Categories of Web User Behaviour Models and Information Retrieval –A Survey

F.Mary Harin Fernandez M.E
Research Scholar
Sathyabama University
Old Mamallapuram Road
Chennai – 103
Tamil Nadu, India

R. Ponnusamy M.Tech., Ph.D.
Principal
Madha Engineering College
Madha Nagar, Somangalam Road,
Kundrathur, Chennai - 69
Tamil Nadu, India

ABSTRACT
The current challenges in the world are search and retrieve accurate information from the massive web. The general term used for searching and retrieving data from the web is 'query' and keyword-matching. The existing structure uses Personalized user information system, recommender system and wordnet ontology. The Personalized user information system used to increase the speed and required response. To extract user likings, the personalized user information system explore the acquirement of user reviews by supervising their browsing behavior. In Recommender system the people rate web pages as interesting and not interesting and it responses according to the relevant feedback. The wordnet ontology uses to retrieve information by means of Synonymy, Antonymy, Hyponymy /Hypernymy , Meronymy / Holonymy, Troponymy and Entailment

General Terms
Web information retrieval, wordnet ontology

Keywords
Personalization, ontology, recommender system, user profiling.UML

1. INTRODUCTION
Information retrieval from the huge web is well challenged in the current world. Many applications used to store and retrieve information. But the user unsatisfied when they get irrelevant data from their search. The development of relevance feedback and word sense disambiguation techniques aim to assist the user in the formulation of a targeted query, and have shown an improvement of the information retrieval (IR) performance [11]. Effectively, relevance feedback techniques require that a user explicitly provides feedback information, such as marking a subset of retrieved documents as relevant documents. On the other hand, the word sense disambiguation techniques use generally an ontology-based clarification interface and required that the user specify explicitly the information need.[11]

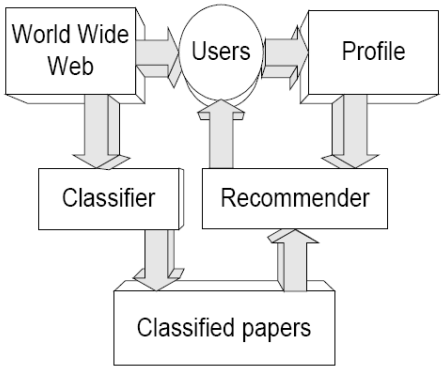
This paper is organized as follows. Section 2 discusses related work in information retrieval. Section 3 discusses the conclusion and future work. section 4 presents references. Finally, Section 5 discusses the comparative analysis.

2. RELATED WORK
2.1 Ontologies in recommender systems
Most recommender systems use a simple binary class approach, using a user profile of what is interesting or not interesting to the user. The Quickstep recommender system uses a multi-class approach, allowing a profile in terms of domain concepts (research paper topics) to be built.

Quickstep Figure [1] is a hybrid recommendation system, combining both content-based and collaborative filtering techniques. This allows profiles that consist of a human understandable list of topics. The classifier assigns each paper a class based on which class vector it is most similar to recommendations are selected from papers classified as belonging to a topic of interest. The profile itself is computed from the correlation between browsed papers and paper topics. This correlation leads to a topic interest history, and a simple time-decay function allows current topics to be computed.

Figure 1 The Quickstep system

The multi-class classification is less accurate than



other binary classification systems, but allows class specific feedback and the use of domain knowledge (via an is-a hierarchy) to enhance the profiling process. The ontology users tended to have more "rounder" profiles, including more general topics of interest that were not directly suggested. This increased the accuracy of the profiles, and hence usefulness of the recommendations.

S.no	Recommender systems	uses
1.	Collaborative recommender systems	Utilize user ratings to recommend items liked by similar people.
2.	Content-based recommender systems	Recommend items with similar content to things the user has liked before. Which recommends funding information from a database.
3.	Personal web-based agents	Track the users browsing and formulate user profiles. Profiles are constructed from positive and negative examples of interest, obtained from explicit feedback or heuristics analysing browsing behavior.

4.	News filtering agents	Recommend news stories based on content similarity to previously rated examples.

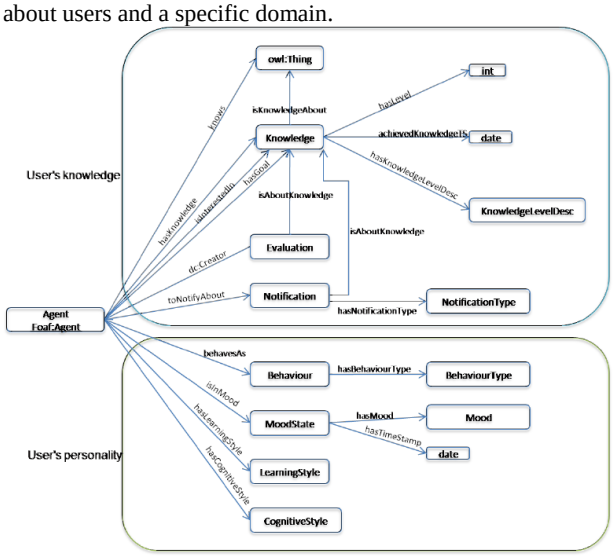
Table 1 Recommender systems

Systems such as CiteSeer [2] use content-based similarity matching to help search for interesting research papers within a digital library. Ontologies are also used to improve content-based search, as seen in OntoSeek [3].Mladenic [4] provides a good survey of text-learning and agent systems, including content-based and collaborative approaches.

2.2 Generic ontology based User Model

A user model is a knowledge source that contains a set of beliefs about an individual on various aspects, and these beliefs can be decoupled from the rest of the system [5]. [6] A user modeling system shows adaptive behaviour concerning its interaction with the user [7]. [4] Introduces Generic User Models [5] as systems which have, among other aspects, two major goals: 1) *generality*: which would allow a model of the user to be usable in a variety of application content domains; 2) *expressiveness:* in that the model is able to express a wide variety of assumptions about the user.

The main goal of user modeling is to understand the user's caracteristics, thus allowing a system to automatically adapt itself specifically to each user. A generic user model is the basic structure of a framework that handles information about users and a specific domain.



.

Figure 2: genOUM concepts and properties

2.2.1 GenOUM

GenOUM [Figure 2] fits in a larger set of ontologies part of the user modeling system :

- ***A user model ontology (GenOUM)*** : To provide a domain independent description and understanding of the user .
- ***A domain ontology*** : To describe the concepts relative to the domain concerning the project.
- ***A system adaptation ontology*** : To adapt the look and functions of the system, i.e. adapt the user's interface.
- ***A content adaptation ontology*** : To adapt the content of the system according to the user preferences, behaviour or knowledge, e.g. give customized answers to users actions or queries.

2.2.3 User's Knowledge

The knowledge and skill of the user can be expressed about anything, but more specifically about the concepts described by the domain ontology. To be really useful, the knowledge is qualified with a knowledge level to specify if the user is for instance a beginner or an expert. A time stamp can be added to follow the evolution of the user's knowledge and skills.[5]

*Example:*Figure 3 states that Franck knows Vietnam, that his knowledge level about Viet- nam on the 16.11.2013 is 20, an advanced knowledge level. The interpretation of the knowledge level '20' depends on the domain and is not defined by GenOUM.

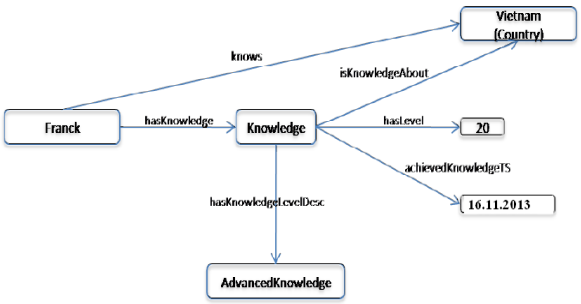


Figure 3: Example of User's knowledge level

2.3 Personalized user behaviour

Web search engines provide a huge number of answers in response to a user query with the consequence that a user cannot always find the results relevant to his/her information needs. In order to overcome this unsatisfying situation, a possible solution is to analyze the behaviour of a user during a search session, namely (1) to study the interactions of a user with search engines, and/or (2) to study the actions that he/she performs visiting Web pages. With the first case, the terms of a user's queries are stored into *query log files*; whereas in the second case, information such as the identifier of a user,*URLs* clicked for each query or actions such as save, print, copy, etc. related to a part of a Web page are stored in Web log files.[9]

All these files are analyzed in order to understand what the user's interests are, and thus to define ad-hoc the profile. This profile is used to improve the quality of a user's search. Ontologies allow giving a semantic organization to the information recorded in query and Web log files. Two typical Information Retrieval problems: query reformulation and results re-ranking to personalize a user's search with the support of personalized ontologies. The use of ontological user profiles obtained by folksonomies can improve the personalized search of the Web[9].

2.3.1 Personalized Information Retrieval

Personalized Information Retrieval (PIR) can be defined as the appropriate information retrieval from a large volume of data or information within a user's context, i.e. preference or profile, and also to present the retrieved information appropriately based on the user's context in generic computing environment where any information could be used by anyone. [10]

Ontology has been a basis for the construction of a user model in several personalized systems ranging from information delivery systems to Intelligent Tutoring Systems.

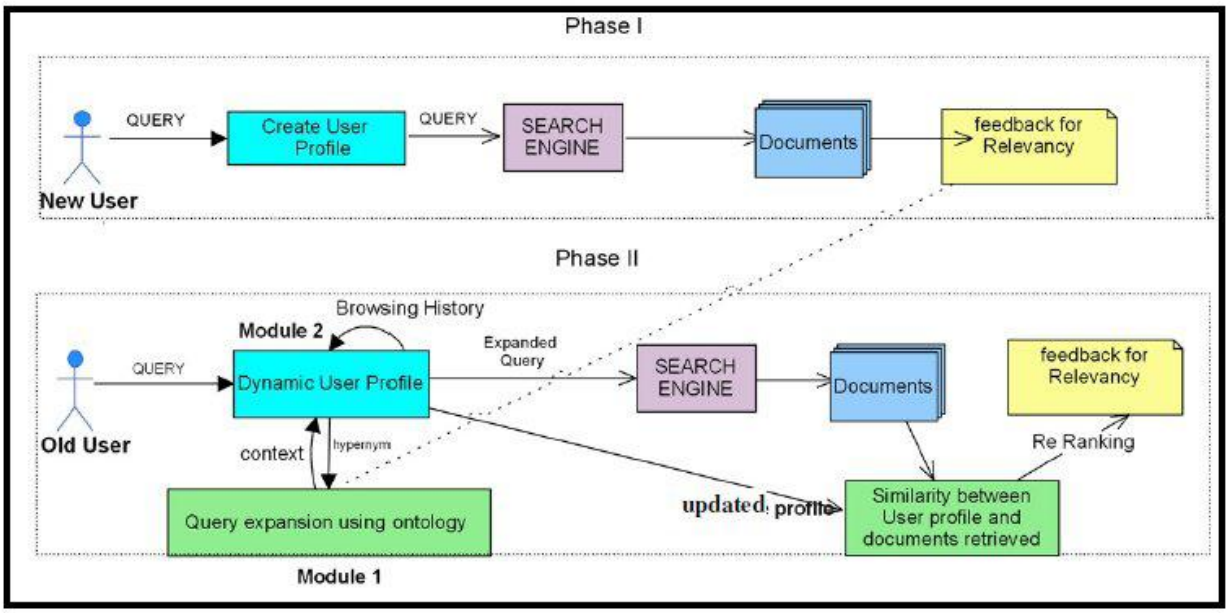


Figure 4. Personalised Information Retrieval using Dynamic User Profile andOntology for Query Expansion

The retrieval models are based on keyword or term matching, i.e., matching terms in the user query with those in the documents.[10]

2.3.2 Open Directory Project(ODP)

ODP is the most widely distributed data base of Web content classified by humans. It is a Web directory where its purpose is to list and categorize web sites. the ODP's concept hierarchy is used profiling component as a fundamental source of a semantic knowledge to represent semantically the user interest. Various methods can be utilized to represent the concept vector of the ODP ontology

2.4 Ontological Behaviour Modelling-UML

To improve the effectiveness of behavior modeling languages through ontological approaches, enabling users and implementers to understand them more uniformly. These approaches specify real-world implications of language sentences more rigorously than informal text, but not directly in mathematics. [12]

A proof-of-concept for ontological approaches is provided by a common semantic basis for UML behaviors. It starts with the existing UML notion of behaviors as classes, where each instance is one occurrence of a behavior in time.

The two relationships of composition (whole-part and part-part) are applied to behavior through a common sense model of time: nested durations for subbehaviors, and time ordering for steps in behaviors, respectively, and summary in Figure 5 and Figure 6.

Events are captured as classes, where each instance is one occurrence of an event in time. This enables them to type step properties and be ordered in time with other steps, Participants in behaviors and associations are treated as parts of a whole, and captured as properties in a composite structure, as summarized in Figure 6.

This enables behaviors to act as links between participants, as associations do, and be used to connect parts of other behaviors. Specialized behavior associations between participants capture the transfer of objects in messaging and object flow, which are distinguished by the kind of source and target of the transfer (objects or behavior occurrences, respectively). Transfers connect elements of behaviors, including steps and participants, through composite structure. Specialized properties identify links connecting objects and occurrences, which are combined with behavior steps to enable transfers to be ordered in time, as in messaging protocols, and inputs and outputs to long-lived behaviors.

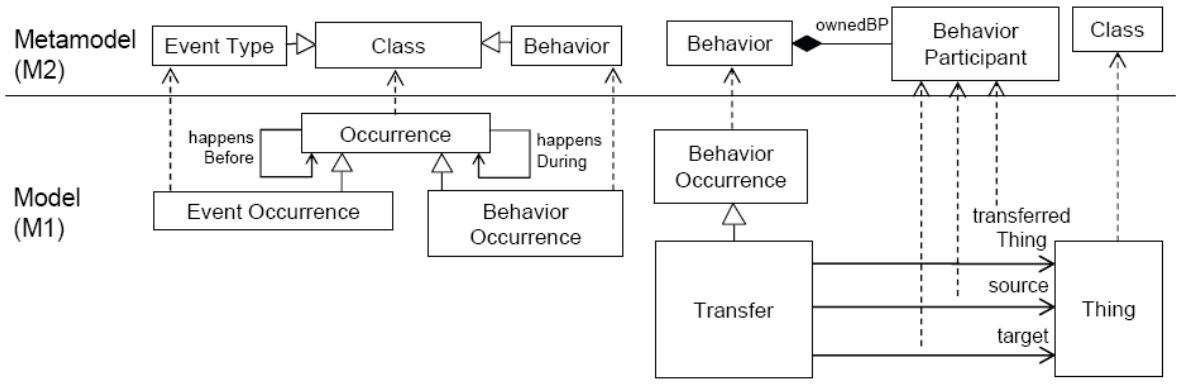


Figure 5: Model Library

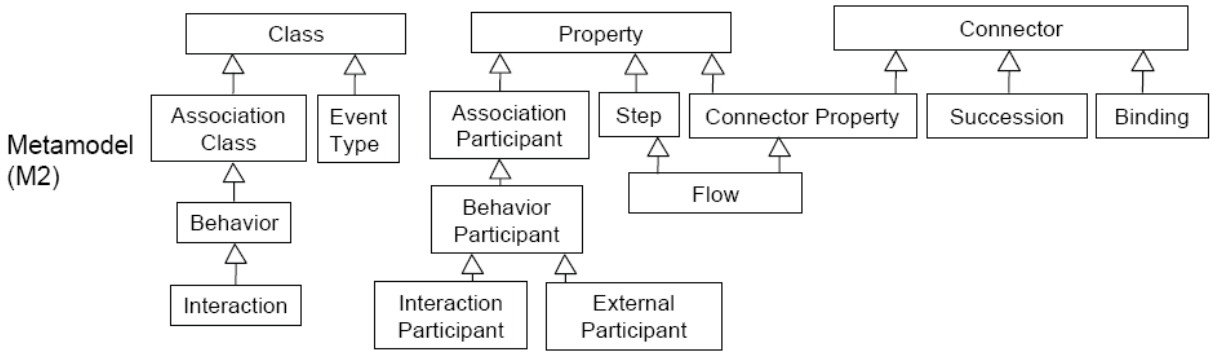


Figure 6: Metaclass Taxonomy

Finally, transfers can be equated (bound) to each other to enable behaviors to coordinate transfers when they use other behaviors

The ontological approach to language specification appears in the above models as simple notions, such as class as category, and properties specifying links between instances, also falling into categories, with both specialized in multiple, thin layers to more sophisticated constructions, such as flows between various kinds of behavior participants. At each stage, the implications of user models for the real world are captured (semantics), sometimes with reusable model libraries. This enables more uniform understanding and implementation of the three UML behavior models, and more expressiveness from their integration.

3. CONCLUSION AND FUTURE WORK

The proposed analysis on user behavior model is based on information retrieval from the web. The current technology uses Personalized user information system, Recommender system and Wordnet ontology. In the context of user behavior model, data acquisition can be user profile or recommender system, but the system should deduce data automatically. The present literature lacks many perfect results as to the extent knowledge-based approaches support real-world systems, where noisy data and conflicting user opinions exist. In future work to retrieve information the generic ontology based model can be used to deduce information automatically.

4. REFERENCES

[1] Stuart E. Middleton, David C. De Roure and Nigel R Shadbolt, "Capturing knowledge of user preferences: ontologies in recommender systems" Proceedings of the 1st International Conference on Knowledge Capture, page 100--107. New York, NY, USA, ACM, (2001).

[2] Bollacker, K.D. Lawrence, S. Giles, C.L. CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications, Proceedings of the Second International Conference on Autonomous Agents, Minneapolis MN, USA, 1998

[3] Guarino, N. Masolo, C. Vetere, G. OntoSeek: Content-Based Access to the Web, IEEE Intelligent Systems , Vol. 14, No. 3, May/June 1999.

[4] Mladenic, D. Text-Learning and Related Intelligent Agents: A Survey, IEEE Intelligent Systems , Vol. 14, No. 4, July/August 1999.

[5] Fabian Cretton and Anne Le Calve' Generic ontology based User Model:GenOUM, SMV technical report series, No 203, Switzerland,June 2008.

[6] Alfred Kobsa and Wolfgang Wahlster, editors. User models in dialog systems.Springer-Verlag New York, Inc., New York, NY, USA, 1989.

[7] Paul de Vrieze, Patrick van Bommel, and Theo van der Weide . A generic engine for user model based adaptation. In Proceedings of the User Interfaces for All workshop, Vienna, 2004.

[8] A. Stewart, C. Nieder, and B. Mehta. State of the art in user modelling for personalization in content, service and interaction - nsf/delos report on personalization. Technical report, Fraunhofer IPSI, Darmsadt, Germany,, 2004.

[9] Silvia Calegari and Gabriella Pasi, " Ontology-Based Information Behaviour to ImproveWeb Search" future internet, ISSN 1999-5903, *PP* 533-558,*October 2010.*

[10] Aditi Sharma," Personalised Information Access Based on Ontology and Collaborative Filtering", International Journal of Engineering Research & Technology (IJERT),Vol. 1 Issue 6, ISSN: 2278-0181, August – 2012.

[11] Mariam Daoud , Lynda Tamine, Mohand Boughanem and Bilal Chebaro," Learning Implicit User Interests Using Ontology and Search History for Personalization" WISE 2007 workshops,PP 325-336,2007.

[12] C. Bock, J. Odell: "Ontological Behavior Modeling," Journal of Object Technology, vol. 10, 2011, pages 3:1-36,2011.

[13] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, M. McDermott, D. McGuinness, B. Parsia, T. Payne, M. Sabou, M. Solanki, N. Srinivasan, K. Sycara: "Bringing Semantics to Web Services: The OWL-S Approach," Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition, pp. 26-42, July 2004. doi:10.1007/b105145.

[14] R. Dumitru, U. Kellera, H. Lausena, J. de Bruijna, R. Laraa, M. Stollberga, A. Polleresa, C. Feiera, C. Busslerb, D. Fensela: "Web Service Modeling Ontology," Applied Ontology, vol. 1, pp. 77-106, 2005.

[15] A. Haller, M. Marmolowski, E. Oren, W. Gaaloul: "A Process Ontology for Business Intelligence," Digital Enterprise Research Institute, Technical Report 2008-04-1, April 2008.

5. THE COMPARATIVE ANALYSIS ON USER BEHAVIOUR MODELS

CATEGORIES	CONCEPT	LIMITATIONS
Recommender systems	➢ Recommender systems use a simple binary class approach, using a user profile of what is interesting or not interesting to the user. ➢ The Quickstep recommender system uses a multi-class approach, allowing a profile in terms of domain concepts (research paper topics) to be built.	➢ Need visualizing the knowledge contained it. ➢ Visualizing the profile knowledge will allo users to build a better conceptual model of system, helping to engender a feeling of co and eventually trust in the system.
Generic ontology based User Model	➢ A user modeling system shows adaptive behaviour concerning its interaction with the user. ➢ A generic user model is the basic structure of a framework that handles information about users and a specific domain.	➢ Need 'testing' status for different conce properties. ➢ There is necessity to prove what is right o what should be totally changed or just and what can be accepted with a 'stable' sta ➢ In the context of user modeling, data ac can be explicit or implicit, but the system deduce information automatically.
Personalized user behavior	➢ Personalized Information Retrieval (PIR) can be defined as the appropriate information retrieval from a large volume of data or information within a user's context. ➢ The retrieval models are based on keyword or term matching, i.e., matching terms in the user query with those in the documents. ➢ The ODP's concept hierarchy is used profiling component as a fundamental source of a semantic knowledge to represent semantically the user interest.	➢ Need to process tagging system for an Web resources with an unstructured list of ➢ Need contextual retrieval and semantic a for information retrieval.
Ontological Behavior Modeling-UML	➢ A proof-of-concept for ontological approaches is provided by a common semantic basis for UML behaviors. ➢ The two relationships of composition (whole-part and part-part) are applied to behavior through a common sense model of time: nested durations for subbehaviors, and time ordering for steps in behaviors, respectively.	➢ Need specialization into the three UML b languages, including such topics as asyn and polymorphic invocations, in exceptions, and more expressive coor constructs for sub-occurrences. ➢ The standards can increase the reliability of communication between users, tools, and implementers, enabling tools to work mor seamlessly with each other and with the pe using them.

Table 2 The comparative analysis of user behaviour models