Agricultural Systems 98 (2008) 159–166

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

Analysing agricultural users' patterns of behaviour: The case of OPTIRas™, a decision support system for starch crop selection

Laura Măruşter *, Niels R. Faber, René J. Jorna, Rob van Haren

Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 24 April 2007
Received in revised form 30 May 2008
Accepted 5 June 2008

Keywords:
Modelling user behaviour
Agricultural users
Decision support systems
Process mining
Cultivar selection

ABSTRACT

Redesigning IT systems for specific user groups encompasses a lot of effort with respect to analysing and understanding user behaviour. The goal of this paper is to provide insights into patterns of behaviour of agricultural users, during the usage of a decision support system called OPTIRas™. This system aids agricultural users in their cultivar selection activities. We analyse logs resulting from OPTIRas™, and we get insights into user's navigational patterns. We claim that the results of our analysis can be used to support the redesign of decision support systems in order to address specific agricultural users' characteristics.

1. Introduction

As on-line services and Web-based information systems proliferate in all domains of activities, it becomes increasingly important to model user behaviour. These systems can therefore more appropriately address user characteristics. Various methods have been developed to model user behaviour in case of generic users (Mobasher, 2006). Particular topics are addressed by research in human–computer interaction (HCI) and user-system interaction (USI), such as the discovering of user behaviour with regard to 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 (Rauterberg et al., 1998; Bollini, 2003; Ernst et al., 2005; Thüring et al., 1995; Lee and Lee, 2003). Research on web log and usage mining topics provides valuables techniques for analysing and modelling user behaviour (Pabarskaite and Raudys, 2007).

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™.

Focusing on agricultural users as a specific IT user group seems to become an important research issue, for instance "due to a need for a higher precision in the use of chemicals and in the care of farm animals" (Thysen, 2000). Jensen (2001) analysed the usage of a web-based information system for variety selection in field crops. The author compared four user groups by constructing measures based on logged information. This analysis reveals interesting similarities and differences in behaviour concerning the four groups. However, no insights could be given about the most typical sequence of navigation patterns, such that it could support the redesign of the system.

In this paper we illustrate a methodology of analysing user behaviour by employing frequency analysis and process mining techniques (van der Aalst and Weijters, 2004b) on user logs. Agricultural users are considered as a specific user group. We provide insights into the patterns of behaviour of agricultural users during the use of a web-based IT system, namely a decision support systems called OPTIRas™. 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. More specific, the investigation of agricultural users' patterns of navigation is centred on answering the following research questions:

* Corresponding author. Tel.: +31 50 3637316.
E-mail address: l.maruster@rug.nl (L. Măruşter).

0308-521X/$ - see front matter
doi:10.1016/j.agsy.2008.06.003

Research question 1: To what extent does the decision support system OPTIRas™ support the decision making phases?
Research question 2: To what extent is the goal of the website fulfilled?
Research question 3: What kind of navigational patterns do agricultural users show?
Research question 4: How should the website be redesigned, in order to address the characteristics of specific agricultural users?

In Section 2 we review the decision making literature in the context of the farming domain, we briefly describe OPTIRas™ decision support system, and we present details about data collection. Section 3 contains the analyses of user behaviour based on frequencies, focusing on user group, user sessions and page requests. In Section 4 we determine farmers' patterns of behaviour with the help of process mining techniques. The discussion of results and implications for redesign are provided in Section 5. Conclusions and future research end this paper in Section 6.

2. Agricultural users and decision making

Concerning the decision making process, a well-known model was formulated by Simon (1977). He explained the human decision process using the three phases of 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, in the *design phase*, the individual formulates one or more alternative solutions to the recognised (sub-)decisions. Eventually, a final solution is formed in the *choice phase*. In this final phase of the decision process, partial solutions are evaluated, in order to meet the outcome criteria. 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 always linearly follow upon each other. Commonly, the decision making 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). In case of specific types of users, decision making models have been also developed. 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). These models also show a trichotomic structure.

Whenever these models are represented as linear (sequential) or nonlinear processes, as a matrix or as an iterative process, they show a three-phase generic structure. The first phase we will call it the *Information gathering* phase, where information is collected about the problem at hand. The second phase is the *Design* phase, where different candidate solutions are formed. Finally, in the *Choice* phase, the selection is made. It should be clear that a decision support system has to address in one way or another these basic phases of the decision making process (Jorna, 2001). In the following, we investigate in which way OPTIRas™ supports these phases.

2.1. 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, that is to say by interacting with the human decision process (Turban and Aronson, 2001).

The OPTIRas™ system was designed to target low-yielding potato farmers, attempting to realise an increase in the yield of these (and other) 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 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 DSS is to ensure that farmers gain insight into PCN infestation and its effect on the growth of the selected cultivar, and into the application of pesticides. Moreover, the DSS provides information about the financial consequences of all these factors.

Web pages do not have all the same function. The function variety depends on the purpose of the web site. In our case, the site's goal is to provide the agricultural user 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 the literature, such a categorization is called a *concept hierarchy*, or a *service-based hierarchy* (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". 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 developed.

The interface of OPTIRas™ DSS consists of seven main pages. We categorize OPTIRas™ 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.

Using the three-phase generic model of the decision making process, we proceed with the following mapping, as we discussed before: *Action* pages correspond to the *Information gathering* phase, where the user is supported to explore the issue (the selection of

Table 1
Categorization of OPTIRas™ pages

Page	Category	Purpose
Field ID PCN history Reorder cultivars	Action	Information gathering
Yield information Crop rotations Report	Target	Design
Details (cultivar) Options Help System pages Help Save	Other	–[a]

[a] There is no mapping between *Other* pages and any decision making phase.

the cultivar), by specifying Field characteristics and the PCN history. *Target* pages correspond to the *Design* phase, where the system provides the user with one or more possible solutions. The *Choice* phase does not have any correspondence in the actual DSS; the final decision is actually taken by the user outside of the system, that is to say the farmer in the field.

Given this mapping, we formulate the following hypothesis: the DSS (in our case OPTIRas™) supports the decision making phases *Information gathering* and *Design* if users visit *Action* and *Target* pages for comparable amounts of times. In Fig. 1 this mapping is illustrated with bold dotted arrows.

In Section 3 we investigate user behaviour while visiting *Action* and *Target* types of pages by focusing on user group, user sessions and page requests. In Section 4 we get insights into the shape of navigational patterns. In order to perform these investigations, we collect data, which is described in the following subsection.

2.2. Data collection

According to Herder and Juvina (2005), users' 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").

Because we want to obtain all three types of information mentioned above, we record data on two types of log files. Each session is mapped to these two types of files: (i) the movement file (MOV), which contains information concerning the movement from one source page to a destination page (such as time stamp, name of source page, name of destination page, and the button used, see Table 2), and (ii) the variable file (VAR), which contains data information concerning actions done within pages (such as time stamp, page name, variable name, field value, and role, see Table 3). The VAR file may provide both *syntactic* and *pragmatic information*.

In Table 2 an example is shown of an OPTIRas™ navigation sequence, recorded in a MOV file. The user sequentially accesses seven types of pages, mostly by pressing the Next button. The corresponding VAR file is presented in Table 3. This session belongs to a grower that prefers Dutch language. We notice that on page *Field*, the user sets some factors, such as soil type, expected cultivar (seresta) and expected yield (42 tonnes/hectare). Further, on page *Pcn* the user fills the PCN infestation estimate (2) and on page *Order* establishes the rank for ordering cultivars based on the relative yield (8). Based on these choices, the user can visualise the yield of first and second crop on page *Yield* (2000 euros/hectare and 1335.69 euros/hectare, respectively). The user invokes the *Crop* page, where information (and structures) about crop rotation is provided, and then returns to the *Yield* page.

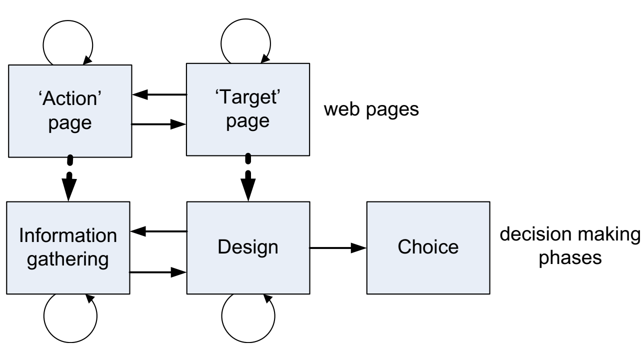

Fig. 1. Mapping generic decision making phases on *Action* and *Target* pages.

Table 2
An example of a MOV log file

Time stamp	FromPage	ToPage	Button
2004-12-22 22:13:29	field	pcn	NEXT_BUTTON
2004-12-22 22:13:35	pcn	order	NEXT_BUTTON
2004-12-22 22:14:00	order	yield	NEXT_BUTTON
2004-12-22 22:14:26	yield	crop	NEXT_BUTTON
2004-12-22 22:16:16	crop	yield	BACK_BUTTON
2004-12-22 22:16:25	yield	crop	NEXT_BUTTON
2004-12-22 22:16:53	crop	details (katinka_l)	CULTIVAR_LIST
2004-12-22 22:17:54	details	allcultivars	DETAILS_TEXT

Table 3
An example of a VAR log file

Time stamp	Page	Variable	Value
2004-12-22 22:13:14	logon	language	nl
2004-12-22 22:13:14	logon	role	grower
2004-12-22 22:13:18	field	soilType	2
2004-12-22 22:13:18	field	fieldName	test
2004-12-22 22:13:29	field	expectedCultivar	seresta_l
2004-12-22 22:13:29	field	expectedYield	42
2004-12-22 22:13:32	pcn	pcnEstimate	2
2004-12-22 22:14:00	order	ordRelativeYield	8
2004-12-22 22:14:01	yield	top_crop1_pi	2000.0
2004-12-22 22:14:01	yield	top_crop2_pi	1335.6859
2004-12-22 22:14:01	yield	top_losses	48.664795
2004-12-22 22:14:38	crop	rotation_0_treatment	4
2004-12-22 22:14:49	crop	rotationToDisplay	1
2004-12-22 22:15:06	crop	rotation_1_cultivar	katinka_l
2004-12-22 22:16:18	yield	top_crop1_pi	2000.0
2004-12-22 22:16:18	yield	top_crop2_pi	1335.6859
2004-12-22 22:16:18	yield	top_losses	48.664795
2004-12-22 22:16:30	crop	showGraphicalOutput	false

The MOV file (Table 2) provides us with information needed to build navigation graphs (see Section 4). The VAR file (Table 3) contains information that we employ to analyse user behaviour (see Section 3).

Users can login in OPTIRas™ either registering with their e-mail address, or 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. This manner of anonymity provides advantages for the users, but brings disadvantages for the analyses. A user may login the first time with his or her e-mail, the second time anonymously using computer A, and the third time again anonymously using computer B; in the analyses, these three sessions will counts as sessions belonging to three separate users, and we cannot see anymore how a user eventually changes his behaviour in time. Moreover, agricultural users can login by specifying their role, i.e. farmer, scientist, extension worker, or other.

OPTIRas™ usage is expected to take place within two peak periods (i) November, December and February (period for purchasing crops, expected to be the peak period) and (ii) April, August and September (period when the yield is actually obtained). In the latter period, agricultural users may compare the obtained yield with the advice given earlier by OPTIRas™.

We inspected the MOV and VAR log files from 18 December 2004 (the date when OPTIRas™ became on-line) till 30 May 2006. We, therefore, have two peak periods: (i) December 2004–February 2005 and (ii) November 2005–February 2006. In total, the behaviour of 501 individual users has been logged and investigated in the analysis.

3. User analysis

In this section we perform a frequency analysis of user behaviour, focusing on user group, user session and page request aspects.

The information used for these analyses comes from the VAR log files.

We concentrate on the nature of the end users. The OPTIRas™ system has been designed for farmers. However, a larger group of agricultural users is actually using OPTIRas™. As mentioned before in Section 2.2, this group can be differentiated into three target groups, namely farmers, extension workers, and scientists. Extension workers are people with a degree in agricultural sciences, who help farmers in their practical work. Scientists also use OPTIRas™, but rather for experimental purposes. Table 4 (based on the VAR files) shows the distribution and the total number of users, sessions and page requests per user group during the logging period. In the following analyses, we take the differentiation between the three target groups into account. As we expected, farmers are the largest group using the OPTIRas™ system, which is reflected also by a large frequency of sessions and page requests. Note that quite a big number of users logged in without a specified role (the *Null* row in Table 4).

In Table 5 (based on the VAR files) the frequencies of sessions, page requests, and page requests per session are shown. For instance 483 users, representing 79.% of the total users, used OPTIRas™ in just one session. Also, 479 users (49.4%) visualised just one page during all sessions, while 211 users (34.7%) have inspected just one page per session. If we look at the frequency on row 8 in Table 5, we see that 3 users have used OPTIRas™ in eight sessions, 1 user visited 8 different pages during all sessions, and 19 users have visited 8 pages per session.

As we can see in Table 5, the majority of users (483, almost 80%) had just one OPTIRas™ session. Also, close to 50% of the users (479) visualised just one OPTIRas™ page. A natural interpretation of this fact is that they logged in just once, and visualised the starting page, and then never reused OPTIRas™. We easily see that as we move from top to bottom in Table 5. The corresponding number of sessions per users, pages per user, and pages per users session rapidly decreases. One remarkable exception is number 5, where the number of pages per user (144, 14.9%) is greater than the previous three frequencies 4, 3, and 2 (83 (8.6%), 73 (7.5%) and 106 (10.9%), respectively). It seems that many users were eager to browse all five OPTIRas™ pages (which is actually the prescribed order of pages).

Table 4
Distribution and total number of users, sessions and page requests per user group (based on VAR file)

User group	Users	%	Sessions	%	Page requests	%
Ext. worker	33	4.8	86	4.4	118	4.7
Farmer	405	59.2	1118	57.7	1727	69.8
Scientist	53	7.7	120	6.2	153	6.2
Other	44	6.4	236	12.2	223	9.0
Null	149	21.8	378	19.5	254	10.3
Total	684	100.0	1938	100.0	2475	100.0

Table 5
Frequencies of sessions, page requests, and page requests per session (based on the VAR file)

Frequency of						
#	Users using # sessions	%	Users using # pages	%	Sessions involving # page requests	%
1	483	79.4	479	49.4	211	34.7
2	62	10.2	106	10.9	58	9.5
3	26	4.3	73	7.5	59	9.7
4	12	2.0	83	8.6	52	8.6
5	6	1.0	144	14.9	54	8.9
6	4	0.7	76	7.8	37	6.1
7	7	1.2	7	0.7	25	4.1
8	3	0.5	1	0.1	19	3.1
9	1	0.2	–	–	14	2.3
10	1	0.2	–	–	8	1.3
11	1	0.2	–	–	6	1.0
>11	2	0.4	–	–	Rest	7.7

Table 6
Frequency and duration of page requests per user (VAR log)

Page	Frequency	%	Duration (min)		
			Mean	Variance	Max.
Logon	608	33.9	0.12	0.01	15.72
Subtotal Logon	608	33.9	–	–	–
Field	318	17.7	0.39	0.15	7.93
Pcn	246	13.7	0.92	0.84	11.62
Order	249	13.9	1.56	2.43	31.15
Subtotal Action	813	45.3	–	–	
Yield	265	14.8	1.55	2.4	34.65
Crop	90	5.0	2.74	7.5	43
Report	4	0.2	–	–	–
Subtotal Target	359	20	–	–	–
Options	3	0.2	–	–	–
Saved	10	0.6	0.08	0.006	0.83
Subtotal Other	13	0.8	–	–	–
Total	1793	100	–	–	–

In Table 6 (based also on the VAR files) we provide an analysis of page requests at the user level (e.g. how often users have accessed a specific page), and duration per page in minutes. As we show earlier in Table 1, we split OPTIRas™ pages in three categories, *Action* pages, *Target* pages and *Other* pages. Although *Logon* is not a specific category, we show it separately because of its high frequency.

We observe that users most frequently visit pages from the category *Action* pages – *Field*, *Pcn* and *Order* (shortcut for Reorder cultivars). Although *Field* page has the highest frequency in this category, the *Order* page is most intensively viewed, with an average of 1.56 min, and a variance of 2.43 (we assume a negative exponential distribution). This is not surprising: the user has to spend time to fill in different parameter values, and based on these, to (re)order the list of cultivars at the left hand side of the screen. In total, *Action* pages are visited in 45.3% of page requests. However, pages from the category *Target* are visualised only in 20% of page requests; this is a disappointing result, given the fact that the main goal of OPTIRas™ is to provide detailed information about yield.

Summarizing the frequency analysis, we obtained two kinds of insights. With respect to *usage*, in most cases (i) farmers used OPTIRas™ only once, in one session (almost 80%), and (ii) farmers visualise only one page (almost 50%). With respect to *goal fulfilling*, (iii) the invocation of *Action* pages (these invocations indicate that the user is pursuing the site's goal) predominates, and (iv) the invocation of *Target* page (these invocations indicate that the user has achieved the site's goal) has a smaller proportion than *Action* page.

These results are used further in Section 5 to provide answers to the first research question. In order to answer the second research question concerning goal fulfilment, we use process mining techniques that can reveal navigational patterns.

4. Mining navigational behaviour with process mining technique

Different methods have been used to investigate the navigational patterns of users, using logged data (Pabarskaite and Raudys, 2007).

In our case, we use process mining techniques to inspect navigational patterns. We search for navigational patterns of all agricultural users and different target groups, as formulated in the following questions:

Question 1: What kind of navigational patterns do agricultural users show?
Question 2: What kind of navigational patterns does each target group show (e.g. growers, extension workers, scientists)?

Process mining techniques make it possible to extracting information from event logs (van der Aalst et al., 2003). For example, the audit trails of a workflow management system or the transaction logs of an enterprise resource planning system can be used to discover models describing processes, organisations, and products. Moreover, it is possible to use process mining to monitor deviations (e.g., comparing the observed events with predefined models or business rules). In this paper we use process mining to inspect navigational behaviour, which, to our knowledge, is a new application of this technique in the domain of farming. To do this, we use the ProM pluggable framework, which offers a wide variety of process mining techniques (ProM, 2007). Process mining can be also used for mining the organisational and case perspectives (van der Aalst and Weijters, 2004b).

The choice of the most appropriate algorithm for a certain problem is not always straightforward. The process mining algorithm we use is the heuristic mining implemented in the ProM framework (Weijters and van der Aalst, 2003). We choose heuristic mining because of its robustness for noise and exceptions (van der Aalst and Weijters, 2004b). The heuristic mining is based on the frequency of patterns. It is therefore possible to focus on the main behaviour in the event log.

4.1. Mining the navigational pattern of all agricultural users

This section concentrates on the navigational patterns exhibited by all users. In Fig. 2 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 MOV log file, see Table 2). The file consists of 763 user sessions, belonging to 501 individual users. The rectangles refer to transitions or activities, in our case to page names, such as *Field* and *Order*. 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 (Fig. 2) shows how many times a page has been invoked. The arcs between two activities A and B are associated with two numbers: (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. The higher the causality measure is, the stronger the relation between A and B is. For instance in Fig. 2, *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. For more details, see Weijters et al. (2006).

The complicated picture from Fig. 2 can be interpreted as follows. We can determine the most common pattern, which consists

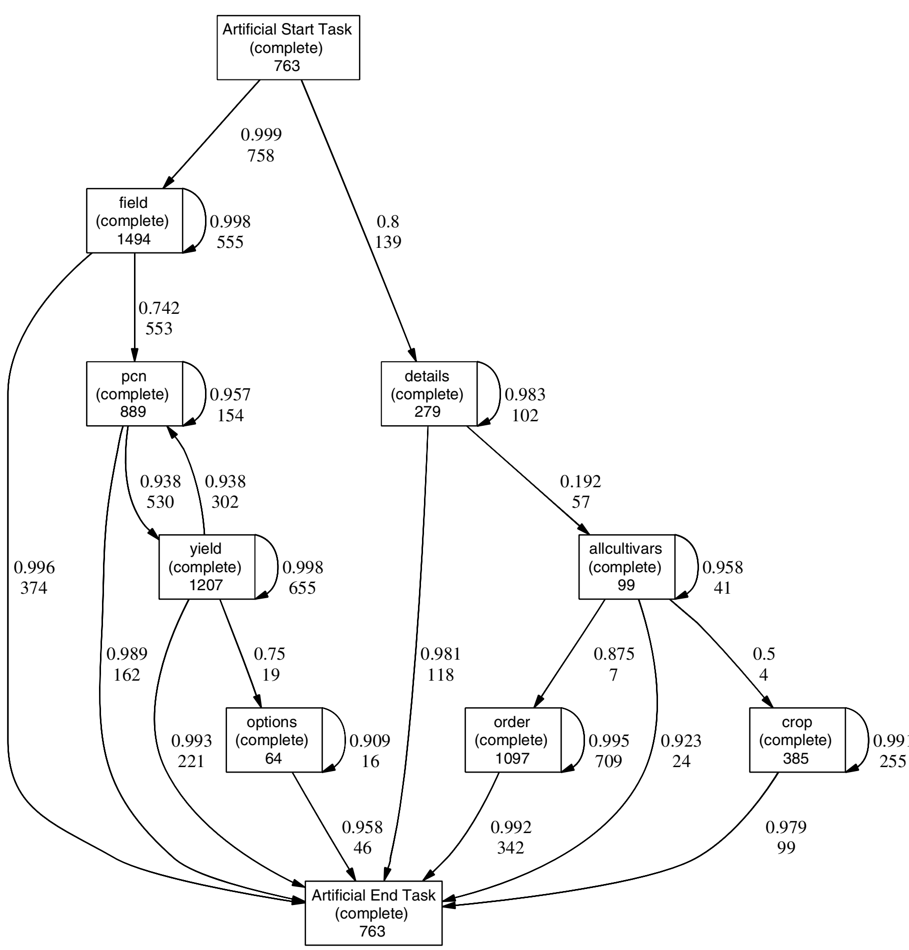

Fig. 2. The mined behaviour corresponding to all agricultural users.

of *ArtificialStartTask* → *Field* (758), *Field* → *Pcn* (553), *Pcn* → *Yield* (530), and *Yield* → *ArtificialEndTask* (221). The sequence *Field* → *Pcn* → *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 change PCN values to observe the impact on calculating the yield (the higher the values of PCN, the smaller the yield). Noticeable is also the highly frequent self-recurrent link to *Yield* (655).

The interpretation of these findings is that agricultural users in general use the prescribed sequence of page invocation. The directed graphs from Fig. 2 reveal that the invocation of *Action* pages predominate. In Fig. 2, the sum of incoming arcs from *Action* pages (Field, Pcn, Order) to activity *ArtificialEndTask* 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 visualised significantly less than *Action* pages. This is a disappointing result, given the fact that the main goal of OPTIRas™ is to provide detailed information about yield. Second, after each page a significant number of sessions stop (in Fig. 2, 374 after page *Field*, 162 after page *Pcn*). However, we also see that whenever users do not give up in early stages, and visualise *Target* pages, they show a relevant activity. They repeatedly recall 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 rises. 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 Fig. 2). This suggests that users are interested in information about different cultivars.

4.2. Mining the navigational patterns of differentiated target groups

In this section we answer the second question mentioned in the beginning of Section 4, namely what are the patterns of behaviour of different target groups. These target groups are farmers, extension workers, and scientists. Whenever a user logs in into OPTIRas, he has to specify his role, e.g. growers, extension worker, scientist or others. The distribution of sessions is as follows: from all 763 sessions, 535 are sessions belonging to growers, 59 are sessions coming from extension workers and 72 are the results of scientists usage. The remaining 97 sessions, which are not labelled, are left out from the analysis.

The run of the heuristic algorithm (see Fig. 3a), depicting the navigational pattern of farmers, provides more insights into the most generic behaviour: there seem to be differences between all agricultural users (Fig. 2) and farmers (Fig. 3a). 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 that the majority of farmers does not reach the *Yield* page, and thus we may doubt whether the DSS fulfils its goal. In addition, the navigational pattern of farmers does not contain loops (see in Fig. 3a. the missing loop between *Pcn* and *Yield*), except from self-loops.

Further inspecting the navigational patterns of the extension workers, we see in Fig. 3b. 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 rather complex behaviour.

The behaviour of scientists is the most complex (see Fig. 3c): 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

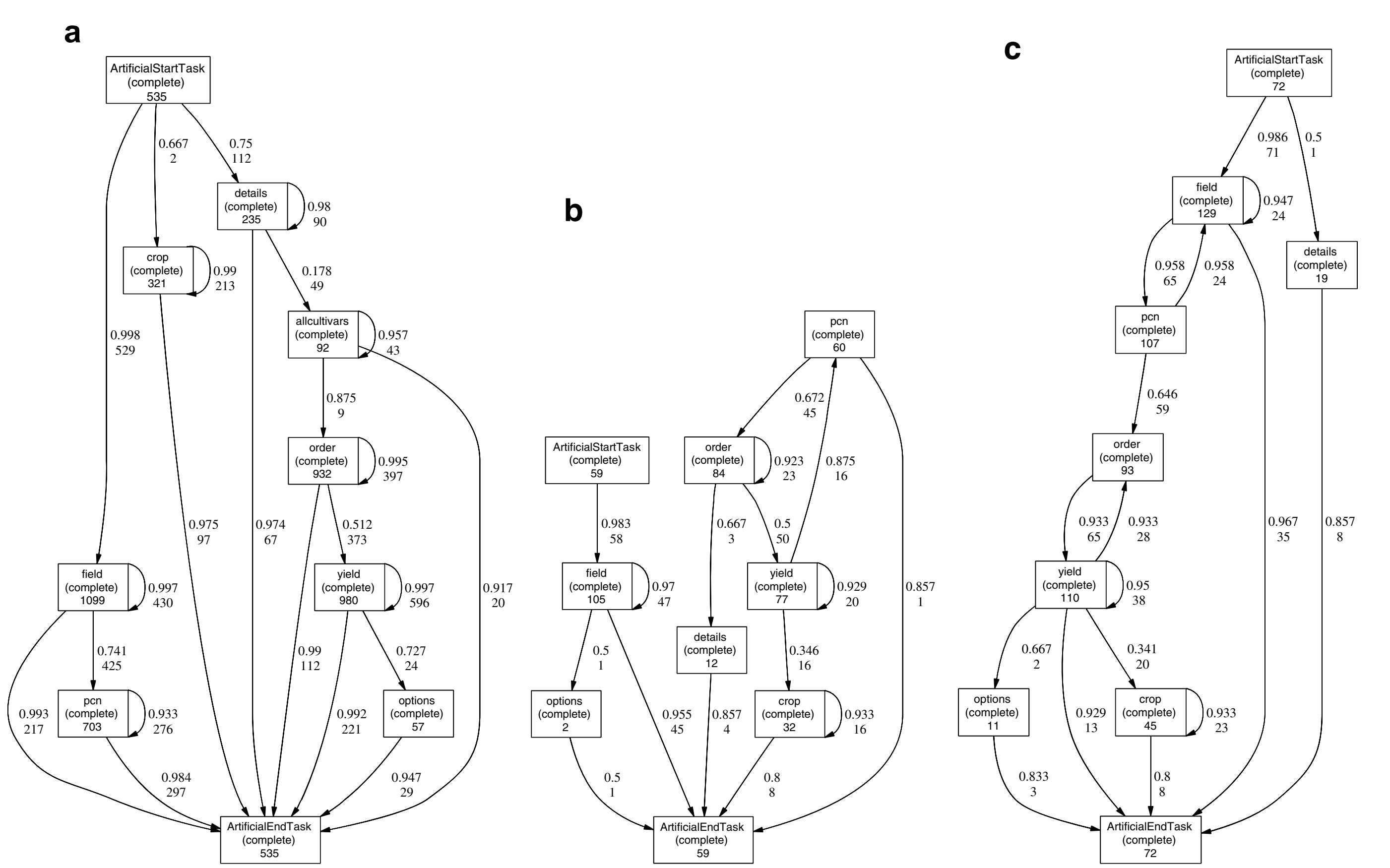

Fig. 3. The mined behaviour of farmers (a), extension workers (b) and scientists (c).

the prescribed behaviour, check the effects of changing value parameters. Also, most users of this group visit *Target* pages (*Yield* and *Crop*), which may suggest that they like to explore the possibilities of the DSS.

5. Discussion of results and implications for redesign

By analysing frequencies within user logs (Section 3), we found that, with respect to usage, agricultural users employed OPTIRas™ only once, in one session (almost 80%, see Table 5). Moreover, in most cases, agricultural users visualise only one page (almost 50%, see Table 5). As we can see from Table 6, the invocation of *Action* pages is dominant (45,3%), while the invocation of *Target* page has a smaller proportion than expected (only 20%). These results show that users do not realise fast enough what the goal of OPTIRas™ is, and what kind of advantages they may have when using this system. This leads to a lack of interest in using the system further. Furthermore, the *Details* page is visited quite often, thus it seems that users are willing to get information about different cultivars. We can now formulate answers to the first two research questions, namely (i) the OPTIRas™ system does not appropriately support decision making phases, and consequently (ii) the goal of the web site is not fulfilled.

The answer to the third research question is obtained by inspecting the navigational patterns of agricultural users presented in Section 4. We found that these users employed the prescribed sequence of page invocation. However, the *Details* page seems to be an alternative to the prescribed path. The directed graph shown in Fig. 2 converges with previous findings, i.e. *Action* pages are dominant, compared with *Target* pages. However, we see that whenever users do not give up in early stages, and visualise *Target* pages, they show a relevant activity. Namely, they repeatedly recall the *Yield* page, they revisit pages not directly linked (the reverse link *Yield* → *Pcn*), and also those that are directly linked. This fact illustrates that whenever users reach a *Target* page, their interest may rise. These findings support the answers to the first two research questions stated above.

The navigational patterns of differentiated target groups reveal a number of differences: in case of farmers, the majority does not reach the *Yield* page, and thus we may doubt whether the goal of the DSS is fulfilled. In addition, the navigational pattern of farmers does not contain loops, except from self-loops. Extension workers and scientists show a more complex behaviour, involving loops and backtracking. Instead of only following the prescribed behaviour, scientists check the effects of changing value parameters. These findings provide the answer to the third research question: various user groups have different navigational patterns.

These answers generate a number of implications for redesign, which was the topic of the fourth research question. First, the system must appropriately support the decision making phases. The designers of OPTIRas™ had the intention to develop an instrument for making optimal choices. In its current version, the user is, at best, encouraged to use the system by visiting pages in the prescribed order. Unfortunately, the user is not supported to perform a decision making process: almost half the users were not going further than the first page. This pinpoints the necessity of letting farmers gather information about the various topics involved in cultivar selection, in the *Information gathering* phase. Supporting this phase without enforcing any prescribed order is a necessary future development. OPTIRas™ is now a DSS with a focus on (optimal) choice for which a *Design* phase in decision making is essential. That is the good news. The bad news is that many users want to explore the search space without being restricted. Currently OPTIRas™ does not support that. Second, we suggest that on first page, the goal of OPTIRas™ should be very explicitly and clearly stated, eventually illustrated with a short example or a drawing. Third, the specification of the goal (yield calculation) should be distinguished from providing information about cultivars, in order to circumvent misleading users. Fourth, depending on the target group, hints for the following steps should be given: e.g. in case of growers, what concrete path should be followed to get the yield overview, in a form of a sitemap.

Our recommendations match with the general recommendations for redesign suggested by Jensen (2001), namely that a system has to comply with requirements: up-to-date, comprehensive, interactive, user-friendly. OPTIRas™ can be characterised as being up-to-date (e.g. the newest information about cultivars are updated at a regularly basis). However, the system is not very comprehensive, it is limited with respect to user-friendliness and it is not at all interactive. Our recommendations will enhance comprehensibility and user-friendliness. The possibilities of enabling interactivity in OPTIRas™ will be tackled in further research.

The methodology presented in this paper and consequently the implications of redesign can go beyond this specific decision support system OPTIRas™. The following requirements are essential in order to apply the proposed methodology to a broader range of information systems. First, the fulfilment of the information systems' goal needs to be explicitly formulated. In case of the present decision support system considered, we have explicitly questioned whether the phases of the decision making process are actually supported by the system (see the first research question from Section 1). However, when the IT system has a different nature, its goal needs to be approached adequately. Subsequently, the goal fulfilment can be assessed. Second, this methodology is especially useful if targeted users can be differentiated in various target groups. Third, the information systems need to have logging functionality available. In case of web-based systems, this functionality is easy to enable. In the case of systems that are not web-based, this functionality can be programed.

6. Conclusion

In this paper we provided more insights into patterns of navigation behaviour, considering a specific group of users, namely agricultural users.

With respect to all agricultural users, they use the prescribed order of pages. We discovered that often users visualise just the first page, and then 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 *Information gathering* 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. 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 shows difficulties in fulfilling the goal of OPTIRas™, while the other two target groups, extension workers and scientists seem to manage better. Given the fact that OPTIRas™ was especially developed to support farmers, it is a clear sign that redesign actions are needed.

The results of our analysis can be used to support the redesign of DSSs in order to address the characteristics of specific agricultural users. First, OPTIRas™ should better support the decision making process. Second, we suggest that in the very first page of OPTIRas™, a user should be confronted with the goal of the system.

Third, depending on the target group, hints for the following steps should be given, therefore 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. Moreover, we plan to let these users participate in the redesigning process and to make OPTIRas™ more interactive.

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 the participating parties in AGROBIOKON for supporting this research. AGROBIOKON is a cooperation between AVEBE, the Dutch potato starch processing industry and HPA, the Dutch Arable Farmers board. AGROBIOKON is supported by the Northern Netherland Provinces the ministry of Agriculture and the Ems Dollart Region (EDR). The project is financially supported by the European Union, the federal state of Niedersachsen and the Samenwerkingsverband Noord-Nederland. Also, we want to thank Iain Milne from Scottish Crop Research Institute for enabling the logging function of OPTIRas™, and the two anonymous reviewers who helped us improving the paper.

References

Balajinah, B., Raghavan, S., 2001. Intrusion detection through learning behavior model. Comput. Commun. 24, 1202–1212.

Bollini, L., 2003. Web interface design based on cognitive maps: generative dynamics in information architecture. In: The Sixth Generative Art Conference – GA2003.

Ernst, N.A., Story, M.-A., Allen, P., 2005. Cognitive support for ontology modeling. Int. J. Hum-Comput. St. 62 (5), 553–577.

Faber, N., Jorna, R., van Haren, R., Maruster, L., 2006. Knowledge and knowledge 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., Pederson, S., 2006. A model of decision-making and informations flows for information-intensive agriculture. Agric. Syst. 87, 192–210.

Herder, E., 2002. Metrics for the adaptation of site structure. In: Proceedings of the German Workshop on Adaptivity and User Modeling in Interactive Systems ABIS'02, Hannover, pp. 22–26.

Herder, E., Juvina, I., 2005. Discovery of individual user navigation styles. In: Proceedings of the Workshop on Individual Differences – Adaptive Hypermedia, Eindhoven, 2005.

Jensen, A.L., 2001. Building a web-based information system for variety selection in field crops-objectives and results. Comput. Electron. Agric. 32, 195–211.

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

Jorna, R., 2001. Decision making, semiosis, and the knowledge aspects: the case of Studas. Semiotica 133 (1/4), 97–119.

Juvina, I., Herder, E., 2005. Individual differences and behavioral aspects involved in modeling web navigation. In: Proceedings of User-centered Interaction Paradigms. Vol. LNCS 3196. Springer, pp. 77–95.

Klein, M., Methlie, L., 1995. Knowledge-based Decision Support Systems: With Applications in Business. John Wiley & Sons, Chichester.

Lee, K., Lee, S., 2003. A cognitive map simulation approach to adjusting the design factors of the electronic commerce web sites. Expert. Syst. Appl. 24, 1–11.

Menasalvas, E., Millan, S., Perez, M., Hochsztain, E., Robles, V., Marban, O., Peña, J., Tasistro, A., 2003. Beyond user clicks: an algorithm and an agent-based architecture to discover user behavior. In: Proceedings of ECML/PKDD.

Mintzberg, H., Raisingham, D., Thorêt, A., 1976. The structure of 'unstructured' decision process. Admin. Sci. Quart. 21, 246–275.

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

Öhlmér, B., Olson, K., Brehmer, B., 1998. Understanding farmers' decision making process and improving managerial assistance. Agric. Econ. 18, 273–290.

Pabarskaite, Z., Raudys, A., 2007. A process of knowledge discovery from web log data: systematization and critical review. J. Intell. Inf. Syst. 28, 79–104.

ProM, 2007. The ProM framework. <http://www.processmining.org>.

Rauterberg, M., Schluep, S., Fjeld, M., 1998. Modeling of cognitive complexity with Petri nets: an action theoretical approach. Cybernet. Syst. 2, 842–847.

Simon, H., 1977. The New Science of Management Decision, Revised Edition. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Song, Q., Shepperd, M., 2006. Mining web browsing patterns for e-commerce. Comput. Ind. 57, 622–630.

Spiliopoulou, M., Pohle, C., 2001. Data mining to measure and improve the success of web sites. Data Min. Knowl. Disc. 5 (1-2), 85–114.

Thüring, M., Hannemann, J., Haake, J.M., 1995. Hypermedia and cognition: designing for comprehension. Commun. ACM 38 (8), 57–66.

Thysen, I., 2000. Agriculture in the information society. J. Agric. Eng. Res. 76, 297–303.

Turban, E., Aronson, J., 2001. Decision Support Systems and Intelligent Systems. Prentice-Hall, Upper Saddle River, NJ.

van der Aalst, W., van Dongen, B., Herbst, J., Maruster, L., Schimm, G., Weijters, A., 2003. Workflow mining: a survey of issues and approaches. Data Knowl. Eng. 47 (2), 237–267.

van der Aalst, W., Weijters, A., 2004b. Process mining: a research agenda. Comput. Ind. 53 (3), 231–244.

Weijters, A., van der Aalst, W., 2003. Rediscovering workflow models from event-based data using Little Thumb. Integr. Comput. Aid. Eng. 10, 151–162.

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