P-News: Deeply Personalized News Dissemination for MPEG-7 based Digital Libraries

Qiuyue Wang[1], Wolf-Tilo Balke[2], Werner Kießling[1], Alfons Huhn[1],

[1] Institut für Informatik, Universität Augsburg,
89165 Augsburg, Germany
{wang, kiessling, huhn}@informatik.uni-augsburg.de

[2] Computer Science Department, University of California
Berkeley, CA 94720, USA
balke@eecs.berkeley.edu

Abstract. Advanced personalization techniques are required to cope with novel challenges posed by attribute-rich MPEG-7 based digital libraries. At the heart of our deeply personalized news dissemination system P-News is one extensible preference model that serves all purposes, preventing impedance mismatches between the various stages: User modeling by structured preference patterns, automatic query expansion including ontologies, preference query evaluation by Preference XPath including nested preferences on categorical data, quality assessment of query results, personalized notification and news syndication.

1 Introduction

The amount of information available in digital libraries via the Internet is overwhelming and the task of extracting all valuable knowledge increasingly time-consuming. Especially in areas with short innovation cycles like IT not only new documents arrive in large bulks every day, but also what is considered to be relevant will strongly differ among various user groups like business-oriented consultants, technology-oriented developers or highly specialized researchers. Users find themselves confronted with a well-known dilemma: spending too much time on going through new, but probably irrelevant information will cost valuable research or working time, whereas spending less time may result in missing some vital information. Many users of digital libraries or subscribers of news services have suffered the troublesome problem of getting 'properly' notified about latest publications. Definitively this should happen in a personalized manner as much as possible. The acquisition and maintenance of user preferences about topics of interest and preferred content characteristics are prerequisites for better solutions than current ad hoc approaches. The P-News[1] project tackles this challenge by applying a highly flexible preference methodology with powerful query capabilities and managed usage stereotypes throughout the process of dissemination.

[1] P-News is funded within the German Research Foundation's strategic research initiative 'Distributed Processing and Delivery of Digital Documents (V^3D^2)'

Previous approaches for news dissemination are mainly focusing on IR-techniques matching a set of (weighted) keywords against the document collection. The introduction of structured documents in XML and related meta-data lead to applying IR techniques on well defined parts of the document like its title or annotations. However, recent work in personalized information systems shows that not only the document is structured, but already the user's query, sets of keywords, notions of relevance, preferences for notification, etc. The contribution of this paper therefore is twofold; first we present an intuitive way for users to express a structure on their information needs and preferences accompanying the entire process of dissemination way beyond Boolean combinations. Secondly we focus on different roles for interaction and providing predefined structures of the information to express common knowledge and thus ease the usage of the system. Unlike common dissemination engines, notifications in P-News are supposed not only to involve matching content-based preferences, but also closely adapting to particular users and situations. Consider a sample scenario:

Example: Cathy is a professor at university and besides research projects also manages a spin-off business. She uses fast Internet access with a PC, but also uses a mobile phone to keep track of current events. Cathy of course wants to know about news related to her research and is interested in specific business news. Assume a suitable document arrives at P-News, e.g. a new research article. The dissemination process first has to recognize from the representation of Cathy's topical preferences that the new article is *relevant*, from her quality preferences that its *degree* of relevance justifies a notification and from her notification preference, how to *syndicate* the document and where to deliver it to. For instance research-related items could always be sent as emails containing the full document. Her preferences as a business woman could also consider current business events as interesting enough triggering notification to the WAP cell-phone. Due to its limited capabilities, the syndication will automatically put up only the headline and a short abstract for delivery.

Please note that Cathy can interact with the system in different roles having different (even contradictory) topical preferences, can have different notions of relevance, i.e. what degree of quality she is willing to accept within these different roles, and can have different preferences on how to be notified in each role. Moreover, in each role there will generally be certain stereotypical preference patterns for user groups. To cater for such deep personalization scenarios we need more powerful techniques than today's publish/subscribe technologies or IR-based keyword searches in XML documents. In this paper we will address all relevant topics towards building such a deeply personalized dissemination engine using a single consistent preference model for all types of preferences. We will show step by step how to tackle each necessary task and present innovative techniques for the application in news dissemination. We will in detail discuss its impact on specialized digital libraries with a focus on the use of categorical metadata and attribute-based searches in an intuitive and cooperative fashion.

The paper is organized as follows: section 2 will revisit related work and MPEG-7 metadata. In section 3 we present the technological innovations for our news dissemination task. Section 4 finally will show how to build these techniques into a running system, and will present a use case interaction lifecycle.

2 News Dissemination

2.1 Related Work in News Dissemination

As a first approach to overcome the problem of having to sift through vast amounts of information, publishing companies have introduced customizable news letters. Users can subscribe to a variety of general terms describing areas of interest, and are periodically informed which new books might be of specific interest (e.g. 'SpringerAlerts' [1]). However, subscription services have still a long way towards personalization, because users might not be offered a specifically interesting category they need, might find that the publisher has chosen too broad/narrow terms as categories, or may have an entirely different understanding of some categories altogether.

The area of news dissemination therefore has moved to employing advanced techniques for keyword matching in the texts of documents. Engines like SIFT [2] show already good results for full-text retrieval featuring IR techniques and prove that the task of finding relevant documents for notification can be efficiently performed even for large numbers of concurrent users. With the advent of XML engines for the search in XML documents like XIRQL [3] or XXL [4] applied the probability-based keyword retrieval to structured XML documents. However, none of these techniques focused on retrieval over structured categorical attributes that is needed for deep personalization, i.e. not only the document structure should be taken into account, but also the structure of query terms (keywords, attributes,…) with respect to each other.

In terms of advanced usability of digital libraries and ease of querying user profile modeling has been proposed, see e.g. [5], and already proved its usefulness through advanced personalization in the field of news dissemination [6]. User models can be automatically expanded and profit from already existing similar user profiles. In this respect also the mining of related information to adapt recommendations more closely to the individual user has been applied [7]. Our work is a direct continuation of these advances; however, we smoothly embed these advances in a powerful preference framework and thus do not have the overhead of managing user profiles, mined data or query terms of different structure. Moreover, we enhance the benefits by using an ontology-based approach to incorporate common domain knowledge into the retrieval process respecting the role of each individual user.

In today's digital libraries compound documents containing text, images, audio or even video files together with adequate annotations or comments are quite common. Acknowledging its necessity, within standardized multimedia description frameworks like MPEG-7 already a simple set of description tools for describing user preferences (*UserPreferences*) [8] has been provided. It enables users to select their preferred multimedia content in terms of attributes related to the creation, classification, and source of the content. Multiple preference components can then be organized into a hierarchical structure, each one carrying a numerical value indicating the relative importance of this preference. The expressiveness of *UserPreferences* descriptions, however, is far more limited than our approach, since only exact matching is supported, even for simple numerical attributes (e.g. media duration time).

2.2 MPEG-7 Annotations and Use in Digital Libraries

The introduction of standardized metadata descriptions facilitates search and retrieval of multimedia content in digital libraries. Currently MPEG-7 [8] is the most complete description standard for multimedia data providing a comprehensive set of standardized tools to describe multimedia data. For example, a video segment can be described in many different aspects, like *MediaInformation* (e.g. storage format, visual coding), *CreationInformation* (e.g. title, creator, classification), *UsageInformation* (e.g. access rights, distributor), structural aspects (e.g. subsegments) and conceptual aspects (e.g. text annotation, semantics). The description tools are specified in the Description Definition Language based on XML schema. Thus MPEG-7 descriptions are complex XML documents. For instance (a) and (b) in Fig. 1 are excerpts from MPEG-7 descriptions of the creation information for two video segments.

With this standard the focus in searching digital libraries or evaluating a document's relevance shifts to attribute-rich search on *categorical data*. Consider for example the information in Fig. 1. A query on all documents that have been created and preferably produced by 'IBLabs' will need more than today's capability of searching the creator tag in e.g. XQuery for the existence of the keywords 'IBLabs' and 'Producer'. Here the evaluation of a nested preference is needed where the keyword 'Producer' has to be the role within the *same* creator tag that also contains the keyword 'IBLabs'. Thus document (a) in figure 1 would be a better match than document (b). However, document (b) should nevertheless still be considered more relevant than other documents because of the keyword 'IBLabs' as 'Editor', which is another role of a creator. So the creator attribute has certain *categories* ('producer', 'author', 'editor') as domain on which users might express preferences. Traditional IR techniques or the MPEG-7 description tools, however, can handle such preferences only up to a certain extent yet, and provide no intuitive (i.e. declarative) way of expressing them. In the following we will show how our system can deal even with such complex preferences.

```
<CreationInformation>
    <Creation>
      <Creator>
        <Role>
          <Name> Producer </Name>
        </Role>
        <Agent>
          <Name> IBLabs </Name>
        </Agent>
      </Creator>
    </Creation>
</CreationInformation>
```

(a)

```
<CreationInformation>
    <Creation>
      <Creator>
        <Role>
          <Name> Producer </Name>
        </Role>
        <Agent>
          <Name> Microsoft </Name>
        </Agent>
      </Creator>
      <Creator>
        <Role>
          <Name> Editor </Name>
        </Role>
        <Agent>
          <Name> IBLabs </Name>
        </Agent>
      </Creator>
    </Creation>
</CreationInformation>
```

(b)

Fig. 1. Example MPEG-7 descriptions

3 Basic Concepts of Deep Personalization for Dissemination

3.1 A Model for Structured User Preferences in XML Libraries

MPEG-7 descriptions are basically XML data. In searching or filtering XML data, traditional IR approaches, keyword sets or vectors, are often unable to refer to the data structure and incorporate semantic relations between the query terms. Query languages for XML, such as XPath or XQuery, can be used to formulate precise queries over data. But they can only express Boolean (or hard) conditions; no ranking or soft conditions are possible. Many efforts are spent on combining the structure and ranking-based search, e.g. XXL [4] and XIRQL [3]. They mainly use vague predicates and probability-based combination function to score structurally matched document fragments. However, numerical ranking approaches are generally less expressive than qualitative ones [21].

In [9], an approach for preference modeling is proposed utilizing strict partial orders featuring an intuitive "I like A better than B" semantics. User preferences are generally considered soft conditions that are evaluated as strict partial orders over the data set. Thus all *best matching* objects, not necessarily exact matches, will be returned. As an essential feature of the approach in [9], a set of predefined *preference constructors* are used to construct arbitrary preferences. For example, *AROUND(x)* is a base preference constructor on numerical attribute values preferring values closest to the stated value *x*. *Pareto* and *Prioritized* are complex constructors for combining preferences of equal importance or with priorities. This set of constructors is extensible and users are enabled to define their own constructors if needed. For P-News we will add a novel preference constructor, called *nested* preference, extending the model to handle complex XML data. Formally let <*path*> be an XML path expression and let *dom*(<*path*>) denote the domain of objects reachable by <*path*>. Then a preference P is defined as P = (<*path*>, $<_p$), where $<_p$ is a strict partial order over dom(<*path*>).

Definition 3.1 Strict partial order relation between sets
Given P = (<*path*>, $<_p$) we define a strict partial order $<<_p$ over the set of all finite subsets of dom(<*path*>) as follows. For all finite subsets X, Y of dom(<*path*>)**:**

$$X <<_p Y \quad \text{iff} \quad (\forall x \in X, \exists y \in Y: x <_p y) \wedge Y \neq \emptyset$$

It can be proved that $<<_p$ is a strict partial order. ■

Definition 3.2 Nested preference
Given a preference P = (<*path*>, $<_p$) and objects O_j, $O_k \in$ dom(<*path**>), let $\{O_j.<path>\}$ and $\{O_k.<path>\}$ denote the sets of selected objects by navigating <*path*> from O_j and O_k, respectively. Then a *nested* preference $P^* = (<path^*>, <_{p^*})$ is defined as: $\forall O_j, O_k \in$ dom(<*path**>): $O_j <_{p^*} O_k$ iff $\{O_j.<path>\} <<_p \{O_k.<path>\}$

Preference XPath [10] has been developed to evaluate preference queries. It extends standard XPath by soft filtering conditions bracketed in '#[' and ']# ' in contrast to hard conditions in brackets ' [' and '] '. A soft condition defines a strict partial order

over the set of elements to be filtered and returns only the best matches. Extending the preference model by nested preferences, we extend Preference XPath as follows:

```
LocationStep : axis nodetest (predicate | pref_pred)*
```

pref_pred is the extension part to the standard XPath, i.e. soft filtering conditions.

```
pref_pred : '#[' preference ']#'
preference : base_preference
           | xpath '{' preference '}'
           | preference 'and' preference
           | preference 'prior to' preference
```

base_preference is defined on atomic attribute values, e.g. strings or numbers. The second case, *xpath '{' preference '}'*, is the extension for nested preferences. The *'and'* and *'prior to'* are for Pareto and Prioritized combinations respectively.

Example (cont.): Assume Cathy prefers videos produced by 'IBLabs'. Using the schema of MPEG-7, such a preference can only be expressed in a nested way:

```
/Mpeg7/Description/MultimediaContent//*
#[CreationInformation/Creation/Creator
{Role/Name is 'Producer' and Agent/Name is 'IBLabs'}]#
```

Here ‘Role/Name is 'Producer'’ and ‘Agent/Name is 'IBLabs'’ are base preferences combined by the Pareto constructor 'and', which means both preferences are equally important. The combined preference induces a strict partial order on objects of CreationInformation/Creation/Creator, which in turn gives a strict partial order on higher level elements, i.e. multimedia segments accessed by /Mpeg7/Description/MultimediaContent//*; thus a nested preference. Evaluating the query on the documents in Fig. 1, for each Creator in (b), there is a better Creator in (a). So video segment (a) is considered better than (b) and the intuitively expected result set {(a)} will be returned.

It is important to note that keyword search on full-text attributes using standard IR methods can be orthogonally embedded in our preference model. For instance, keyword search using the vector space model can be implemented as a basic $rank_F$ constructor [9] on the full-text attribute, possibly combined with preferences on other attributes [11]. Thus, P-News caters for arbitrary queries on XML data.

3.2 Structured Preference Patterns of User Groups

For meta-data-based document retrieval in digital libraries users will always assume a certain amount of common knowledge within the system. In recent years ontologies as a way of representing common knowledge or shared vocabulary within a domain have spread widely [13]. With ontologies we can model complex semantic relationships and exploit them for subsequent structured querying. As shown in [12] expanding queries along certain ontology-based patterns will result in improved querying, because the choice of useful preferences for relaxation can be limited down to a sensible applicable set. While an ontology can be as complicated as arbitrary semantic graphs including various relationships and inference rules, P-News does not attempt entirely

ontology-driven querying, but instead uses ontologies in their most basic incarnation: concept hierarchies. Such ontological information poses a partial order on the string-valued domain set, e.g. the specified term is preferred to its synonyms and hyponyms, which in turn are preferred to its hypernyms, while the hypernyms are still considered better than other values. When a preference query is evaluated, the partial order induced by the ontology structure is respected. The basic technique is to use the EXPLICIT [9] or a user-defined preference constructor to expand the original query.

Since different interest groups often show different but within the group sufficiently similar interests, useful default values can be assumed for all preferences not explicitly provided by a user. We refer to what is typically considered relevant in different user groups as default preference patterns. These patterns are predefined and can evolve over the usage cycle of group members by the feedback given. As in dissemination frameworks unnecessary notifications have to be avoided, integrating this common knowledge into user queries in an unstructured way would often confront users with lots of useless results. With the preference query model, we can use the "prior to" preference combination to integrate the user group's preference pattern into user's query, however in a lower level of priority. Figure 2 shows the structure of query terms induced by query expansion with ontologies and preference patterns.

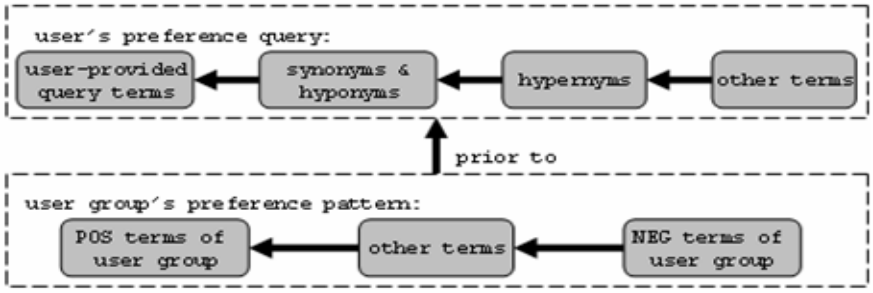

Fig. 2. Query expansion with ontologies and preference patterns

Example (cont.): Consider that our sample user Cathy might research concepts of object-oriented programming in Java. Our system has a simple ontology as concept hierarchy in IT domain that models the terms 'Java', 'C++', 'Smalltalk' as subtypes of 'object-oriented languages', which in turn is a subtype of 'programming languages'. Note that 'C++' and 'Smalltalk' are not synonyms for 'Java', but unlike systems developers focusing on Java programming, researchers, who are interested in the basic concepts of Java might with high probability also be interested in C++ concepts. Thus, if Cathy has a POS preference on 'Java', and a video talk on 'C++' arrives, we might consider a notification based on the specific group or pattern Cathy belongs to. However, we can always ensure by not relaxing above the 'object-oriented languages' node in our ontology that query results have still enough in common with Cathy's query. Moreover, individual preferences always override preferences of the group. Now assume that a significant number of Cathy's group members have stated to be no longer interested in 'Smalltalk'. After query expansion with ontologies and group preferences, Cathy's original query (represented in the compact algebra notation of [9]

due to the limited space), P = POS(keyword, {'Java'}), is expanded into P' = EXPLICIT(keyword, {'C++' < 'object-oriented languages', 'Smalltalk' < 'object-oriented languages', 'object-oriented languages' < 'Java'}) prior to NEG(keyword, {'Smalltalk'}). Hence, we get a new single preference query that can simply be evaluated like before, but now takes common interests of a user group into account.

3.3 Selecting Best Matches by Assessing Result Quality

The knowledge maintained in the structure of preferences can also be used to assess the quality of retrieved results. To distinguish relevant from non-relevant objects declarative query languages offer capabilities that will stop the relaxation at a certain degree of generalization or will only relax within a certain range of objects, e.g. in the previous example Cathy's constraints were only relaxed to languages that are object-oriented. P-News takes quality assessment beyond mere numerical thresholds to a relaxation directed by the needs of each individual user even for categorical data. To assess the quality of each object returned by a query, the preference structure of the query is compared with the actual matches of attributes/keywords like shown in [18]. Our model uses different linguistic quality levels to express the relevance of an object ranging from 'sufficient', 'acceptable', 'good', 'very good' to 'perfect match'. The user can individually assign his/her perception of these levels for each base preference, e.g. in the case of keyword matching using standard IR distance measures or in the case of categorical data using the tolerated discrepancy for relaxation. Within complex preferences these basic measures are then aggregated again according to the specifications of the user, e.g. *median*, *maximum* or *minimum* (see [22] for details).

Example (cont.): Assume Cathy prefers videos produced by 'IBLabs', but would also rather prefer files with a size of about 250 MB. She only wants to be notified of results having 'very good' quality or better. Evaluating the nested keyword matching Cathy can easily define acceptable quality thresholds. For the numerical attribute, file size, Cathy may state a tolerable deviation of 10%. For each 10% more or less the quality level drops one step. Assume that P-News has to compare documents (a) and (b) from Fig. 1 for possible notification with file sizes {(a), 310 MB} and {(b), 240 MB}. In terms of quality on two base preferences this leads to {(a), perfect, good} and {(b), very good, very good}. Now Cathy again can express her exact needs in terms of overall quality. She might e.g. define a *maximum* of 'very good' as sufficient and gets notified about both documents or she might only be willing to accept documents with a *minimum* of 'very good' and gets notified about document (b) only. In any case she is enabled to intuitively specify her notion of relevance.

4 The P-News System for Personalized Dissemination

Having presented the underlying technologies to build a running system let us now take a closer look on the prototype system's architecture and a lifecycle of interaction.

4.1 P-News MPEG-7 Library and System Architecture

In P-News MPEG-7 annotations for multimedia content are stored in XML repositories, associated with a query engine for evaluating Preference XPath queries [10]. As in our running example, we set our use case in the domain of IT technology. The test library consists of about 90 videos (ca. 24 GB) from Computer Chronicles [14] and colloquium series of computer science department in the University of Washington. All videos have been manually annotated using MovieTool [15]. The annotations focus on a set of *controlled-term* attributes provided by the MPEG-7 standard, whose values are from a predefined vocabulary. This use of controlled vocabulary leads to a standardized annotation and subsequent querying with categorical data that shows a strictly typed structure and thus makes user queries comparable within user groups.

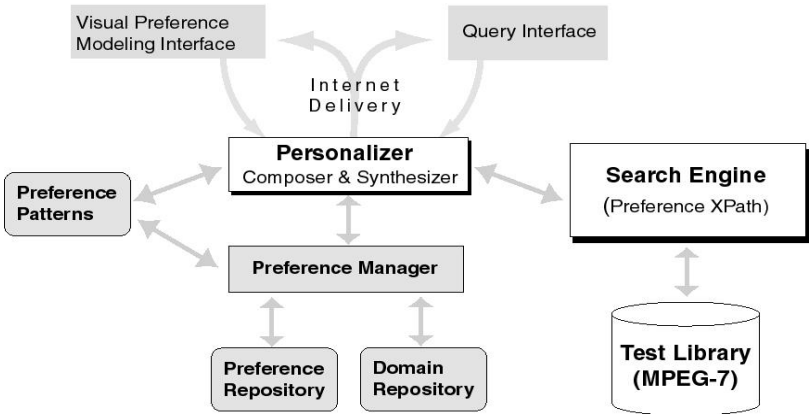

Fig. 3. General architecture of the P-News system

Fig.3. sketches the general architecture of the P-News system. The central component of our architecture is called *personalizer*. It composes the user query by integrating a user's preferences. Then it expands the query using the preference pattern of the user's group (respecting his/her current role) and poses it to the retrieval engine. The result is evaluated and all data whose quality assessment allows for notification are syndicated into the preferred format, layout, etc. and delivered to the appropriate client device. To adapt his/her stored preferences or enter new preferences our architecture also provides a visual preference modeling interface for users to graphically construct structured preferences. These components are built on top of a *preference manager* that manages the different kinds of preferences. For the query composition and expansion the *user provided preferences* and d*omain-specific knowledge* grouped into *preference patterns* are used. Furthermore, each user's notion of relevance (i.e. the expected quality of the results) is represented within the patterns stating individually how much relaxation is still acceptable. Notification preferences of each user are specified for content syndication. Please note, though all these different preferences are used for specific tasks, their basic structure is always defined by our underlying preference model. Arbitrarily complex preferences thus can be stored in the same

repository individually characterized by the user they belong to, and the group, in which they are applicable. We store all the preferences in XML format; our repository [16], however, distinguishes between user preferences, the group specific preference patterns, and the ontologies stored in the domain repository. Using Preference XPath, the preference manager chooses all applicable preferences for the personalizer.

4.2 The User Interaction Lifecycle

User Registration and Preference Modeling. Our use case focuses on a digital library of IT-related content annotated in MPEG-7. The system maintains a set of predefined preference patterns for different user groups. Each new user registering with the system is provided a default preference pattern from the user group that he/she has been assigned to. The user now can view, edit, construct or remove his/her individual preferences using a visual interface, before they are stored in a preference repository.

Query Composition. As discussed in section 3.1, complex preference queries can include keyword search, attribute matching etc. In MPEG-7, there is a predefined vocabulary associated with each *controlled-term* attribute, which can be viewed as a simple ontology specifying all the valid terms and subsume relationship between terms for the attribute. For each user group there is an ontology representing their view of the world, i.e. a taxonomy of all the categories, topics and keywords in IT domain, which is applied to the full-text attributes. When scheduling user queries P-News expands them with the predefined vocabularies or ontologies and group-specific preference patterns as discussed in section 3.2.

Query Evaluation. According to timing information specified for each query, i.e. StartTime, Interval and ExpireTime, queries are activated periodically and evaluated over the set of new data, i.e. the data that come into the system between the last processing time of the query and the current time. To improve the efficiency and/or scalability, the common parts of the activated queries, e.g. common data sets, common XPath expressions, are identified and the computation is shared among them. Existing work on multiple XML query evaluation have been adapted for this purpose [17].

Quality Assessment. Before making a decision on notifying a user, P-News analyzes the returned results in terms of quality as discussed in section 3.3, and filters objects, for which soft constraints have been too far relaxed. The preference structure of the query is constructed and compared with respective structure of the matches in the result set. P-News enables users to specify the acceptable distances for relaxation on each base preference and computes quality values for complex preferences again according to user preferences. Thus an overall quality value for the object in terms of the query can be computed inductively and all irrelevant objects discarded.

Notification and News Delivery. When P-News decides for notification, the system sends a message summarizing new media data that might be of interest. This message is syndicated according to the user-specified form and sent to the preferred device (see [19] for details on personalized multimedia content delivery). By default P-News notifications are simple e-mails listing the interesting documents and containing links to the full multimedia content. When the user follows a link, the P-News server will automatically adapt the content to user's technical device characteristics. Cur-

rently we use SMIL (Synchronized Multimedia Integration Language [20]) files to deliver video data. SMIL also offers a set of attributes facilitating limited adaptations.

Relevance Feedback. Our current implementation features only a limited approach to exploit relevance feedback. For example, P-News assumes a user to be interested in a particular video (and related keywords/topics), when he/she follows the link to open a video after reading the respective abstract. This behavior is recorded on the server and can be used for modifying the user's preference as well as his group's preference pattern. Deriving good patterns from feedback will be part of P-News ongoing work.

5 Summary and Outlook

In this paper we addressed the problem of deep personalization in news dissemination systems. The key technology in implementing the P-News system is one coherent extensible preference model that serves all purposes, preventing an impedance mismatch between various stages. Enabling the user to apply individual preferences in every single step throughout the dissemination process, P-News facilitates a tailored notification about relevant documents in a digital library. We focused on multimedia documents described by MPEG-7 metadata to allow users to express their preferred content, notion of relevance and delivery preferences in an intuitive way. We presented an extension by nested preferences that are essential in structured querying of XML documents, using our unique Preference XPath. Moreover, in addition to exploit the document structure to gain better result sets, we also allowed for expressing preference structures on users' preferred keywords or categorical attribute values. Since users should not be burdened with all the extensive modeling of preferences within stereotypical interest groups, an ontology-based approach for automatic query expansion with typical preference patterns has been realized. Finally, we enabled users also to specify their individual quality preferences to avoid unnecessary notifications as far as possible. Merging these techniques into the workflow of dissemination the P-News system essentially extends the expressiveness of dissemination in digital libraries catering for both textual information and categorical metadata descriptions.

Our future work will concentrate on the experimental evaluation of the system, user case studies and the relevance feedback used to reflect current changes of interest within the user groups. Also managing users changing between different groups needs some deeper research. A detailed analysis of each individual user's interaction with the delivered results can be expected to allow for monitoring dynamic changes in each group's profile. Finally, to keep up with standards, we will migrate from Preference XPath to Preference XQuery.

Acknowledgements

Part of this work was funded by the German Research Foundation DFG within an Emmy-Noether Grant. The authors would like to thank José Maria González Pinto and Timotheus Preisinger for their helpful comments and work on the P-News system.

References

1. SpringerAlerts, http://www.springer.de/alert
2. Yan T.W., Garcia-Molina H.: The SIFT information dissemination system. ACM TODS, Vol. 24, Issue 4 (1999) 529–565
3. Fuhr N., Großjohann K.: XIRQL: A Query Language for Information Retrieval in XML Documents. Proc. Int. ACM SIGIR Conf. (SIGIR 2001) 172–180
4. Theobald A., Weikum G.: The Index-based XXL Search Engine for Querying XML Data with Relevance Ranking. In Proc. Int. Conf. on Extending Database Technology (EDBT 2002) 477–495
5. Amato G., Straccia U.: User Profile Modeling and Applications to Digital Libraries. In Proc. Europ. Conf. on Digital Libraries (ECDL 1999) 184–197
6. Esteban A.D., Gómez-Navarro P.G., Jiménez A.G.: Evaluating a User-Model Based Personalisation Architecture for Digital News Services. In Proc. Europ. Conf. on Digital Libraries (ECDL 2000) 259–268
7. Hwang S.-Y., Lim E.-P.: A Data Mining Approach to New Library Book Recommendations. In Proc. Int. Conference on Asian Digital Libraries (ICADL 2002) 229–240
8. MPEG-7 Overview, http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm
9. Kießling W.: Foundations of Preferences in Database Systems. In Proc. Int. Conf. on Very Large Databases (VLDB 2002) 311–322
10. Kießling W., Hafenrichter B., Fischer S., Holland S.: Preference XPATH, A Query Language for E-Commerce. In: Buhl H.U., Huther A. Reitwiesner, B. (eds.): Information Age Economy. Physika-Verlag, Heidelberg (2001) 427–440
11. Leubner A., Kießling W.: Personalized Keyword Search with Partial-Order Preferences. Proc. Brazilian Symposium on Database Systems (SBBD 2002). 181-193
12. Balke W.-T., Wagner M.: Cooperative Discovery for User-Centered Web Service Provisioning. In Proc. Int. Conf. on Web Services (ICWS 2003) 191–197
13. Fensel D.: Ontologies: Dynamics Networks of Meaning, In Proc. of the Semantic Web Working Symposium (2001), Stanford, CA, USA
14. Computer Chronicles, http://www.archive.org/movies/computerchronicles.php
15. MovieTool, http://www.ricoh.co.jp/src/multimedia/MovieTool/
16. Holland S., Ester M., Kießling W.: Preference Mining: A Novel Approach on Mining User Preferences for Personalized Applications. Proc. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD 2003) 204-216
17. Chen J., DeWitt D.J., Feng T, Wang Y.: NiagaraCQ: A scalable Continous Query System for Internet Databases. In Proc. Int ACM SIGMOD Conf. (SIGMOD 2000) 379–390
18. Balke W.-T., Wagner M.: Through Different Eyes – Assessing Multiple Conceptual Views for Querying Web Services. Proc. Int. World Wide Web Conf. (WWW 2004)
19. Wagner M., Balke W.-T., Kießling W.: An XML-Based Multimedia Middleware for Mobile Online Auctions. Proc. Int. Conf. on Enterprise Information Systems (ICEIS 2001), 934–944
20. SMIL, http://www.w3.org/AudioVideo/
21. Chomicki J.: Preference Formulas in Relational Queries. ACM Transactions on Database Systems, 28(4), December 2003, pp.427–466.
22. Kießling W., Fischer S., Döring S.: COSIMA B2B - Sales Automation for E-Procurement. 6th IEEE Conference on E-Commerce Technology (CEC04), San Diego, USA, July 2004