A PROCESS MINING APPROACH TO ANALYSE USER BEHAVIOUR

Laura Măruşter, Niels R. Faber, René J. Jorna and Rob J. F. van Haren
Faculty of Economics and Business, University of Groningen, PO Box 800, Groningen, The Netherlands

Keywords: Modelling user behaviour, decision support systems, process mining, web personalization.

Abstract: Designing and personalising systems for specific user groups encompasses a lot of effort with respect to analysing and understanding user behaviour. The goal of our paper is to provide a new methodology for determining navigational patterns of behaviour of specific user groups. We consider agricultural users as a specific user group, during the usage of a decision support system supporting cultivar selection - OPTIRas^{TM}. Combining process mining techniques with insights from decision making theories, we provide a method of analysing logs resulted from usage of decision support systems. For instance, farmers show difficulties in fulfilling the goal of OPTIRas, while other agricultural users seems to manage better. The results of our analysis can be used to support the redesign and personalization of decision support systems.

1 INTRODUCTION

As the on-line services and Web-based information systems proliferate in all domains of activities, it becomes increasingly important to model user behaviour and personalization, so that these systems will appropriately address user characteristics. Particular topics are addressed by research in human-computer interaction (HCI) and user-system interaction (USI), such as the discovering of user behaviour or navigation styles (Herder and Juvina, 2005; Menasalvas et al., 2003; Balajinah and Raghavan, 2001), developing metrics involved in modelling and assessing web navigation (Juvina and Herder, 2005; Herder, 2002; Spiliopoulou and Pohle, 2001), cognitive models for improving the redesign of information systems (Bollini, 2003; Ernst et al., 2005; Lee and Lee, 2003). Various methods have been developed to model web navigation in case of generic users (see for instance (Mobasher, 2006)).

Furthermore, it becomes increasingly important to address specific user groups (Song and Shepperd, 2006). By investigating navigational patterns of these groups, the (re)design of the systems used by specific user groups can be made more effective. Although various methods have been developed to model user behaviour of generic users, no research specifically targeted, as far as we know, the navigational patterns of agricultural user groups. Our contribution particularly focuses on agricultural users as a specific user group. By analysing agricultural users' patterns of behaviour, we aim to support the redesign of web-based information systems, illustrated in particular for the redesign of the decision support system OPTIRas^{TM}.

The goal of this work is to illustrate a new methodology of analysing user behaviour using process mining techniques (van der Aalst and Weijters, 2004) and insights from decision making theories. By considering agricultural users as a specific user group, we investigate agricultural user's patterns of behaviour during the use of a web-based IT system, namely a decision support systems called OPTIRas^{TM}, that aids farmers in their cultivar selection activities (AGROBIOKON, 2006). The results of the analysis provide recommendations concerning the redesign of the decision support system's website in order to address specific agricultural users' characteristics. Therefore, in our analysis we use insights from decision making theories.

The organisation of our paper is as follow. In Section 2 we provide an introduction into decision making, decision support system OPTIRas^{TM}, and data collection issues. In Section 3 we determine agricultural user's patterns of behaviour with process mining techniques. The conclusions and implications for redesign are presented in Section 4.

Măruşter L., R. Faber N., J. Jorna R. and J. F. van Haren R. (2008).
A PROCESS MINING APPROACH TO ANALYSE USER BEHAVIOUR.
In *Proceedings of the Fourth International Conference on Web Information Systems and Technologies*, pages 208-214
DOI: 10.5220/0001526002080214

2 AGRICULTURAL USERS AND DECISION MAKING

2.1 Decision Making and Decision Support Systems

A well-known decision making model was formulated by Simon (Simon, 1977). He explained the human decision process using three phases, namely intelligence, design, and choice. In the *intelligence phase*, an individual explores the issue about which he is making a decision, and determines relevant issues. Subsequently, the individual formulates one or more alternatives solutions to the recognised (sub)-decisions in the *design phase*. Eventually, a final solution is formed in the *choice phase*. In this final phase of the decision process, the partial solutions are evaluated against criteria of the outcome that have to be met. Partial solutions that best meet these criteria are selected. The selected partial solutions are combined into the overall decision. Because information that is used in the three phases is not always complete, the phases do not linearly follow upon each other. Commonly, the decision process is characterised by many iterations in which additional information about the issue is collected (*intelligence*) or more alternatives are explored (*design*). Mintzberg refers to a similar trichotomy: identification, development, and selection (Mintzberg et al., 1976). Specific decision making models have been developed for different kinds of users. For instance, in case of farmers, decision making models have been developed by (Johnson et al., 1961; Öhlmér et al., 1998) and (Fountas et al., 2006).

Whenever these models are linear (sequential) or iterative, they all show a common structure: first, information is collected about the problem at hand, second, different candidate solutions are formed, and finally, a choice is made. It is evident that a decision support system has to address in one way or another the basic phases of the decision making process. In the following, we investigate in which way the decision support system OPTIRas™ has addressed these phases.

2.2 OPTIRas™

A decision support system (DSS) is perceived as a computer system that aids people in making a decision regarding a specific domain (Klein and Methlie, 1995). This aid is provided by the DSS by connecting to the human decision process (Turban and Aronson, 2001).

OPTIRas™ system was designed to target low-yielding potato farmers, attempting to realise an increase in the yield of farmers. OPTIRas™ is a decision support system for cultivar selection that supports a farmer in selecting cultivars relative to cultivar characteristics. A variety of properties, relating to yield, resistance against pests and diseases, and storage, characterise a potato cultivar. Potato Cyst Nematodes (PCN), a severe potato disease, causes each year an average loss of 100-150 euros per hectare per year in the North East region of The Netherlands. This is about 10-15% of the net revenues of cropping (before taxation and investment calculations). Reducing the PCN infestation level to economic acceptable levels requires the right combination of sampling data, cultivar, growth frequency, field choice, and nematicide usage. The goal of the decision support system is to ensure that farmer gains insight into PCN damage levels and the financial consequences of selected cultivar and the application of pesticides.

The interface of OPTIRas™ DSS consists of seven main pages: *Field ID*, *PCN History*, *Reorder Cultivars*, *Yield Information*, *Crop Rotation*, *Report* and *Option*. Moreover, there are also pages offering *Help*, *Details* about a specific cultivar, but we do not use them in our analysis.

Web pages do not have the same function, and this depends on the purpose of the web site. In our case, the site's goal is to provide the farmer with information about different yield scenarios, given the chosen factors. Therefore, the following types of pages can be distinguished: (i) the pages reflecting the site's goal, e.g. containing information about the yield, (ii) the pages that can lead to fulfil the site's goal, and (iii) the other pages.

In literature, this categorization is called *concept hierarchies*, or *service-based hierarchies* (Spiliopoulou and Pohle, 2001). Concept hierarchies are used in market-basket analysis or to study the segmentation of companies clients. Spiliopoulou and Pohle developed a service-based concept hierarchy to determine the success of a web site, which distinguishes between action pages and target pages (Spiliopoulou and Pohle, 2001). According to their definition, "an *action page* is a page whose invocation indicates that the user is pursuing the site's goal. A *target page* is a page whose invocation indicates that the user has achieved the site's goal".

As we stated in the previous section, a DSS system should support the user in all phases of the decision making process. For instance, in terms of Simon's decision making phases, 'Action' pages would correspond to the *Intelligence* phase, where an individual explores the issues about which he is making a decision, while 'Target' pages would correspond to the *Design* phase, where candidate solutions are de-

Table 1: Categorization of OPTIRas^{TM} pages.

'Action' pages	'Target' pages	Other pages
Field ID	*Yield information*	*Details*
PCN history	*Crop rotations*	*Options*
Reorder cultivars	*Report*	*Help*
		Logon
		Help
		Save

veloped.

We categorize OPTIRas^{TM} pages as presented in Table 1. In the following, we consider that whenever a user has visited one of the target pages (*Yield information*, *Crop rotations*, *Report*), the user has reached the goal of the web site.

Based on Simon's decision making phases, we proceed with the following mapping:

1. 'Action' pages corresponds to the *Intelligence* phase: the user is supported to explore the issue (the selection of the cultivar), by specifying Field characteristics and the PCN history;
2. 'Target' pages corresponds to the *Design* phase: the system provides the user with one or more possible solutions;
3. the *Choice* phase does not have any correspondence in the DSS; the final decision is actually taken by the user outside of the system.

Assessing whether the DSS is "really" supporting the targeted users, or in other words, whether the site goal is fulfilled, we expect that both decision making phases, *Intelligence* and *Design*, should be performed, to the same extend. This translates into visitations of both 'Action' and 'Target' pages, to a comparable extend. In Section 3 we will investigate how the navigational patterns look like, e.g. how 'Action' and 'Target' types of pages are visited.

2.3 Data Collection

According to Herder and Juvina (Herder and Juvina, 2005), user's navigation behaviour can be modelled using *syntactic information* ("e.g. which links are followed, what does the navigation graph look like, what is the time that users spent on each page"), *semantic information* ("i.e. what is the meaning of the information that the user encountered during navigation"), and *pragmatic information* ("i.e. what is the user using the information for, what are the user's goals and tasks"). We choose to record per session *syntactic information*, e.g. information concerning the movement from one 'source' page to a 'destination' page (time stamp, name of source page, and name of destination page, see Table 2).

Table 2: An example of the log file.

Time stamp	FromPage	ToPage
2004-12-22 22:13:29	field	pcn
2004-12-22 22:13:35	pcn	order
2004-12-22 22:14:00	order	yield
2004-12-22 22:14:26	yield	crop
2004-12-22 22:16:16	crop	yield
2004-12-22 22:16:25	yield	crop
2004-12-22 22:16:53	crop	details
2004-12-22 22:17:54	details	allcultivars

Table 2 shows an example of an OPTIRas^{TM} sequence of navigation, where the user sequentially accesses 7 types of pages. The log file (Table 2) provides us *syntactic information*, namely information needed to build navigation graphs (see Section 3).

Analysing users as a homogeneous group is not a suitable basis for re-designing decision support systems that target heterogeneous end users. OPTIRas^{TM} system was designed for farmers. However, a larger group of agricultural users is actually using OPTIRas^{TM}, that can be differentiated into three target groups, namely farmers (called also growers[1]), extension workers, and scientists. Extension workers are people with a degree in agricultural sciences, who help growers in their practical work. Scientists also use OPTIRas^{TM}, but rather for experimental purposes.

Agricultural users can login in OPTIRas^{TM} registering with their e-mails, but they can register also with an anonymous ID. The system provides this opportunity for the sake of user-friendliness, but also to protect users' privacy. The anonymous ID takes into account the IP number. In this way advantages for the users are provided, but it implies disadvantages for the analysis. A user may login the first time with his/her e-mail, the second time anonymously using computer A, and the third time again anonymously using computer B; in the analysis, these three sessions will count as sessions belonging to three distinct users, and we cannot see anymore how a user eventually changes his behaviour in time. Moreover, users can login by specifying their role, i.e. grower, scientist, extension worker, or other. OPTIRas^{TM} is expected to be used in particular within two peak periods (i) November/December and February (period for purchasing crops, expected to be the peak period) and (ii) April and August/September (period when

[1] In this article, the terms 'farmer' and 'grower' are used interchangeably.

the yield is actually obtained). Farmers may compare the obtained yield with the advice given earlier by OPTIRas™ .

We have inspected only the log files starting from the 18th of December 2004 (the date when OPTIRas™ became on-line) and stopped at with the 30th of May 2006. Thus we have two peak periods: (i) December 2004 - February 2005 and (ii) November 2005 - February 2006. In total, 763 user sessions belonging to 501 individual users (see discussion above about distinct users) have been logged and included in the analyses.

3 MINING NAVIGATIONAL BEHAVIOUR WITH PROCESS MINING TECHNIQUES

Developing navigation patterns and performing web personalization is often done by using data mining techniques such as Clustering, Classification, Association Rule Discovery, Sequential pattern Discovery, Markov models, hidden (latent) variable models, and hybrid models (Mobasher, 2006; Herder, 2002; Juvina and Herder, 2005; Chang et al., 2006). Mostly of these techniques assume web navigation as a sequential process. However, the decision making process exhibits a distributed nature (Holsapple, 1988), which should be captured by the web-mining technique.

Process mining techniques allow for extracting information from event logs and can be used to discover models describing processes, organisations, and products (van der Aalst et al., 2003). Because these techniques are able to model parallel events, monitor deviations (e.g., comparing the observed events with predefined models or business rules), and rely on robust mathematical formalism, they are a good candidate for providing insights into navigational behaviour for decision making processes. However, the choice of the most appropriate process mining algorithm for a certain problem is not straightforward. A suitable algorithm is the heuristic mining, implemented in the ProM framework. ProM offers a wide variety of process mining techniques (available at www.processmining.org) (ProM, 2007). We choose heuristic mining because for its robustness for noise and exceptions (Weijters and Aalst, 2003). The heuristic mining is based on the frequency of patterns, therefore it is possible to focus on the main behaviour in the event log.

We provide insights into two kinds of navigational patterns: (i) navigation patterns of all agricultural users show, and (ii) navigational patterns of differentiated target group (e.g. growers, extension workers, scientists).

3.1 Mining the Navigational Pattern of All Users

This section concentrates on the navigational patterns exhibited by all users. In Figure 1 the result of applying the ProM heuristic mining plug-in is shown. The focus is to grasp the patterns of navigation from one page to another (we use the FromPage field of the log file, see Table 2). The file consists on 763 user sessions, belonging to 501 individual users. The rectangles refer to transitions or activities, in our case to page names, such as *Field*, *Order*, etc. There are two special activity names, ArtificialStartTask and ArtificialEndTask that refer to a generic start or end page. The term 'complete' refers to a finished action (in terms of a finite-state machine, an activity can be in the states 'new', 'sent', 'active', 'suspended', 'complete', etc.). The number inside the rectangle shows how many times a page has been invoked. The arcs between two activities A and B are associated with two numbers [2]: (i) a number between 0 and 1, which represents the causality/dependency measure between A and B, and (ii) an integer number that represents the frequency of occurrences of A and B next to each other. For instance in Figure 1, *Field* page is directly followed 553 times by the *Pcn* page, with a dependency measure of 0.742, and 374 times by *ArtificialEndTask*, with a dependency measure of 0.996.

The complicated picture from Figure 1 can be interpreted as follows. We can determine "the most common pattern" which consists of *ArtificialStartTask* → *Field* (758), *Field* → *Pcn* (553), *Pcn* → *Yield* (530), and *Yield* → *ArtificialEndTask* (221). The sequence $Field \rightarrow Pcn \rightarrow Yield$ is actually the "prescribed" order of pages in OPTIRas™ , e.g. the order in which pages are presented in OPTIRas™ .

We notice the reversed link *Yield* → *Pcn*, which may suggest that users are changing PCN values to observe the impact on calculating the yield (the higher the values of PCN, the smaller the yield). Noticeable is also the high frequent self-recurrent link to $Yield$(655).

The interpretation of these findings is that farmers in general use the prescribed sequence of page invocation. The directed graphs from Figure 1 reveal that invocation of 'Action' pages are predominating. In Figure 1, the sum of incoming arcs from 'Action' pages (Field, Pcn, Order) to activity ArtificialEnd-

[2]For more details, see (Weijters et al., 2006).

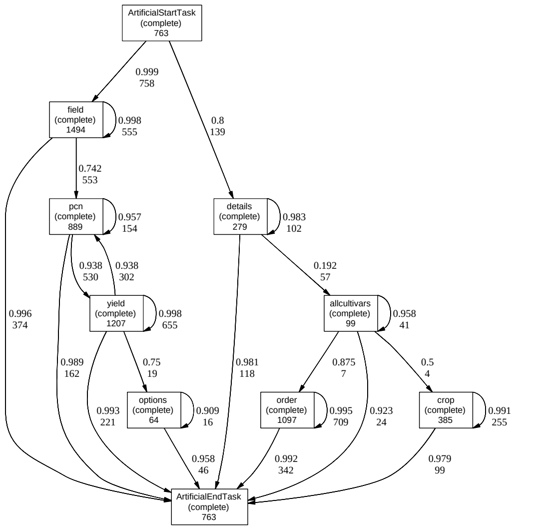


Figure 1: The mined behaviour corresponding to all agricultural users.

Task is 374+162+342=878. This value is much larger than the sum of incoming arcs from 'Target' pages (Yield, Crop) to the activity ArtificialEndTask, which is 221+99=320.

The first conclusion is that pages from the category 'Target' are visualized significantly less than 'Action' pages. This is a disappointing result, given the fact that the main goal of OPTIRas^{TM} is to provide detailed information about yield. Second, after each page a significant number of sessions stop (in Figure 1, 374 after page *Field*, 162 after page *Pcn*). However, we also see that whenever users do not 'give up' in early stages, and visualize 'Target' pages, they show a relevant activity. They recall repeatedly the *Yield* page, and they revisit pages which are not directly linked in the prescribed order (the reverse link *Yield* → *Pcn*). This fact illustrates that whenever users reach a 'Target' page, their interest may rise. The *Details* page is visited unexpectedly often: 279 times, and even forms a path containing only one page (e.g. the path *ArtificialStartTask*, *Details* and *ArtificialEndTask* (see Figure 1). This suggests that users are interested in information about different cultivars.

3.2 Mining the Navigational Patterns of differentiated Target Groups

In this section we investigate the patterns of behaviour of different target groups. These target groups are growers, extension workers, and scientists. Whenever a user logs in into OPTIRas^{TM}, he/she has to specify his/her role, e.g. growers, extension worker, scientist or others. The distribution of sessions is as following: from all 763 sessions, 535 are sessions belonging to growers, 59 are sessions coming from extension workers and 72 are results of scientists usage. The rest of 97 sessions which are not labelled, are left out from the analysis.

The run of the heuristic algorithm (see Figure 2a.) depicting the navigation pattern of farmers provide more insights into the most generic behaviour: there seems to be differences between all agricultural users (Figure 1) and farmers (Figure 2a.) For instance, in case of farmers, the most used path is constituted by *Field* and *Pcn* pages, without the *Yield* page. This fact could be interpreted as the majority of farmers do not reach the *Yield* page, and thus we may doubt whether the DSS fulfils its goal. Also, the navigational pattern of farmers does not contain loops (see in Figure 2a. the loop between *Pcn* and *Yield*), except from self-loops.

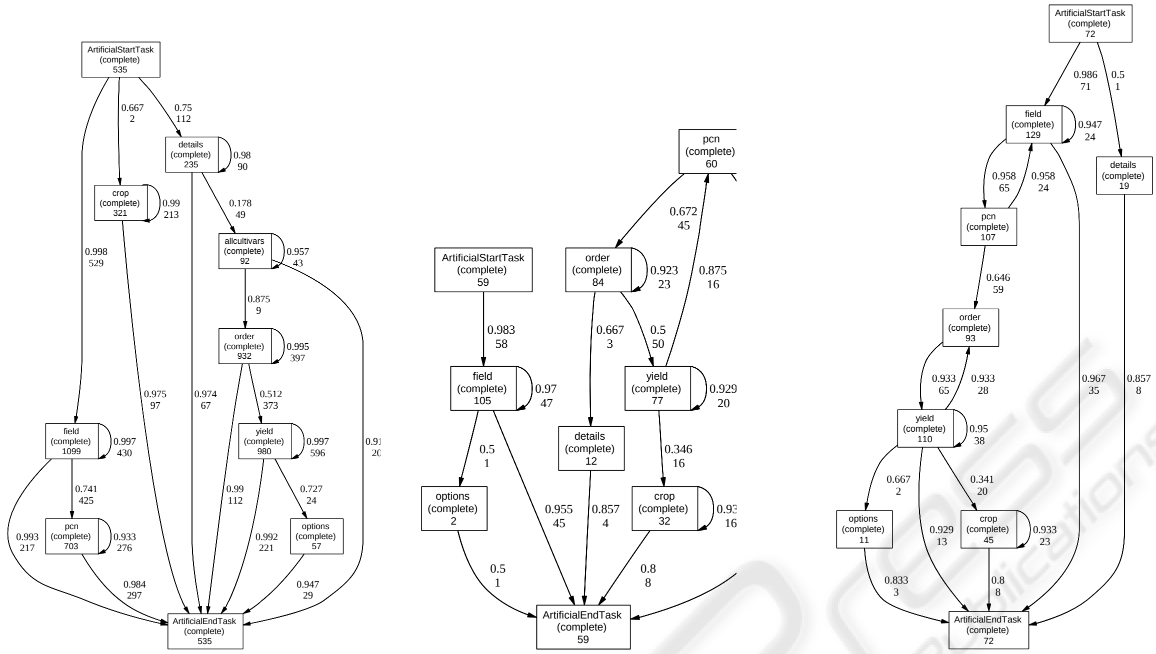
Inspecting further the navigation patterns of the extension workers, we see in Figure 2b. that the most common behaviour is to stop after visiting the *Field* page. The other path starting with the *Pcn* page is less easy to be interpreted (without any incoming arc); however, we can note the loop between *Pcn*, *Order* and *Yield*, which illustrates a rather complex behaviour.

The behaviour of scientists is the most complex (see Figure 2c.): there are two loops, e.g. (i) *Field* and *Pcn* and (ii) *Order* and *Yield*. This illustrates the fact that scientists, instead of only following the prescribed behaviour, are checking the effects of changing value parameters. Also, most of users of this group visit 'Target' pages (*Yield* and *Crop*), which may suggest that they like to explore DSS possibilities.

4 CONCLUSIONS

In this article we provided a new methodology for determining navigational behaviour of agricultural users, by combining process mining techniques and insights from decision making theories. To our knowledge, this combination, together with applying the approach of process mining in case of decision support systems supporting farming activities are new. This methodology can be used to redesign the decision support system, by addressing the characteristics of agricultural users.

We analysed logs resulting from a decision support system called OPTIRas^{TM}. With respect to all agricultural users, they use the prescribed order of pages, e.g., *Field*, *Pcn*, *Order*, *Yield*. We discovered that often users visualise just the first page, and then



a. growers b. extension workers c. scientists

Figure 2: The mined behaviour of farmers (a.), extension workers (b.) and scientists (c.).

leave the system. With respect to goal fulfilling, the invocation of 'Action' pages predominates, while the invocation of 'Target' pages have a smaller proportion than expected. This corroborates with our assumption that users of OPTIRas™ spend time in the *Intelligence* phase of the decision making process, exploring the alternatives in the domain of cultivar selection. This shows that users do not realise fast enough what the goal of OPTIRas™ is, and what kind of advantages they may have when they use this system. This leads to a lack of interest, which implies that users stop using the system. The fact that often the *Details* page is visited in the beginning of farmer's sessions suggests that in some cases, users are more interested in information about different cultivars, rather than cultivar selection based on Yield or Crop figures.

By analysing logs of different target groups, we found that the most common behaviour of farmers (growers) shows difficulties in fulfilling the goal of OPTIRas™, e.g. visiting 'Target' pages, while the other two target groups, extension workers and scientists seems to manage better. Given the fact that OPTIRas™ was especially developed to support farmers (growers), it is a clear sign that redesign actions are needed.

This methodology can be used to support the redesign of DSSs in order to address specific agricultural user's characteristics. First, OPTIRas™ should better support the decision making process, i.e. by letting users gather information about the various topics involved in cultivar selection, instead of presenting itself as an instrument to make optimal choices. Second, we suggest that in the very first page of OPTIRas™, a user should be confronted with the goal of the system. This should be very explicitly and clearly stated, perhaps illustrated with a short example. Third, depending on the target group, hints for the following steps should be given: e.g. in case of growers, what path should be followed to get the yield overview. For this, the use of a sitemap is regarded to be appropriate.

In future research we plan to investigate in more detail the navigation behaviour of different farmer groups, e.g. top-farmers, quality, quantity and normal farmers (for farmer categories, see (Faber et al., 2006)), especially to relate their navigational patterns with their preferred learning styles. Information about the behavioural patterns of various agricultural user groups may enhance the successful design and use of DSSs. Second, we plan to let these users participate in the redesigning process and to make OPTIRas™ more interactive. Third, we intend to incorporate in our analysis also semantic and pragmatic information.

As a final remark, we would like to state that, although the present research focused on decision support systems and agricultural users, the methodology presented in this paper can be used in case of any information system enhanced with logging functionality, and with groups of users that, based on some criterion, can be differentiated into subgroups.

ACKNOWLEDGEMENTS

We thank Iain Milne from Scottish Crop Research Institute for enabling the logging function of OPTIRas^{TM}. AGROBIOKON is supported by the Northern Netherlands Provinces (SNN), the Product Board for Arable Products (HPA), and AVEBE.

REFERENCES

AGROBIOKON (2006). Optiras - rassenkeuze adviessysteem voor zetmeelaardappelen. http:// optiras.agrobiokon.eu/.

Balajinah, B. and Raghavan, S. (2001). Intrusion detection through learning behavior model. *Computer Communication*, 24:1202–1212.

Bollini, L. (2003). Web interface design based on cognitive maps: Generative dynamics in information architecture. The 6th Generative Art Conference - GA2003.

Chang, S. E., Changchien, S. W., and Huang, R.-H. (2006). Assessing users' product-specific knowledge for personalization in electronic commerce. *Expert systems with Applications*, 30:682–693.

Ernst, N. A., Story, M.-A., and Allen, P. (2005). Cognitive support for ontology modeling. *International Journal of Human-computer Studies*, 62(5):553–577.

Faber, N., Jorna, R., van Haren, R., and Maruster, L. (2006). Knowledge and knoweldge use for sustainable innovation: the case of starch potato production. In *Proceedings of ISMICK'06*, Stellenbosch, South Africa.

Fountas, S., Wulfsohn, D., Blackmore, B., Jacobsen, H., and Pederson, S. (2006). A model of decision-making and informations flows for infomation-intensive agriculture. *Agricultural Systems*, 87:192–210.

Herder, E. (2002). Metrics for the adaptation of site structure. In *Proc. of the German Workshop on Adaptivity and User Modeling in Interactive Systems ABIS'02, Hannover*, pages 22–26.

Herder, E. and Juvina, I. (2005). Discovery of individual user navigation styles. In *Proc. of the Workshop on Individual Differences - Adaptive Hypermedia, Eindhoven 2005*.

Holsapple, C., W. A. (1988). *Distributed Decision Making: A Research Agenda*. ACM SIGOIS Bullentin.

Johnson, G., Halter, A., Jensen, H., and Thomas, D. (1961). *A Study of Managerial Processes of Midwestern Farmers*. Iowa State Press, Ames, Iowa.

Juvina, I. and Herder, E. (2005). Individual differences and behavioral aspects involved in modeling web navigation. In *Proc. of User-centered interaction paradigms*, volume LNCS 3196, pages 77–95. Springer.

Klein, M. and Methlie, L. (1995). *Knowledge-based decision support systems: with applications in business*. John Wiley & Sons, Chichester.

Lee, K. and Lee, S. (2003). A cognitive map simulation approach to adjusting the design factors of the electronic commerce web sites. *Expert Systems with Applications*, 24:1–11.

Menasalvas, E., Millan, S., Perez, M., Hochsztain, E., Robles, V., Marban, O., a, J. P., and Tasistro, A. (2003). Beyond user clicks: an algoritm and an agent-based arhitecture to discover user behavior. In *Proceedings of ECML/PKDD*.

Mintzberg, H., Raisingham, D., and Thorêt, A. (1976). The structure of 'unstructured' decision process. *Administrative Science Quarterly*, 21:246–275.

Mobasher, B. (2006). *The Adaptive Web: Methods and Strategies of Web Personalization*, chapter Data Mining for Personalization. Berlin: Springer-Verlag.

Öhlmér, B., Olson, K., and Brehmer, B. (1998). Understanding farmers' decision making process and improving managerial assistance. *Agricultural Economics*, 18:273–290.

ProM (2007). The ProM framework. http:// www.processmining.org.

Simon, H. (1977). *The new science of management decision*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, revised edition.

Song, Q. and Shepperd, M. (2006). Mining web browsing patterns for e-commerce. *Computers in Industry*, 57:622–630.

Spiliopoulou, M. and Pohle, C. (2001). Data Mining to Measure and Improve the Success of web Sites. *Data Mining and Knowledge Discovery*, 5(1-2):85–114.

Turban, E. and Aronson, J. (2001). *Decision Support Systems and Intelligent Systems.* Prentice-Hall, Upper Saddle River, N.J.

van der Aalst, W., van Dongen, B., Herbst, J., Maruster, L., Schimm, G., and Weijters, A. (2003). Workflow Mining: A Survey of Issues and Approaches. *Data and Knowledge Engineering*, 47(2):237–267.

van der Aalst, W. and Weijters, A. (2004). *Process Mining*, volume 53 of *Special Issue in Computers in Industry*. Elsevier Science Publishers,Amsterdam.

Weijters, A. and Aalst, W. (2003). Rediscovering workflow models from event-based data using Little Thumb. *Integrated Computer-Aided Engineering*, 10:151–162.

Weijters, A., van der Aalst, W., and de Medeiros, A. A. (2006). Process mining with the heuristics miner-algorithm. BETA Working Paper Series WP 166, Eindhoven University of Technology.