Combining Adaptive Hypermedia Techniques and Ontology Reasoning to produce Dynamic Personalized News Services

Owen Conlan, Ian O'Keeffe and Shane Tallon

Knowledge and Data Engineering Group, Department of Computer Science,
Trinity College, Dublin 2, Ireland
{Owen.Conlan, Ian.OKeeffe, stallon}@cs.tcd.ie
http://kdeg.cs.tcd.ie

Abstract. Applying traditional Adaptive Hypermedia techniques to the personalization of news can pose a number of problems. The first main difficulty is the fact that news is inherently dynamic, thus producing an ever shifting pool from which content can be sourced. The second difficulty arises when trying to model a users interests and how they may be related to the available news items. This paper investigates the use of ontologies as a means of providing semantic bridges between available news items from RSS [1] news feeds and the interests of a user. Specifically, it investigates the combination of AH techniques with the ideas of loose and strict ontologies as the basis for personalization. This combination is highlighted through the design, development and evaluation of the Personalized News Service (PNS), which is based on the APeLS architecture [2].

1 Introduction

The personalization of information for each user is an area of research which provides an alternative to the "one-size fits all" [3] view of today's World Wide Web. Such personalization of information allows the users of a system to have tailored experiences where they are only presented with information which is of interest to them. The majority of personalization systems [4] [5] [2] work on a closed world model, where the information which is personalized is defined and marked-up with appropriate metadata before adaptation occurs. This step is often seen as necessary as is ensures there are semantic relationships between the content in the information space, the adaptation logic and the models (usually the user model) upon which the adaptation is based [6] [2]. Alternatively, many older Adaptive Hypermedia Systems [4] [5] explicitly refer to content in their adaptive logic in order to facilitate personalization.

Applying such approaches to the personalization of news content, however, would prove difficult. News is highly dynamic and users' interests in it can be fickle. The application of a closed world approach does not lend itself to the dynamism of news as the definition, classification and mark-up of individual news items would prove

cumbersome. More significantly, the development of appropriate adaptive logic would need to be carried out on a continuous basis. Therefore, there exists a potential semantic gap, brought about by the dynamism of news, between the expression of a users interests and the ever changing news domain. For example, if, as a consumer of news, I said in 1998 that I was interested in Bill Clinton, what does that mean? It could be interpreted that I have a general interest in the office of President of the USA. Alternatively, it could have been a passing interest in the impeachment proceedings initiated in that year. Then again it could be a specific interest in his family. This example is indicative of the problem faced in applying current AHS techniques to such a dynamic and complex domain as news. This paper investigates the application of ontologies to help bridge the semantic gap between news items and user interests.

Ontologies provide a structured, semantically rich way of modeling a domain. They are frequently defined using the eXtensible Markup Language (XML) and can be built using definition languages such as OWL [7], and DAML+OIL [8]. Ontologies count classes, inheritance, relationships between classes and instances as some of their major components. The ability to reason over relationships defined in an ontology and, therefore, relate instances to their abstracted types is the primary benefit to using ontologies. From the example above, we may be able to reason that Bill Clinton (an instance) is the President (a class) of the USA (an instance of class Country) through the relationships in an ontology. Furthermore, we may be able to determine that he is married to (a relationship) Hilary Clinton.

This paper proposes the application of ontologies as a means of achieving semantic precision between a user's interests and the news items available. It investigates the issues surrounding building different types of ontologies, modeling user interests and achieving effective personalization. Specifically, it presents the Personalized News Service, a service developed in Trinity College Dublin, as the embodiment of these principles. Section two gives an overview of existing personalized news/information systems; section three follows with a discussion on designing ontologies for personalized news; section four describes user modeling issues and the architecture of the PNS; and section five presents the evaluation results of the trial of PNS.

2   Overview of Existing Systems

This section will briefly review a number of existing personalized news/information systems. The goal of this overview is to provide a backdrop upon which design decisions in the Personalized News Service may be based.

In Merialdo et al. [9] deals with the adaptation of video. The approach taken is not very different to adaptive insertion/presentation of text found in many AHS, as the video is indexed or annotated with metadata providing a good source of uniqueness of the data stream. Of more interest to this work is the fact that the project deals with a very large scope of news, including international politics, national politics, international society, national society, economy, culture and sport. The user model provided by the system combines a level of interest in a category (or multiple categories), and uses a labeling mechanism to annotate a story with an associated importance level. By

combining the two, and using a probabilistic formulation, a simple yes or no answer can be given to determine a users interest in a particular story. Articles of interest are compiled together so the user can view the combined video articles. A basic approach of providing a level of interest in particular topics was used to model the users. Yet the simple user model did not detract from the quality of personalized news that was provided for the users. The domain models, were also relatively simple, being divided into several categories. Each article was then associated with a category and given a weighting from 1 to 100. This approach allowed for a good probabilistic chance that the user would not receive information which they would deem unsuitable to their needs.

In Jokela et al. [10], structured content, in the form of domain ontologies, are used. News content is provided with a semantic structure, which is then compared to a user profile to establish the relevancy of news articles. It is noted however, that interests can change and shift over time, and to provide a mechanism to adapt to such changes, user feedback is employed. In gathering user feedback, the ontologies can change, and become even more powerful than static ontologies. SmartPush aims to get rid of some of the shortcomings available in commercial systems, such as the lack of customizable ontologies and relative depth for experts. The SmartPush system applies weights to semantic relationships between objects, culminating in more power and expressiveness in the system. The domain model, was represented as an ontology, providing semantic meaning to the domain model. As the concepts carried weights, to distinguish between important and unimportant concepts, the domain model had more power over the information provided by the content.

SeAN [11] is an adaptive system which starts by classifying documents into a tree, made up of sections and subsections. The hierarchical nature of the system is in parallel with the same kind of structure associated with newspaper editorial systems. SeAN attempts to be able to personalize the detail level of a news document based on the user model. The user model is an ontology, which is rather different to most ontologies, as most ontologies represent content or conceptual domains. The user model is broken down into different dimensions, providing an altogether different view of a user than is usual. These dimensions are Interests, Expertise, Cognitive Characteristics and Lifestyle. Behavior tracking is used to a great extent in the system also. Such behavior as the time spent reading a news article, false positives and misses are taken into account, and these instances are learned from. User axes such as preferences, cognitive style, and domain knowledge form the basis of the user modeling approach. Stereotypes such as these can provide the system with enough knowledge to base its first few adaptations, with adaptability becoming more focused as more use is made of the system. Due to the domain being represented as a hierarchical structure of articles, from high level concepts to lower level niche topics, adaptive presentation is used where only information relating to the user model is provided, and other redundant information is not presented to the user.

PIN [12] is an adaptive system which uses neural networks to learn user profiles. User profile learning is done quickly and easily using this method. User profiles grow and adapt to new interests of the user. User feedback also helps with the dynamism of the user models. Adaptive Navigation forms the basis of the adaptive techniques used in PIN. Links are sorted based on relevance, in decreasing order. Users are modeled

by associating concepts from the domain, with interest weights provided for those concepts. Concepts have semantic meaning, providing more weight for the information which is of interest to the user. User Models are also updated "on the fly" as the system is being used, providing an adaptive user model, which is of better use than a static model.

3   Designing Ontologies for Personalized News

As part of this research two different forms of ontologies were proposed as a basis for personalization. These were termed the *strict* and *loose* ontologies. This section describes the rationale behind their design and the impact of gathering and using these ontologies. Personalization is facilitated through using these ontologies as semantic bridges between the general interests of a user and the specific instances of the domain from whish the news is taken.

3.1   Influences from State of the Art

Ontologies provide the semantic relationships between objects and instances in a particular domain. In order to ensure the use of ontologies is viable they should be as easy to maintain as possible. Lightweight ontologies make this feasible. Using a tool such as Protégé [14] allows the manual maintenance and updating of an ontology in a relatively easy manner. Classes, relationships and instances can be added and deleted as needs be.

Jokela et al. [10] make use of an ontology to describe the domain. The domain ontology provided a structured view of the domain, with concepts carrying weights to determine their importance within the domain. The weights allow even more inferences to be made about the suitability of a news article when combined with the user model. The weighting in the domain model, however, is a little restrictive as it assumes all users place the same importance on the relative relationships. In using an ontology to represent a user model such as SeAN [11], an overall, structured, and weighted view can be placed on the user's interests. The structure and the semantic meaning which an ontology can provide gives a view of a domain which would be similar to that of a human's view. For example, knowing which are the more important concepts in a domain, the important relationships between concepts and the weight the concepts carry within the domain are examples of a persons view of that domain. However, there is a potential for high overhead in the maintenance of an ontology per user as the user model could grow quite large. There is also the possibility that the relationships in personal ontologies will become so idiosyncratic that they cannot be reconciled by the adaptive mechanisms used.

The Metasaur system [14] performs automated ontology building. The system allows the insertion of objects into a data dictionary, which is analyzed and incorporated into a domain ontology. This kind of automatic creation provides the developer with a much less time consuming method of creating a domain ontology. This method also removes the restricted ontology problem, which is when all the concepts in the

domain are not available for markup. The disadvantage with this method, is that while it provides an ontology, there is little or no semantic meaning within the ontology. The domain is trawled, and links are inferred from objects, which lead to relationships being provided between those objects. Without much semantic meaning however, objects which have a higher importance than others can only be inferred by the number of related objects.

3.2 Defining the types of Ontologies

There are primarily two methods of constructing ontologies, being manual creation and automatic creation. Manually building an ontology requires the identification of concepts and properties within the domain of the ontology. There is also the need to populate this ontology with instances of the classes (and possibly properties), which are generally temporal instances specific to the ontologies content domain. The primary type of ontology defined by manual ontology construction is a strict ontology, or an ontology rich with semantics.

The other method of ontology construction is automatic or semi automatic ontology creation, similar to that applied in the Mercureo [15] and Metasaur [14] systems. Semi automatic ontology construction can be implemented by trawling hyperlinked web resources attempting to determine concepts and relationships. The relationships derived however, have little meaning associated with them, as the nature of relationships through trawling are difficult to obtain. This can result in what can be termed a *loose* ontology, or an ontology with little semantic meaning between the relationships.

The ontology can then be considered in two forms, a strict ontology and a loose ontology. The strict ontology provides meaning for things and their relationships. such as a driver is a sub class of a person, as is a team technical director. A loose ontology may be derived by stripping the meaning from a strict ontology, leading to a mass of things and relationships, but little or no explicit meaning for those things and relationships. Instances may also be included in the loose ontology. With so little semantic meaning in the loose ontology a position could be argued for not counting it as an ontology at all. It could be considered a linked taxonomy.

3.3 Building the Experimental Ontologies

Strict and loose ontologies present significantly different challenges in their creation, i.e. loose ontologies may be created in an automated or semi automated fashion, while currently the only way to create a strict ontology is to build it manually. Since news is a highly dynamic domain there are significant arguments for an automated approach to building ontologies, and thus basing personalization on loose ontologies. As part of this work two ontologies were developed – one strict and one loose. The first ontology provides a domain model which is semantically rich in meaning and relationships. For the experiment this strict ontology is created manually using Protégé and exported into OWL. The other, loose, ontology provides a domain model which has no semantic meaning and is created from the strict ontology by stripping out the class

and relationship information. The goal of creating two ontologies is to investigate the relative benefits of rich semantics versus weak semantics. For example, the strict ontology will note that Michael Schumacher is a Driver and that Ferrari is a Team. It will also note that there is a reciprocal relationship of drivesFor/isDriverOf between these instances. In the loose ontology Michael Schumacher is a Thing, as is Ferrari and they are related. As may be gathered from this example the domain modeled for the experiment is that of Formula One.

One important point to note about the strict ontology is that the classes and relationships are not likely to be subject to temporal change. In other words, the structure of the domain does not change much with time. Formula One has not changed significantly in structure in the last fifty years, i.e. there were still drivers who drove for teams, team cars had to have tyres, an engine and a chassis etc. What has changed, and is the main source of dynamism in this news domain, are the instances (and possibly number of instances). For example, the driver line ups of teams change year on year. In this sense, ontologies are an effective way of describing a domain, especially in terms of sport, as structurally is remains quite consistent.
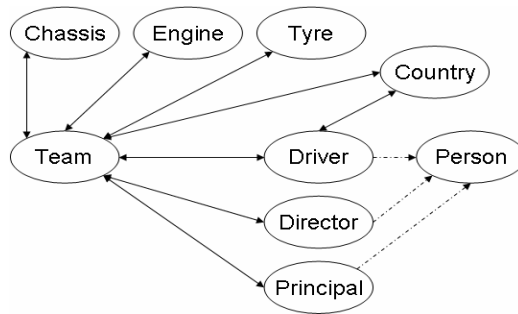

Fig.1. Strict Ontology Classes and Relationships

Figure 1 shows the classes provided by the strict ontology and which classes are related to each other. The broken line between Person to the Driver, Principal and Director imply that the three classes are a subclass of the class Person. The relationships (which are not explicitly named in Figure 1) between classes are symmetric. For example, if a driver drives for a team, then the team employ that driver. Transitive relationships may be inferred from the strict ontology, i.e. you may wish to determine which Drivers drive on which Tyres.

4 Design and Implementation of the Personalized News Service

This section details the architecture Personalized News Service (PNS) that supports the use of ontologies as the basis for news personalization. It begins with an overview of the how ontologies and the user's interests, and their level of interest, are related. It then presents the suite of services that comprises the architecture of the PNS.

4.1 Modeling the User's Interest and Level of Interest

Separating the interests of a user from their level of interest provides independence for the user model. This independence can be very useful when presented with users who would define these levels differently and also when presented with different domains. By having the interests separated from the level of interest there is the opportunity to provide users with control over their level of interest. From a domain perspective it enables there to be multiple definitions of medium interest. In the case of Formula One a user with a medium interest in Michael Schumacher may be interested in his Team and Team Mate also. In this example, the user model would simply state that the user has a medium interest in Michael Schumacher. The (separate) level of interest model would define interest in a Driver as being interested in his Team and Team Mate. The benefit of separating interest from the level of interest means that the user model does not become cluttered with unnecessary instance information. Continuing the example, the user model does not say that the user is also interested in Rubens Barrichello and Ferrari. This instance information is temporal and may become stale (e.g. Rubens is no longer Michael's team mate).

User modeling is achieved by presenting the user with a web-based instrument, derived from the domain ontology, that lists the instances of the domain by type and asks the user to grade their interest on the scale: none, low, low, medium and high.
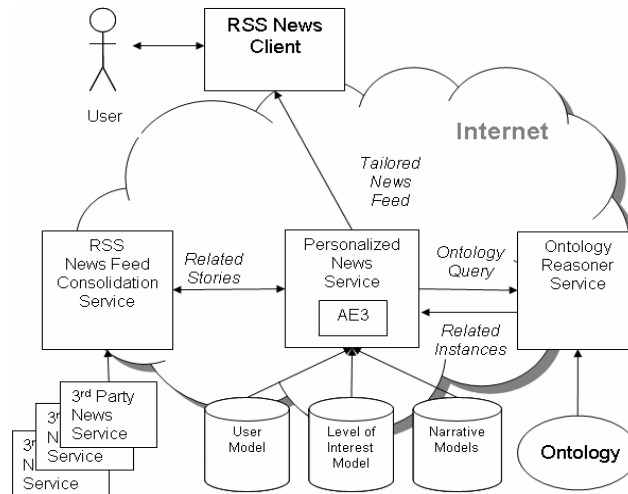
4.2 Architecture

Fig.2. Service Architecture

The Personalized News Service is actually the combination of three services – the RSS News Feed Consolidation Service, the Ontology Reasoner and the Personalized News Service itself. All threes services are implemented as Java web services and have associated Web Service Description Language (WSDL) descriptions. The RSS

News Feed Consolidation Service is the most basic of the three. It provides a means of registering third party news feeds, such as those offered by most commercial news sites, and of performing keyword queries on the articles available from those services.

The Ontology Reasoner Service is based on Jena [16] and exposes a WSDL interface over which RDQL (RDF Data Query Language) [17] queries may be passed using SOAP. Jena provides an open source ontology reasoner and can be used to reason over OWL DL [7] ontologies. Jena allows the checking of consistency of ontologies, classification of ontologies and answering a subset of RDQL queries. The Ontology Reasoner Service simply exposes this functionality.

The central service is that of the Personalized News Service. This service is built on top of the Adaptive Engine (AE3) used by the Adaptive Personalized eLearning Service (APeLS) [2]. The AE3 provides specific adaptive functionality that is suitable for this form of personalization. The model-driven approach used by the engine, and its ability to strategically reconcile these models at runtime through narrative, means that it can be used to generate queries (in RDQL). Furthermore AE3 has built in support for accessing and invoking external web services. Using this functionality the RDQL query, which has been adaptively composed, using the narrative to examine the user's interests and level of interests, is passed to the Ontology Reasoner Service. In this way AH techniques are used in conjunction with ontology reasoning. Through this step the reasoner can determine what other instances would be of interest to the user. For example, if the user model states that the user is highly interested[1] in Michael Schumacher the following steps are carried out: 1) The Personalized News Service determines what a high interest in a Driver means, i.e. what relationships are of interest; 2) It assembles an appropriate RDQL query to request the associated instances along those relationships; 3) Invokes the Ontology Reasoner Service; 4) Receives a result set with a list of related instances. In the case of Michael Schumacher this may include Ferrari (Team), Felipe Massa (Driver associated with Ferrari), Bridgestone (Tyre) etc.

Once the appropriate related instances have been determined the Personalized News Service then invokes the RSS News Feed Consolidation Service to search for appropriate articles. These articles are then assembled, again using narrative, into a tailored news feed which may be accessed from any appropriate RSS Reader.

5 Evaluation of PNS

The evaluation of the Personalized News Service was carried out by conducting a user trial. Users were presented with three news feeds: a personalized news feed based on their interests, level of interest and using the strict ontology; a personalized news feed based on their interests, level of interest and using the loose ontology; and a consolidated non-personalized news feed that showed all of the articles that were available. The trial was conducted with ten users over a period of four weeks with

---

[1] When reconciling level of interest using the loose ontology semantic relationships are meaningless (as they don't exist in the loose ontology). Conceptual distance (one relationship away, two relationships away, etc.) was used instead for the loose ontology.

varying degrees of interest ion Formula One. The news feed, upon which the personalized services operated, was taken from www.itv-f1.com. The objective of the trial was to see if the strict ontology and loose ontology based services produce dramatically different personalized user experiences.

Each user was asked to complete the web-based instrument to solicit their interests and levels of interest in Formula One. Once this step was completed they were asked to access three feed URLs, corresponding to the three services above, daily using their preferred RSS news reader. The feeds were identified as Feed One, Feed Two and Original Feed. Feed One corresponded to the personalized service using the strict ontology and Fees Two to that using the loose ontology, however, the users were not told this.

Following the trial each user was interviewed, with a specific set questions relating to usability and quality of experience being asked. All of the users found the services easy to use as, for them, is simply meant completing a web-based instrument and then pointing their RSS reader at the generated feeds as most users had experience of using RSS news readers before. In general the two personalized news feeds presented articles that the users believed appropriate. Occasionally, through looking at the static news feed, the users identified articles that the loose-ontology service missed. When asked if they perceived differences between Feed One and Feed Two, only one person felt there were Major differences, with the remainder of the users perceiving only minor or no differences. The user who experienced major differences was also the user with the lowest overall interest in Formula One.

6   Conclusion

The evaluation showed that end users perceived little difference in experience and satisfaction between the two personalized services. This is an interesting result as the level of semantic matching carried out when using the strict ontology is much higher, compared to that of the loose ontology. The strict ontology, however, is much more time consuming to generate and maintain as automated approaches are not viable. This is an encouraging result as systems such as Mercureo and Metasaur produce ontologies that are not dissimilar to the loose ontology used in the experiment, pointing to the viability of this approach.

The overall approach of using ontologies to bridge the semantic gap between user model and the content available proved successful. The ability to keep the user model minimal and only containing the items the user was centrally interested coupled with their level of interest meant the user model was not full of tangentially interesting concepts. The danger of such concepts in a dynamic domain such as news is that they may become *stale* quickly.

The approach of using the narrative to compose the query leaves open the opportunity to extend the service to cater for other axes of adaptivity. For example, the number, quality and source of news articles could all be personalized. Another feature of narrative as it is supported by the AE3 is the possibility to invoke services as part of a service oriented approach. This capability means that the adaptive systems can be broken up into logical (and reusable) services.

References

1. Really Simple Syndication (RSS) Specification v2.0, Available online at http://blogs.law.harvard.edu/tech/rss
2. Owen Conlan, Vincent Wade, Catherine Bruen,, Mark Gargan: 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, London, UK, Springer-Verlag (2002) 100–111
3. Conklin J. (1987). Hypertext: An Introduction and Survey. IEEE Computer, 20(9), 17-41
4. Paul De Bra, Ad Aerts, Bart Berden, Barend de Lange, Brendan Rousseau, Tomi Santic, David Smits, Natalia Stash, AHA! The Adaptive Hypermedia Architecture, in Proceedings of the ACM Hypertext Conference, Nottingham, UK, pp 81-84
5. Peter Brusilovsky, Elmar Schwarz, Gerhard Weber, ELM-ART: An Intelligent Tutoring System on World Wide Web, in Third International Conference on Intelligent Tutoring Systems, ITS-96, Montreal, 1996
6. Marcus Specht, Milos Cravcik, Leonid Pesin, Roland Klemke, Authoring Adaptive Educational Hypermedia in WINDS, in Proceedings of ABIS2001, Dortmund, Germany, October, 2001, pp 1-8
7. Web Ontology Language (OWL) Reference, Available online at http://www.w3.org/TR/2004/REC-owl-ref-20040210/
8. DAML+OIL (March 2001) Reference Description, Available online at http://www.w3.org/TR/daml+oil-reference
9. Bernard Merialdo, Kyung Tak Lee, Dario Luparello, Jeremie Roudaire, Automatic Construction of Personalised TV News Programs, in ACM Multimedia '99, October 99, Orlando, Florida, USA, pp 323-331
10. Sami Jokela, Marko Turnpeinen, Teppo Kurki, Eerika Savia, Reijo Sulonen, The Role of Structured Content in a Personalised News Service, in Proceedings of the 34th Hawaii International Conference on System Sciences 2001, pp 1-10
11. Liliana Ardissono, Luca Console, Iliara Torre, An Adaptive System for the Personalised Access to News, in AI Communications 14: pp 129-147, 2001
12. Ah-Hwee Tan, Christine Teo, Learning User Profiles for Personalised Information Dissemination, in Proceedings of International Joint Conference on Neural Network 1998, pp 183-188
13. The Protégé Ontology Editor and Knowledge Acquisition System, Available online at http://protege.stanford.edu
14. J Arjona, R Corchuelo, A Ruiz, M Toro, Automatic Extraction of Semantically-Meaningful Information from the Web, in Adaptive Hypermedia 2002, LNCS 2347, pp 24-35
15. Trent Apted, Judy Kay, Andrew Lum, Supporting Metadata Creation with an Ontology Built from an Extensible Dictionary, in Adaptive Hypermedia 2004, LNCS 3137, pp 4-13
16. Jena Semantic Web Framework, Available online at http://jena.sourceforge.net/
17. RDQL, A Query Language for RDF, Available online at http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/
18. Yannis Kalfoglou, John Domingue, Enrico Motta, Maria Vargas-Vera, Simon Buckingham Shum, myPlanet: An Ontology-Driven Web-Based personalised news service, in Proceedings IJCAI 2001 Workshop on Ontologies and Information Sharing, Seattle USA, pp 140-148