

Strategies for personalizing the access to news servers

Liliana Ardissono, Luca Console, Ilaria Torre

Dip. Informatica, Università di Torino
Corso Svizzera 185 - 10149 Torino (Italy)
Email: {liliana,lconsole,ila.torre}@di.unito.it

Abstract. In this paper we discuss how we applied personalization techniques to the problem of accessing news servers on the internet. We describe the user modeling techniques that we adopted and we show how imposing a (shallow) structure to the database of news allowed us to define strategies for personalizing the detail level of the presentation.

Introduction

The possibility of providing access to news via the web is attracting a growing attention. This is, in fact, an interesting opportunity for all companies operating on the web, not only for those in the communication sector (such as newspapers, radio or television companies). Almost all portals provide access to news (paradigmatic is the recent evolution of the sites of the internet search engines) and also many companies provide specialized news services, usually related to their commercial fields or to the interests of their customers. The primary goal of these services is to attract web surfers and, more important, to ensure that they will visit the site regularly and frequently.

Personalization in the access to news, i.e., personalized forms of information retrieval and filtering, is one of the keys for the success of these services. In fact, a personalized service could allow a user to find easily the pieces of information that (s)he needs or that are of interest for her/him; this is the added value that can lead her/him to visit a site frequently. Indeed, information retrieval is one of the main areas of application for personalization techniques. Simple forms of personalization are implemented in many sites (based on preferences directly expressed by the users), while several more complex approaches have been proposed in the recent literature; e.g., (Billsus & Pazzani 1999; Joachims, Freitag, & Mitchell 1997; Nill, Fink, & Kobsa 1999; Burke, Hammond, & Young 1997; Milosavljevic & Oberlander 1998; Domingue & Scott 1998).

In this paper we discuss an approach to the personalization of news that presents some significant differences with respect to the other approaches in the literature.

The main peculiarity of our approach is that we aim at achieving two different forms of personalization. On the one hand, and similar to the approaches in the literature, we aim at selecting the topics and news that are most relevant to each user (personalized selection of news). On the other hand, we aim at customizing the detail level of the presentation of the news corresponding to each topic (or even of each news item) to the user's preferences (personalized presentation of the content of news). This is a significant difference with respect to most other approaches. Finally, we aim at introducing personalized advertisements in each one of the pages presented to the user. This means that the selection of the advertisements depends on (i) the user's preferences and priorities and (ii) the topic of the news items presented in the page.

The approach we present in this paper exploits user modeling techniques to capture the users' preferences and, in order to achieve the various forms of personalization that we aim at, it manages different types of information (we call them different *dimensions*) in the user model (this topic will be discussed in the third section of the paper).

Moreover, in order to be able to personalize the detail level in the presentation of news, we introduced a (shallow) structure for the database of news (see next section) and we defined a specialized agent for selecting what has to be presented to each user (briefly presented in the fourth section of the paper). Personalized information is presented in a hypertextual format characterized by the fact that the user can always modify the choices of the system, both as regards the news (and topics) selected by the systems, and the detail level of the presentation. This is important for at least two reasons: on the one hand, we believe that a system should never impose its choices to the user, who must have the possibility of accessing every piece of information; on the other hand, the fact that the user can make changes to the presentation decided by the system allows the system itself to learn about the user's preferences. Thus, observing the user's behavior is the basis for refining or revising user models dynamically (this topic is discussed in the section on user modeling).

Structuring news for personalization

In this section we discuss how the goal of personalizing the detail level of the presentation of news influenced the architecture of our system. In fact, this goal requires that the page corresponding to a news item is generated dynamically, making it more or less detailed according to the content of the user model. Clearly, there are different ways of modifying the detail level of some news item. One expensive approach is to store different versions of each news item; alternatively, one could insert annotations in the text of the news item (using some markup language) and then produce different versions interpreting the annotations; for each portion of the news item the annotations could specify a sort of level of detail for that portion or the target (in terms of classes of readers) to which it is directed. A more complex approach is to use techniques for generating automatically summaries from detailed news; see e.g., (Dale *et al.* 1998). While the first approach is not feasible in practice, the last requires the availability of a lot of knowledge on the topic of each news and can be adopted only on very specialized domains. The use of annotations seems to be an interesting option. The approach we introduce takes this idea one step further: we decompose news (which are viewed as complex entities with several attributes) and we impose a (shallow) structure on the database of news.

Most of the approaches in the literature assume that the repository of news is not structured, i.e., that it is simply formed by a set of text files, organized in a taxonomy of categories and, possibly, with indexing for information retrieval. This makes the administration of the repository very simple and, moreover, makes the approaches very flexible since the techniques adopted for personalization (personalized filtering) can be applied to any repository of text files. However, in this way it is difficult to define strategies for personalizing the presentation of a document. Basically, either a document is selected and shown to the user (possibly with some labels specifying the estimated interest level for the user) or it is not.

On the other hand, we aim at structuring news in such a way that different detail levels for the presentation of a news item can be obtained as different aggregations of the components of the item itself. Conceptually, a news item corresponds in our view to a chunk of information, concerning e.g., an event, to be conveyed to a reader. This may include different components: a title, a text (and, possibly, an abstract), pictures, graphs but also commentaries or interviews (e.g., an editorial or an interview to some of the people involved in the event or to some experts on the topic reported by the news item).

Thus we structured the repository of news according to the following principles:

- News are organized in a taxonomy of sections such as politics (with subsections such as internal and foreign politics), sport, economics, technology, etc.; see also

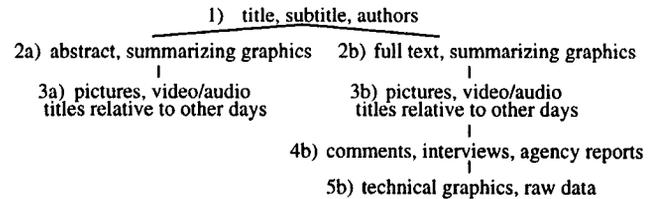


Figure 1: Detail levels in the presentation of news

(Balabanović 1998; Billsus & Pazzani 1999).

- We introduced the concept of “news” as the main structured entity in our database. News have a set of attributes that define their components: title and subtitle; author(s); abstract; text; a set of graphics summarizing the content of the text; photos, video, audio clips; commentaries, interviews, agency reports; raw data and/or detailed charts/graphics, and so forth. Some of the attributes are optional or can be multi-valued (e.g., more than one photo or clip may be associated with a news item); moreover, the same object (e.g., a photo or a clip) may be associated with more than one news item.
- Finally, the database is an historical one, so we can store information concerning several days. In particular, the same news item can be present in the database on different days, possibly with different attributes.

For example, let us consider a news item concerning the financial status of a company *ACME* (the item belongs to the section “economy”). Besides the title and a text, one may have other attributes such as graphs describing the financial status of *ACME*, an interview with the CEO of the company, interviews with financial experts (some of the interviews may possibly have the form of a video or audio clip). Moreover, we can regard “Financial Status of Company *ACME*” as a news item for which different pieces of information may be available on different days (for example, one may have a report on one day and interviews the day after).

This structured organization of news leads to some interesting advantages. The main one is that different detail levels for the presentation of news can be obtained as different aggregations of their attributes. For example, at a low detail level one may include only the title and abstract; a more detailed level may include also the text (possibly replacing the abstract); an even more detailed level may include also graphs or commentaries or interviews.

More specifically, we defined a fixed number of alternative aggregations (called *presentation formats*) and we introduced a partial order between them, which corresponds to an ordering as regards the detail level provided by the aggregations. In our first prototype we considered the formats and ordering shown in figure 1. In other words, we have a tree defining alternative aggregations of attributes; moving from a node to its

descendants correspond to adding detail to the presentation: for example, level 1 can be made more detailed in two different ways, either adding the abstract of a news item or the full text.

The main problem with our approach is that it requires more efforts from the manager of the news server. In fact, filling in the repository requires adding the news items to the database and thus decomposing information according to the attributes of news. If news are stored simply as text files, our approach can still work; however, in this case only the selection of the topics (sections) and news to be presented can be personalized, while the detail level in the presentation of news cannot be varied. The extra work for filling in the database, however, is not very different from the work required by the software systems that are being used in the editorial offices of some newspapers. These systems require that the author of a paper submits her/his work to a specific section and, in case there are photos or extra items, s(he) must specify the paper to which they are related (this is used to define automatically the layout of the pages).

It is worth noting, in conclusion, that the adoption of a structured database is not incompatible with the adoption of other approaches, such as the use of indexing techniques (that can be used to provide a further classification of news items) or the use of annotations in the news item or the automatic generation of summaries. We shall return to this point in the conclusions.

User modeling for news presentation

The selection of the news to be presented and of their detail level relies on the adoption of user modeling techniques. In particular, we distinguish between two phases: (i) the generation of an initial model for a new user and (ii) the refinement of the model.

In both phases we use rather conventional user modeling techniques: stereotypical descriptions of classes of users in the first phase, rules activated by events captured during the user's navigation in the site in the second phase. The stereotypes can provide a first rough prediction on the user's preferences and interests and are activated on the basis of a few data provided by the user in a registration form. The user model initialized by the stereotypes is then refined/reviced by periodically applying a set of rules, which take into account the user's behavior during the navigation of the hypertext (i.e., the fact that (s)he deleted or selected specific sections, or that (s)he changed the detail level proposed by the system for some specific sections or news items). The user models are stored into a database (if the user agrees) for future reference; in this way, the model of a user is refined and evolves across time, becoming more and more accurate as the user connects to the server.

Although the techniques we adopted are conventional, the approach we propose for modeling users is innovative as regards the kind of information we use and as regards the way a user model is structured.

A number of motivations led us to elaborate our approach.

First of all the system we are presenting in this paper is part of a bigger project on adaptive interaction via the web. In the project we are considering different tasks and applications ranging from the access to news servers to electronic commerce and personalized configuration of products/services (see other papers in this workshop). One of the goals of the project is to investigate if it is possible to re-use knowledge for user modeling across multiple tasks and applications.

Second, in each one of the applications, we aim at different forms of personalization; for example, in all cases we distinguish between the personalization of the presentation format and the personalization of the content (information or service to be provided). These different types of personalization require different types of information in the user model.

Another motivation, which is more strictly related to the application we are discussing in this paper, is that the various forms of personalization we aim at in our news server require that different aspects of the user are modeled. For example, the selection of the (sub)sections and news depends on the user's interests and capabilities; the detail level is also related to her/his expertise and receptivity (a parameter we use to evaluate the amount of information that a user can read); finally, the selection of the advertisements that are added to the pages must be related to her/his priorities (life-style).

Thus, we have that multiple forms of personalization have to be supported by the user models in order to achieve the goals we are pursuing. The solution we propose to address this problem is decomposing the user model across multiple (independent) *dimensions*. Let us analyze the specific case of the news server, in which we decomposed the user models into four dimensions: user interest (in the topic of each (sub)section of the news database), expertise (again related to each topic), receptivity (a dimension through which we estimate the amount of information that a user can read), life-style. We have defined a group of stereotypes (and a group of refinement rules) for each one of the dimensions (we shall refer to these groups as *families* of stereotypes or rules); the user is classified independently in each group and then the initial user model is created by composing such independent predictions.

The decomposition allows us to achieve the goals discussed above. On the one hand, it allows us to cope with the complexity of the problem of dealing with multiple features of the user model (a unique set of stereotypes would cause a combinatorial explosion in the number of classes taking into account all combinations of features; consider e.g., the combinations of interest, expertise and life style). On the other hand, some dimensions can be re-used across multiple tasks and applications, especially as regards the stereotypes. For example, we share knowledge about life styles and about receptivity between the application in this pa-

per and adaptive systems for electronic commerce (see (Ardissono *et al.* 1999) and another paper in the proceeding of this workshop).

Let us analyse more precisely the structure of stereotypes and refinement rules (see (Ardissono, Console, & Torre 1999; 2000) for more details). First of all stereotypes are divided into families, each one corresponding to one of the dimensions discussed above. Each stereotype contains two groups of slots: classificatory slots (which correspond to features asked to the user in an initial registration form) and prediction slots. Stereotypes in different families make use of different classificatory data¹ and make different predictions. For example, the stereotypes in the family “interests” use classificatory data such as the user’s age, gender, type and field of job and purpose of the access to the server, to make a first prediction on her/his interest level in each (sub)section of the server. On the other hand, the stereotypes in the class “receptivity” make predictions on the related user feature by exploiting classificatory data such as the user’s education level, job and familiarity in reading Web pages (derived from her/his frequency of access to the Web).

We use a heuristic scheme (partially based on probabilities) for computing a degree of match between stereotypes and the data provided by the user. The slots of a stereotype S have the following form:

$$F: \langle v_1, X_1 \rangle, \dots, \langle v_n, X_n \rangle,$$

that is: for each linguistic value v_i of the feature (slot) F , we consider a weight X_i which is the probability of observing the value v_i for F in the individuals belonging to the stereotypical class S .

The degree of match $p(S)$ between a stereotype S and a user is computed using the classificatory slots and the a-priori probability distribution of the stereotypes, estimated in the population of readers.

As regards the predictive slots, given $p(S)$ and the pairs $\langle v_i, X_i \rangle$, associated with each predictive feature (slot) F , we can associate a probability $p(S) * X_i$ with each linguistic value v_i of F . The predictions of the stereotypes are then merged (and normalized); see (Ardissono, Console, & Torre 1999; 2000) for the details about the approximate reasoning techniques that we adopted.

These predictions constitute the initial model for each new user.

The rules for the revision of the user model are also divided into families. The rules take into account the user’s behavior during the navigation in order to revise/refine her/his model. A number of events is captured by the system during the user’s navigation of the hypertext; for example, the fact that (s)he selects (suppresses) some sections (or news items) that were (were not) included by the system, the fact that (s)he changes the detail level of some news items with re-

spect to the level decided by the system, the number of news items (s)he reads in each section, the advertisements that (s)he selects, etc. Different types of events are relevant for rules belonging to different families. For example, changes in the life style may depend on the advertisements selected by the user, while changes to the interest and expertise in the topic of a (sub)section are related to the fact that the user changes the detail level of the news presented by the system in that (sub)section or that (s)he explores more (sub)sections or news than those presented by the system.

Basically, the rules have the following format: the antecedent is a condition on the events captured by the system (e.g., the fact that in at least 70% of the cases the user increased the detail level for news in a specific section), the consequent is a prediction on some features of the user model: more specifically, on the probability distribution associated with the linguistic values of a feature, e.g., the user’s interest on one of the topics in the news server.

The rules are activated periodically; in the current prototype they are activated at the end of each session so that changes to the user model will be active the next time a user connects to the server. The effect of the activation of the rules is a new prediction to be inserted in the user model. For those features for which the user model already contains some information, the new prediction is merged with the information in the user model (for each feature involved in the prediction and for each linguistic value of that feature, we take the average between the probability predicted by the rule and that in the user model, see again (Ardissono, Console, & Torre 1999; 2000)). This is a conservative choice leading to smooth changes in the user model. Indeed we wanted to avoid abrupt changes even though in some cases this may lead to a slow convergence to the user actual interests, especially for those aspects for which the prediction generated by the stereotypes was not very precise.

Extracting relevant information

In this section we briefly discuss how the information to be presented to a user is selected from the database of news, given the information in the user model; a more detailed discussion can be found in (Ardissono, Console, & Torre 1999; 2000). Basically, three decisions have to be made: (i) which sections and subsections to include, (ii) which news to present in each section and at which detail level and (iii) which advertisements to include in each page presented to the user. In the current prototype we associate the same detail level to all the news in the same (sub)section. This is a first choice to simplify the system; we are investigating how to keep track of the preferences for specific news within a (sub)section; this requires keeping track of the specific topic of each news item (this can be determined, e.g., using indexing techniques, as in (Billsus & Pazzani 1999)).

The selection of the information to be presented relies on a knowledge base formed by three different sets of

¹The knowledge we use in the stereotypes has been partially derived from segmentations of the Italian population (Eurisko).

rules and on a heuristic scoring approach.

A first group of rules exploits the user's interest and expertise (as specified by the user model) to score, for each (sub)section, all the possible detail levels which could be used to present the contents of the (sub)section. We noticed in the section "Structuring news for personalization" that we defined a fixed number of alternative presentation formats, corresponding to different detail levels. Thus, the output of this set of rules is a score for each one of these formats (including a special "empty" format corresponding to the fact that the news item should not be presented to the user), for each (sub)section.

A second group of rules uses information about the user's receptivity and the scores computed by the first group of rules to decide which (sub)sections should be presented and their detail level (i.e., which one of the predefined formats has to be used). The part of the user model concerning receptivity provides an hint on the detail level that is appropriate for the user. For each (sub)section, the detail level is thus decided by balancing (i) the distribution of scores of the various formats and (ii) the information about the user's receptivity. Basically, the (sub)sections are ranked given the scores computed by the first group of rules; we exclude all those (sub)sections for which the empty format has the highest score or the cumulative score of the formats corresponding to low detail (i.e., which include only the title or up to title and abstract) is over a threshold; on the other hand, we include all those (sub)sections for which the cumulative score of the most detailed formats are over a threshold; all the others are ranked according to the distribution of the scores. At this point, the information about the user's receptivity is used to decide how many and which (sub)sections have to be included, how many news items in each subsection and the appropriate detail level for each one of these (sub)sections.

Finally a third group of rules selects the advertisements to be included in each page, based on the user's life style, and the topic of the page. In this way advertisements are tailored to both these aspects.

The result of the selection is passed to another module (agent) which is in charge of generating the hypertextual pages to be sent to the user's browser.

Hypertextual presentation

We chose a simple hypertextual format for the presentation. The most important aspect is that the user has the possibility of changing each one of the choices made by the system, e.g., removing (sub)sections selected by the system, adding other (sub)sections, removing/adding news, changing the detail level.

The hypertext is formed by two types of pages: *index pages* and *news pages*.

An index page is associated with each (sub)section and contains: (i) the list of the (sub)sections selected by the system, if the section is divided into subsections; (ii) the titles of the news selected by the system for that section, otherwise. A "delete" box is associated with

each item of the lists and allows the user to hide the (sub)section/title. The page contains also a button to visualize the (sub)sections/titles that were not selected by the system (this button allows the user to explore these sections/news). Finally, it contains one or more banners for the advertisements selected by the system.

A news page is associated with each news item and contains the pieces of information according to the presentation format selected by the system.

An example of a page generated by the system is reported in figure 2. The page in the figure is an index page for the subsection "Economia Nazionale" ("National Economy"). Three areas (frames) can be recognized: on the left, a list of the titles of the news in the section ("delete" buttons can be used to suppress each item, while a menu — the hotword "Visualizza altre notizie", i.e. "Display other news" — allows the user to explore the news items not selected by the system); when the title of a news item is selected, the news page corresponding to the item is presented in the central frame; advertisements are located in the right frame. In the example the news item is presented with limited detail; however, the attributes that are not displayed (e.g., "commenti/interviste" — "comments/interviews" in the figure) are included as links and the user can select each link to display the corresponding piece of information. On the other hand, "delete" buttons can be used to suppress the pieces of information (attributes) that are displayed.

Conclusions

In the paper we described an approach to the use of personalization techniques for accessing news servers. A main peculiarity of our approach is that we introduced a structure for the database of news. Such a structure allows us to define strategies for personalizing the detail level of the presentation. As we noticed, the structure we imposed is not incompatible with the use of other techniques, especially those exploited in other approaches to (personalized) information filtering. In particular, it could be interesting to couple our approach with indexing techniques or with techniques for annotating news. The former could be used to group news based on the indices, that is at a lower level of granularity than the one provided by the section/subsection hierarchy, and then to select the groups that are more interesting for a user or to vary the detail at the level of these groups. The latter technique could be used to define new presentation formats.

We used rather conventional user modeling techniques. However, we introduced a conceptual decomposition of user modeling along different dimensions. We believe that this approach can be useful for coping with the complexity of the problem and for sharing knowledge across multiple applications. Some of the dimensions we are considering, in fact, may be useful (possibly in combination with other dimensions) in other domains and applications; indeed, we used the same knowledge about life styles also in adaptive electronic commerce



Figure 2: An example of a page presenting the news in section “economy”.

(Ardissono *et al.* 1999). This possibility will be further investigated in future work.

A first prototype of the system has been implemented as a distributed architecture. We recently started a first qualitative test with the aim of evaluating and refining the knowledge bases. The performance we obtained are quite interesting as regards the ability of the system to tailor the presentation to different typologies of users. More extensive and systematic testing will be performed in the next months.

References

- Ardissono, L.; Goy, A.; Meo, R.; Petrone, G.; Console, L.; Lesmo, L.; Simone, C.; and Torasso, P. 1999. A configurable system for the construction of adaptive virtual stores. *World Wide Web* 2(3):143–159.
- Ardissono, L.; Console, L.; and Torre, I. 1999. Exploiting user models for personalizing news presentations. In *Proc. 2nd Workshop on adaptive systems and user modeling on the World Wide Web*, 13–20.
- Ardissono, L.; Console, L.; and Torre, I. 2000. On the application of personalization techniques to news servers on the www. In *Proc. AI*IA 99, Lecture Notes in Computer Science*. Springer Verlag.
- Balabanović, M. 1998. Exploring versus exploiting when learning user models for text recommendation. *User Modeling and User-Adapted Interaction* 8:71–102.
- Billsus, D., and Pazzani, M. 1999. A personal news agent that talks, learns and explains. In *Proc. 3rd Int. Conf. on Autonomous Agents (Agents '99)*, 268–275.
- Burke, R.; Hammond, K.; and Young, B. 1997. The FindMe approach to assisted browsing. *IEEE Expert* 32–39.
- Dale, R.; Green, S.; Milosavljevic, M.; and Paris, C. 1998. Dynamic document delivery: Generating natural language texts on demand. In *Proc. 9th Int. Conf. and Workshop on Database and Expert Systems Applications (DEXA '98)*.
- Domingue, J., and Scott, P. 1998. KMi Planet: a web based news server. In *Asia Pacific Computer Human Interaction Conference (APCHI'98)*.
- Eurisko. Sinottica.
- Joachims, T.; Freitag, D.; and Mitchell, T. 1997. Web-Watcher: a tour guide for the World Wide Web. In *Proc. 15th IJCAI*, 770–775.
- Milosavljevic, M., and Oberlander, J. 1998. Dynamic hypertext catalogues: Helping users to help themselves. In *Proc. the 9th ACM Conference on Hypertext and Hypermedia (HT'98)*.
- Nil, A.; Fink, J.; and Kobsa, A. 1999. Adaptable and adaptive information for all users, including disabled and elderly people. *New review of Hypermedia and Multimedia*.