IMAP - Intelligent Multimedia Authoring Tools for Electronic Publishing: Extended Abstract*

Sarit Kraus[1], Alexander Kröner[2], and Lea Tsaban[1]

[1] Department of Mathematics and Computer Science
Bar-Ilan University, Ramat-Gan, 52900, Israel
{sarit,lea}@macs.biu.ac.il
[2] German Research Center for Artificial Intelligence (DFKI) GmbH
Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany
kroener@dfki.de

Abstract. IMAP provides software tools that support the authoring of electronic presentation by helping the author in multimedia content selection and layout design. IMAP consists of a Content Manager and a Layout Manager. In this paper we present the usage of IMAP tools for the authoring of on-line newspapers. The content selection is based both on the author's specifications and the user's interests, where the interrelations between objects play an important role in the evaluation of the set of objects. Layout management relies on a set of layout requirements taken from a layout profile and an author-defined style sheet. As was demonstrated in our experiments, integrating of both techniques yields interesting newspapers, whose layout can be customized by the reader.

1 Introduction

The long-term goal that we address in IMAP is the development of software tools for content and layout management that will simplify and accelerate the creation of electronic presentation by automating the complex, time-consuming tasks of multimedia content selection and layout design. In this contribution, we focus on the authoring of personalized online newspapers.

Personalization of presentations has become a main issue in recent years (e.g., [1, 2]). Part of this research focuses on electronic access to news (e.g., [5]). However, most of these works focus on identifying the user's fields of interest, i.e. the construction of a *user profile*. Then the articles are scored using the user profile and the highest scored documents are included in the newspaper. We focus on the interest of the user in the overall newspaper while enabling the author to influence the content of the newspaper via constraints, and maintaining the satisfaction of the user.

An appropriate layout for the personalized content selection should reflect the user's personal interests as well. From the manifold approaches to automated

* This research was supported by a Grant from the G.I.F., the German-Israeli Foundation for Scientific Research and Development.

layout (e.g., [8,4,9]), we chose constraint techniques, which had already turned out to be a valuable means for creating Web pages (cf. [3,6]). We show how to extend such an approach so that the layout reflects constraints imposed by the given browsing environment, and may further be customized by the user according to his preferences.

2 Content Management

The input of the Content Management (CM) system includes: (i) A set of candidate documents for the newspaper. (ii) A set of constraints on the selection of documents provided by the *author* (editor) of the newspaper, e.g., "At least a certain percent/number of the documents should be from a certain subject and sub subject." (iii) A *reader profile*: The profile of preferences of the reader. (iv) A number K: The number of documents the newspaper should include. The output of the CM is a list of K documents for the newspaper, ranked according to the level of the reader's interest.

The evaluation of a set of documents depends on the estimation of the agent's interest in the documents and the measure of the satisfaction of the set of documents of the author's constraints. We score each document according to the reader profile, using the correlation between the two. The similarity between a reader profile and a document profile is calculated by referring to the keywords that appear both in the reader's profile and the document's profile (see [7]). Given a set of K documents, and a set of constraints we compute the satisfaction measure of each of the constraints by the documents. The score assigned to a constraint increases as the set of chosen documents comes closer to fully satisfying the constraint. We use the following function to predict the reader's satisfaction from a set of documents: $Eval(constraints, K\ documents) = Par \cdot \sum_i w_i + (1 - Par) \cdot \sum_j c_j \cdot v_j$ where: (i) Par is a number between 0 and 1, indicating the level of importance we give to the reader's preferences. This parameter is determined by the author. (ii) i goes through the K documents in the set considered for the newspaper. (iii) w_i is the normalized relevancy measure of the i'th document according to the reader profile. (iv) j goes through all the constraints. (v) c_j is the weight of the j'th constraint (vi) v_j is in the interval $[0,1]$, according to the level of satisfaction of the j'th constraint.

We asked 31 readers to use the software we developed, with 1225 documents we downloaded from CNN (see *http://www.cnn.com/*). We built reader profiles for the readers by explicitly asking about their interests. By checking the similarity between each pair of readers we verified that the readers were not too similar. Each reader received three different newspapers, one with no constraints, and the other two with two different sets of constraints.

The first set: (i) At least 1 document from the subject "World" and the sub subject "Middle East" should appear. Importance: 6 (ii) No more than 20% of the documents will be from each subject. Importance: 6 (iii) Each document that appears will not be lower than a relevancy of 0.004. Importance: 1 (i) At least one document from each subject shall appear. Importance: 10

The second set: (i) At least 1 document from the subject "World" and the sub subject "Middle East" should appear. Importance: 10 (ii) At least 10% of the documents should be from the subject "US" and the sub subject "News". Importance: 8 (iii) The documents that appear will not be lower than a relevancy of 0.004. Importance: 1 (iv) The documents that appear will not exceed a similarity of 0.7. Importance: 10

For each newspaper, we asked the readers to give each document a score from 1 to 5, where: 1 means "not interesting at all" and 5 means "very interesting". We also asked the reader to answer the following questions about the newspaper in the same manner: (i) How interesting did you find the newspaper? (ii) How varied did you find the newspaper? (iii) How good did you find the newspaper? (according to the way the reader defines "a good newspaper").

The answers of the readers show that the first set of constraints is better than the second set. Our results show that with respect to all the measurements, the first set of constraints is significantly better than the second set and also better than the "empty" set. We also checked which portion of the documents received a score higher than 3 from the reader, i.e. "interesting" or "very interesting". We found that the documents of the newspaper with the good set of constraints received significantly higher scores than the documents of the other two newspapers. Significance was tested using the *one-tailed paired t-test* with $p = 0.05$.

3 Layout Management

The Layout Management (LM) aims at achieving a layout for the selected content, which meets a set of layout requirements imposed by an author-defined style sheet as well as by a *layout profile*, which describes the reader's technical requirements and preferences.

In IMAP, layout is created by the *DesignComposer* (DC), which computes layout for a given XML file, a layout profile and a style sheet (for details, see *http://www.dfki.de/~kroener/DC.htm* and [6]). This software is embedded in a *layout server*, which provides the reader with a portal to the documents selected by the CM. The server uses the DC to create layout in two steps: first, the layout of the portal page is created. Then, when a user tries to access one of the offered documents, a layout for the requested document is created on-the-fly before delivery. If the creation fails, e.g., because of exceeding the generation time bound, the non-optimized XML file (which is in IMAP an XHTML file) is provided as fall-back solution.

The layout profile is created by a so-called *Client Analyzer* (CA), which is implemented as an applet that is submitted by the layout server to the client side. In the first place, the applet collects layout-related data, such as the browser window's dimension, the installed fonts, and the display's DPI. Furthermore, the CA enables the reader to communicate with the layout server and thus to customize layout. Therefore it provides a GUI, which enables several kinds of interaction.

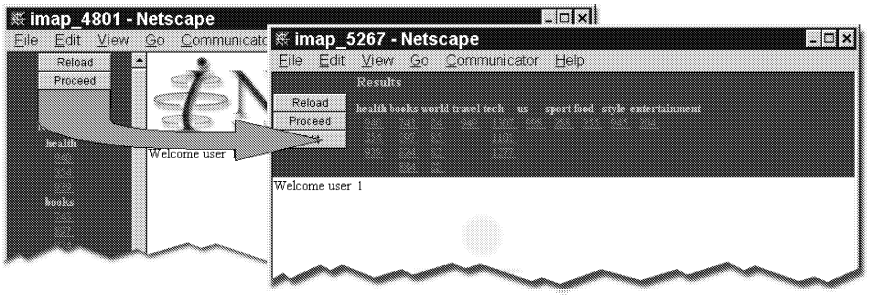

Fig. 1. Two layout alternatives of the portal page; the layout profile is identical

First of all, the reader may submit his profile to the server. In that case, the layout server decides using thresholds (e.g., about changes in the window dimension) if the changes in the current profile require a re-generation of layout. Furthermore, the reader may request a layout alternative. Such a request affects that the server evokes the DC again with the current layout, which in turn starts searching for a new layout for the last requested document (see Fig. 1). The CA allows the reader also to edit parameters included in his profile, which have been prepared by the author of the applied style sheet for customization. This way, the reader may modify style aspects (e.g., colors and font sizes) and/or the content selected for layout, just depending on which modifications the author has granted.

References

1. L. Ardissono, L. Console, and I. Torre: *On the application of personalization techniques to news servers on the WWW.* In Proc. AI*IA 99, 2000.
2. P. Baumgartner and A. Blohm: *Automated deduction techniques for the management of personalized documents.* In Proc. of MKM 2001.
3. A. Borning, R. Lin, and K. Marriott: *Constraint-Based Document Layout for the Web.* In Multimedia Systems **8.3** (2000), 177–189.
4. S. Feiner: *A Grid-based Approach to Automating Display Layout.* In Proc. of the Graphics Interface 1988, 192–197, Morgan Kaufmann, Los Altos, CA.
5. T. Kamba, H. Sakagami, and Y. Koseki: *ANATOGONOMY: a personalized newspaper on the World Wide Web.* In International J. of HC Studies, **46**, 1997, 789–803.
6. A. Kröner: *Adaptive Layout of Dynamic Web Pages.* In DISKI - Dissertationen zur künstlichen Intelligenz **248**, infix (2001), ISBN 3-89838-248-6
7. L. Tsaban: *Intelligent Multimedia Authoring Tools For Electronic Publishing.* Master thesis, Bar-Ilan University, Ramat-Gan, 2001.
8. L. Weitzman and K. Wittenburg: *Automatic Generation of Multimedia Documents Using Relational Grammars.* In Proc. of the ACM Multimedia 1994, 443–451, CA.
9. W.H. Graf: *Constraint-Based Graphical Layout of Multimodal Presentations.* In M.T. Maybury and W. Wahlster, Readings in Intelligent User Interfaces, 1997.