What Readers Want to Experience: An Approach to Quantify Conversational Maxims with Preferences for Reading Behaviour

Hanna Knäusl[1], Bernd Ludwig[1]

[1]*Chair for Information Science, University Regensburg, Regensburg, Germany*
{hanna.knaeusl, bernd.ludwig}@ur.de

Abstract: Searching information on web pages is a tedious task for users as web servers provide complete web pages and do not tailor their content to the user's current information need. This leaves an enormous amount of workload for the user and influences his emotional attitude towards the whole task even if a search engine has filtered pages that are relevant to a user query. In this paper, we propose an approach to adapt the response to queries to user preferences for his reading experience in order to leverage the problem of information overload. With these preferences, it is possible to select the most preferred content from a web page. In our view, the preferences are a quantitative way to express conversational maxims. We present our experimental approach to learn these preferences from annotated browsing sessions and introduce a decision strategy for the selection of content on the basis of the learned preferences.

1 Introduction

Imagine you are thinking about what you want to cook tonight, but you don't have any good ideas. On top of this, you are run out of supplies and short on time, as well. All you need is a good tip how to cook a quick dinner with the few ingredients left at home. For such a tip, you consult your favorite cooking portal in the Internet. What do you get there? A long list of recipes, pictures, videos, and health care advices which is overwhelming you as you are just looking for a few simple cooking instructions to get rid of your hunger. Obviously, you are unsatisfied with the given information and its presentation and most probably you will end up at McDonalds.

However, after too much fast food you start to feel bad and plan to pay more attention to your nutrition. Therefore, you spend more time on choosing healthy dishes and want to learn more about nutrition. In this situation, you consult your cooking portal again. This time, you are disappointed as the recipes do not provide more information you are interested in. On the other hand, maybe the information is there, but it is just too difficult to find it quickly.

1.1 Context- and Task-Awareness

As it becomes obvious from this example, information seeking processes are always influenced by many situational context factors. It is also obvious that the problem for you is not related to the content available, but to the way in which the portal presents it to you. It does not pay any respect to your precise information need in situations that vary to a large extent. Actually, it *cannot* pay this respect as all pages are static and cannot be adapted to your information need. This is valid for sections of text as well as for multi media elements on the page.

Therefore, it is our hypothesis that information seeking can be supported substantially by information systems that generate the responses to a concrete information need at run time and take important context factors into account. However, generating content from a larger pool of information elements implies a selection process to take place. Which and how many parts of the text, which and how many images, which and how many multi media data should be selected to be presented to the user?

In order to address this problem, in our research, we are targeting towards the development of an algorithm that, instead of selecting complete pages, can – at least approximately – recombine elements of content by taking quantitative versions of the GRICEan maxims of conversation (Grice, 1975) into account. In order to do that, the algorithm has to be capable of making selections similar to those of a human reader. Therefore, we are interested in developing a quantitative model of the maxims of conversation that can be

applied by a decision making algorithm for inclusion or exclusion of content elements. While the principle of quality can be formalized by applying information retrieval techniques, the principles of quantity, relation, and manner depend on the context of the user and the current situation in which the information need has arisen.

For an information system to be capable of selecting content elements (such as text, images, or videos), the original static documents have to be split up into sets of these elements (see Fig. 1). Their (discourse) logical coherence can be reconstructed by annotating coherence relations as introduced by prominent linguistic research such as the work on attentional and intentional state by (Grosz and Sidner, 1986). We apply the two coherence relations introduced by (Grosz and Sidner, 1986) for our purposes:

- **satisfaction precedence** (*x* SP *y*): content element *x* *satisfaction preceedes* content element *y* if *x* should be presented earlier than *y* as *x* provides explanations for *y*.

- **dominance** (*x* DOM *y*): *x* *dominates* *y* if *y* provides content that supports *x* as it is complementary or in some other way helpful for the user to understand the content of *x*.

In essence, this approach results in a graph-like model for each (original) document (see Fig. 2). Given an information need, the relevance of each element can be computed by calculating relevance scores for it (e.g. TF-IDF similarity scores) from the textual content of the element or from lingustic tags assigned to it – depending on the type of the element.

With the elements of the original document ranked in the way described above, and assuming the other maxims formalized quantitatively as well (details will follow below), an information interface can simulate human reading behaviour in a continuous loop until the information need is sufficiently satisfied. In each iteration, among the elements ranked best, the interface selects those optimizing the valuation of the other maxims (see below) and presents them to the user.

Our hypothesis is that the selection of appropriate elements should be guided by the readers's (or user's) preferences for elements of content as well in quantity (how many text blocks, images, or graphics are sufficient?) as in manner (which type of information is preferred: text, lists, tables, or graphics?).

1.2 A Probabilistic Model for the Maxims of Conversation

The idea presented above relates any quantification of the maxims of conversation to user preferences for the way to consume a document and for its content. Consequently, we model the process of deciding on how many and which elements of content to include in an answer to a user's information need as a sequence of context-aware recommendations (see (Jannach et al., 2011), chap. 12.2 for an introduction).

An example of such a process is the following trace of a user consuming the document in Fig. 1:

Read introduction – scan the headlines – read text for section 2 – look at the image mentioned in this section

This trace can be understood as a sequence of decisions made by the user that lead to a comfortable reading experience. By taking these decisions the user selects content elements optimizing quantity, relation, and manner of the selected elements.

Mathematically, such an optimal selection can be modelled as a stochastic process

$$\alpha_1, \alpha_2, \ldots, \alpha_{t-1}, \alpha_t$$

where the α_i are actions such as those in the example above (more details will follow later). This process is conditioned by context factors C_i (e.g. the type of the search task, the current user satisfaction, the degree to which the the information need is satisfied, time constraints for the search, to mention some examples).

From this viewpoint, an optimal selection maximizes the *a posteriori* probability

$$P(\alpha_1, \alpha_2, \ldots, \alpha_{t-1}, \alpha_t | C_0, C_1, \ldots, C_{t-2}, C_{t-1})$$

The iterative process of stepwise presentation of the relevant information elements makes the assumption plausible that α_i is dependent only on α_{i-1} (the preceeding user activity) and on C_{i-1} (the set of context factors before α_i takes place). Therefore, the above probability can be approximated by

$$\eta \cdot \prod_{1 \le i \le t} P(\alpha_i | \alpha_{i-1}, C_{i-1})$$

with $\eta^{-1} = P(C_0, C_1, \ldots, C_{t-2}, C_{t-1})$ (the *a priori* probability of the sequence of contexts that develops during the search process).

When presenting the information elements iteratively, the information interface can find the optimal next step by choosing the action α_i that maximizes

$$P(\alpha_i | \alpha_{i-1}, C_{i-1})$$

with α_{i-1} known from the previous step and C_{i-1}. As we will discuss in more detail later, the different α_{i-1} encode different degrees of quantity and manner of the information presented and therefore are a basic building block of our quantative model of the maxims of conversation.
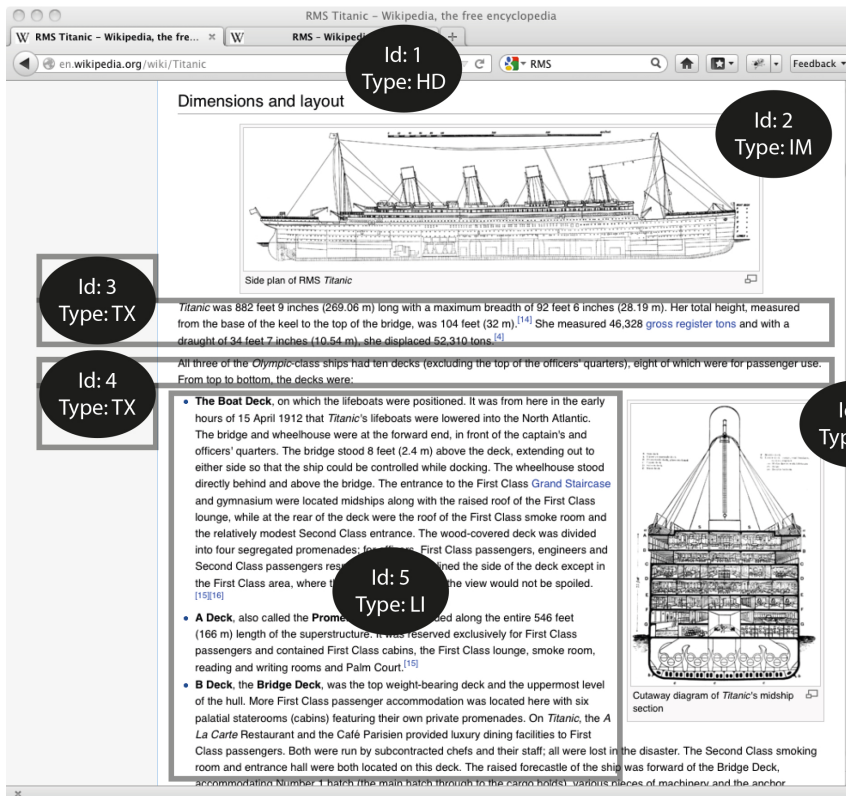
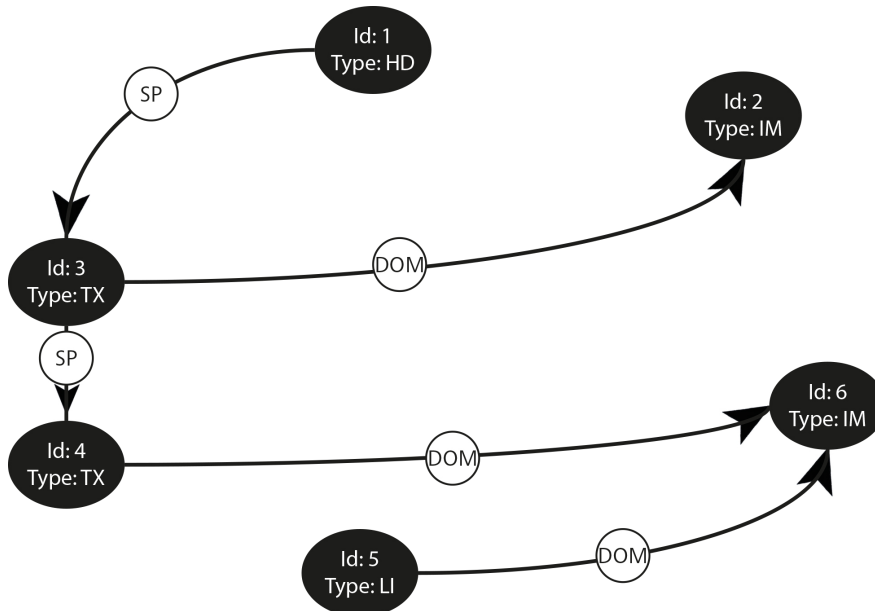Figure 1: Splitting up a document into a set of content elements

Figure 2: Discourse relations between content elements

In remainder of this paper, we explain our document model and our quantitative model of conversational maxims in more detail, define the parameters that have to be estimated from training data, describe our experiments how to collect the necessary data and their evaluation. From these results we derive a decision procedure that is capable to select relevant content elements taking the information need and conversational maxims into account. Finally, we discuss the conclusions we reached and outline future work.

1.3 A Graph Model for Multimodal Documents

As outlined in Sect. 1.1, our approach relies on the idea to split documents into a set of multimodal content elements (see Fig. 1). In order to preserve the structure of the document's content, the coherence relations between the single content elements have to be reconstructed and stated explicitly as a graph (see Fig. 2). Given an information need, all elements can be ranked by some information retrieval approach that we apply as a "black box" (i.e. we rely on the computed ranking without analyzing it further). Constructing an excerpt of the original document that satisfies the information need can therefore be interpreted as a search problem in the graph. A path has to be found that contains enough relevant information and, in parallel, is optimal with respect to the applied conversational maxims and therefore constitutes an optimally user-, task-, and context-aware response to the information need. The content elements on the path constitute the excerpt from the complete document. We will discuss this planning problem later in Sect. 5. Before that, in the next sections we will discuss the data required to formalize the idea of an optimal fulfillment of conversational maxims.

1.4 A Data Model for Task- and Context-Awareness

In the formal model introduced in Sect. 1.1, context factors play a prominent role as they may influence the user's preferences for the activities and content elements and therefore modify his behaviour. Consequently, quality, quantity, relevance, and manner of information have to be interpreted differently if the context factors change. We conclude from this observation that conversational maxims are not fixed and static rules, but – put formally – functions of the context factors. Some of these factors may change during the search process, e.g. the current user satisfaction, the degree to which the information need is satisfied, the fact how many times a content element has been

presented already, time constraints for the search, the current type of search, or the user's emotional state. Other factors may be static during the search process, e.g. the information need, the domain of the information need, or personal aspects of the user (age, gender, interests, preferences, reading behaviour).

Another crucial property of a search task is the concrete information need. In our experiments, users state it by typed queries (*specified information need*) to Wikipedia or by using the available navigation links (*non specified information need*).

We formalize all considered properties of a search task in a user model for the context dependent user preferences for his activities during search and for the possible types of content elements. As these preferences are nondeterministic in nature, the user model by the probability distribution

$$P(\alpha_i | \alpha_{i-1}, C_{i-1})$$

for a particular user. In our study, we considered the type of the search task as the only context factor C which is constant during the search process. This simplifies the user model to

$$P(\alpha_i | \alpha_{i-1}, C)$$

Our hypothesis is that these user preferences actually exist and differ given a set of context factors. Furthermore, in order to assess the user activities α_i at run time, they have to be observable from sensory input. In the current literature, many results can be found that reading behaviour during browser sessions is actually observable from his browser interactions and gaze pattern analyses. For our work presented here, we rely on these results. In order to discuss them in detail and to identify open issues we revise the state of the art in the next section and then draw conclusions for our work.

1.5 State of the Art

A main prerequisite for our approach to be implementable effectively is the observability of the current user activity. As the example in Sect. 1.2 suggests, it is crucial for an adaptive information system to have a reliable guess about the user activity at any point of time. As activities such as reading cannot be observed automatically by a desktop computer, tablet, or smartphone, the only tractable solution is indirect observation from the interactions occurring between user and web browser during the search process.

Therefore, in this section we revise current research results on user behaviour and observability in information seeking applications that the prediction of cognitive user activities from his interaction behaviour is indeed tractable.

(Rodden et al., 2008; Guo and Agichtein, 2010b) describe general patterns between the users' eye movement and mouse movement. They identified three patterns of mouse-eye coordination: *incidental* mouse usage, *bookmarking* and *follow*. The *bookmarking* and *follow* patterns are potentially useful to provide implicit feedback on search results. In case of *bookmarking* the mouse position can indicate a point of interest. In the bookmarking pattern, it is even possible to predict preferred elements of content directly from mouse events.

(J. Huang, 2012) found out that it is possible to distinguish between different cursor behaviour patterns *reading*, *scrolling*, *hesitating* and *click* to get an even better idea of how gaze position and mouse movement are aligned. They show that the user himself, the time and the search task (to a lesser extent) also contribute to the cursor-gaze alignment. This research provides valuable insight that and how eyetracking can be substituted by tracking the user's mouse interactions. (Guo and Agichtein, 2010a) take a similar direction. They analyzed the eye-mouse coordination which they found to be highly correlated.

We conclude from these results that it is tractable to predict gaze from mouse events in the browser.

(Hauger et al., 2011) report on their experiments that from user interactions with a web browser they can predict the sections of a web page the user looked at and even the paragraphs of hypertext the user had read on the web page. This study indicates that it is tractable to automatically observe which content elements have been consumed by the user at any time.

Predicting the task type from user interaction has been investigated by (Kellar and Watters, 2006). The authors provide significant statistical results that the current task type during a search process can be predicted from web browser interactions.

There is also some work on user preferences for elements of content. (Cutrell and Guan, 2007) show that indeed there is a preference in text snippet length in response to users's search task. Users with precise information needs prefer shorter snippets as well as users running out of time for the task. When browsing users spend more time on reading longer textual parts. However, (Cutrell and Guan, 2007) analyze search engine result pages and again do not address the user preferences for elements to be included in response to an information need. (Buscher et al., 2012) present results that the relevance of text passages may be predicted from eye movements.

(Agichtein et al., 2006) distinguish different types of user behaviour: *Query-text*, *Browsing* and *Click-through*. For each type they predict user preferences for the the presentation and ranking of search engine result pages. These results provide another hint for the influence of the task type on user preferences, but again do not address preferences for content elements.

1.6 Advancing the State of the Art

In summary, while according to the results of the reviewed literature, it turns out feasible to predict user behaviour from web browser interactions, a precise analysis of the reading activities in a search process and the preferences for content elements is still an open issue. Furthermore, this lack of knowledge blocks the development of adaptive user interfaces for information systems that can predict the user behaviour at any time and provide just the content needed for the expected next activity.

Such an interface would be helpful for many purposes. Most prominent among them is the need to address the problem of information overload. Another motivation comes from the growing number of smart phones and mobile applications. With the limited display size, it becomes necessary to split content into several pages in an intelligent way. Our work provides a contribution to these issues by understanding the presentation of content as a variant of the problem of text generation. We apply principles from discourse analysis in order to find an intelligent way to split content.

In the work presented in this paper, we focus on a key issue in this context. We are looking for a quantitative model of the qualitative principles from discourse analysis, which is obligatory to implement a decision procedure that can split and select content "intelligently". For the development of this quantitative model, our paper contributes two aspects to the research field of human-computer interaction and theory of (natural language) discourse. The first contribution is the description of a data collection of several hours of information seeking processes in the German edition of Wikipedia. Analyzing this data we can show that there are correlations between browser usage and eye movement. Furthermore, even user preferences correlate with eye movement.

As a second contribution we present how those results can be considered in a decision procedure that can select relevant information elements and generate web pages from these elements tailored to the user, the context, and the task type of his search.

2 Data Collection

In order to collect data to learn the preferred reading activities and content elements from, as the first

action	label	description
Read	RE	User is reading text
Scan	SC	User scans content e.g. headlines, lists or whole page
Look at	EX	User gazes something (e.g. a picture or other graphical element)
Navigate	NV	User performs navigational tasks

element	label
Headline	HD
passage within an article	TX
List	LI
Introduction	IN
Picture	PI
Info Box	IB
Charts, tables etc.	IG
Navigation within Wikipedia	WI
Other navigation	ON

Figure 3: Annotation labels for the user activities and for the gazed elements

issue we had to fix the possible set of events that could be assigned to the random variables α_i (see Sect. 1.2) and the possible types of search tasks.

To understand the importance of this issue, one has to consider the influence of the information need on the task type as studied by (Lindley et al., 2012). They show that since the Web is waved into everyday life there is indeed a variety of motivations for web use, especially in casual leisure scenarios. From these observations they conclude that a wide range of different tasks with different intents is performed in the web potentially influencing the user's preferences we are interested in.

As will be described in the Sect. 2.2, for our experiments we took these results into account by constructing tasks of three different types for the study participants. Another important variable to be controlled is the application domain. In our experiments, we restrict the search tasks to a small set of topics (see Fig. 4) and limit navigation to the pages of the German Wikipedia. With this restriction, we can reach similar levels of quality and similar structure among all pages visited by the test persons during each search process. In particular, the restriction implies that the set of user activities can be fixed as as the types of content elements is limited to those defined for Wikipedia (see the overview in Fig. 3).

2.1 The Probabilistic Model Revisited

For the reading activities, we constructed a typology (see the upper table in Fig. 3) and another typology for content elements on Wikipedia pages (see the lower table in Fig. 3). Reading activities and content elements occur always in pairs as the example in Sect. 1.2 illustrates. With the new names, the trace reads as:

$$RE\text{-}IN - SC\text{-}HD - RE\text{-}TX - EX\text{-}IM$$

The set of all pairs that can be combined from activities and content elements constitutes the set of events which can be assigned to the random variables α_i (in Sect. 2.3, we illustrate that we use this set as a set of labels in order to annotate video data and generate a training sample). For the example in the trace above we have $\alpha_0 = RE\text{-}IN$.

2.2 Study Design

Eye tracking is widely used to investigate users consumption of content presented in the web. In our experiments, we record gaze patterns in order to analyze which of the displayed items are of relevance for the user. As eye trackers can only be used in a lab situation, for our intention to predict the user preferences during every day usage of web browsers, we have to predict gaze patterns from data obtainable from the browser. Therefore, in accordance with the literature discussed above, in parallel to gaze patterns we track mouse and keyboard events in the web browser.

Participants had to complete six tasks each. The the tasks were pre-defined search tasks of different types (learn, lookup, browse). To avoid learning effects, the tasks were randomized. There was a limited time to complete the tasks. The participants were seated in a separate room on their own. The tasks are presented on the screen after they had filled the questionnaire. With such a setup, they were not interrupted during the whole experiment and therefore not biased by the examiner. The tasks present a short epsiode around the actual search problem to create a kind of context. Each task takes about 3.5 minutes on average. We have recorded data from 25 participants (all of them students or researchers in our department) – each of them conducting six tasks.

Up to now we have analyzed around 30 minutes of data whose evaluation will be presented in Sect. 3.

2.3 Data

During the experiments, we collected the following kind of data:

- Eyetracking Videos:

  All sessions with test persons are recorded automatícally by the eyetracking software. The tool records the screen content, stores it in a video file, and adds an additional layer indicating the

Lookup: You are watching a report about the sinking of the Titanic. During a commercial break you left the room and when you came back you had missed the information about how many passengers were on board of the Titanic that night the catastrophe happened. So you decide to search in Wikipedia for that information.

Learn: You are from Berlin but one week ago you started to study in Regensburg. Next weekend your parents are visiting. You are planing to show them the city and for that reason you want to know more about Regensburg's history and check Wikipedia to find out.

Browse: You have some minutes left till your class starts but you are already sitting in the lecture hall. So you have the chance to browse in Wikipedia for about ten minutes and look at whatever you like.

Figure 4: Examples of the tasks assigned to study participants.

| | Mean length of saccades | | scroll frequency | |
	long	short		
TX	25	79	14	90
PI	35	5	25	15
IN	6	0	0	6
IB	3	0	1	2
IG	10	0	4	6
WI	2	8	0	10
LI	16	7	9	14
HD	20	18	10	28
	$\chi^2 = 76.7655$		$\chi^2 = 43.8878$	
	df = 7		df = 7	
	$p: = 6.3 \cdot 10^{-14}$		$p: = 2.247 \cdot 10^{-7}$	

Figure 5: Influence of saccade distance and scroll frequency on the preferred elements.

user's eye focus at any point of time during the sequence. We annotate these videos manually with two kinds of labels: First one concerning the user action, second one the focussed element of content. So every frame gets two labels, e.g. for reading a passage of text (LE:AA) and for gazing a picture (BE:BI). The full set of labels for actions and the elements of content is presented in Fig. 3.

- Browser Logs:

  `mousemove`, `click`, `scroll`, `keydown` events were recorded to track the user's interaction behaviour.

- Eyetracking Logs:

  The eyetracking software provides data about the frequency and duration of saccades (i.e. the user is moveing his eyes) and fixations (i.e. the user is focussing something with his eyes).

For all kinds of data, each item is assigned a time stamp that we use to align data from different sensors chronologically.

3 Evaluation of the Data

Logged items occur at varying frequency for the different sensors used. Therefore, it is impossible to align individual items chronologically. Instead, we divide the time span for each experiment in constant frames of 500 milliseconds. To each of the frames the set of items is assigned whose timestamp falls into the frame. The items in the set mostly are of different type and occur in varying frequency per frame. Therefore, in order to characterize each frame for each sensor, we construct a probability distribution for each frame and each sensor.

3.1 Predicting Focussed Elements

With this data available, we can now formulate our hypothesis formally. We expect that some of the browser events and eyetracking events are good predictors for the user action taking place in the same frame and for the focussed element of content (assessed by the labelled eyetracking videos).

In the current state of our work, we processed the data for 28 min 46 sec of *browsing tasks* for which we have already annotated the screen recordings produced by the eyetracking software. From the logged mouse and keyboard events, we dropped those occurring on average more rarely than 10% of the frame length. For all others, we built dichotomous random variables. The relative frequencies were split into a group below and another group above average. The eyetracking logs were preprocessed in the same way.

In order to evaluate the predictive power of the new random variables for the preferred user action and the preferred content element respectively, we constructed contingency tables and performed χ^2-tests. Due to space constraints, in the figures below, we only list the variables leading to significant results of the χ^2-tests. The numbers in Fig. 5 indicate that there is a strong influence of the average distance between two fixated points on the screen and the element of content focussed at that time. Short distances indicate that the user is very probably reading text, while longer distance give evidence for the user to just scan roughly and quickly over the content presented. The data shows that in these cases users very often look at headlines or images of the individual sections of a web page. They rarely look at the info boxes (a

Contingency table of the variables *user action* and *mean length of saccades*

	length of saccades		scroll frequency	
	long	short	high	low
EX	45	5	27	23
NV	2	8	0	10
RE	31	79	14	96
SC	39	25	22	42
	$\chi^2 = 59.608$		$\chi^2 = 35.3898$	
	df = 3		df = 3	
	$p: = 7.1 \cdot 10^{-13}$		$p: = 1.0 \cdot 10^{-7}$	

Figure 6: Influence of saccade distance and scroll frequency on the preferred user activity.

kind of table of contents) or the navigation bars.

The relative scroll frequency in a frame seems to be another good indicator for the user to be reading text (see the numbers in Fig. 5). Similarly to the saccade distance, in cases with high scroll frequency the user looks briefly at headlines or images.

3.2 Predicting the Current User Activity

In Fig. 6, we give an overview of the most significant factors that modify the user preferences for activities within a frame. The data indicates a strong correlation between the element type discussed up to now and the actions performed during the information seeking process. Headlines and images correlate with the *scan* and *look at* actions. For scanning the average saccade length is high while it is short for *read*.

For the scroll frequency, there is a similar correlation to be observed between *read*, *text*, and *low* scroll frequency as for the mean saccade distance. On the contrary the scroll frequency tends to be high when users just look at pictures and images.

The evaluation results in Fig. 5 and Fig. 6 are in line with other results presented in the literature and provide further evidence that cognitive activities of users while they are consuming documents can be predicted from web browser interactions.

3.3 Predicting Transitions between User Activities

While the analysis presented so far in the preceeding sections confirms previous results about the observability of user activities during web browsing, a step towards the dynamic composition of content may be achieved by analyzing the transitions between user activities recorded in our experiments.

In Sect. 2.1, we have introduced for this purpose a set of labels with which we annotate the recorded

search processes chronologically (see Sect. 2.3). The result of the annotation is a sequence of labels for user activities spanning the whole recorded search process (an exemplary excerpt of such a sequence can be seen in Sect. 2.1). From these sequences, the probabilities $P(\alpha_i | \alpha_{i-1}, C)$ introduced in Sect. 1.4 and 2.1 can be estimated by the relative frequencies of the conditional events

$$(\alpha_i = x | \alpha_{i-1} = y)$$

from the labelled sequences where x and y are events as defined in Sect. 2.1. In total, with four cognitive activities taking one out of nine element types as argument, theoretically there are 36 possible events (e.g. RE-TX – *read text*). Some of them can never occur (such as RE-PI – *read a picture*).

We performed these calculations for the three possible values of the context factor C, i.e. for the three task types *browse*, *learn*, and *lookup*. The resulting conditional probability distributions are sketched as transition graphs in Fig. 7, 8, and 9. In order to keep the graphs readable, edges are omitted if the transition probability $P(\alpha_i | \alpha_{i-1})$ is smaller or equal to 0.03.

The benefit of these distributions is the following: if C and α_{i-1} can be observed from the user behaviour as described in the preceeding sections, an hypothesis for α_i (i.e. the next user activity) can be calculated by:

$$\max_x P(\alpha_i = x | \alpha_{i-1} = \hat{\alpha}, C)$$

($\hat{\alpha}$ is the observed last user activity).

We argue in this paper, that with x assigned one of the 36 events and the a posteriori probability of x calculated as above, we are able to quantify conversational maxims based on observing the user's preferred reading experience. The combinations of user activities and content elements encode different selections for quantity and manner. *read* and *examine* require more detail, while *scan* requires much less. Depending on the type of perferred content elements, the spread in quantity and manner becomes even greater. Scanning headlines stands for extremely few quantity while reading text is on the opposite side of the scale. The manner of presentation can be varied by changing the type of content elements appropriately. Scanning text is a different manner to consume information than scanning info graphics that contain the same information as text, but present it visually.

This said, in our view selecting the optimal combination of cognitive activity and related content element, provides means to decide about the best way to satisfy conversational maxims at any time of a search process. The decision is based on the user's preferred reading experience as learned from the training material. Learning these preferences was the main purpose of our experiments. As the preferences are estimated
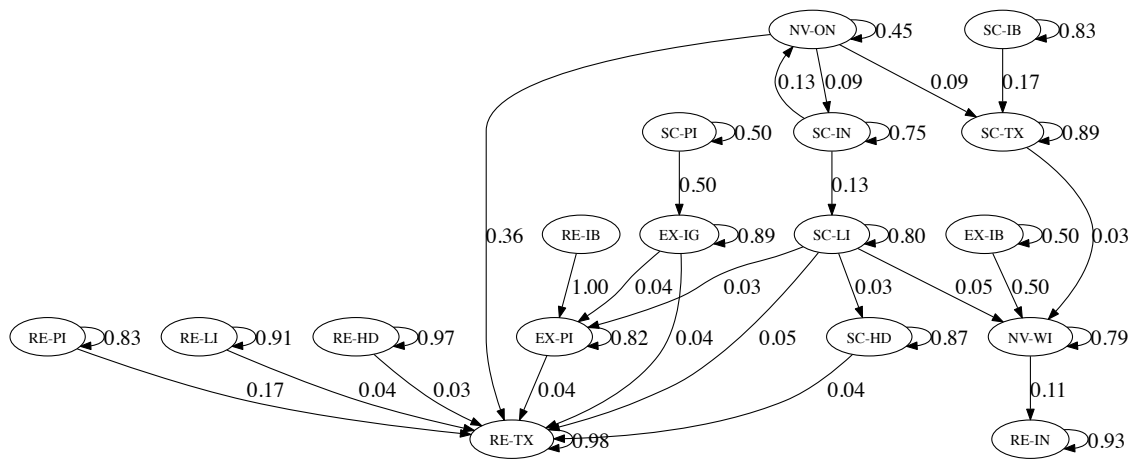
Figure 7: Most probable transitions between user activities in the task type *browse*
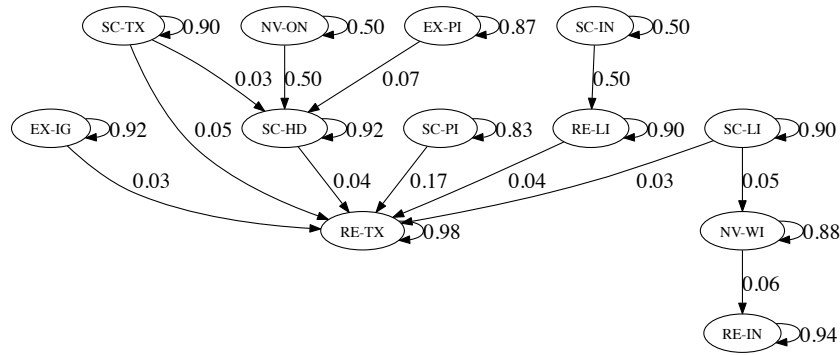

Figure 8: Most probable transitions between user activities in the task type *learn*
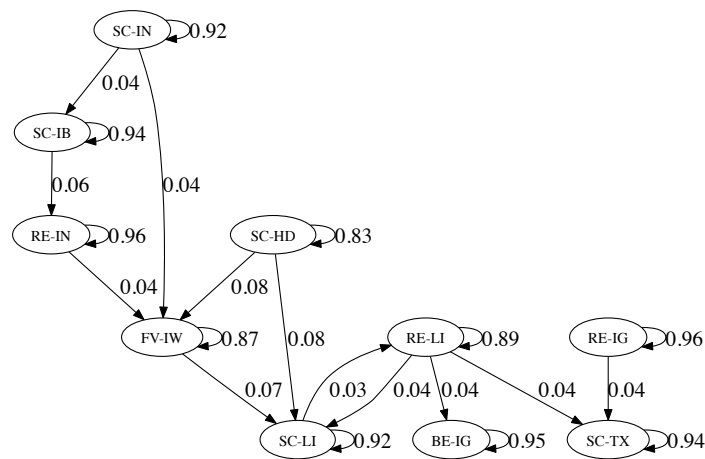

Figure 9: Most probable transitions between user activities in the task type *lookup*

in terms of probabilities from relative frequencies, it is even possible to constantly adapt them to the user's behaviour as observed while he is using the information system.

4 Predicting Activities in Complex Contexts

The decision procedure as presented up to now underlies two limitations:

- The context is limited to one time-invariant factor.
- The procedure computes just one user activity in advance.

In a very simplistic interface, the second limitation can be overcome by presenting a set of content elements matching the most probable next user activity (e.g. a set of headlines for SC-HD) and then asking the user if he needs more information. In this case, planning a sequence of events is substituted by a reactive approach as introduced in Sect. 1.1. More sophisticated solutions would try to predict even a series of user activities, an approach that implies planning capabilities to be available. We will discuss planning of sequences in Sect. 5.

Here, we focus on the first limitation. In Sect. 1.4, we introduced a set of context factors that are potentially relevant for user preferences. Other factors come from the current state of the search process and the coherence structure of the (original document):

- The number of times a content element has already been presented in the current search process. The higher this number, the lower the user's preference for looking at it again.
- The fact whether a content element is satisfaction preceeded in the sense of (Grice, 1975) by another element that has not been presented so far. Skipping elements that are necessary for understanding the document is dispreferred as it is opposed to the maxims of quality and relevance.

The current values for these context factors can be computed effectively during the search process as the values for the other context factors in Sect. 1.4 can. However, while for the simple context consisting just of the task type of the search process computing the a posteriori probability for α_i conditioned on the task type can be achieved by maintaining one probability distribution for each type (as have we done in Fig. 7, 8, and 9), for sets of context factors this approach becomes intractable as the number of distributions grows exponentially with the number of values for the context factors. Therefore, we need a more sophisticated way to compute $P(\alpha_i|\alpha_{i-1},C_{i-1})$ if we assume

C_{i-1} to be a set of context factors whose values may even change at any step of the search process. For that purpose, we apply techniques from probability theory – in particular from BAYESIAN filtering (see (Thrun et al., 2005), chap. 2.4 for an introduction). First, we rewrite

$$P(\alpha_i|\alpha_{i-1},C_{i-1}) = \frac{P(\alpha_i,\alpha_{i-1},C_{i-1})}{P(\alpha_{i-1},C_{i-1})}$$

Define: $\eta^{-1} := P(\alpha_{i-1},C_{i-1})$
$$= \eta \cdot P(C_{i-1}|\alpha_i,\alpha_{i-1})$$
$$\cdot P(\alpha_i|\alpha_{i-1}) \cdot P(\alpha_{i-1})$$

Next, α_i cannot influence the context that results from the *preceeding* activity α_{i-1}. So, we do not have to condition C_{i-1} on α_i.
$$= \eta \cdot P(C_{i-1}|\alpha_{i-1})$$
$$\cdot P(\alpha_i|\alpha_{i-1}) \cdot P(\alpha_{i-1})$$

Finally, for the sake of tractability, we assume – approximatively – the factors in C_{i-1} to be independent from each other. Let C_{i-1} consist of the $1 \leq k \leq K$ factors C_{i-1}^k:
$$= \eta \cdot \prod_{k=1}^{K} P(C_{i-1}^k|\alpha_{i-1})$$
$$\cdot P(\alpha_i|\alpha_{i-1}) \cdot P(\alpha_{i-1})$$

As we see, the initial a posteriori probability has been decomposed into a number of factors:

- The probability $P(C_{i-1}^k|\alpha_{i-1})$ for a certain value for each of the K context factors as an effect of the observed last user activity,
- the probability $P(\alpha_i|\alpha_{i-1})$ of a certain activity in step i to follow the activity observed in the last step, and
- the probability $P(\alpha_{i-1})$ of a certain activity to have occurred as the last user activity.

All these probabilities can be estimated from training data.

Computing the most probable activity in step i of the search process is the most general way to predict what the user prefers to do next. In this way, even for complex contexts conversational maxims can be applied effectively for the automatic stepwise composition of pages to present from an arbitrarily structured and complex original document.

5 Planning Sequences of User Activities

In Sect. 1.2, we introduced the idea to evaluate sequences of user activities and content elements by calculating their a posteriori probabilities. As known

from the theory of MARKOV chains (see e.g. (Manning and Schütze, 1999), chap. 9), this idea may be applied to find an optimal sequence by computing

$$\max_{\alpha_1,\dots,\alpha_t} \prod_{1 \le i \le t} P(\alpha_i | \alpha_{i-1}, C_{i-1}).$$

This algorithm is computationally intractable as it has to compare *all* sequences of length t which grow exponentially in number with t. However, there is an efficient approximative solution to this problem: the VITERBI algorithm (see (Manning and Schütze, 1999), chap. 9.3.2). It allows to compute the desired probabilities iteratively using a greedy technique:

$$\delta_\alpha(1) = P(\alpha | C_1)$$
$$\delta_\alpha(t+1) = \max_\beta \delta_\beta(t) \cdot P(\alpha | \beta, C_t)$$

where α and β range over all defined combinations of user activities and content elements. To give a simple example, for $C_i = lookup$ ($i \ge 0$) we have

$$\delta_{\text{SC-IN}}(1) = 0.25$$
$$\delta_{\text{SC-IN}}(2) = \max_\beta \delta_\beta(1) \cdot P(\text{SC-IN} | \beta, C_1)$$

According to the probabilities given in Fig. 9, we have to compute $\delta_{\text{SC-IN}}(2)$ for $\beta \in \{\text{SC-IN}\}$ as there is no other edge ending in SC-IN. This leads to

$$= \max_{\beta \in \{\text{SC-IN}\}} \delta_\beta(1) \cdot P(\text{SC-IN} | \beta, C_1)$$
$$= \delta_{\text{SC-IN}}(1) \cdot P(\text{SC-IN} | \text{SC-IN}, C_1)$$
$$= 0.25 \cdot 0.92.$$

While for details of the algorithm the reader is referred to (Manning and Schütze, 1999), we want to observe that the probability distribution $P(\alpha | \beta, C_t)$ is the key in VITERBI's algorithm. It is just the distribution that we estimated in Sect. 3.3. We conclude that our data evaluation provides the main ingredient for the problem stated in the introduction of this paper. With the additional help of VITERBI's algorithm it is now possible to recombine selections of content elements from an arbitrarily long document into a user-, task-, and context-adpative excerpt of the initial document.

6 Conclusions

In this paper, starting from the more practical problem of information overload when users navigate web pages, we made the claim that this overload could be reduced if web pages were tailored to the current information need and situational context factors of the current navigation process. We postulated that this tailoring has to be performed by taking conversational maxims (such as those specified by (Grice, 1975)) into account.

6.1 Quantified Conversational Maxims

Such a postulate requires firstly to segment documents into reasonable content elements while preserving their coherence relations and secondly to understand how the conversational maxims can be used as rules for selecting and eliminating content elements from the original document. To the best of our knowledge, such a formal model of conversational maxims has not been presented in the literature so far.

The approach taken in this paper to find a quantitative reformulation of the conversational maxims is to learn preferences of a user's reading behaviour in experiments where test persons have to carry out search tasks in a controlled environment. These preferences can then be interpreted as a user-adpative model of conversational maxims as they provide evidence what kind of and how much information the user prefers to consume in which modalities.

As user preferences are non-deterministic in nature, we choose a probabilistic model to formalize them. In order to assess the necessary probabilities we labelled screen recordings collected in the experiments with user activities and the content elements focussed by the test persons. In parallel to these labels, we logged interaction and eye tracking data. The complete data set allows us to firstly reconstruct the reading experience of the test persons in terms of label sequences documenting the search process. Secondly, we can understand how to predict the reading experience from the test person's web browser interactions. A statistical analysis of our data indicates with high significance that these predictions are highly reliable – these findings are backed up by other researchers.

6.2 Preference-based Presentation

From the label sequences, we derive a probabilistic model of the user's cognitive activities. Our claim is that by selecting content elements from a document in a way that fits the learned probabilistic model optimally we can implement the conversational maxims in a task-, context- and user-adaptive way. Currently, we are building a prototype system in the aforementioned eHealth domain that applies the data we got from the experiments. The system is intended as an evaluation platform in the next series of experiments that we have scheduled.

Finally, in this paper we presented a decision procedure that can perform the necessary computations effectively and efficiently. With this algorithm available, we can provide evidence that – as claimed initially – it is possible to automatically tailor multimodal documents to the current information need and

context factors and thereby address the problem of information overload.

On the basis of this decision procedure, we presented a second algorithm that is capable of planning optimal sequences of cognitive activities. With this algorithm, it is practicable to generate multi-modal documents of arbitrary length from a set of content elements with known coherence relations between them. We note that our approach does not apply any domain knowledge as e.g. that by (Zukerman et al., 2008). We see approach as a shallow method in which the domain knowledge is encoded implicitly in the graph of coherence relations. This graph has to be defined by a domain expert. In our view, our approach is best suited for document collections of a well-defined domain of restricted size. Examples could be mobile city guides on smart phones or the eHealth application mentioned at the beginning of this paper.

REFERENCES

Agichtein, E., Brill, E., Dumais, S., and Ragno, R. (2006). Learning user interaction models for predicting web search result preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 3–10, New York, NY, USA. ACM.

Buscher, G., Dengel, A., Biedert, R., and Elst, L. V. (2012). Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Trans. Interact. Intell. Syst.*, 1(2):9:1–9:30.

Cutrell, E. and Guan, Z. (2007). What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 407–416, New York, NY, USA. ACM.

Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Syntax and semantics*, volume 3. New York: Academic Press.

Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Guo, Q. and Agichtein, E. (2010a). Ready to buy or just browsing?: detecting web searcher goals from interaction data. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 130–137, New York, NY, USA. ACM.

Guo, Q. and Agichtein, E. (2010b). Towards predicting web searcher gaze position from mouse movements. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA '10, pages 3601–3606, New York, NY, USA. ACM.

Hauger, D., Paramythis, A., and Weibelzahl, S. (2011). Using browser interaction data to determine page reading behavior. In *Proceedings of the 19th international conference on User modeling, adaption, and personalization*, UMAP'11, pages 147–158, Berlin, Heidelberg. Springer-Verlag.

J. Huang, R.W. White, G. B. (2012). User see, user point: gaze and cursor alignment in web search. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 1341–1350, New York, NY, USA. ACM.

Jannach, D., Zanker, M., Felfernig, A., and (Autor), G. F. (2011). *Recommender Systems: An Introduction*. Cambride University Press.

Kellar, M. and Watters, C. (2006). Using web browser interactions to predict task. In *Proceedings of the 15th international conference on World Wide Web*, WWW '06, pages 843–844, New York, NY, USA. ACM.

Lindley, S. E., Meek, S., Sellen, A., and Harper, R. (2012). "it's simply integral to what i do": enquiries into how the web is weaved into everyday life. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 1067–1076, New York, NY, USA. ACM.

Manning, C. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge.

Rodden, K., Fu, X., Aula, A., and Spiro, I. (2008). Eye-mouse coordination patterns on web search results pages. In *CHI '08 extended abstracts on Human factors in computing systems*, CHI EA '08, pages 2997–3002, New York, NY, USA. ACM.

Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics*. MIT Press.

Zukerman, I., Makalic, E., Niemann, M., and George, S. (2008). A probabilistic approach to the interpretation of spoken utterances. In Ho, T. B. and Zhou, Z.-H., editors, *PRICAI*, volume 5351 of *Lecture Notes in Computer Science*, pages 581–592. Springer.