

# Evaluating Scrutable Adaptive Hypertext

Marek Czarkowski

Department of Computer Science,  
University of Sydney, Australia  
marek@cs.usyd.edu.au

**Abstract.** Adaptive hypertext systems personalise documents to meet the individual's particular preferences, knowledge and goals. There is a debate over how much control should be given to the user as well as how much transparency there should be to the inner workings of the system. Some adaptive systems make the user model available to the user. We propose transparency and control should extend beyond this by involving the user in the personalisation process and granting them power to change it. Our previous evaluations of scrutable systems have revealed users have difficulty understanding and controlling personalisation. We have developed SASY with a focus on improving scrutability support tools. This paper describes our design for the evaluation of SASY.

## 1 Introduction

Adaptive hypertext systems personalise documents to meet the individual's particular preferences, knowledge and goals. This offers benefits including an improved user experience, efficiency in information retrieval and navigation support [3].

There is a growing debate within the field of adaptive systems as to how much control should be given to the user and how much transparency there should be to the inner workings of the system [7, 10]. There is an argument that users must understand, to some degree, the workings of an adaptive system in order to trust it to perform tasks on their behalf [10]. In addition, users must feel as though they have ultimate control over the system if and when they choose to exercise it. There is also literature that shows that systems that expose the user model to the student promote learner reflection and enhance learning [2, 11, 12]. In contrast, an evaluation of a tool to recommend relevant conference events [8] found users do not always exercise control.

We propose providing transparency and control should extend beyond making the user model available to the user by involving the user in the personalisation process and granting them power to change it. Our motivation for increasing transparency and control over personalisation is based on several key drivers:

- Increasing legal requirements for access to personal data [9].
- In non-critical applications, such as movie recommendation systems, people trust machines to personalise information. The user is not too concerned if the personal-

isation is faulty, for example, recommends movies the user has no interest in. However, for more critical tasks the user would want more transparency and traceability, for example, a system that invests a user's money,

- Empowering users to correct the misconceptions a system holds about them which impacts personalisation.
- Support a user's sense of curiosity and exploration of the personalisation.
- Allow users to develop comfort in understanding how the system works [10].

We have developed SASY, a personalised system that allows users to scrutinise, or inspect, the personalisation to understand why it occurred and how to control it. The design was informed by the development and evaluation outcomes of the Tutor series of systems which achieved some success in supporting scrutability [5,6]. From the evaluation of Tutor3, we concluded that although users seem comfortable with the notion of personalisation driven by a simple user model, being able to control the personalisation is foreign. Furthermore, despite the compelling reasons for scrutinising personalisation, we have found users are not typically willing to do so. A key concern was that some users had difficulty finding the scrutability tools when needed.

We had two main goals in mind when developing SASY. Firstly, we wanted to test scrutability support in a more genuinely adaptive environment. In Tutor3 the user model was entirely driven by answers to profile questions and was otherwise static. SASY is adaptive rather than adaptable as the user model is updated behind the scenes based on observations about the user. The user model contains attributes that are inferred by the system and these can also be changed by the user. Secondly, the interface design of SASY addressed the issue that some users of Tutor3 failed to find and activate the scrutability support tools.

This paper introduces SASY and describes our evaluation that is in progress.

## 2 Overview of SASY

SASY provides a generic framework for the delivery of personalised content over the World Wide Web. SASY is a web application that presents personalised content to end-users through a standard web browser. Authors create content by publishing a set of XML documents that conform to a schema called Adaptive Tutorial Mark-up Language (ATML). ATML (and hence SASY) is domain independent, but has additional features which simplify the creation teaching material. An ATML document is an HTML 4.0 document with additional tags that allow the author to define personalisation rules for adaptive content.

SASY builds a profile of each user's interests, background, goals, etc. from answers the user provides to a brief questionnaire. The profile is also populated with beliefs that are inferred about the user as they use the system. For example, accessing a certain page may cause SASY to update the user profile to capture interest in the subject matter. When a user requests to view a page, SASY evaluates personalisation rules embedded in the ATML page against their user profile. Thus, SASY determines

what adaptive content should be included in the user's view of the page and what should be omitted. This is a standard Adaptive Hypertext technique [3], similar to the approach employed by AHA! [4].

What makes SASY unique is the built-in tools for scrutinisation that allow users to see how content is personalised and change the personalisation should they choose to do so. Each page includes a summary of the personalisation that was performed. The user can also access a detailed explanation of the personalisation. In this mode, SASY highlights content that was included or removed from the user's personalised view of a page. It also indicates the user profile attributes and their values that caused content to be included or removed. The user can not only view and update their user profile, but also view an explanation of how each profile attribute was set. This is particularly useful for user profile attributes that have been inferred by SASY through observations about the user.

As part of our evaluation of SASY, we have developed scrutable, personalised content for a number of different domains: UNIX Security course, Holiday Planner, Television Guide and a Museum Guide. We describe the first three in the evaluation design.

### 3 The User's view of SASY

To use SASY the user logs in and selects from a list of available topics. Each topic is effectively a separate application, with its own set of content. On selecting a topic, SASY shows the profile page. The first time it is accessed it displays a questionnaire that is used to build a profile of the user's interests, background, goals, etc. For example, the Holiday Planner profile asks potential vacationers whether they seek Adventure or Relaxation. Otherwise, it shows the questionnaire and any other profile attributes that SASY has inferred about the user. The user can click a hyperlink to pop-up an explanation of how and why each profile attribute was set. For inferred attributes, this tells the user what event caused SASY to create this belief about them.

Following the profile page, the user is taken to the Home page. Figure 1 shows the Home page from the personalised Holiday Planner. This is a typical personalised page where adaptive content has been included and/or removed, depending on whether the condition to show the adaptive content matched the user's profile.

Every page contains a panel on the right hand side titled "Personalisation", from which the user invokes tools to scrutinise the personalisation. It includes labels to indicate the number of adaptive content items have been removed and included to create the user's personalised view of the page. For example, the page in figure 1 has had five items removed and one item included. Clicking the hyperlink "Click to highlight removed/included items on this page" reloads the same page but this time shows the raw view of the page without any personalisation applied (Figure 2). Content that was removed through personalisation is highlighted in yellow, included content is highlighted in grey. For example, in Figure 2, the content starting with the title "Wine Country Touring" is highlighted in yellow because it was removed. The page instructs the user to hold their mouse over a highlighted item to see an explanation of why the item was included or removed. In Figure 2, the user had moved their mouse

over the “Wine Country Touring” section to pop-up the text “Not shown to you because your profile says: You’re seeking adventure”. That is, the content was removed by the personalisation because the user’s profile states they are seeking adventure.

Below the “Click to highlight removed/included items on this page” link is a list of user profile attributes that affected personalisation on the page. In Figure 1, the personalisation considered the user is single, seeking adventure and have a low budget. Clicking on a user profile attribute, e.g. the hyperlink “You are single”, pops up a window that explains why SASY holds this belief about the user.

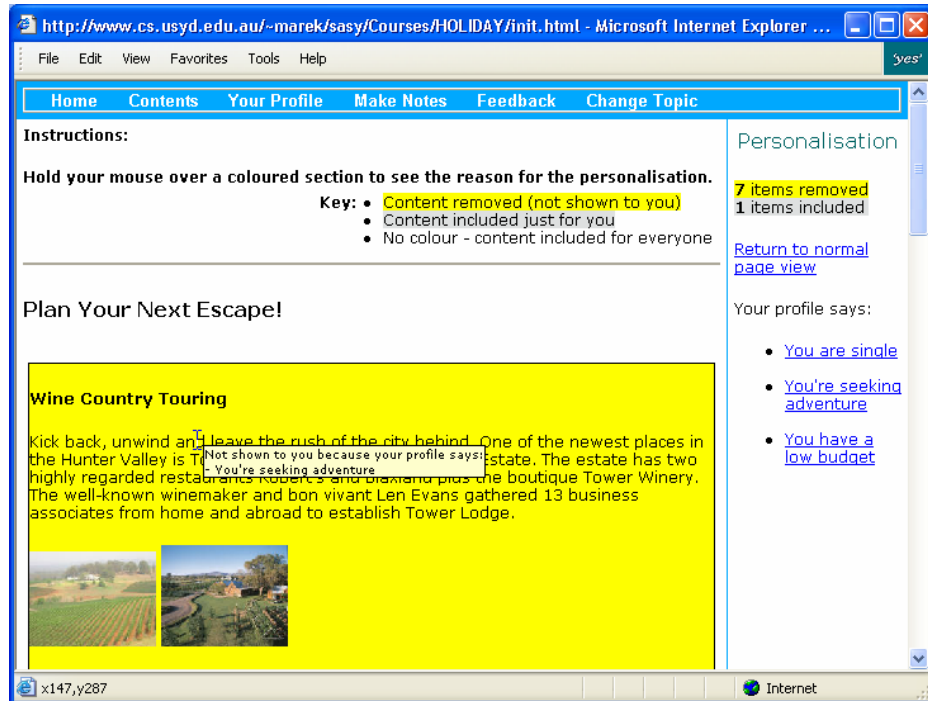


**Fig. 1.** A typical personalised page in SASY. SASY removed and included adaptive content items to create view of the page.

## 4. Evaluation Design

We wish to evaluate SASY to determine whether users are able to:

- Understand personalisation is driven by their user profile that may be updated by SASY through their interaction with the system.
- Scrutinise the personalisation on a page and understand why content was included or removed to create their personalised view.
- Demonstrate control over the personalisation by changing values in their user profile to change the personalisation.



**Fig. 2.** Typical personalised page in SASY in highlight mode, showing all the content available on the page and explaining why content was included or removed to create the personalised view (Figure 1). Content that was removed is highlighted in yellow, included content is highlighted in grey.

The difficulty in measuring how effectively users scrutinise and control personalisation is that we know from evaluations of the Tutor systems [5, 6] users will not scrutinise often. This is understandable since scrutinisation is not the user's main purpose for using the system. In Tutor evaluations, participants noted the default personalisation seemed appropriate hence they were not motivated to scrutinise.

To counter this difficulty, we base our evaluation of SASY around the most common scenarios where the user might be motivated to scrutinise:

- A user believes the personalisation to be faulty because it produces unexpected results.
- A content author wishes to debug the adaptive content they have created.
- A user is curious as to what the system believes about them or how a page was personalised and wants to explore alternatives.

The evaluation comprises of three separate experiments across different subject matter domains to reduce the effect of a particular domain on the results. Since the personalisation engine is the same in each experiment, results can be directly compared. Each evaluation captures both quantitative results (time to complete task, task

correctness) and qualitative feedback to gauge user satisfaction and capture any concerns raised by users.

## 5. Evaluation 1 – Personalised TV Guide

This evaluation is inspired by the article “My Tivo thinks I’m Gay” [1]. It describes a situation where Tivo, an adaptive movie recommendation system, consistently recommends movies with a homosexual theme, much to the dismay of its owner who is not interested in this genre. The owner has no way to directly correct the beliefs Tivo holds about them.

This laboratory based experiment models this scenario with a Personalised TV Guide system, where the user has a need to understand the personalisation process and be able to exercise control over it. Unlike Tivo, SASY supports the user in doing this.

### 5.2 Aim

To measure how effectively SASY supports the user to:

- Scrutinise a page to determine why adaptive content is included/removed in relation to their user profile.
- Explain how/why a belief held by the system was instantiated. In this case the belief is inferred by the system through the user’s interaction with the system.
- Demonstrate control over the personalisation by altering their profile to change how content is included and removed.

Additionally we wish to evaluate:

- The effect of the displaying user profile attributes that affected personalisation on each page. We suspect it discourages the use of the content highlighting feature as reading the profile attributes is a quick form of scrutinisation.
- The effect of user training in the form of reading through an online overview.

### 5.3 Participants

Fifteen randomly selected participants, a small number based on Nielson [13]. It is assumed that all participants have basic familiarity with web browser based applications and interfaces. Users are informed their identity is anonymous throughout the evaluation.

## 5.4 Method

Participants randomly divided into groups which have different system variants:

- Group 1 – by default the system will display user profile attributes that affected personalisation in the right hand column of the page.
- Group 2 – by default the system will not display user profile attributes.
- Group 3 – same as Group 1 but are given a system introduction to read as training.

All groups have access to online help documentation. During the evaluation user actions are logged to allow us to measure task performance and accuracy. Participants perform tasks described in a worksheet that is presented in an online application separate to SASY. The online worksheet presents one task at a time and logs the time each task is started and completed. The worksheet asks participants to:

- Access the TV Guide and complete a questionnaire to determine their interest in television program genres: Sports, Business & Finance and Current Affairs. SASY builds a user profile and displays a recommended TV viewing schedule.
- Inspect and change the personalisation such that their viewing schedule only includes current affairs programs.
- Use the system to read about any programs they are interested in. The system is crafted so that regardless of the user’s selection, SASY will set an attribute in their profile stating they are a member of “Special Interest Group 1”, which will cause religious programs to be included in the viewing schedule. This is intentionally obscure so that profile attributes displayed on the page will not directly indicate an interest in religious programs.
- Review the viewing schedule to explain why it now includes religious programs and change the personalisation to return the schedule to its previous state. Participants will need to use the system feature that highlights personalised content and shows the user profile attributes that caused content to be included or removed.
- Provide qualitative feedback to evaluate user satisfaction and usability. Participants will select answers based on a Likert scale.

## 5.5 Expected Results

We expect participants will be able to scrutinise the TV viewing schedule and change their user profile to control how content is included/removed by the personalisation.

We expect users will rely heavily on the user profile attributes shown in the personalisation panel but will be also able to use the highlight feature to complete tasks.

## **6. Evaluation 2 – Personalised Holiday Planner**

In real world applications of adaptive systems, the content author is required to ascertain whether the content is correct and ready for publication. This evaluation asks the participant to assume the role of an editor, and validate that the personalisation in a Holiday Planning system is correct. This task is conceptually different from those in Evaluation 1 in that the user must be able to answer ‘how was this page personalised to me’ as well as ‘how would this page be personalised to someone else’.

### **6.2 Aim**

Determine whether participants can understand complex personalisation. In Evaluation 1, content is either removed or included based on a single user profile attribute. Here we use complex personalisation rules involving two or more attributes.

### **6.3 Participants**

Ten randomly selected participants. It is assumed that all participants have basic familiarity with web browser based applications and interfaces.

### **6.4 Method**

Participants randomly divided into two groups:

- Group 1 – not provided with an introductory system guide.
- Group 2 – given a system introduction to read as training.

The Holiday Planner presents a personalised holiday recommendation based on the user’s vacation style preference (adventure, relaxation, family oriented), budget (low, high) and status (single, couple, family). Participants are asked to validate the personalisation is correct in the Holiday Planner. Unknown to participants, the system has the following errors:

- Simple error: a user with a status of ‘single’ is shown children’s holidays.
- Complex error: a user with a status of ‘couple’ and a ‘low’ budget is shown expensive holidays.

All participants have access to online help and user actions are logged to determine task completion time and correctness.

### **6.5 Expected Results**

We expect that users will quickly find the simple errors but may struggle with the complex ones.

## 7. Evaluation 3 - Personalised UNIX Security Course

Adaptive systems are commonly used as a means of providing personalised instruction. In this class of systems, the user is focused on learning and may not be interested in how the personalisation works unless they perceive value from scrutinising. For example, to see what content they have missed out on or to correct a misconception the system holds about them. In this evaluation, we allow the users to freely use the system to learn UNIX security concepts. However, we have planted adaptive content in the teaching material that would annoy most learners and we wish to evaluate whether we can provoke users to scrutinise content and change default personalisation to remove the annoying content.

### 7.2 Aim

- Determine whether learners can be provoked to scrutinise content to remove annoying material that is distracting to their learning focus.
- Is there a relationship between scrutinisation and the students learning outcomes?

### 7.3 Participants

Computer Science Students from the University of Sydney, Australia. Part of the syllabus of several subjects requires students learn about UNIX Security. Over one hundred students are expected to use the UNIX Security course.

### 7.4 Method

The students are all asked to read a system introductory guide that explains the scrutinisation feature of the system.

Students all sit a pre-test to access their pre-knowledge of UNIX security concepts. Students are then allowed free use of the system and will sit a post test at the end of their session to capture knowledge gains and provide qualitative feedback regarding their use of the scrutinisation tools. The course content includes annoying content in the form of jokes about the UNIX operating system.

### 7.5 Expected Results

We expect that users will scrutinise the system to remove the annoying content. This will test the notion that users always trust the default personalisation.

Users may also be curious about other aspects of the personalisation. We expect some users will further scrutinise the system out of curiosity or to reflect on their learning.

## 8. Conclusion

We have found that providing an effective user interface to support the process of scrutinising personalisation is very challenging. Although users seem comfortable with the notion of personalisation, being able to control the personalisation is new and not expected. However, based on user feedback from previous evaluations, users would like to be able to scrutinise an adaptive system. Building effective tools for this is challenging as we must support casual users who will not scrutinise often.

We have described the design for our evaluation of SASY that is currently in progress.

## References

1. "Oh no! My TiVo thinks I'm gay", Jeffrey Zaslow. The Wall Street Journal, Dec. 04, 2002.
2. Bull, S., Brna, P. and Dimitrova, V. (Eds.) (2003). Proceedings of Learner Modelling for Reflection Workshop, Volume V of the AIED2003 Supplementary Proceedings, 193-298.
3. Brusilovsky, P. (2001) Adaptive hypermedia. User Modeling and User Adapted Interaction, Ten Year Anniversary Issue (Alfred Kobsa, ed.) 11 (1/2), 87-110
4. Calvi L., Cristea, A. (2002). Towards Generic Adaptive Systems: Analysis of a Case Study. De Bra, P., Brusilovsky, P., Conejo R. (Eds.), Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Malaga, Spain, May 2002, Springer:LNCS 2347, 79-89.
5. Czarkowski, M., Kay, J. (2000). Bringing scrutability to adaptive hypertext teaching. ITS'2000, Intelligent Tutoring Systems, Gauthier, G., Frasson, C., VanLehn, K. (Eds.), Intelligent Tutoring Systems, Springer, 423-432.
6. Czarkowski, M., Kay, J. (2003). How to give the user a sense of control over the personalization of adaptive hypertext? Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, User Modeling 2003 Session, 121-132.
7. Höök, K., Karlgren, J., Waern, A., Dahlbäck, N., Jansson, C.J., Karlgren, K., and Lemaire, B. (1996). A Glass Box Approach to Adaptive Hypermedia. Journal of User Modeling and User Adapted Interaction, special issue on Adaptive Hypermedia, 6(2-3), 157-184.
8. Jameson, A., Schwarzkopf, E. (2003). Pros and Cons of Controllability: An Empirical Study, De Bra, P, P Brusilovsky, R Conejo (Eds.), Proceedings of AH'2002, Adaptive Hypermedia and Adaptive Web-Based Systems, Springer, 193-202.
9. Kobsa, A. (2002). Personalized hypermedia and international privacy. Communications of the ACM archive, V 45 , Issue 5 (May 2002), Special Issue: The adaptive web, 64-67.
10. Maes, P., Schneiderman, B. (1997). Direct Manipulation vs. Interface Agents: A Debate. Interactions, Volume IV Number 6, November 1997, ACM Press, 42-61.
11. Mitrovic, A. & Martin, B. (2002). Evaluating the Effects of Open Student Models on Learning, in P. DeBra, P. Brusilovsky & R. Conejo (eds), Proceedings of Adaptive Hypermedia and Adaptive Web-Based Systems, Springer Verlag, Berlin Heidelberg, 296-305
12. Kay, J. (2001). Learner control. User Modeling and User-Adapted Interaction, Tenth Anniversary Special Issue, 11(1-2), Kluwer, 111-127.
13. Nielsen, J. (1994). Estimating the number of subjects needed for a thinking aloud test. International Journal of Human-Computer Studies, 41(1-6), 385-397.