User Models and Domain Ontologies for Generating Personalized Questions and Answers

Marta Gatius
Computer Science Department
Technical University of Catalonia
Barcelona, Spain
e-mail: gatius@lsi.upc.edu

***Abstract*—The incorporation of user models has become more popular in recent years. This paper describes how user models, domain ontologies and natural language generation techniques can be integrated to generate personalized questions and answers related to a particular scenario. For this purpose, a general organization of different types of knowledge involved is proposed, that facilitates its adaptation to different domains, user models and languages. In this proposal, the domain conceptual knowledge is represented by a domain ontology, that provides a framework for representing the user's knowledge and interests. Ontologies provide a powerful formalism because they allow the representation of detailed descriptions of concepts, attributes and relations, thus facilitating richer inferences about user knowledge and interests. Additionally, in the approach presented, the domain concepts in the ontology are also related to a general syntactic-semantic taxonomy that facilitates the generation of more appropriate questions and answers for each user. The proposed organization has been applied for generating personalized questions and answers, that help the user when searching information about university courses**.

Keywords-user models; personalized content generation; domain ontologies, multilingual language generation.

I. INTRODUCTION

The integration of user models and domain ontologies is a relatively new direction of research to personalize different types of contents and applications. Although user models have been used since the 80's, during the last years the interest in adapting the contents to the different types of users is increasing in many areas (as described in [1] and [2]). The huge amount of contents and systems in the web, where user population is diverse, have also increased the interest on the incorporation of users models in different types of systems, such as web communication systems, intelligent web systems, information extraction, e-learning and e-health.

User's models can be used for personalized content generation and can also be incorporated in adaptive systems, that behave different for different users. User's models can incorporate diverse information from the user. Most models use one of the following features about the user: background (i.e., profession), knowledge, interest, goals and individual traits (personality and cognitive factors). Information on the context of the user's work can also be modeled. Additionally, more than one of these features can also be combined.

The use of a specific user's feature is usually related to the type of the application that incorporates the model. Thus, user's background is used most frequently for content adaptation (as in the work on medical domain described in [3]), user's interests are usually incorporated in information retrieval and recommended systems (such as the dialogue system described in [4]), user's knowledge is useful in educational systems, user's goals are incorporated in intelligent interfaces, while individual traits are considered in several commercial systems (such as the system described in [5]).

This paper describes the use of information about the user's backgrounds to generate personalized questions (and their answers) in a specific domain. Information about the user profession or experience is considered when classifying the users in different groups or stereotypes. Specific questions and answers for each group are generated semi-automatically from the domain ontology.

Stereotypes have been used for more than 25 years because they are simple and can be powerful for several domains. Stereotypes have been used in domains where the distinction of two or three user groups is easy and useful, such as the medical domain, where users can be classified as professionals (and subclassified as doctors, nurses and students) and patients (and subclassified considering different knowledge levels). Information necessary to classify the user is provided explicitly, usually by the user.

There are already several research works that integrate user models and domain ontologies [6][7][8]. Ontologies provide a framework for representing the user knowledge (or interest) about different domain concept, or about the attributes describing the concept. Ontologies provides a more powerful formalism than other forms of semantic

 ISBN: 978-1-61208-369-8

representation, used in previous work on user adaptation (such as concept networks), because they allow the representation of detailed descriptions of concepts, attributes and relations, thus facilitating richer inferences about user knowledge and interests.

The approach described in this paper differs from other works, such as the above mentioned, in that it is focused on the optimization of the generation of questions and answers in several languages. For this purpose, it uses a general syntactic-semantic taxonomy (described in [9]), that establishes general relations between the ontology concepts and the linguistic structures supporting questions and answers about them.

The integration of user models and language generation techniques has also focused several research works, such as the relevant information-systems Ilexand and PEBA-II (described in [10]), that dynamically generates pages according to the user profiles. However, those works used other language generation techniques, not based on ontologies.

The organization of the conceptual and linguistic knowledge that is proposed in this paper was already incorporated in an adaptive dialogue system (described in [11]), where information on how well the communication is doing is used by the system to select the most appropriate respond to a user's intervention: a directed (explicit) question or a more open message. The approach presented in this paper incorporates a user model that uses the domain ontology to represent the user background related to the domain concepts.

In the approach presented, the process to generate the personalized questions and answers consists of four steps: The study of the different types of users, the representation of the domain concepts, the relation of the attributes describing concepts to the syntactic-semantic taxonomy and, finally, the generation of questions and answers.

The work described is being done in the framework of the Spanish project SKATER, for Scenario Knowledge Acquisition by Textual Reading. The personalized questions and answers are generated from the ontology describing a particular scenario, previously built from text.

The next section describes how the proposed approach has been applied for a particular scenario, in which the user is seeking specific information about a particular university course. Then, the last section describes the conclusion and future work.

II. GENERATING PERSONALIZED QUESTIONS AND ANSWERS

As mentioned in the introduction, several systems incorporate user models and ontologies to adapt contents and presentations to different types of users. For example, in several scenarios, the generation of personalized questions, dialogues and summaries can help the adaptation of the web contents and systems to the user's needs.

Domain ontologies have been incorporated in many different types of applications because they provide a great flexibility to represent knowledge and are easy to adapt to different domains. Furthermore, they are especially appropriate for representing conceptual knowledge in multilingual systems, because they provide a common semantic representation for the several languages supported.

Recently, domain ontologies have also been used for personalizing presentation in several communication systems (such as [12] and [13]). Ontologies favor the generation of personalized content because they provide a framework for representing user's domain knowledge (or interest). In more simple semantic representations the user's knowledge about a domain concept is represented as a binary value (known-not known) or as a weighted value. Ontologies allow the representation of more detailed information on the user's knowledge about a specific concept, such as that related to the different attributes describing the concept and the relations with other concepts. Additionally, they also support inference and other reasoning mechanisms.

This section describes how domain ontologies are used to generate personalized questions and answers for several user's stereotypes. In a particular domain, groups of users or stereotypes are distinguished because they have different interests and/or knowledge. This information is represented as an overlay of the domain concepts in the ontology.

The approach presented has been applied to the scenario in which the user searches for specific information about a specific university course. In this particular scenario, the representation of the domain concepts and the interests of the different type of users is simple, as shown in Figure 1. There are only two domain concepts involved and two different types of users: teachers and students. The domain concepts are described by a set of attributes, and user domain interests are represented as a binary value associated with each of the conceptual attributes. When generating personalized questions, only the conceptual attributes interesting for each specific user type are considered.

This approach uses of a syntactic-semantic taxonomy that defines general relations between conceptual attributes and linguistic knowledge to facilitate the generation of the questions for different domains and languages.

The process of generating personalized questions and answers for a particular scenario consists of the four steps:

- Study of the needs of the different types of users.
- Description of the domain concepts involved in the scenario**.** Each concept is described by a set of attributes and those attributes are associated with the specific types of users interested on their values.
- Relating conceptual and linguistic knowledge.
- Generation of the questions and answers from the conceptual attributes for each type of user.

The next subsections describe the adaptation of these four general steps to generate the personalized questions and answers about university courses.

 ISBN: 978-1-61208-369-8

A. The Study of the User needs

Most web universities incorporate the description of the degrees and master courses the university offers. Those descriptions are usually accessed by both, teachers and students. In this particular scenario, teachers and students can be considered different types of users because the specific information they need about a course is not exactly the same. Although course descriptions in web universities usually include information interesting for both, teacher and students (i.e., the number of the credits of the course), they also can include formal details only interesting for teachers (such as the course code and the responsible unit) and other more practical information mostly accessed by students (i.e., the teaching language and the teacher email)**.**

B. Representing the Domain Concepts

The description of most courses in a web university usually includes similar information. There are two main concepts involved in this scenario: **Course** and **Teacher**. A partial description of those concepts is shown in Figure 1. As shown in the Figure 1, each conceptual attribute is associated with the specific group of users interested on its value: teacher, student or both (represented by **T/S**). The questions (and the corresponding answers) for a group of users are generated by selecting the specific attributes associated within the particular group**.**

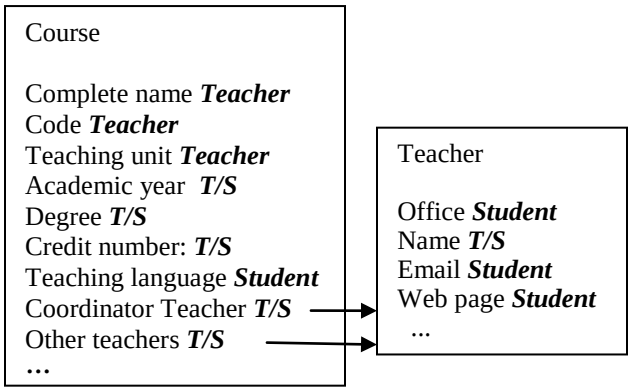

Figure 1. Concepts describing course information

C. Relating Conceptual and Linguistic Knowledge

The concepts and attributes describing the domain have to be associated with the corresponding lexical entries for each language. Additionally, in order to generate more natural language expressions, the conceptual attributes are also associated with the corresponding category in a syntactic-semantic taxonomy. For example, the attribute *Credit number* describing the course is associated with the class *quantity*, the attribute *Coordinating teacher* with the class *who_does* and the attribute *Office address* with the class *where*. Each category in the syntactic-semantic taxonomy is associated with the specific linguistic patterns needed for generating question and answers in different languages**.**

D. Generating Questions and Answers

Two general grammar rules are used to generate questions and answer from the conceptual representations:

Rule 1. questionattribute -> questionpattern conceptname

where *questionpattern* corresponds to the linguistic pattern associated with the sintactico-semantic class of the attribute (i.e attributes in the class *quantity* are associated with the pattern *how many*).

Rule 2. answerattribute -> attributename conceptname verbe attributevalue

Using these two general rules, the questions and answers for the conceptual attributes associated with each user's group are automatically generated. Then, the resulting sentences have to be manually supervised.

```
% generatext (Lex,Para,Lex).
generatext([],_,[]).
generatext([Leftcat1|Rest],Param,[[Leftcat1,Text1]|TextF]):-
        ruletext(Leftcat1,Param,Text1),
        generatext(Rest,Params,TextF).

%ruletext(catlgory,Lex).
ruletext(Leftcategory,Parameters,Text):-
         rule(Leftcategory,Parameters,Rightpart),
         obtainlexentries(Parameters,Rightpart,Text).

% obtainlex(Param,categoriesrightpart,Lex).
obtainlexentries(P,[],[]).
obtainlexentries(P,[Cata|R],[L1|L]):-
    lex(Cata,L1),obtainlexentries(P,R,L).

%rule(Leftcategory,Parameters,Rightpart).
%questionattribute - questionpattern  conceptname

rule(questionattribute,
parameters(language(L),user(U1),concept(C),attribute(A)),
[questionpattern(language(L),user(U),concept(C),attribute(A)),
conceptname(language(L),concept(C))]):-
      member(U1,U).

% answerattribute -
%  attributename conceptname verbe attributevalue

rule(answerattribute,
parameters(language(L),user(U1),concept(C),attribute(A)),
[attributename(language(L),user(U),concept(C),attribute(A)),
of(language(L)),
conceptname(language(L),concept(C)),
verbe(language(L)),
attributevalue(language(L),user(U),concept(C),attribute(A))]):-
           member(U1,U).
```

Figure 2. Grammars for generating questions and answers

 ISBN: 978-1-61208-369-8

The general grammar rules described can be implemented following different formalisms. Figure 2 shows their implementation in Prolog language. The unification mechanism of Prolog is appropriate for working with such conceptual grammars, because the general categories can be augmented with features representing specific information, such as the user group, the language as well as the concept and attribute identifiers.

As shown in Figure 2, Rule 1 and Rule 2 are implemented in Prolog using a predicate (*rule*) with three parameters (*Leftcategory*, *Parameters*, *Rightpart*), where *Leftcategory* is the category in the left part of the rule. The category *Leftcategory* is augmented by the features appearing in the second parameter (the predicate *parameters)*: language (represented by **L**), user (**U1**), concept (**C**) and attribute (**A**). The third parameter of the predicate *rule*, *Rightpart*, is a list containing the categories in the right part of the rule and these categories have been augmented with the corresponding features.

For example, Rule1 is represented as the following predicate:

```
rule(questionattribute,
 parameters(language(L),user(U1),concept(C),attribute(A)),
[questionpattern(language(L),user(U),concept(C),attribute(A)),
 conceptname(language(L),concept(C))])
```

The features associated with the left-part of the rule, *questionattribute*, are *language (L)*, *user(U1)*, *concept(C)* and *attribute(A)*. These four attributes have to be unified with the same features associated with the category *questionpattern*, and the category *conceptname*. That is, the variables associated with the corresponding features (**L** for *language*, **C** for *concept* and **A** for *attribute*) have to be the same, while the variable **U1**, associated with the feature *user*, has to be a member of the variable **U**, associated with the same feature in the category *questionpattern* and that represents a list of user groups interested on the conceptual attribute**.** Figure 3 shows examples of the generated questions and answers for the type of user *students* using Rule 1 and Rule 2.

Q1. How many credits does the course has?
The course has 6 credits
Q2 Who is the coordinator teacher of the course?
The coordinator teacher of the course is Dr. John Smith
Q3 Which is the teaching language of the course?
The teaching languages of the course are Catalan and Spanish
Q4 Where is the coordinator teacher office?
The coordinator teacher office is at U Building Office number 322.

Figure 3. Examples of question and answers for students

E. Considering other Scenarios

Other scenarios where a personalized presentation of web information can help the users have also been considered. For example, in the scenario where users access the web for finding information about a particular medical specialists two different types of users can be easily distinguished: professionals (doctors, nurses and medicine students) and no professionals.

The main concept in this domain is the concept doctor. A fragment of this concept description is shown in Figure 4. A more complete representation of the concepts involved in this scenario is described in [14].

Doctor

Name
Specialization
Visiting at equipment
Visiting timetable
….

Figure 4. A fragment of the concept Doctor

Although the information the two types of users need about the medical specialist can be the same (i.e., the complete name, the specialization and the place of work), the presentation of this information may differ. For example, professionals only need the name of the specialization, while information about the specific parts of the body related to the specialization can be useful for no professionals.

The approach described could also be adapted in scenarios where users can be classified in different groups considering their language level. For example, distinguishing between language natives and not natives can be useful in many different scenarios. The specific lexical words and syntactic patterns selected when generating questions and answers for each type of user would be different.

III. CONCLUSION AND FUTURE WORK

This paper is about how user models, domain ontologies and natural language generation techniques can be combined to generate questions and answers adapted to the different types of users. The approach described is based on the use of a domain ontology for representing the user's background related to the domain concepts. The ontology concepts are also related to a general syntactic-semantic taxonomy that facilitates the generation of the questions and answers in several languages. This modular organization of the knowledge in separated and declarative knowledge bases facilitates its adaptation to different domains, user models and languages.

Future work will also include working with more complex user models, such as those using dynamic information about the user's knowledge and interests that can be combined with user's background. This information obtained dynamically could also be represented as a layout of the domain ontology. The resulting user's model can be incorporated for different purposes, such as for educational

 ISBN: 978-1-61208-369-8

applications. Furthermore, the same organization can be used to generate other language resources, such as personalized summaries, dialogues and domain-restricted grammars.

ACKNOWLEDGMENT

This work has been supported by the Spanish project SKATER(TIN2012-38584-C06-01).

REFERENCES

[1] P. Brusilovsky and E. Millán, "User Models for Adaptive Hypermedia and Adaptive Educational Systems", *The Adaptive Web*, LNCS 4321, 2007, pp. 3 – 53.

[2] P. Brusilovsky, "Adaptive Hypermedia", *User Modeling and User-Adapted Interaction*, 2001, vol. 11, pp. 87-110.

[3] S. Williams, P. Piwek, and R. Power, "Generating monologue and dialogue to present personalised medical information to patients", in.Proc. of the Eleventh European Workshop on Natural Language Generation, 2007, pp. 167 – 170.

[4] M.A. Walker et al. , "Generation and evaluation of user tailored responses in multimodal dialogue", *Cognitive Science*, September–October 2004, vol. 28(5), pp. 811–840.

[5] M. Belk, P. Germanakos, P. Zaharias, and G. Samaras, "Adaptivity Considerations for Enhancing User-Centric Web Experience", in Proc. Of ACHI, January 2012, pp- 348-353.

[6] D.G. Sampson, M.D. Lytras, G. Wagner, and P. Diaz, "Ontologies and the Semantic Web for E-learning" (Guest editorial), *Journal of Educational Technology & Society*, 2004, vol. 7, pp. 26-28.

[7] O. Stock, and al., "Adaptive, intelligent presentation of information for the museum visitor in PEACH", User Model and User-Adapted Interaction, 2007, vol. 17, pp. 257–304.

[8] F. Cena, S. Likavec, and F. Osborne, "Property-Based Interest Propagation in Ontology-Based User Model", UMAP, LNCS 7379, 2012, pp. 38–50.

[9] M. Gatius, "Improving knowledge representation for speed-up the generation of grammars for a multilingual web assistant", in Proc. Of Semdial, December 2013, pp- 188-189.

[10] M. Milosavljevic and J. Oberlander, "Dynamic hypertext catalogues: helping users to help themselves", In Proc. Of the ninth ACM conference HYPERTEXT, ACM, New York, NY, USA, 1998, pp. 123-131.

[11] M. Gatius and M. González, "A Flexible Dialogue System for Enhancing Web Usability", in Proc. Of WWW, April 2009, pp- 1167-1168.

[12] D. G. Sampson, M. D. Lytras, G. Wagner, P. Diaz, M. E. Foster, and M. White, "Assessing the impact of adaptive generation in the comic multimodal dialogue system", In Proc. Of the IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, 2005.

[13] H. Bosch, D. Thom, G. Heinze, S. Wokusch, and T. Ertl, "Dynamic Ontology Supported User Interface for Personalized Decision Support", in Proc. Of CENTRIC, 2012, pp. 101-107.

[14] M. Gatius and P. Paliwal, "Generating Domain-Restricted Resources for Web Interaction in Several Languages: Hindi, English and Spanish", in Proc. Of INFOCOMP, November, 2013, pp. 94-98.

 ISBN: 978-1-61208-369-8