to the user's beliefs.
- Those related to the main features to consider for the messages, as follows:
  - Family and love;
  - Habits;
  - Interests;
  - Physical location;
  - Preferences.
- Those related to the strolling programme.

Most of this information is directly related to part(s) of the discourse, as illustrated in the next section, so that they can serve as personalisation features in the production of the discourse. In Figure 2 an extract of the main classes is shown.

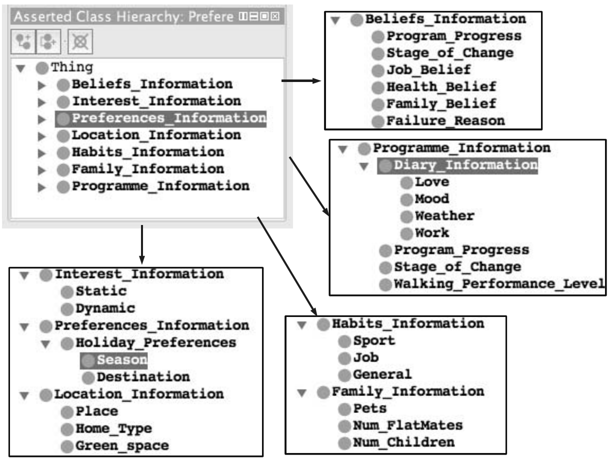

Figure 2. User Ontology (snapshots)

5 Composition as Ontology Querying

In creating a synergy between the two ontologies, we aimed at using a general method, which could be reproduced in other domains. Therefore, we experimented with the creation of a sort of stereotypes, or better "views", on the most important features needed for the message composition. Let us explain in detail. For each user characteristic that is relevant to the message, a class definition is created, on the basis of such characteristic. For example: in Beliefs_Information we might have information on the user's attitudes towards various topics, such as work (Job_Belief). Also, the user can daily input the diary values related to this aspect of his life. By means of a class restriction, we can make sure that the system, the reasoner used on the ontology in this case, automatically classifies the user as belonging to a class, characterised by the feature under consideration. For example, the class User_JobVImp is defined by the axiom:

$$\exists \text{ has\_belief.(Job\_Belief AND } \exists \text{ has\_value.}\{\text{High}\})$$

that is: "define the class User_JobVImp as the set of the users for whom the property has_belief is referred to Job_Belief, and has value High". At the moment, over 30 axioms have been defined to classify a user according to the relevant characteristics, or those useful for personalisation. Each class defined in such a way can be associated with the discourse segments defined in the Discourse Ontology. Moreover, one can combine the two views, so that the association is more restrictive. Once the User model is instantiated, the reasoner will classify the user according to the given criteria. When the message needs to be generated, the system will automatically fetch all the discourse segments that are appropriate and relevant for that user. As explained before, the structure of the discourse is expressed by means of the has_part property, while the possible refinement of a concept is expressed by means of the sub-class relationship. Every time a message needs to be created, the system extracts, starting from the root, each leaf in the has_part relationship. Then, for each of these parts, by enforcing the established constraints in connection with the user ontology, the "class" is isolated in the refinement hierarchy that best relates to the user. This means that, among various alternatives, once the leaf is identified in the has_part relationship, the message will be chosen that satisfies the highest number of constraints, as this will be the most personalised one. This allows managing the conflicts, in the choice among alternatives for the concept to utilise, by exploiting the hierarchy. The criterion can therefore be paraphrased as "choose the candidate that is closest to the hierarchy leaves". Similarly, when more candidates are identified (it may well be that more than one concept is acceptable), the message is chosen which is an instance of the highest number of the candidate concepts.

For example, suppose that the user to be considered is in Stage of Change "Contemplation": this tells us that he is concerned about his health, so the precondition for the Health class will be considered as satisfied. Also, suppose that the mood value for that day is low, so the class Positive Consequence is selected. Finally, suppose the user considers his work as very important in his life, and that the diary job value for that day is also low. Then, these four values in the user model determine the choice, in the example of the Argumentative Message, of three classes: Positive Consequence, Health, and Job. Once the classes which will form the message are identified, the system looks for that specific message which is an instance of all the classes, if it exists. If not, the message instance of the highest number of classes is selected. When many candidates have the same value, the current strategy selects one at random, though some further work is in progress on a strategy which takes into account the dialogue history and other factors to improve this choice.

Figure 3 shows the choice of a complete message. Note that in case of incomplete information, for example if the system does not have information on how far the user is in the programme, the more general class is chosen (Performance $> 80\%$, with the message "*You're doing a great job!*" in the figure).

6 Conclusions

The main hypothesis of this work is that the messages are predictable enough to be suitable for generation by a template style mechanism for natural language generation, a very standard domain hypothesis [7]. The work here

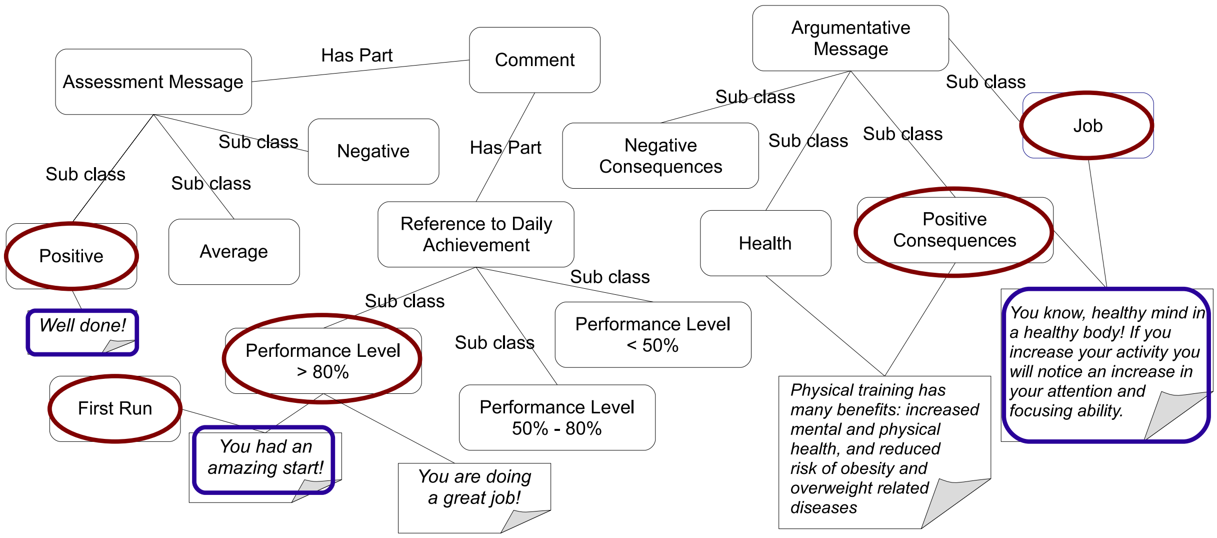

Figure 3. Message Composition

is however more than a simple schemata: while in pure schemata approaches, the system is not able to justify its choices, in our case it is the ontology that allows this capability. In an ontology each concept is associated with an explicit "meaning", therefore by the very same link between the discourse and the user ontologies, the system is able to "explain" why a given message has been composed in a given way, and leaves space to a possible refinement of the system itself, where the composition strategy can take this information into account. Moreover, a better refinement of the knowledge base will allow expressing also other types of constraint, such as those based on the history of previous transactions: in these terms, our system is highly scalable. Our approach, therefore, keeps the simplicity and efficiency of the schemata approach, while at the same time it allows keeping track of the rationale of each message, thus allowing the possibility to extend the system by implementing a recover function for misunderstandings from the users, or for failed communicative goals.

Both ontologies have been built in the web ontology language OWL, while the querying system has been implemented as a stand-alone application using Java and the JENA API to interact with the ontologies. Work is still in progress for a complete population of the ontologies, as well as the refinement of the strategies for resolving conflicts.

An important aspect still to be covered in this work is evaluation. Research has shown that evaluation may lead to surprises [8], so we are not in the position to advance claims in this respect. The first prototype of the system (the one with the constraint satisfaction approach) has been evaluated with a sample set of users, with a positive outcome, so we hope that adding more personalisation can improve on this result.

References

[1] E. del Hoyo-Barbolla, R. Kukafka, M. Arredondo, and M. Ortega. A new perspective in the promotion of e-health. *Stud Health Technol Inform.*, 124:404–412, 2006.

[2] F. Grasso, A. Cawsey, and R. Jones. Dialectical Argumentation to Solve Conflicts in Advice Giving: a case study in the promotion of healthy nutrition. *International Journal of Human Machine Studies*, 53(6):1077–1115, 2000.

[3] D. Heckmann, T. Schwartz, B. Brandherm, M. Schmitz, and M. von Wilamowitz-Moellendorff. Gumo - the general user model ontology. In L. A. et al, editor, *User Modeling 2005, 10th International Conference, LNCS 3538*, pages 428–432, 2005.

[4] M. Kreuter and V. Strecher. Do tailored behavior change messages enhance the effectiveness of health risk appraisal? results from a randomized trial. *Health Education Research*, 11(1):97–105, 1996.

[5] A. Morandi and R. Serafin. A personalized motivation strategy for physical activity promotion in diabetic subjects. In A. Cawsey, F. Grasso, C. Paris, S. Quaglini, and R. Wilkinson, editors, *2nd workshop on Personalisation for eHealth, User Modeling 2007*, 2007.

[6] J. Prochaska, W. Velicer, J. Rossi, M. Goldstein, B. Marcus, W. Rakowski, C. Fiore, L. Harlow, C. Redding, D. Rosenbloom, and S. Rossi. Stages of Change and Decisional Balance for 12 Problem Behaviors. *Health Psychology*, 13(1):39–46, 1994.

[7] E. Reiter and R. Dale. *Building Natural Language Generation Systems*. Cambridge University Press, 2006.

[8] E. Reiter, R. Robertson, and L. Osman. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence*, 144:41–58, 2003.