Ontology and Rule based Retrieval of Sound Objects in Augmented Audio Reality System for Museum Visitors

Marek Hatala, Leila Kalantari, Ron Wakkary and Kenneth Newby
School of Interactive Arts and Technology
Simon Fraser University
Surrey, BC, Canada, V3T 2W1
Phone: +1-604-268-7431

{mhatala, lkalanta, rwakkary, newby}@sfu.ca

ABSTRACT

ec(h)o is an "augmented reality interface" utilizing spatialized soundscapes and a semantic web approach to information. The initial prototype is designed for a natural history and science museum. The platform is designed to create a museum experience that consists of a physical installation and an interactive virtual layer of three-dimensional soundscapes that are physically mapped to the museum displays. The source for the audio data is digital sound objects. The digital objects originate in a network of object repositories that connect digital content from one museum with other museums collections. The interface enables people to interact with the system by movement and object manipulation-based gestures without the direct use of a computer device. The focus of this paper is the retrieval mechanism for the sound objects for the museum visitor. The retrieval mechanism is built on the user model and conceptual descriptions of the sound object and museum artifacts in the form of ontologies for sound and psychoacoustics, topic ontology and Conceptual Reference Model for museum information. The retrieval criteria are represented as inference rules that represent knowledge from psychoacoustics, cognitive domain and composition aspects of interaction. The system will be demonstrated in exhibition space in Nature Museum in Ottawa in January 2003.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *information filtering, search process, selection process.* H.5.1 [**Information Interfaces and Presentations**]: Multimedia Information – *augmented reality, audio output*

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

Augmented audio reality, user model, ontologies, inference rules

1. INTRODUCTION

Audio museum guides have existed for some time as a means of overcoming the scheduling inflexibility of group tours by museum docents. While beneficial in many respects, the audio guides are limited by their linear sequence and non-interactive structure. Bedersen [3] developed a prototype utilizing portable mini-disc players and an infra-red system to allow museum visitors to explore at their own pace and sequence. As museum visitors approached artifacts on display, relevant audio information would be triggered on the mini-disc player and heard through headphones. *Hyperaudio* [16] provided visitors with palmtop computers and developed specific user models for adaptive systems within a museum setting. *MEG* [2] is a portable digital museum guide for the Experience Music Project in Seattle that allows visitors 20 hours of audio and video on demand. Visitors make their selections either by use of the keyboard within the PDA device or by pointing the device at transmitters located adjacent to artifacts.

In the previous works, the relationship of the digital content to the artifacts is either pre-planned and fixed, or the digital content is not networked and limited to the local device, in some cases both limits are true. *ec(h)o* employs a semantic web approach to the museum's digital content thus it is networked, dynamic and user-driven. The interface of *ec(h)o* does not rely on portable computing devices, rather it utilizes a combination of gesture and object manipulation recognized by a vision system.

The dynamic and user-driven nature of ec(h)o requires a highly responsive retrieval mechanism with a criteria defined by psychoacoustics, content and composition domains. The retrieval mechanism is based on user model that is continually updated as user moves through the exhibition and listens to the audio objects. The criteria are represented by rules operating on the ontological descriptions of sound objects, museum artifacts and user interests.

The capturing of the user interests is in the center of the research of several disciplines such as information retrieval, information filtering and user modeling [21]. Most of the systems were developed for retrieval of documents where document content is analyzed and explicit user feedback is solicited to learn or infer the user interests. In the context of ec(h)o there is no direct feedback from the user. Ec(h)o can be categorized as a personalized system as observes user's behavior and make generalizations and predictions about the individual user based on their interactions [11][18]. Our is an unobtrusive approach to observation of user behavior, similar to the certain approaches to monitoring user browsing patterns [12][14] or user mouse movement and scrolling behavior [8].

The paper is organized as follows. We first describe how echo works and present an overall ec(h)o architecture. Next we provide details of the semantic object descriptions, retrieval criteria, user model and describe retrieval mechanism. Finally, we provide details on the current stage of the system implementation and conclude by highlighting our contributions.

2. EC(HO) ARCHITECTURE

The platform for ec(h)o is an integrated audio, vision and location tracking system installed as an augmentation of an existing exhibition installation. The platform is designed to create a museum experience that consists of a physical installation and an interactive layer of three-dimensional soundscapes that are physically mapped to museum displays and the overall exhibition installation.

Each soundscape consists of zones of ambient sound and “soundmarks” generated by dynamic audio data that relates to the artifacts the visitor is experiencing. The soundscapes change based on the position of the visitor in the space, their past history with viewing the artifacts, and their individual interests in relation to the museum collection. To achieve this type of audio experience the overall system must be integrated with a position tracking system that has a frequent update cycle and a high level of spatial resolution. A pattern of the user movement can indicate the type of the museum visitor [20] as well as user intentions [17].

When the user stops in front of an artifact, she is presented with three sound objects spatially positioned to the left, center and right. By way of a gesture-based interaction, the visitor can interact with a single artifact or multiple artifacts in order to listen to related audio information. The audio delivery is dynamic and generated by agent-assisted searches inferred by past interactions, histories and her individual interests. The source for the audio-data is digital objects. In the case of ec(h)o, we developed a large sample set of digital objects that originated from the partner museums. These digital objects were used to populate the network of object repositories.

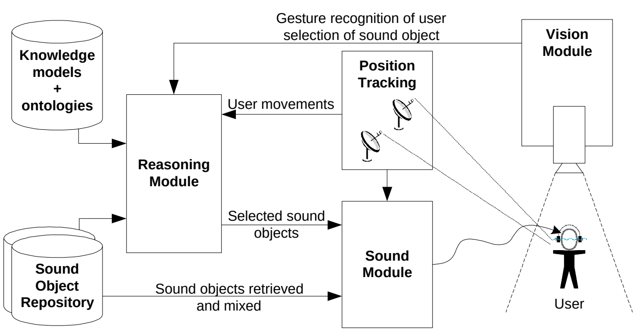

Figure 1 ec(h)o high level architecture

The ec(h)o architecture (Figure 1) consists of four independently functioning modules: position tracking module, vision module, sound delivery module, and reasoning module. Two main types of events trigger the communication between the modules: user’s movement through the exhibition space and user’s explicit selection of the sound objects..

3. SOUND OBJECT RETRIEVAL MECHANISM

One of the main goals of ec(h)o is to achieve an enhanced experience for the museum visitors without inserting an extra layer of technology between the visitor and the museum exhibit. Two mechanisms contribute to an accurate retrieval of sound objects in ec(h)o: the user model and ontology descriptions of objects. As mentioned above user’s interaction space is limited to three sound objects. This poses extreme requirements on the retrieval mechanism as there is no recourse once the ‘bad’ choices are made.

3.1 Semantic Description of Objects

We have identified the following information as essential for ec(h)o:

- the content description of the user interests (user model), sound objects and museum artifacts
- psychoacoustics and sound characteristics of the sound objects
- sequencing models of an interaction

3.1.1 Ontologies for Describing Content

The interaction model is based on the semantic description of the content of the objects. We have developed an ontology where a sound object is described using several properties. As an ability to link to other museums is an important feature of ec(h)o our ontology builds significantly on the standard Conceptual Reference Model (CRM) for heritage content developed by CIDOC [5]. The CRM provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. To describe sound objects we use CRM `TemporalEntity` concept for modeling periods and events and `Place` for modeling locations. We describe museum artifacts using the full CRM model.

The content of the sound object is not described directly but annotated with three entities: concepts, topics, and themes. The concepts describe the domains that are expressed by the sound object such as evolution, behaviour, lifestyle, diversity, habitat, etc. Since the collections in individual museums are different so are the concept maps describing these collections. A topic is a more abstract entity that is represented by several concepts, such as botany, invertabretes, marine biology, etc. To facilitate the mappings between topic ontologies in individual museums we have mapped the topics to the Dewey Decimal Classification [7] whenever possible. Finally, themes are defined as entities supported by one or more topics. For example, the theme of bigness in invertebrates and marine biology.

Table 1 shows content related properties[1] with their domains and ranges.

---

[1] To enable the system to relate sound objects to exhibition artifacts exhibition ontology defines exhibition artifacts as a subclass of an content object. Effectively this provides an exhibition object with the same content descriptive properties as sound objects.

Table 1 Content related properties

Property	Domain	Range
hasTheme	SoundObject	Theme
hasTopic	SoundObject	Topic
hasPrimaryConcept hasSecondaryConcept	SoundObject	Concept
relatesToTemporalEntity	SoundObject	CRM_TemporalEntity
relatesToPlace	SoundObject MuseumArtifact	CRM_Place
describesArtifact	SoundObject	MuseumArtifact

3.1.2 Psychoacoustics and Sound Characteristics

The auditory interface of ec(h)o follows an ecological approach to the sound composition. It provides the basic mechanisms of navigation and orientation within the information space. Three areas are taken into account: psychoacoustic, cognitive, and compositional problems in the construction of a meaningful and engaging interactive audible display. Psychoacoustic characteristics of the ecological balance include spectral balancing of audible layers. Cognitive aspects of listening are represented by content-based criteria. Compositional aspects are addressed in the form of the orchestration of an ambient informational soundscape of immersion and flow that allows for the interactive involvement of the visitor.

Table 2 shows the psychoacoustics ontology that defines the characteristic of the sound objects that are used by the composition rules.

Table 2 Psychoacoustic properties for the Sound Object

Property	Range Values
hasSpectralDensityCenter	<number>
hasSpectralDensityWidth	<number>
hasBandwidth	<number>
relatesToEnvironment	Physical_Environmnet
relatesToEvent	CRM_Event
hasSource	SourceTypeValue (i.e. AnimalSound, HumanEnvironmnetSound)

3.2 The User Model

In the core of the ec(h)o's reasoning module is a user model [21] that is continually updated as user moves through the exhibition and listens to the audio objects.

Figure 2 shows an interaction schema of the user model with other modules. There are two main update sources in the system. First, as the user moves through the exhibition the speed of the movement and stops or slowing down at different artifacts provide updates to the user model. The user behaviors are computed based on the speed and homogeneity of the user movement. The stops and slowing down in front of an artifact are interpreted as an interest in topics represented by the artifact. The user interests and intentions influence the presentation of soundmarks. For example, soundmark radius and volume is increased for those artifacts that correspond with current user interests. Another example can be reducing the number of soundmarks in the exhibition if user's recognized intention is to quickly cross the room.

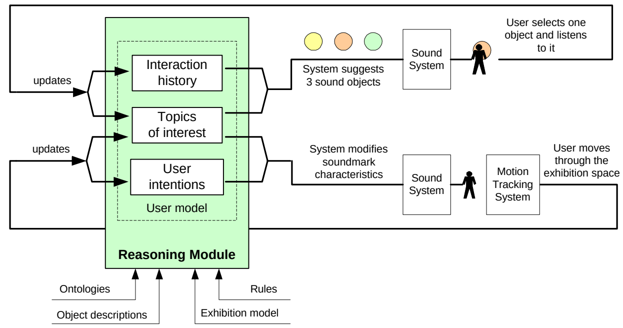

Figure 2 Interaction of user model with other modules

The second source of updates to the user model considers user's direct interaction when user selects a sound object to listen to. In the model this maps to an increased user interest in topics presented by the sound object and updates to the user's interaction history. We describe the user model and retrieval mechanism in detail below.

3.2.1 Interaction History

Interaction History is a record of how the user interacts with the ec(h)o-augmented museum environment. Two types of events are stored in the interaction history: the user's movement and user's selection of objects. The user path through the museum is stored as discrete time-space points of locations on the path. A time-space point is represented as a fact:

```
(user-location (user-id john) (x-position 10.7)
   (y-position 11.5) (time 172.0))
```

The user model correlates the user locations with the exhibition physical model to calculate the relative location of the user to the artifacts/exhibitions. Also, the speed of the user and how much time the user spends in front of the artifact is determined and used to infer a type of user's behavior.

Second type of information stored in Interaction History is user's selections in the form of URLs of sound objects selected by the user.

```
(user-selection (user-id john)
   (sound-object http://echo/narratives/123.mp3)
   (in-front-of artifact-1) (time 184))
```

This information is essential for several tasks ranging from simple avoidance of the delivery of redundant narratives to updating user interests.

3.2.2 User Behavior

The user behavior in the museum context is well studied in curatorial science [6] and used in several systems personalizing the user experience [19][20]. Several categorizations were used, for example one user may go through almost every artifact that is one his/her way, and another user may be more selective and chooses artifacts that have certain concepts. Our categorization of user types is based on Sparacino's work [20] and it classifies users to three main categories. These categories were validated by our own research of site studies and interviews with staff at our partner museums:

- The *greedy* type wants to know and see as much as possible. He is almost sequential, and does not rush.

- The *selective* type explores artifacts which represent certain concepts, and wants to dig into those concepts only.
- The *busy* type does not want to spend much time on a single artifact and wants to stroll through the museum to get a general idea.

In ec(h)o, the user behavior is not static but is updated every minute. The rules consider the location data from user history accumulated within 3 minute interval and topics of previously selected sound objects.

3.2.3 User Interests

The interests for the user are represented as a set of facts where each fact represents a single interest and its relative level

```
(user-interest (user-id john)
   (concept evolution) (level strong))
```

As described in the previous sections, each artifact/exhibition is associated with a set of concepts. The sound objects address a set of particular concepts as well. The interaction of the user and artifacts and sound objects is stored in the Interaction History that together with the user behavior type are used to infer the user's interests. The following principles for the user interest inference are implemented using the reinforcement learning approach [13]:

- If a greedy type user slows/stops in front of an artifact, we can infer that the user is interested in any of general concepts represented by the artifact. If the user continues with his *greedy* behavior in front of that artifact, his interests is updated with related concepts from sound objects selected (not necessarily closely related).
- Interests of a *selective* user do not get easily overwritten. If a *selective* user is moving slowly in front of an artifact he is not interested in[2], one of his previous interests is overwritten by a concept that is 'close' to his previous interests. If a *selective* user stops in front of an artifact he is not interested in, one of his previous interests is overwritten by a concept that is represented strongest by the artifact.
- If a *busy* user slows/stops in front of an artifact, several of his interests are overwritten by general concepts that are also represented strongly by the artifact.
- If a user's behavior is not categorized yet, *User Interest*s can be any general concepts that are strongly represented by the artifact the user slows/stops in front of.

We limit the number of concepts represented in the user model as user interests to 6 to reduce the error in retrieved objects [15].

This is an example of a rule that computes interests of a *greedy* user who just stopped in front of an artifact:

```
(defrule get-greedy-user-interest
  (time ?current-time)
  (user-behavior (user-id ?user)
                 (behavior greedy))
  (or (is-slow ?user ?current-time)
      (stopped ?user ?current-time))
  (just-came-in-front ?user ?a)
  (object-model (object-name ?a) (x-position ?x)
                (y-position ?y) (radius ?r))
                (has-concept ?object ?i strong)
=>
  (assert
     (user-interest (user-id ?user)(concept ?i)))
```

[2] There is no overlap between artifact concepts and user interests.

3.3 Sound Object Retrieval

We have identified the following requirements for the retrieval of appropriate sound objects:

1. Content-relevant to the viewed artifact
2. Content-relevant to the user interests
3. Inviting to explore other areas
4. Plausible from the psychoacoustics perspective

In addition to the criteria for an individual objects the following criteria apply to the sequence of the objects offered to the user:

5. Provide for exploration of a subject in depth
6. Provide for the fluidity in experience both in content and sound experience
7. Provide a mix of informational and entertaining objects

3.3.1 Retrieval Process

The retrieval process in ec(h)o can be broken into four steps as illustrated in Figure 3.

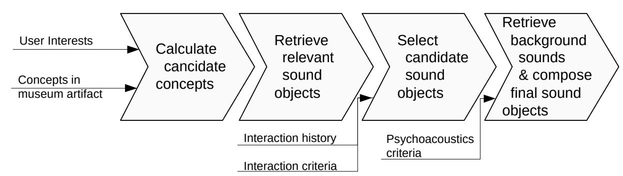

Figure 3 Retrieval process

First, the system determines the candidate concepts as an overlap between user interests and concepts represented by the museum artifacts. The candidate concepts are ranked by a combination of the level of the interest of the user and how strongly they are represented by the artifact.

In the second step the candidate concepts are used in the simple pattern matching algorithm to retrieve semantic descriptions of the information sound objects. The temporal and location properties of the artifact are used to narrow the search to sound objects that are closely related to the presented artifacts.

While the first two steps considered objects as independent acts, the rules in the next step, the content related composition criteria are applied. The criteria consider the next object in the context of the previous objects the user listened to before. The selection is based on theme, topic, concepts, and described artifacts. The relative weight of each type of composition criteria depends on the user type. For example, for the "greedy" user concepts,topics, and 'described artifacts' are of equal criterion to enable the system to offer a wide range of audio objects. For the selective user, the artifact criterion is dominant since the user is very selective among the artifacts on display.

Finally, the background sound objects for each information objects are retrieved using psychoacoustics criteria and the psychoacoustics rules are applied to finalize the choice of the sound objects. For example, if neither event nor environment is specified then use both place and temporal information to infer environment type and use it for selecting the background sound object.

3.3.2 Implementation

The reasoning module is fully implemented with all features described in the previous section and embedded into the Tomcat environment. The Figure 4 shows the implementation schema with Jess Inference engine in the center of the reasoning module. DAMLJessKB[3] converts DAML+OIL ontologies to Jess facts. Reasoning module is connected with other modules through the UDP socket connections and communicates with other 'museums'[4] via SOAP based protocol [10]. We have developed a web-based tool for editing of ontological descriptions of sound objects that generates forms based on the ontology definition by direct querying of the ontologies loaded into the Jess inference engine.

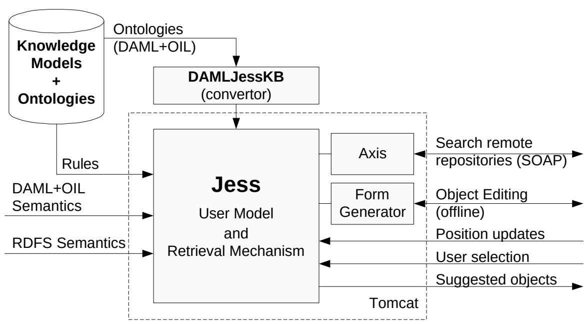

Figure 4 Implementation schema of the reasoning module

The use of a forward chaining inference engine has proved itself to be an efficient mechanism for responding to the dynamic nature of the user input. The system loading time is relatively long as a lot of initial inference is performed on the ontologies and object descriptions. After the startup phase the amount of the inference is limited to the updates from user input resulting in fast responses. Although more extensive testing still needs to be done the pattern of this behavior makes us optimistic with regard to the scalability of the system.

4. NETWORK OF MUSEUMS

One of the main features of the ec(h)o system is that it enables the user to experience the richness of the museum collections located not only in the visited museums but also from the other linked museums. For example, a visitor standing in front of a bear specimen in Nature Museum in Ottawa can listen to the sound object about the role of the bear in the mythology of aboriginal tribes on the West Coast[5] retrieved from the Museum of Anthropology in Vancouver.

Two aspects are critical for fluid retrieval and access of sound objects from other museums: protocol compatibility and semantic mapping between conceptual structures. We addressed the protocol compatibility issue by reusing the infrastructure and protocols our group developed for connecting learning object repositories [9][10]. The only difference is that instead of learning object metadata we share the sound object semantic descriptions.

As different museums can have different conceptualization of the topics covered by their stories the problem of mapping between these conceptualizations need to be addressed. First, we use the standard Conceptual Reference Model for describing temporal and spatial entities which allows us to relate sound objects to time and space. Second, we use Dewey Decimal Classification as an intermediary for mapping between museum specific topic maps. Although this does not provide for an exact mapping our solution is acceptable in the museum setting where the exploration aspect[6] of the user experience dominates the in-depth learning aspect.

5. CONCLUSIONS

In this paper we presented retrieval mechanism used in an augmented audio reality system for museum visitors named ec(h)o. Each visitors experience is tailored to the user interests. The user interests are inferred from the user movement through the exhibition as well as from the visitor's interaction with the sound objects. The sound objects are retrieved based on their relevance to the user interests, narrative criteria and psychoacoustic criteria. Ec(h)o uses ontologies to describe concepts, temporal and spatial characteristics, psychoacoustic and sound characteristics of sound objects. The retrieval mechanism is represented in form of the rules that capture contextual, sound, psychoacoustic and composition criteria for plausible user experience.

The system is a result of convergent research streams from research in object repositories, interaction design, auditory display, knowledge representation, and information retrieval. The ontologies combined with the rule based inference proved to be a powerful implementation platform well suited for this type of the systems. We believe this has enabled us to extend works cited through the paper in several directions. First, it extends the work of the Alfaro et al. work [1] by building rich model of the concepts represented by the sound objects. In ec(h)o, the content presented to the user is not pre-processed for possible linkages as in the systems using Rhetorical Structure Theory [22]. Our approach replaces pre-processed linkages with a retrieval mechanism based on composition and interaction criteria formulated in the form of the rules and applied to semantically described independent objects. This allows ec(h)o to create a network of museums sharing objects and providing richer user experience.

6. ACKNOWLEDGEMENTS

Work presented in this paper is supported by Canarie Inc. grant under E-Content program. Authors would especially thank to the Mark Graham and his colleagues in the Nature Museum in Ottawa for the enthusiastic support to this project. We would also like to

[3] http://plan.mcs.drexel.edu/projects/legorobots/design/software/DAMLJessKB/

[4] Currently we emulate a museum network by seeding independently operating repositories on separate computers. As the project is funded by the Canarie who operates the broadband internet in Canada our assumption is that the connectivity between museums will be at the superior level.

[5] Assuming the user model indicates the user is also interested in history.

[6] Providing that information is relevant to the temporal and spatial aspects of a museum artifact.

thank our colleagues and participants in several workshops that contributed to the development of the project, namely Dale Evernden, Doreen Leo, Gilly Mah, Robb Lowell, Mark Brady, Jordan Williams and Phil Thomson.

7. REFERENCES

[1] Alfaro, I., Zancanaro, M., Cappelletti, A., Nardon, M., Guerzoni, A. Navigating by Knowledge. In Proceedings of International Conference on Intelligent User Interfaces. Miami, Florida. January 12-15, 2003.

[2] Andolesk, Diane, Michael Freedman. “Artifact As Inspiration: Using Existing Collections And Management Systems To Inform And Create New Narrative Structures.” Archives and Museum Informatics Museums and the Web 2001, 2001.

[3] Bederson, B. “Audio Augmented Reality: a prototype automated tour guide.” Conference companion on Human factors in computing systems, 1995.

[4] Bordwell, D.: Film Art: An Introduction (5th edition). McGraw Hill, 1997.

[5] Crofts, N., Dionissiadou, I., Doerr, M., Stiff, M. (eds.) Definition of the CIDOC object-oriented Conceptual Reference Model , July 2001 (version 3.2.1)

[6] Dean, D. (1994). Museum Exibition: Theory and Practic. London, Routledge.

[7] Dewey Decimal Classification, http://www.oclc.org/dewey/

[8] Goecks J and Shavlik J (2000) Learning users’ interests by unobtrusively observing their normal behavior. In: Proceedings of the 5th International Conference on Intelligent User Interfaces, New Orleans, Louisiana, January 2000, pp. 129–132.

[9] Hatala, M and Richards G. Global vs. Community Metadata Standards: Empowering Users for Knowledge Exchange, in: I.Horrocks and J.Hendler (Eds.): International Semantic Web Conference 2002, Springer, LNCS 2342, pp. 292-306, 2002.

[10] Hatala, M., Richards, G., Eap, T., Willms, J. The eduSource Communication Language: Implementing an Open Network for Learning Object Repositories and Services, Will appear in Special Track on Engineering e-Learning Systems held at ACM Symposium on Applied Computing (SAC) 2004, Nicosia, Cyprus, March 2004

[11] Kobsa, A., Fink, J. User Modeling for Personalized City Tours. Artificial Intelligence Review 18, 2002, pp. 33-74, Kluwer Academic Publishers.

[12] Liberman, H. Letizia: An Agent that Assists Web Browsing. IJCAI-95, pp. 924-929.

[13] Mitchell, T. (1997). Machine Learning. New York, NY, McGraw-Hill.

[14] Mladenic, D. Personal WebWatcher: Implementation and Design, Technical Report IJS-DP-7472, Department for Intelligent Systems, J.Stefan Institute, October, 1996.

[15] Mostafa,J., Mukhopadhyay, S,.and Palakal, M. Simulation Studies of Different Dimensions of Users' Interests and their Impact on User Modeling and Information Filtering, Information Retrieval 6 (2): 199-223, April 2003

[16] Sarini, M., Strapparava, C. Building a User Model for a Museum Exploration and Information-Providing Adaptive System. Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia, HYPERTEXT'98, 1998.

[17] Schlieder, Ch., Vogele, t., Werner, A. Location Modeling for Intentional Behavior in Spatial Partonomies Ubicomp 2001, Workshop "Location Modeling for Ubiquitous Computing", Atlanta, Georgia, USA, pp.63-70

[18] Seo Y-W and Zhang B-T (2000) A reinforcement learning agent for personalized information filtering. In: Proceedings of the 5th International Conference on Intelligent User Interfaces, New Orleans, LA, January 2000, pp. 348–351.

[19] Serrell, B. (1996). The question of visitor styles. In S. Bitgood (Ed.). Visitor Studies: Theory, Research, and Practice, Vol. 7.1. Jacksonville AL: Visitor Studies Association, pp. 48-53.

[20] Sparacino, F. (2002). The Museum Wearable: real-time sensor-driven understanding of visitors' interests for personalized visually-augmented museum experiences. In: Proceedings of Museums and the Web (MW2002), April 17-20, Boston.

[21] Wahlster, W. and A. Kobsa (1989). User Models in Dialog Systems. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg Œ Berlin, Springer Verlag.

[22] Zancanaro, M., Stock, O., Alfaro, I. Using Cinematic Techniques in a Multimedia Museum Guide. Using Cinematic Techniques in a Multimedia Museum Guide. In Proceedings of Museums and the Web 2003. Charlotte, North Carolina. March 19-22, 2003.