TourismKM: a new Web Semantic based approach for E-Tourism

Romain Picot-Clemente, Christophe Cruz, Christophe Nicolle

TourismKM: a new Web Semantic based approach for E-Tourism

Romain Picot-Clémente, Christophe Cruz, Pr. Christophe Nicolle

Abstract—**since 1970, technology innovations transformed radically the tourism processes and strategies, essentially with the Internet apparition. Nowadays, E-Tourism systems are well implanted into the Internet, with a lot of offers and online possibilities (reservations, virtual visit, etc.). Nevertheless, with the great evolution of the Internet, users' needs and attends changed too, they are now looking for more personalized and adapted contents. Here is the lack of the most E-Tourism systems, they are not, or a little, user centered. Based both on techniques of the Semantic Web and Adaptive Hypermedia Systems, this paper presents a solution for dynamic adaptation of the offer according to semantic criteria characterizing the Internet, in order to improve customer relationship management. A case study about tourism in the Côte-d'Or department in Burgundy is presented.**

Index Terms—**Adaptive web system, adaptive hypermedia system, domain ontology, e-tourism.**

I. INTRODUCTION

For the past 40 years, technologies have revolutionized the exchanges between customers and suppliers pushing enterprises to continuously adapt their sales strategies. The tourism sector is one of the most striking examples of this trend. Since 1970, from reservation systems on computers to today's online reservations, innovations in technology have radically transformed the tourism processes and strategies [1]:

- With the development of Internet, the number of tourist offers seems to grow exponentially.
- The link between customer and supplier has been shortened, removing intermediaries. Customers have the opportunities to directly access to the offers.
- The low cost strategy seems to have become the only criterion of quality.

However, these innovations have significantly reduced customer support in their choices and consultancy services and customization of the offer as they exist in travel agencies. Managing customer relationships is a strategic thinking based on a watchful and careful analysis of the client, establishing a relationship of trust, involvement of staff and partners, continuous monitoring for continuous innovation. A new technology is booming: profiling (or behavioral targeting). This new area of web marketing is to analyze the requests of users to better understand and target more efficiently expectations and demands. Through the research of a surfer you can not only know what type of service he likes, if he has children or even if he has responded positively to an electronic newsletter. The impact of these new fields of study is likely to exceed the current investigations and to help to adapt efficiently demands of users in the area of tourism. Based both on techniques of the Semantic Web and Adaptive Hypermedia Systems, this paper presents a solution for dynamic adaptation of the offer according to semantic criteria characterizing the Internet. The Semantic Web is an evolving development of the W3 Consortium in which the meaning (semantics) of information and services on the web is defined (http://en.wikipedia.org/wiki/Semantic_Web). The Adaptive Hypermedia Systems have traditionally attempted to deliver dynamically adapted and personalized presentations to users through the sequencing of reconfigurable pieces of information [2].

Actually, hypermedia applications (web sites, e-Learning platforms, etc.) are not, or very basically, user centered and adaptive. The lack of user centered adaptation generates a problem called "lost in hyperspace", the user loses himself navigating link to link in the system. Moreover, a phenomenon called "cognitive overload" could occur due to the huge amount of information proposed to the user. This project is inscribed in the context of the customer relationship management enhancement in e-tourism. It consists in the creation of an adaptive system which considers various features mainly deduced from the users' behavior in order to provide them with an efficient search engine.

This paper is made up of three parts. The first part presents the context. The second part presents a quick state of art on Semantic Web and Adaptive Hypermedia Systems and the third part presents our proposal.

II. TOURISM AND TECHNOLOGY, A CASE STUDY

The tourism in France has an important place, it is highly regarded. Indeed, it generates over 72 billion Euros of turnover, representing over 6% of the French GDP (against 3.5% for agriculture and agribusiness and 1% for automotive industry). There are about 200 000 tourist companies and nearly 2 million direct and indirect jobs generated by tourism.

Since 2006, the company "Côte-d'Or Tourisme", which aims to promote tourism, is conducting research to develop a website to unite all tourism offers of the department of Côte-d'Or. This department is located in France in the heart of

Romain Picot-Clémente is with the Laboratory LE2I (electronics, Image and computer Science) of Burgundy University and with the company "Côte-d'Or Tourisme", 21000 Dijon, France (e-mail: recherche-tic@cotedor-tourisme.com).

Christophe Cruz is with the Laboratory LE2I (electronics, Image and computer Science) of Burgundy University, 21000 Dijon, France (e-mail: christophe.cruz@u-bourgogne.fr).

Pr. Christophe Nicolle is with the Laboratory LE2I (electronics, Image and computer Science) of Burgundy University, 21000 France (e-mail: cnicolle@u-bourgogne.fr).

Fig. 1. Evolution of attendance - Technical trends

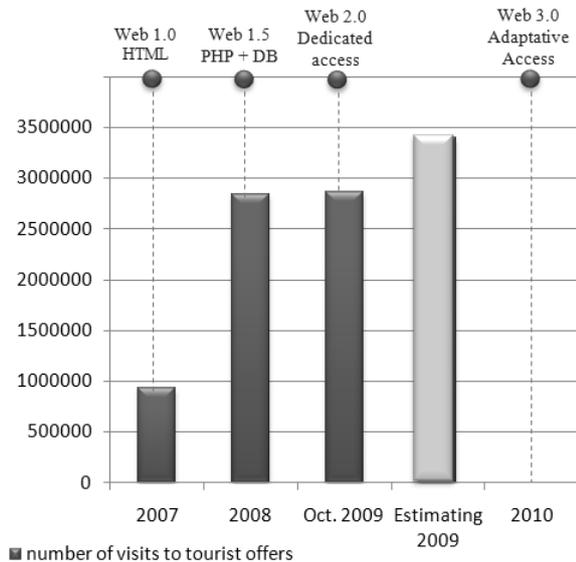Fig. 2. Monthly evolution of attendance

Burgundy (Bourgogne). It has an area of 8 763.21 km2 with a population of about 500 000 inhabitants. The vineyards of Burgundy (Marsannay, Gevrey Chambertin, Chambolle Musigny, Aloxe Corton, Meursault, ...) are located in this country area with the "Yonne" and "Saône et Loire". In 2008, the department of Côte d'Or received over 1.8 million tourists (data from hotels, campgrounds and bed & breakfasts in France). Fig. 1 shows the impact of technical developments of the website on the Internet crowd. In 2007, the site is developed in HTML with some Java applets for animation. The tourism offer is not structured. Suppliers may not directly fill the website. In 2008, almost 4,000 suppliers are federated in a database. Filling processes are developed. However, the use of these new solutions shows a real digital divide between the suppliers. Oversight mechanisms to help providers have been established. As a consequence, the site traffic has multiplied by 3. The access to a complete set of up to date data seems to be a factor of attractiveness.

In 2009, the company wants to offer web sites dedicated to specific audiences. This first step results from an analysis of profiling users. The website becomes a portal, providing access to a variety of websites: nearby tourism (http://www.tamtam21.com/), tourism dedicated to vineyards (http://www.divine-comedie.com/), nature tourism (http://www.bouger-nature-en-Bourgogne.com/). Professionals also have their website dedicated to the filling process and support process to help them on technical aspects.

The new features of Web 2.0 (Social Web) are available such as personalized access, blogs, forums, etc. The number of attendance on these websites has increased by about 15% during this last year. The addition of Social Web features to tourist offers shows real improvements for some Internet users (comments, reviews, tips ...).

Fig. 2 depicts more precisely the attendance of the website, month by month, since 2007. The first evolution of the web site (federation of suppliers in the database) was deployed in December 2007. From that point, the attendance rate increases until October 2008. In December 2008, the curve decreases until March 2009. During this period, the agency Cote-d'Or Tourism begins uploading new web sites with
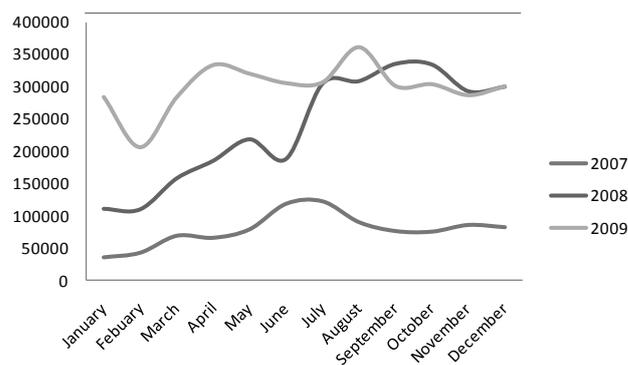
dedicated access. Unfortunately, the real measure of the impact of these new Web2.0 technologies is influenced by the economic context of that period.

Currently, the company wants to develop a customer relationship management adapted for its specific customers. The company continues its research work in the field of Semantic Web and Adaptive Systems. The idea is to develop a site that dynamically profiles the user's behavior by providing him with a combination of tourist offers according to his profile. For example, if the user wishes to find a restaurant that would welcome him with his children, the system will provide first restaurants with children's menus which are located near a public garden or park.

III. BACKGROUND

The Adaptive Web Systems are commonly called Adaptive Hypermedia System (AHS). In the following section the main properties of an AHS are introduced as well as the semantic Web technologies.

A. Adaptive Hypermedia Systems

The research in adaptive hypermedia systems has been very prolific these last 10 years. It results in new terms, models, methodology and systems which prove to be useful and powerful. Some systems have been developed during these 10 last years, giving principally solutions for e-Learning which is considered as the first application domain, KBS Hyperbook [3], APeLS [4] [5] ,Interbook [6], WINDS [7], MOT [8], RATH [9], etc. Moreover, attempts to define reference models have been made as shown in these projects, Tower [10], AHAM [11], Munich [12] and LOAS [13].

Each system and model brings its own architecture and methods. Most of them are based on a set of several layers also called models which separate clearly the different tasks. It appears that all of them have at least in common a user model, a domain model and an adaptation model. The domain model contains the whole domain knowledge (information, documents, etc.), the user model contains the user's knowledge, his interests, etc., and the adaptation model manages the user model, matches the domain model with the user model, and provides adaptation (e.g. Fig. 3). There are three main types of adaptation, according to research:

• Content adaptation: consists in hiding/showing or highlighting (or not) the information.

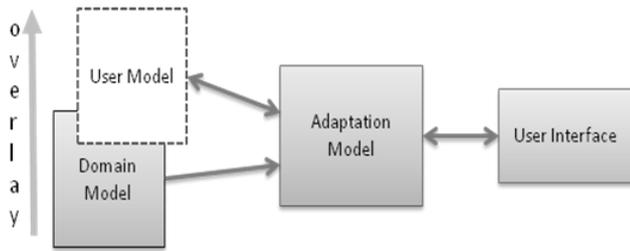• Navigation adaptation: consists in modifying the

Fig. 3. Overview of an adaptive hypermedia structure

Fig. 4. Web 3.0 applications are Web 2.0 applications based on linked data and web semantic technologies

hypermedia structure by suggesting links or driving the user to a defined destination.

• Presentation adaptation: consists in focusing (or not) on pieces of content or links, and in adapting to the hardware or page setting preferences.

The adaptation model is the central master piece of the Hypermedia Adaptive System because it uses the others models in order to process the adaptation.

In addition, most of the model definitions are driven by the Oriented Object Model (OOM). This design model is well adapted to define the architecture and the interactions between the different system components. Nevertheless, the use of OOM for the definition of the user and domain models to define the knowledge of the domain is not adapted in our opinion. Actually, the knowledge management and ontology domain which has been enriched during thirty years of research in artificial intelligence and description logic [14] is of value in AHS in order to model the knowledge. The semantic Web takes also advantages of this previous research in AI by defining languages and associated tools allowing the processing of knowledge as well as assertions on the knowledge.

B. The semantic Web

The semantic Web led to a standardization of languages like OWL for ontology creation. In the context of computer sciences, an ontology defines a set of representational primitives which allows the knowledge modeling of a specific domain. Ontologies are widely used to capture and organize knowledge on a particular domain. Ontologies formally define relations between terms. A typical ontology for the web has a taxonomy and a set of inference rules. The taxonomy defines object classes (concepts) and relations between them and inference rules provide efficient reasoning on the terms. Thus, associating concepts with documents, keywords, etc. it is possible to reason directly on them. Beyond the possibility to create complex models quite easily, their standardization allows their sharing between applications. However, the challenges in Ontology development and management are great. Ontologies providing the semantics for the Semantic Web have to be developed, managed, and endorsed by committed practice communities [15]. Whatever is the domain of application (meteorology, bank transactions, biology, tourism, etc.), concept definition is required. In addition, the Semantic Web is not just for the Web. Actually, it represents a set of technologies that will work equally well on internal corporate intranets. The Semantic Web will resolve several key problems facing current information technology architectures as Web services [16]. Moreover, it is

not easy to precisely define the nature of Web 3.0 technologies, but the outline of the emerging application has become clearer. The Web 3.0 applications are based (e.g. Fig. 4) on the Resource Description Framework (RDF) that provides a means to link data from multiple web sites or databases. With the SPARQL query language, a SQL-like standard for querying RDF data, applications can use native triple stores and extract RDF data from traditional databases. In RDF, if two data elements are recognized with the same URI, then they can be joined in a merged graph.

RDF Schema (RDFS) and the Web Ontology Language (OWL) provide the ability to infer relationships between data in different applications or in different parts of the same application. These Semantic Web languages allow for the assertion of relationships between data elements. The term "linked data" is often used to describe the evolving RDF development space and "Semantic Web" is increasingly being used to describe coupling linked data with RDFS and OWL [17].

C. The contribution of the semantic Web

The semantic Web technologies can be easily integrated as a data format which allows to define linked data. Actually, our adaptive hypermedia system is based on the semantic Web technologies, and especially on domain ontologies and their inference rules.

Thus, the domain model is defined as an ontology of domain where each concept represents an abstract or concrete domain entity (accommodation, hotel, camping, activity, etc.) and where relations between concepts are defined (hierarchical links, possession links, etc.).

The user model is constituted of several domain dependant and independent parts. The first part is represented using an overlay model on the domain model; it is an overlay enabling the definition of user interests on the concepts of the domain. The second part is a set of <attribute-value> pairs which is a centric user definition and independent of the tourism domain (age, sex, etc.).

The inference rules are defined into the adaptation model to reason over ontology models, the domain model and the overlay part of the user model. The rules are first order rules well known as Horn clauses and use the syntax SWRL, a semantic Web rule language that combines OWL and RuleML. The first order logic is simple and powerful and gives solutions to complex problems. The adaptation is processed with the use of an inference engine allowing the assertion of content, presentation and navigation from the domain rules on the domain ontologies.

From the realized inferences on the user and domain models, an adaptation model can provide different adaptations. Following the system creator's will, the adaptations are done differently using multiple datamining algorithms in the adaptation model. They consist in diverse algorithms like clustering algorithms to create clusters of concepts or users, based on same features for example. Content and navigation adaptation are provided according to the results of these algorithms. For the presentation adaptation, only domain independent user features are used, it has no impact on the content (only on the appearance).

IV. APPLICATION

This section presents the developed application. It is based on an ontology, built from an Oracle view on a Tourism database. This section is articulated in three parts. The first part presents an overview of the existing use of the database and an overview of the adjustments made for adaptation. The second part depicts the construction of the ontology. The last part describes briefly the process of building the user model, done by the adaptation model.

A. Overview of the improvement

The system is developed by "Côte-d'Or Tourisme" which aims to promote the tourism in its region called "Côte-d'Or" in France. This company collects a great quantity of data using a website where tourism providers fill in their offers. These data are joined and centralized in an important database. This database is full of relevant information and its structure developed and updated many times is full of knowledge about the domain. Then, using this database, the offers are shown on the website in the same way for each customer (e.g. Fig. 5).

To provide personalized and adapted content, navigation and presentation to customers, processes are inserted between the Database and the Output. Tourism providers don't change their filling process and data are yet stored in the database. Then, data from the database are used to create an ontology, in order to model domain knowledge. This creation process is explained in the next section, it uses views on the database
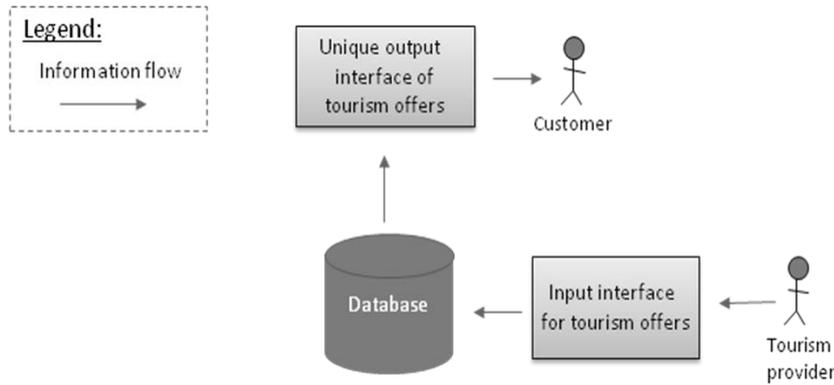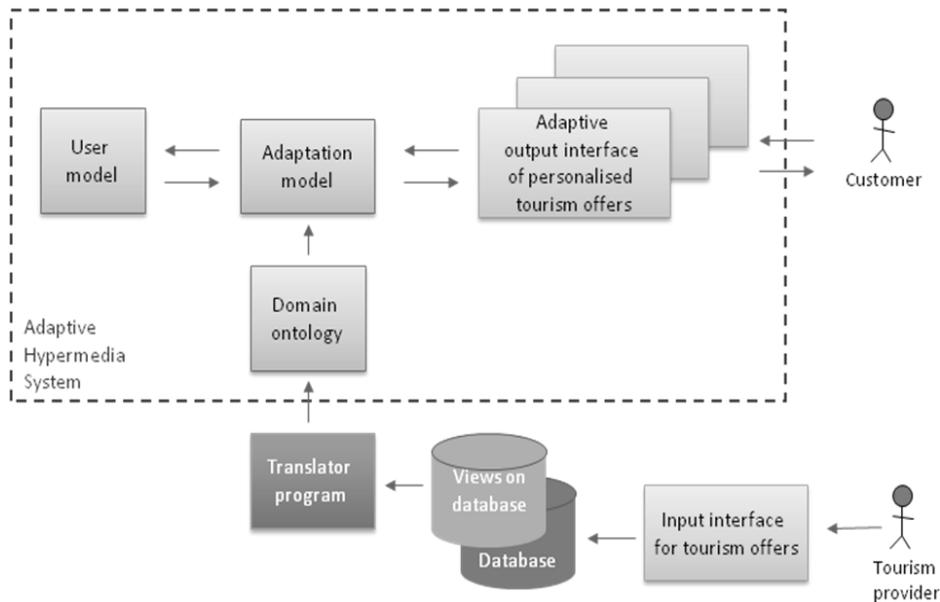

Fig. 5. Current system work

Fig. 6. Improved system work

and a translator program. As usual in adaptive systems, a user model to represent users (customers) and an adaptation model are defined. The adaptation model uses knowledge from the ontology to provide personalized and adapted output according to the user models initialized following the customers behavior with the interface. These processes are shown in the Fig. 6.

B. Ontology construction process

As explained in the previous section, the database is full of relevant information used to create the domain ontology. In addition, people of the company have also a good knowledge about the domain of tourism and the tourists. This section explains the ontology construction process.

In order to create and develop our ontology, we processed in three steps. First, the database was used as a starting point to create the ontology of domain for the domain model in the AHS. Second, the following step consists in enriching the ontology with axioms not present in the database but well known by the professional of the domain. Most of the time, it is materialized by semantic links between concepts which are most of the time lost by the relational model. In addition, the hierarchical structure is refined with the help of the professional in order to define a more appropriate hierarchy related to the links previously added. Finally some axioms that permit to define necessary conditions are added. For instance, the concept "gîte de France" is a restaurant and a hotel:

$$\text{GiteDeFrance} \equiv \text{Restaurant} \sqcap \text{Hotel}$$

Equation 1. Description logic definition of GiteDeFrance

Thirdly, the last step consists in populating the ontology with instances from the database. In order to formalize the knowledge, we have distinguished three types of tables in the database:
- "Data" tables corresponding to touristic objects like hotels, camping, etc.
- "Const" tables corresponding to constant tables like touristic labels, languages, etc.
- "Relation" tables corresponding to association tables between "Data" and "Const" tables.

For instance, the table "rel_speak_res" models the languages spoken by the restaurants. It connects the touristic object hotel with the constant table Language, and each row defines a language spoken by a hotel. A table "Relation" could also connect two "Data" tables.

In order to facilitate the population process, views are created due to the complexity of the base and the useless tables and attributes. These views are listed in a centralizing table where their type is specified ("Data", "Const" or "Relation"). This is done to simplify the base and select useful data for the ontology construction. It is a manual process done with help from experts. All the view names are formatted to be explicit (for instance "res" becomes "restaurant").

A translator program has been developed to translate this set of views in ontology. Each "Data" and "Const" view becomes a concept in the ontology and each "Relation" view
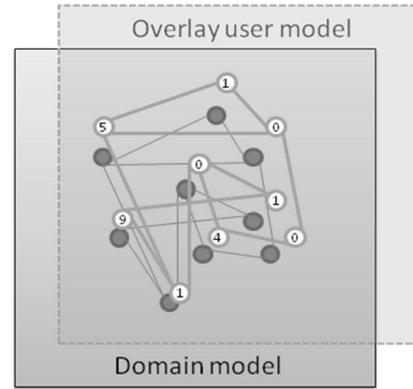

Fig. 7. Overlay part of the user model

becomes a relation between two concepts in the ontology. Then the program populates the ontology. This process consists in adding instances to concepts and relations to the ontology (for instance, adding the individual "Hotel Ibis" to the concept "Hotel"). This population is done by using the row of each view. Actually, we can consider that a data indexation is done on the database by the ontology. The phases of translation and population are automatic phases.

C. Interaction between user and adaptation models

User features are captured in a user model. These features are either domain dependent or independent. As said before, independent elements are <attributes, value> pairs and the dependent part is an overlay on the domain model. The overlay consists in a weighting of domain concepts in order to model interests on them (e.g. Fig. 7).

The user model is initialized and kept up to date by the adaptation model following user interactions with the system. The proposed links are all (almost) relied to one or several concepts of the domain ontology. Basically, a click on a link generates a weight increasing of the associated concepts in the user model. Domain independent elements are principally collected by direct questions to the user (age, sex, job, etc.).

The adaptation model comprises a set of rules about ontology concepts and relations. These rules are defined by domain experts. Following the user interests on concepts, rules are triggered. They can generate a user model update, which can also trigger other rules, etc. The adaptation model proposes information to the user from the domain model, corresponding to the user overlay model. It is a content adaptation if the content is directly shown to the user, or it is a navigation adaptation if links are proposed. Presentation adaptation is based principally on the non-overlay part of the user model. Indeed, the user context (hardware type, bandwidth, etc.), page setting preferences, etc. are in this part.

V. CONCLUSION

This paper presented a new Web Semantic based approach for E-Tourism. This paper has shown that technological developments in the field of tourism have transformed the relationship between customer and supplier. In addition, experiments conducted within the agency Côte-d'Or Tourism have shown that the association of suppliers in a database contributes to increase the attractiveness of a tourism website. Finally, a Web 3.0 solution has been proposed in this article.

This solution allows the customer profiling and dynamically adapts responses according to the website profile. This targeting allows a better combination of tourist offers. This solution is based on a semantic modeling: data from the database and user profiles. An adaptive hypermedia system has been deployed to support this architecture. It is based on three models (user model, domain model, adaptation model) using semantic web tools. Indeed, ontology models are used to model complex and sharable knowledge structures, and to realize rules based inference.

Beyond the use of domain, user and adaptation models, the trend is to use additional models like presentation, goals, context or other models but no generic model for adaptive hypermedia systems has already integrated them. It will be the object of a future article.

REFERENCES

[1] Buhalis D. and Law. R. 2008. Progress in tourism management: Twenty years on and 10 years after the internet: The state of eTourism research, Tourism Management, 29(4), pp.609–623, 2008.
[2] Levacher K., Hynes E., Lawless S., O'Connor A., Wade V. 2009. A Framework for Content Preparation to Support Open-Corpus Adaptive Hypermedia, International Workshop on Dynamic and Adaptive Hypertext: generic Frameworks, approaches and Techniques. June 29th – July 1st, Torino, Italy, pp.1-11, 2009
[3] Henze, N., 2000, "Adaptive hyperbooks: Adaptation for project-based learning resources.", PhD Dissertation, University of Hannover, 2000.
[4] Conlan, O., Hockemeyer, C., Wade V. and Albert, D. 2002a. "Metadata driven approaches to facilitate adaptivity in personalized eLearning systems", Journal of the Japanese Society for Information and Systems in Education, 1, pp. 38-45, 2002a.
[5] Conlan, O., Wade, V., Bruen C. and Gargan M. 2002b. "Multi-model, metadata driven approach to adaptive hypermedia services for personalized eLearning", in AH'02 Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, P.D. Bra, P. Brusilovsky and R. Conejo (Eds), London: Springer-Verlag, pp. 100-111, 2002b.
[6] Brusilovsky, P., Eklund J. and Schwarz, E. 1998. "Web-based education for all: A tool for developing adaptive courseware", Computer Networks, 30(1 7), pp. 291-300, 1998.
[7] Sprecht, M., Kravcik, M., Klemke, R., Pesin, L. and Huttenhain R., "Adaptive learning environment for teaching and learning in WINDS", in AH'02 Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, P.D. Bra, P. Brusilovsky and R. Conejo (Eds), London, UK: Springer-Verlag, pp. 572-575, 2002.
[8] Cristea A. and Calvi L. 2003b, "The three layers of adaptation granularity", in UM'03: 9th International Conference on User Modeling, J.G. Carbonell and J. Siekmann (Eds), Berlin/Heidelberg: Springer, p. 4-14, 2003.
[9] Hockemeyer, T. Held and D. Albert, 1998. "RATH-A relational adaptive tutoring hypertext WWW-Environment based on knowledge space theory", in CALISCE'98: Proceedings of the Fourth International Conference on Computer Aided Learning in Science and Engineering, C. Alvega°rd (Ed), Sydney: UniServe Science, pp. 417-423, 1998.
[10] De Bra, P., Houben, G-J. and Kornatzky Y. 1992. "An extensible data model for hyperdocuments", in Proceedings of the ACM Conference on Hypertext, New York, NY: ACM, pp. 222-231, 1992.
[11] De Bra, P., Houben, G-J. and Wu H. 1999. "AHAM: A dexter-based reference model for adaptive hypermedia", in Hypertext'99: Proceedings of the Tenth ACMConference on Hypertext and Hypermedia: Returning to our Diverse Roots: Returning to our Diverse Roots, J. Westbomke, U.K. Wiil, J.J. Leggett, K. Tochtermann and J.M. Haake (Eds), New York, NY: ACM, pp. 147-156, 1999.
[12] Koch N. and Wirsing M., 2002. "The Munich reference model for adaptive hypermedia applications", in AH'02 Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, P.D. Bra, P. Brusilovsky and R. Conejo (Eds), London: Springer-Verlag, pp. 213-222, 2002..
[13] Cristea A. and De Mooij A. 2003a, "LAOS: Layered WWW AHS authoring model and their corresponding algebraic operators", in WWW'03 Proceedings of World Wide Web International Conference (Alternate Paper Tracks), New York, NY: ACM, 2003a.
[14] Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., Patel-Schneider, P. F., 2007, The Description Logic Handbook, Theory, Implementation, and Applications, 2nd Edition Edited, Cambridge University Press, 2007.
[15] Shadbolt N., Tim Berners-Lee, T. and Hall, W. 2006, The Semantic Web Revisited, IEEE Intelligent Systems, 2006.
[16] Daconta, M.C., Obrst, L. J., Smith K. T. 2003. The Semantic Web: A guide to the Future of XML, Web Services, and Knowledge Management, Wiley, 2003.
[17] Hendler, J. 2009. "Web 3.0 Emerging", Computer, vol. 42, no. 1, pp. 111-113, Jan. 2009

Romain Picot-Clémente is a PhD student at the laboratory LE2I (electronics, Image and computer Science) since January 2009. He is also research engineering at the Tourism Agency "Côte-d'Or Tourisme". He obtained a Master in Image, Computer Science and Engineering in June 2008.

Christophe Cruz is associate professor at the IUT of Dijon (Technological Institute of the University of burgundy). He obtained a PhD in Computer Science at the University of Burgundy in 2004. He works on ontology modelisation.

Christophe Nicolle is full professor at the IUT of Dijon (Technological Institute of the University of Burgundy). He obtained a PhD in Computer Science at the University of Burgundy in 1996. His research area is the interoperability of information systems and more specially the problem of semantic heterogeneity.