

Adaptive User Modeling for Filtering Electronic News

Michael Shepherd, Carolyn, Watters, Ajitha T. Marath
Faculty of Computer Science
Dalhousie University
 Halifax, Nova Scotia, Canada B3H 1W5
 {shepherd | waters} @cs.dal.ca
ajitha.t.marath@alcatel.com

Abstract

A prototype system for the fine-grained filtering of news items has been developed and a pilot test has been conducted. The system is based on an adaptive user model that integrates stereotypes and artificial neural networks. The stereotypes are based on newspaper sections and sub-sections, along with editor specified and user specified keywords. Eight subjects trained the system over six days of news papers (986 news items) and then tested the system on a seventh day (171 news items). Five users were simply asked to 'read the news' while three users developed 'corporate' profiles with explicit information needs. The evaluations suggests that such an integrated adaptive user model did, in fact, reflect the difference between the two different types of task. In both cases, the results also reflect the quality of the training of the adaptive neural network by the user in creating the user profile.

1. Introduction

During the past few years, electronic news has drawn a lot of attention not only in the research world but also in the commercial world. There is a huge amount of news available and it is part of the fabric of most people's lives. Readers are not interested in all news items equally on any given day, i.e., some sort of selection or filtering is done by the reader. Although it would seem desirable to personalize the news, there are many challenges that must be met before this is possible. News has with a short shelf-life [20], thus making it extremely challenging to customize content to an individual's interests.

In designing news filtering or personalization systems, we must understand what people expect to gain from "getting the news". Two behavioral theories that have been applied to news reading are *uses and gratification* and *play or ludic* [9]. The research described in this paper integrates two different user models, stereotypes and artificial neural networks (ANN) and evaluates this

integrated model for filtering news for both the *uses and gratification* and *play or ludic* news reading behavior.

The *uses and gratification* theoretical perspective is based on the assumption that the reader has some underlying goal, outside the reading itself, that reading the news satisfies. Such news reading is an example of extrinsically motivated behavior in that there is some reward to be gained by engaging in the activity [7]. That is, "... an assumption that media use, including news reading, serves some ulterior purpose external to the communication behavior itself." [9] As an example, a reader may read the paper specifically to check the performance of the stock market. This perspective implies that optimal content and form can be determined once the particular goal is known. So we expect that a profile of user interest could be used to predict and select items for individual readers based on his or her information goal(s).

Success in "getting the news" is, however, a very vague concept and seldom do any *a priori* queries exist against which results can be measured. Task analysis and profile definition may not be appropriate to behavior that is not specifically goal oriented and that has a large social component. News reading/gathering is more than a task of getting information; it is a task for which the "getting" is part of the reward.

The *ludic*¹ or *play* theory of news reading [19] is an example of intrinsically motivated behavior [7] in that the activity appears to be spontaneously initiated by the person in pursuit of no other goal than the activity itself. This theory asserts that, "... the process of news reading is intrinsically pleasurable, and that intrinsic pleasure is at the root of a mature, orderly, and highly ritualized form of news reading as well as a more casual, spontaneous, and unstructured form of news reading." [9]

The research described in this paper investigates the personalization of news selection through a filtering process. The research does not address the issue of presentation or interface design, rather it deals with the

¹ Throughout this paper we use the term *ludic*, although Stephenson's original term was *ludenic*.

issue of content selection, i.e., the selection of news items that the system determines should be of interest to the user. In order to accommodate the changing interests of the user and the ever-changing news itself, an adaptive user profile representation was developed through the combination of stereotypes and ANN. This user profile was “trained” through direct user feedback for both *ludic* and *uses and gratification* tasks. Section 2 of this paper presents background material on news filtering and on knowledge acquisition and modelling. Section 3 describes the approach used in this research that combines stereotypes and ANN in a single system. Section 4 discusses the results of a pilot user study of this approach and Section 5 summarizes the results.

2. Background

2.1. Filtering News Articles

Electronic news delivery systems have been available for several years. Systems, like Infogate² and Crayon³ retrieve news items using very coarse-grained profiles from one or more designated sites, like *The Wall Street Journal* or *The New York Times*, and present the items in an integrated format. For example, the reader may choose the business section from *The Wall Street Journal* and the arts section from the *The New York Times*.

The creation of fine-grained personalized editions, where individual stories are chosen from a variety of sites for a particular user, is a very different problem. It is virtually impossible to predict what items a reader will read in today’s news based on a history of the items a reader has read over the previous few days [1]. There is also a concern that a very narrowly defined user profile will defeat the social and context function of news by filtering out all news except that identified by the profile. The reader will be exposed to no new items of potential interest and the reader may not receive the information necessary to participate fully as a citizen in the local, national, and international community [2]. Several research projects have focused on fine-grained filtering of news articles. [3, 4, 5, 11, 13, 14, 18]. Results from these studies indicate that fine-grained filtering of news items is very difficult and suggest that personal profiles need to be offset by community interests for ludic news reading behavior.

2.2. Knowledge acquisition and modeling

The process of acquiring the relevant information from the user in the context of a given task or otherwise is

fundamental for any user model to be successful. The knowledge acquisition can be implicit or explicit, long term or short term based, depending on the task domain. The acquired knowledge should be incorporated as needed into the existing user model without causing conflicts and contradictions. In order to achieve this, usually default reasoning and evidential reasoning are used.

Stereotypes are based on default reasoning [8] that allows the modeling process to maintain hypothetical knowledge about a user in the absence of evidence to the contrary. This leads to inconsistencies and the user modeling system must be able to re-examine the assumptions and the conclusions when these inconsistencies appear. Evidential reasoning systems assign probability values to sets of different possibilities, permitting the combination of evidence carried by a number of different sources with no need of any *a priori* distributions [6]. ANN systems exhibit such evidential reasoning behavior in that the activation value at any node is a combination of the input values of the nodes connected to it and the weights on the connecting arcs.

2.2.1. Stereotypes. Stereotypes are based on the premise that the system’s beliefs about the user can be represented in terms of a set of assumptions or default values. The stereotypical knowledge can be organized hierarchically, from general to specific, permitting inheritance among the different stereotypes. A given stereotype inherits the features of all of its ancestors [16, 17].

Usually, the modeling process starts from the root of the hierarchy. If one or more assumptions are verified by certain facts or observations during interaction, the stereotype that contains these assumptions is activated and then added into the current user model. As the interaction progresses, more and more stereotypes are activated and the user model becomes more and more complex.

While the simplicity and ease of use are some of the attractive features of stereotypes, the abilities of this approach are severely limited when reasoning under uncertainties for fine-grained individualization and for feature identification.

² <http://www.infogate.com>

³ <http://www.crayon.net>

2.2.2. Artificial neural networks. Artificial neural networks (ANN) are quite popular for their abilities as pattern classifiers [10]. They are characterized by a set of input nodes representing the possible inputs in the problem domain, a set of output nodes capable of representing an output state selected/identified by the system, and a learning algorithm for collecting and updating the information from the problem domain at the time of training. The learning or self-adapting ability of these systems allows them to store and recognize the input states in an abstract way and thereby generate the appropriate output.

In the context of user modeling, each user can, in principle, have a unique profile characterized by a set of attribute values. As the training progresses, the network state, characterized by link weights, represents the characteristics or attributes of the user or user class, i.e., the profile information.

A well-trained neural network can recognize incomplete input patterns, even in the absence of full knowledge about a user's attributes and assumptions. This is done in a much simpler way in these systems than the conventional stereotype-based systems.

3. Integrating Stereotypes and ANN

The adaptive user modeling system developed in this research combines stereotypes and ANN to model individual user profiles. Stereotypes provide a means for the development and presentation of the individual user profile information. A user profile based on stereotypes consists of those stereotypes that are of interest. With a stereotype-based user profile system, the issues to be addressed include accommodation of changing user interests, i.e., 'adaptability' and 'quantification' of level of interest for each news item. Artificial neural networks can be used in this context to overcome the bottleneck of identifying the weighting factors. An artificial neural network that encapsulates stereotypical knowledge is used in an adaptive user profile implementation that reflects the changing needs and interests of the user in reading news articles.

If stereotypes are not used, then the input layer of the ANN would consist of a typically large vector of words extracted from the text of the news items and this input layer would be fully connected to the next layer of the ANN. The use of stereotypes simplifies the design of the ANN and reduces the processing load at run time [12].

3.1. Stereotype Component

Categorical information about a news item provides the basis for defining stereotypes in the domain of electronic news. For example, a news item can be categorized as a 'sports' item based on its contents, while based on the

coverage, it can be further classified as 'international' or 'national' and so on. While a sports person is very much interested in sports, a business person may not be very interested in all sports items but still may be interested in 'local' or 'national' sports. The newspaper used in this research is *The Halifax Herald Ltd.* and it has 17 major categories and 79 sub-categories. Using these as the basis for the stereotype hierarchy, there are 79 sub-stereotypes.

The other information available for each news item that is made part of stereotype knowledge is the set of keywords that best describe the news item content. For each news item and each user, there are editor-specified keywords and user-specified keywords. The user-specified keywords are collected from each subject (user) at the time of training data collection. The incorporation of user-specified keywords makes the stereotype more powerful in representing the tastes of each individual user. As each news item is in one of the 17 major categories, these sets of keywords are associated with that category and thus with that stereotype. Similarly, the editor-specified keywords are associated with the major categories.

Figure 1 is a screen dump of the training interface. In this day's paper, there are only 9 of the 17 base categories or stereotypes. The article selected for review by the user is, "Market meltdown has many victims." It is in the base category of Business, sub-category Canada. The editor assigned keywords are "stock, market, summary". The subject is requested to evaluate this article as either interesting or not interesting. If the article is not evaluated, the default value is not interesting. The subject is also request to add keywords from the article itself to be associated with the base category. The article is viewed by scrolling the bottom window.

3.2. Adaptive Neural Net Component

The time-dependent evolving nature of the user profiles requires the system to be 'adaptable'. The varying component ('user-specific keywords') within each stereotype itself leads to variation of influencing capacity for each stereotype in an individual's profile. The neural network based 'adapting component' enables the system to accommodate the 'changing needs' of the user over time and also offers a solution to the 'quantification problem' due to its ability to derive conclusions even with an imperfect and partial input information.

The neural network simulator used for this implementation is an adapted version of the PDP++ Software Version 1.2 released by CNBC group in CMU [15]. The learning algorithm used for self-adaptation in the current system is the Backpropagation algorithm for supervised learning meant for multi-layer feed-forward neural network architectures.

The resulting ANN consists of three layers of simple processing units – the input layer, one hidden layer and

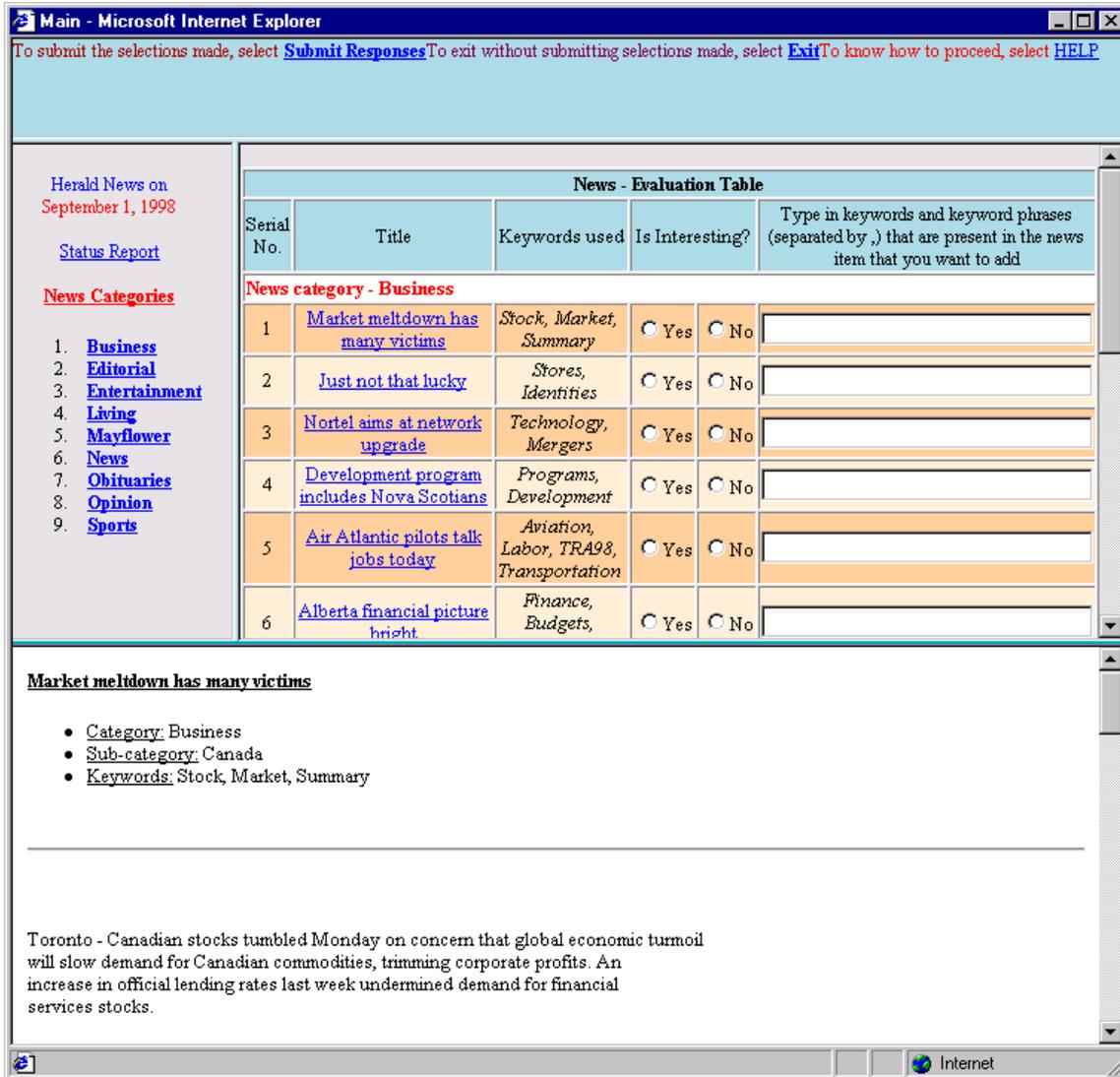


Figure 1. Training interface.

an output layer. The input layer consists of 114 nodes representing the 79 sub-categories within the 17 major categories, two nodes for each of the 17 major categories representing the presence or absence of user-specified keywords and editor-specified keywords within each previously viewed news item, and one node representing whether the user had found the article to be of interest. The hidden layer maps the categorical information of the various news items; the layer consists of 17 nodes with each node having links from only those nodes in the input layer that represent sub-categories of that category. The input layer and hidden layer together represent the category and sub-category information as well as keyword related information. The output layer has a single node with two possible states indicating 'interesting' or 'Not

interesting' output states. The hidden layer is fully connected to the output layer.

The ANN representing the profile for each of the users was trained by that user on 986 news items in *The Halifax Herald Ltd.* during a 6 day period.

4. Results

Each of 8 subjects in the pilot study was presented with news items from one day following the six training days. This test set consisted of 171 news items. As in the case of the training data collection, the feedback collected was in the form of 'interesting' or 'not interesting' (Figure 2).



Figure 2. Testing interface

The input pattern for each news item in the test set was created automatically using stereotype extracting routines. Figure 2 is a screen dump of the interface for the testing of the system and collection of the user evaluations. In this instance, the day's newspaper has 11 of the 17 base categories or stereotypes.

Five subjects (1 through 5) were asked to train the system (create profiles) based on *ludic* behavior, i.e., the task was to simply "read the news". Three subjects (11, 12 and 13) participated in the study with *uses and gratification* behavior. These subjects were asked to provide the responses, for both the training phase and the test phase, representing three different corporate interest profiles; E-commerce, unemployment and job

opportunities in Canada, and North American stock market, respectively.

During training, user feedback is binary; either the news item is of interest or it is not of interest. Similarly, during testing, the user evaluates a news item as being either of interest or not of interest. The ANN system, however, does not operate in binary mode. The activation state of the output layer of the ANN is in the range of 0 to 1. For the system to determine if an input news item is of interest or not, a threshold must be established to activate the output layer. In this research, the threshold value of 0.5 was chosen arbitrarily. An optimum threshold value based on high recall (rather than high precision) was also determined for each subject by examination of the activation values for the news items in the test phase.

4.1. Results of ludic scenario

Table 1 summarizes the results for the five subjects browsing in a ludic fashion, showing results for both thresholds of 0.5 and optimum threshold values. The system is represented by the letter ‘S’ and the user by the letter ‘U’. Tables 1 and 2 have the following information:

- **output activation states** that were the thresholds for filtering. For each subject there are two sets of data given, one for the threshold level > 0.5 and one for the optimal threshold level as determined by viewing the results.
- S_{yes} represents the set of news items that the system determined would be of interest to the user based on the activation state at the output layer being greater than the given threshold for these items.
- U_{yes} represents the set of news items that the user determined actually were of interest.
- $S_{yes} \cap U_{yes}$ represents the set of news items that both the system and the user determined to be of interest.
- **Recall** the proportion of news items that the user determined to be of interest that the system also determined to be of interest.
- **Precision** the proportion of news items that the system determined to be of interest that the user also determined to be of interest.

As stated previously, Subjects 1 through 5 represent general readers of news, with no specific *information need*. Among them, training data sets from subjects 1 and 3 were characterized by a richness of *user specified*

keywords pertaining to the news items that are of interest to them. As Table 1 shows, the system was picking up 31% and 53% of the news items *that are interesting to the user* in the case of subject 1 and 3, respectively, with a threshold of > 0.5 for the activation state. The analysis also shows that there is an optimal threshold value for the activation state that substantially improves the recall of the news items of interest to the subject (100% or close of interesting news items to the subject). However, by lowering the activation threshold to the optimum threshold, the precision of the system is also lowered. The analysis of the data showed that as the keywords specified by the user during the training phase become more and more unique for the individual interesting items, the number of not interesting items picked up by the system for the subject decreases.

When we do the same type of analysis for subjects 2 and 4 we can see that with a threshold of > 0.5 the system fails to pick up *interesting* items for the subject. The performance of the system was poor for both subjects. This can be explained partly on two facts: (1) the number of *user specified keywords* supplied by the subject as part of training set data was very low in the case of these subjects. As a result, many of the news items, some of them interesting to the subject and others not interesting, were represented by a single input state, leading to inadequate representation of input information. (2) real-life news reading is not necessarily driven by an *information need* for the subject; the behavior of the subject may not be predictable or explainable. Subject 2,

Table 1. Summary for ludic behavior

| Output activation state | S_{yes} | U_{yes} | $S_{yes} \cap U_{yes}$ | Recall | Precision |
|-------------------------|-----------|-----------|------------------------|--------|-----------|
| Subject 1 | | | | | |
| > 0.5 | 21 | 45 | 14 | 0.31 | 0.67 |
| > 0.0175 | 132 | 45 | 45 | 1.00 | 0.34 |
| Subject 2 | | | | | |
| > 0.5 | 3 | 78 | 1 | 0.01 | 0.33 |
| > 0.0155 | 47 | 78 | 5 | 0.06 | 0.11 |
| Subject 3 | | | | | |
| > 0.5 | 26 | 19 | 10 | 0.53 | 0.39 |
| > 0.032 | 74 | 19 | 15 | 0.79 | 0.20 |
| Subject 4 | | | | | |
| > 0.5 | 0 | 41 | 0 | 0.00 | 0.00 |
| > 0.015 | 66 | 41 | 10 | 0.24 | 0.15 |
| Subject 5 | | | | | |
| > 0.5 | 14 | 27 | 4 | 0.15 | 0.29 |
| > 0.015 | 171 | 27 | 27 | 1.00 | 0.16 |

for example, chose a large number of news items from different areas in the test data set as *interesting* (typical ludic behavior). Even after lowering the threshold for the output node activation state, the system was not able to improve the performance substantially for these two subjects.

4.2 Results of uses and gratification scenario

Table 2 summarizes the results for those subjects (11-13) representing uses and gratification behavior, i.e., the corporate profiles. Note that there were far fewer news articles considered to be of interest in this scenario than in the ludic scenario. Among these subjects, training data sets from subjects 12 and 13 were characterized by richness of keywords pertaining to the news items that are of interest to them. As Table 2 shows, the system identified 67% of the news items that were of interest to the subject in the case of both subject 12 and 13, with a threshold of > 0.5 for the activation state. The analysis also shows that there is an optimal but lower threshold value (less confidence) for the output node activation state so that the system is capable of returning all of the items of interest. However, for both activation thresholds, precision is quite low

5. Conclusions

Based on the pilot study evaluation of the prototype system, the following inferences can be made in the context of prototype and the user model based on the integration of stereotypes and ANN.

Adaptive user modeling based on this integrated approach may be promising especially for readers with specific *information need*. In the case of ‘casual’ readers, the system may not exhibit a consistent behavior; this is consistent with the behavioral theories of news reading discussed above. In either case, the system can perform reasonably well in filtering out items not of interest to the subject.

Like in any other neural network based system, identification of an orthogonal parametric space to represent the input patterns is very important for the success of the system. In the case of electronic news this becomes quite challenging because of the ‘always new’ and changing nature of the ‘news’; in the current system, an attempt is made to define the parametric space with user keywords and categorical information. As a result the uniqueness and completeness of the keywords become very important in determining the accuracy and completeness of the input patterns.

For improved filtering and increased confidence level (activation threshold), the information content of individual patterns representing the news items should be increased. This can be achieved by providing more accurate and more complete information related to the keyword contents of individual items to the system.

The incorporation of stereotype information within the neural network architecture is promising, and certainly reduced the complexity of the design of the ANN and the processing load at run time. More information can be added to the stereotype by providing a set of scale points for providing the *level of interest* that may provide more information about the input pattern to the neural network. Collecting the information related to the extent of interest for each news item in the training data set (instead of collecting Boolean state values *interesting or not*

Table 2. Summary for uses and gratification behavior

| Output activation state | S_{yes} | U_{yes} | $S_{yes} \cap U_{yes}$ | Recall | Precision |
|-------------------------|-----------|-----------|------------------------|--------|-----------|
| Subject 11 | | | | | |
| > 0.5 | 0 | 2 | 0 | 0.00 | 0.00 |
| > 0.0045 | 94 | 2 | 2 | 1.00 | 0.02 |
| Subject 12 | | | | | |
| > 0.5 | 25 | 3 | 2 | 0.67 | 0.08 |
| > 0.0128 | 54 | 3 | 3 | 1.00 | 0.06 |
| Subject 13 | | | | | |
| > 0.5 | 4 | 3 | 2 | 0.67 | 0.50 |
| > 0.008 | 10 | 3 | 3 | 1.00 | 0.30 |

interesting) using a nominal scale with scale points ranging from ‘very interesting’ to ‘not at all interesting’ can provide more accurate information to the system about the level of interest of the subject in the individual items and thereby improve the knowledge of the system about the user’s interests. The nominal scale also provides a means for assigning *weighting factors* to the various keywords present in each interesting news items thereby improving the performance system based on this approach.

In the context of keyword information collection from the subject, a more transparent way of keyword collection may provide more consistent and accurate information than collecting it explicitly from the user. For example, we could extract the keywords automatically instead of requesting the subject to identify them at the time of training the system. The information regarding the *level of interest* collected from the subject could be mapped to a weighting factor for the keywords present in the current item, thereby providing information related to priorities among the entire set of keywords for a subject in the profile. At the time of performance evaluation, the keywords from the input news item could be extracted. The vector consisting of the weighted keywords that are part of the user profile can be applied against the keyword set for the current item and calculate the *relevance* or *level of interest* of the current item for the current subject based on the keywords present. This *relevance factor* can contribute to the input state representing the current item for the trained system with profile information. This also accommodates the temporal evolution of the user’s interests to some extent.

In summary, this pilot study suggests that an integrated adaptive user model, based on stereotypes and ANN, does reflect the differences between tasks of a ludic nature and tasks of a uses and gratification nature. This inference is drawn on the substantially lower numbers of items determined to be of interest in the uses and gratification task by both the system and the user. Also, the performance of the system directly reflects the quality of the information provided by the user during the training phase; those profiles with more user-specified keywords did better during the test phase. In any event, the fine-grained filtering of news items is still an open research question, although it may be easier to satisfy tasks of a uses and gratification nature than of a ludic nature.

6. References

- [1] R.B. Allen, “User models: theory, method, and practice”, *Int. J. of Man-Machine Studies*, 1990, 32: 511-543.
- [2] K. Asp. “Mass media as molders of opinion and suppliers of information”. In C. Wilhoit & C. Whitney (eds.), *Mass Communication Review Yearbook*, 2, (pp. 332-354). Sage Publications: Beverly Hills. 1981.
- [3] W. Bende. (1994/1995). “Read all about it in the Daily You”. *Communicating Business*, Forward Publishing: London. Winter, 1994/1995
- [4] D. Billsus and M.J. Pazzani. “A Hybrid User Model for News Classification”, In Kay J. (ed.), *UM99 User Modeling Proceedings of the Seventh International Conference*, Springer-Verlag, Wien, New York, 1999, pp. 99-108.
- [5] P.R. Chesnais, M.J. Mucklo and J.A. Sheena. “The Fishwrap personalized news system”, MIT Media Laboratory. In *Proceedings of the 1995 IEEE Second International Workshop on Community Networking Integrating Multimedia*. 1995.
(<http://fishwrap-docs.www.media.mit.edu/docs/dev/CNGLue/cnglue.html>).
- [6] Fabio Cuzzolin and Ruggero Frezza. “An evidential Reasoning Framework for Object Tracking”. <http://www.dei.unipd.it/~cuzzolin/Spie/SPIE.html>
- [7] E.L. Deci and R.M. Ryan. The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, 1987, 53: 1024-1037.
- [8] J. Doyle. “A Truth Maintenance System”, *Artificial Intelligence*, 1979, 12: 231-272
- [9] Dozier D. & Rice R., “Rival theories of electronic newsreading”, in R. Rice (ed.), *The New Media*, Sage Publications: London. 1984, pp. 103-128.
- [10] R. Hecht-Nielsen. *Neurocomputing*, Addison-Wesley Publishing Company, 1989
- [11] T. Kamba, K. Bharat and M.C. Albers. “The Krakatoa Chronicle - An interactive, personalized, newspaper on the Web”. In *Proceedings of the Fourth International World Wide Web Conference*. 1995.
(<http://www.w3.org/pub/Conferences/WWW4/Papers/93/>)
- [12] Bill Kules. “User Modeling for Adaptive and Adaptable Software Systems”, 2000.
<http://www.otal.umd.edu/UUGuide/wmk/>
- [13] K. McGillivray. *Personalizing Electronic News through User Modelling*. Honours Thesis. Dept of Mathematics, Statistics, and Computer Science, Dalhousie University, Halifax, Canada, 1995.
- [14] M. Ohkubo, N. Kobayashi and T. Nakagawa. “Design of an information skimming space”, In *Proceedings of ACM Multimedia '93*. 1993. 365-371.
- [15] The PDP++ software, 1998.
<http://www.cnbc.cmu.edu/PDP++/PDP++.html>
- [16] E. Rich. “User Modeling via Stereotypes”, *Cognitive Science*, 1979, 3: 329-354.

- [17] E. Rich. "Users are individuals: individualizing user models", *Int. J. Man-Machine Studies*, 1983, 18: 199-214.
- [18] M. Shepherd, J.F. Duffy, C. Watters and Nitin Gule. 2001. "The Role of User Profiles in News Filtering", *Journal of the American Society for Information Science and Technology*, 2001, 52(2): 149-160.
- [19] W. Stephenson. *The Play Theory of Mass Communication*. Chicago: The University of Chicago Press: Chicago. 1967.
- [20] J. van Cuilenburg. "The information society: some trends and implications", *European Journal of Communications*, 1987, 2: 105-121.