Reading Scanpaths

Barbara Rösch, Rainer Hammwöhner

Universität Regensburg, Lehrstuhl für Informationswissenschaft
Universitätsstraße 31, 93053 Regensburg, Germany
{barbara.roesch, rainer.hammwoehner}@ur.de

Abstract

Research Questions: Our aims are on the one hand to find differences in users' information behavior reading Wikipedia articles and on the other to evaluate different methods commonly used to compare scanpaths.
Approach: We conducted eye-tracking experiments and exported scanpaths from raw data to identify gaze patterns of different user groups, considering the parameters pre- and post-knowledge, learning style and subject of study.
Method: We applied string distance measures and cluster analysis to investigate scanpath sequences, fixation times and the distribution of the visual attention.
Results: Despite encountering difficulties in comparing scanpaths, there are indications that the users' knowledge influences their gaze patterns.

Keywords: Scanpath, Eye-tracking, Cluster analysis, Distance metric

1 Introduction

In today's networked society the online-encyclopedia Wikipedia is used anytime and anywhere for private as well as professional purposes. Kim et al. (2013) reported that undergraduate and graduate students use Wikipedia for entertainment as well as for information seeking in academic context. The German version contains 1,787,400 articles and 2,035,351 registered users (Wikipedia 2014). Every day millions of users query Wikipedia with various

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 290–301.

information needs and intentions (Mesgari et al. 2015). The online-encyclopedia is a widely used subject of research especially in the areas of natural language processing, information retrieval, ontology building and information extraction (Medelyan et al. 2009). A review of research concerning the content quality and the size of Wikipedia articles gives Mesgari et al. (2015). The perception, the use and the motivation of college students was examined by Lim (2009) in order to understand their information behavior on sites of Wikipedia applying social cognitive theory. All students used Wikipedia, in particular for getting background information. The users' information needs were satisfied, but they had a moderate perception of the information quality (Lim 2009).

Experiments using eye-tracking technology for implicit relevance feedback while reading text passages were conducted by Buscher et al. (2012). They stated that gaze-based feedback is appropriate to find out whether the read text was relevant to an individual user during web search.

Our aim is to detect individual characteristics of users' interaction with the elements presented in Wikipedia articles in order to support the personalization of information presentation and their reception. To observe users' behavior while reading a Wikipedia article eye-tracking methodology was used.

2 The role of scanpaths

The common output of eye-tracking experiments consists of fixations and saccades. It is assumed that the location and the duration of the fixations are related to cognitive information processing (Just & Carpenter 1980). Beside these synchronic indicators diachronic variables like scanpaths or saliency maps are underused (Le Meur & Baccino 2012). Eye-tracking evaluation software allows processing the raw data by computing the fixation start and end time and their positions or by providing graphical displays, like heatmaps or visualized scanpaths. However those tools do not include automatic analysis and comparison of different scanpaths. The comparison of scanpaths is of growing interest, but there is little known about groupwise similarity measures and classification algorithms (Grindiger et al. 2010).

Due to our research context, we are interested in comparing the fixation sequences of users reading Wikipedia articles. A scanpath is defined as the order of the fixation of certain elements (for example text, image or subtitle) in the presented stimulus and thus serves as a surrogate for the order of the users' interaction with the information and for the relevance of the presented elements. Another aim is to identify patterns of recurring sequences or contact points as well as transitions, for example between corresponding textual and pictorial information.

The fixation times of the elements and the distribution of the visual attention are of further interest. The challenging task we are facing with is to detect similarities, to compare them and to find groups among the data, which will deepen our knowledge of users' information behavior.

3 Experimental settings

This approach of reading scanpaths is related to a project highlighting the relationship between textual and pictorial elements in Wikipedia articles. A particular interest is in the role of images, their functions regarding attention, comprehension, and retention as well as the spread of information within an article, i.e. information contained in the text, images or combinations of both. Another important research topic is the impact of different user groups on the information behavior and the reception of the articles' information. In this study the parameters 'visual' or 'verbal' learning style, pre- and post-knowledge and subject of study were collected. Currently, we are investigating the differences in users' interaction with articles, e.g. fixation times on textual and pictorial elements or sequences of several elements. The aim is to identify specific patterns and to compare and categorize them in order to get hints for the optimal design for specific information needs and user groups.

Wikipedia serves as a test environment because of its accessibility and its standardized format of text and images. Spoerri (2007) found out that the major topics of interest were 'entertainment' and 'politics/history' among the 100 most visited articles. The topic 'history' is well structured with a huge number of articles containing different image types like maps, portraits and black and white photographs. For these reasons we chose six articles about Bavarian history in the 20th century (King Ludwig III., Kurt Eisner, Kapp-

Putsch, Hitlerputsch, Bavarian history and Kingdom of Bavaria) in cooperation with an expert from the local chair of history. The articles were shortened to be comparable and all hyperlinks were removed. However, the layout of the textual and pictorial elements remained unchanged.

To get insights into the users' behavior an eye-tracking study with 91 participants, students of information science and history, was conducted in the eye-tracking laboratory at the chair of information science. The two groups were chosen to get different levels of prior knowledge. Data analysis after the experiment confirmed that the prior knowledge differs significantly. Students of information science had less knowledge about the topic before the experiment. The SMI 50 Hz remote eye-tracking device recorded the students' eye-movements – saccades and fixations – while reading an article. This method gave us information about the fixated elements, the sequences and the fixation times. All participants were instructed to read two of the six articles mentioned above and were given the fictitious task to prepare the topic for a course at the university. Afterwards they were asked to answer several questions about retention and comprehension. Additionally, they had to fill in a demographic questionnaire, a learning style test and a pre-knowledge test.

4 Methods

As we are interested in scanpaths in this context, we need to use appropriate measures enabling us to evaluate our data statistically. Thus we tested different methods for the analysis of scanpaths, which can be modeled as graphs or random processes. To characterize them as graphs, variables like the sequences, the length of saccades or the angles of the fixations and saccades can be used. We focus on the comparison of the raw sequences of *areas of interest* (AOIs) with different metrics like the Levenshtein distance or the Needleman-Wunsch algorithm. For the analysis of n-grams metrics like Dice, Jaccard or the longest common subsequence were applied. Those measures result in a matrix of dissimilarities and can be clustered hierarchically or partitioning in a second step to explore existing patterns of scanpaths. Besides the sequences more abstract indicators like entropy, the distribution of the spatial attention or the fixation times were analyzed as well.

4.1 String comparison measures

Scanpaths are coded as strings (see section 5) and thus require specific methods like the Levenshtein distance or the Needleman-Wunsch algorithm. The first one compares two strings transforming them into one another by applying the operations substitution, insertion, and deletion (Levenshtein 1966). Each step needed for this transformation increases the count for its Levenshtein value by 1. The higher the generated value, the more the strings differ. An output of 0 represents identical scanpaths. The values computed by the Levenshtein distance for each pairwise scanpath comparison can be transformed in a distance matrix.

The second measure to compare strings is the Needleman-Wunsch algorithm originally used in bioinformatics for DNA sequences (Needleman & Wunsch 1970). This measure was implemented in the ScanMatch toolbox for scanpath comparison (Christino et al. 2010). The alignments gap and substitution were associated with certain costs during the global sequence comparison. Based on these costs a score between 0 and 1 is computed, normalized by the length of the sequences. The maximum value of 1 indicates the similarity of two strings. For further analysis the scores are converted into a symmetric matrix of dissimilarities.

4.2 Clustering processes

Based on computed dissimilarity matrices, cluster analysis can be applied to detect structures and find group similarities within the data. There are hierarchical and nonhierarchical cluster processes.

The hierarchical clustering method identifies and combines the two closest observations, then the second closest pair and so on. This repetitive process clusters the observations stepwise into larger units (Hair et al. 2010). The range spans from the lowest level of one observation per cluster to the highest level containing all observations in one cluster. This pairwise clustering of distances operates with Ward's method calculating the sum of squares within the clusters summed over all observations (Backhaus et al. 2011).

The graphical output is commonly displayed as a dendrogram, a treelike graph. This visualization points out the number of clusters for each hierarchical level and may help to find the appropriate number of clusters.

Nonhierarchical clustering algorithms are partitioning processes (like k-means or k-medoids). In contrast to hierarchical methods, the number of clusters has to be set in the beginning. PAM, *partitioning around medoids*, also known as k-medoid distributes values of a distance matrix into a predefined number of clusters (Kaufman & Rousseeuw 1990). Clustering via grouping with medoids was proven to be a very robust method (van der Laan 2002). In contrast to k-means, which only uses with the Euclidean or Manhattan distance, PAM allows clustering with any distance metric.

The validity of clusters is measured by silhouette based on tightness and separation (Rousseeuw 1987). The silhouette width indicates whether an object is well-clustered. The average silhouette width is computed for each single cluster and for the entire dataset. The corresponding plot graphically displays the silhouettes and their values. The appropriate number of clusters can be selected by maximizing the average silhouette width. The values can reach a maximum of 1 and a minimum of -1.

5 Data preparation

The first step of data analysis was the preparation of raw eye-tracking data, fixations and saccades, in order to get scanpaths. Therefore, every article was divided into several areas, which are specific for the structures of the articles like the header, paragraph one, image one, subtitle one, introduction or table of contents. Every element was defined as an AOI in the eye-tracking software *BeGaze* for data analysis. This division into AOIs allows the assignment of fixations and their export for every person.

As we were interested in sequences of fixated areas, repeated fixations in one area, as they result by reading a text, were not considered. The different AOIs were labelled with a combination of letters and numbers, e.g. A1 corresponds to the header, B1 to the introduction, C1 to the table of contents and so on. By this sequence we got a string for every person (e.g. A1B1C1A1D1B1). We then analyzed and compared the different scanpaths of each article using the measures explained above, the Levenshtein distance and the Needleman-Wunsch algorithm.

Figure 1 shows the first part of the assigned AOIs of the article ‘King Ludwig III.’ and an idealized section of a scanpath. This fictitious person starts reading the introduction, switches between the table of contents and the introduction, reads paragraph one, takes a look at the first image and then reads paragraph two.

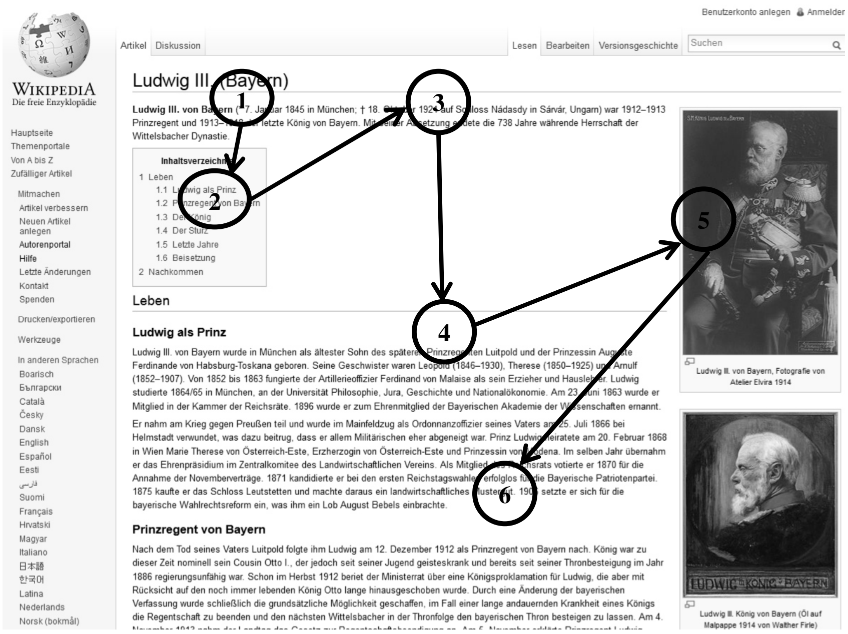

Figure 1. Idealized scanpath (article 'King Ludwig III.')

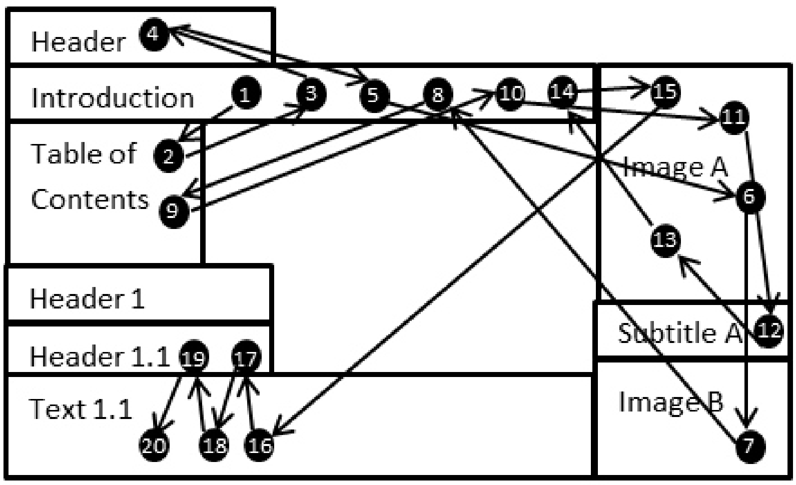

Figure 2. A user's recorded scanpath over a scheme of the article 'King Ludwig III.'

The collected data reveal that the users' gazes jump back and forth between the presented elements while reading a Wikipedia article. Figure 2 shows the first 20 contact points of a test person scanning the article 'King Ludwig III.'

In these sequences only the fixated area but not the fixation times on the corresponding AOIs were considered in the first approach. In a further step the fixation times were computed and normalized for every AOI.

6 Data analysis

6.1 Sequences

Scanpaths were compared to each other resulting in dissimilarity matrices both for the Levenshtein distance and the Needleman-Wunsch algorithm for each article. The matrices were clustered hierarchically as well as with partitioning methods. Figure 3 shows a dendrogram of the hierarchical clustering using the Levenshtein distance of the article 'King Ludwig III.' with 28 scanpaths. To ascertain the optimal number of clusters silhouette was applied. The validation of the cluster performs best with the division into two clusters. Nine scanpaths were assigned to cluster one and nineteen scanpaths in cluster two. The average silhouette width for the entire data set is 0.34. Cluster 1 has an average silhouette width of 0.05, cluster 2 of 0.47.

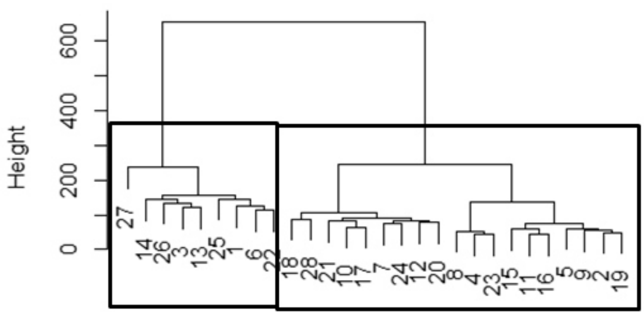

Figure 3. Dendrogram of the hierarchical clustering with the Levensthein distance of 28 scanpaths

The cluster procedure PAM was applied to the same data set of dissimilarities with two clusters. Cluster one contains 21 (silhouette width 0.43), cluster two contains seven scanpaths (silhouette width 0.09). The average silhouette width is 0.34.

The hierarchical cluster analysis of the scanpaths applying the Needleman-Wunsch algorithm showed the average silhouette width of 0.15 (cluster size 8 and 20). The average silhouette of PAM was 0.16 (cluster size 9 and 19). All six articles were analyzed that way.

The application of the Levenshtein metric and the Needleman-Wunsch showed no significant results. The average silhouette widths for cluster evaluation indicate only weak relations. Despite that fact, we have to keep in mind that the scanpaths are heterogeneous and vary strongly, concerning length and sequences. Two scanpaths that match perfectly do not exist. The analyzed scanpaths contain a lot of noise in the form of repetitive sequences.

The hierarchical clustering proved a two-cluster solution. For the following steps we used k-medoids with two predefined clusters.

6.2 Visual attention

In a next step more abstract variables were analyzed, like the fixation times and the distribution of the visual attention.

The spatial distribution of the users' gazes was investigated in order to identify AOIs, for example the beginning, the center or the end of an article. The average distribution of the AOIs was computed by the mean and clustered exemplarily for the article 'King Ludwig III.'. The analysis of the mean distribution shows two clusters with the average silhouette of 0.66 (cluster one: 14 values, silhouette width 0.63 – cluster two: 14 values, silhouette width 0.68). The silhouette coefficient indicates a reasonable structure among the clusters. To evaluate the group assignment extern differences in the collected parameters were tested. The clusters differ significantly according to the t-test for the parameter post-knowledge ($t(26) = 2.32$; $p = 0.028$). There were no significant differences between subject of study, learning style and pre-knowledge. This could be an indication that difference in knowledge after reading an article is caused by the different strategies people use to deal with the presented information.

6.3 Fixation times

Furthermore the normalized fixation times over all AOIs (i.e. fixation time on images, subtitles, introduction, table of content, header and text) and articles were clustered by k-medoid. Two clusters of 30 and 134 values, respectively, were identified. Their average silhouette is 0.90, the silhouette width for cluster one is 0.86, for cluster two it is 0.91. The silhouette coefficient indicates a strong structure among the clusters, but all scanpaths of the article 'Kapp-Putsch' were assigned to one cluster. The differences between the two clusters were analyzed by the nonparametric Wilcoxon rank sum test with the variables mentioned above. They differ significantly by the parameter pre-knowledge ($W = 147$; $p = 0.0019$). The article 'Kapp-Putsch' differs from the other articles concerning the fixation times. This analysis shows that the users' pre-knowledge influences his behavior.

7 Conclusion

In this article we illustrated our initial investigation of differences in users' information behavior reading Wikipedia articles. Additionally, we evaluated methods frequently used to compare scanpaths and further variables.

We applied different low-cost techniques for comparison and clustering of strings. The comparison measures, the Levenshtein distance and the Needleman-Wunsch algorithm, showed no significant dissimilarities among the scanpaths and the silhouette coefficient evaluating the clusters showed only weak structures. However, the analysis of the fixation times and visual attention revealed first indications that the pre- and post-knowledge influences the users' gaze patterns.

The articles and the probands were very homogeneous. The scanpaths contained a lot of repetitive transitions, making them difficult to compare. The analysis of more abstract variables, the fixation time and the spatial distribution of the visual attention delivered better results and need to be followed up. Despite those difficulties, there are hints that the pre-knowledge influences the fixation times on the elements of an article and that the spatial distribution of visual attention correlates with the knowledge after reading.

The performance of preliminary evaluations of n-grams metrics was weak and received no further attention. However, the initial analysis of the entropy of scanpaths was promising and should be pursued.

Future work will also be the analysis of smaller substrings of the scanpaths and combining the sequences with the fixation times. Furthermore, we will conduct experiments in different settings and with different stimuli to ascertain our findings concerning users' scanpaths. Moreover, there might well be parameters influencing gaze behavior which have not been considered here yet.

Our findings provide information on the processes of perceiving a Wikipedia article. We can use to these insights for further research concerning individual characteristics of users' interaction with textual and pictorial elements in order to support the personalization of information presentation.

References

Backhaus, K., Erichson, B., Plinke, W., Weiber R. (2011). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Heidelberg: Springer.

Buscher, G., Dengel, A., Biedert, R., Van Elst, L. (2012). Attentive Documents: Eye Tracking as Implicit Feedback for Information Retrieval and Beyond. *ACM Transactions on Interactive Intelligent Systems*, 1 (2), Article 9.

Christino, F., Mathôt, S., Theeuwes, J., Gilchrist, I. D. (2010) ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, 42 (3), 692–700.

Grindinger, T., Duchowski, A. T., Sawyer, M. (2010). Group-Wise Similarity and Classification of Aggregate Scanpaths. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 101–104). New York: ACM.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. (2010). *Multivariate Data Analysis. A Global Perspective*. New Jersey: Pearson.

Just, M. A. & Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review*, 87 (4), 329–354.

Kaufman, L. & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons.

Kim, K.-S., Sin, S.-C. J., He, Y. (2013). Information Seeking through Social Media: Impact of User Characteristics on Social Media Use. *ASIST 2013, Nov. 1–6, 2013*, Montreal, Quebec, Canada.

Le Meur, O. & Baccino, T. (2012). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research and Methods*, 45 (1), 251–266.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics*, 10 (8), 707–710.

Lim, S. (2009). How and Why Do College Students Use Wikipedia? *Journal of the American Society for Information Science and Technology*, 60 (11), 2189–2202.

Medelyan, O., Milne, C., Legg, C., Witten, I. H. (2009). Mining meaning from Wikipedia. *International Journal of Human-Computer Studies*, 67 (9), 716–754.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. A., Lanamäki, A. (2015). "The Sum of All Human Knowledge": A Systematic Review of Scholarly Research on the Content of Wikipedia. *Journal of the American Society for Information Science and Technology*, 66 (2), 219–245.

Needleman, S. B. & Wunsch, C. D. (1970). A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology*, 48, 443–453.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

Spoerri, A. (2007). What is popular on Wikipedia and why? *First Monday*, 2 (4). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1765/1645 <3.3.2015>.

Van der Laan, M. J., Pollard, K. S., Bryan, J. (2003). A new partitioning around medoids algorithm. *Journal of Statistical Computation and Simulation*, 73 (8), 575–584.

Wikipedia (2014). Seiten- und Benutzerstatistik. http://de.wikipedia.org/wiki/Spezial:Statistik <18.12.2014>.