

# Toward Data-Driven Analyses of Electronic Text Books

Ahcène Boubekki  
ULB Darmstadt/TU Darmstadt  
boubekki@dipf.de

Ulf Kröhne  
DIPF Frankfurt/M  
kroehne@dipf.de

Frank Goldhammer  
DIPF Frankfurt/M  
goldhammer@dipf.de

Waltraud Schreiber  
KU Eichstätt  
schreiber@ku-  
eichstaett.de

Ulf Brefeld  
TU Darmstadt/DIPF  
brefeld@cs.tu-  
darmstadt.de

## ABSTRACT

We present data-driven log file analyses of an electronic text book for history, called the *mBook*, to support teachers in preparing lessons for their students. We represent user sessions as contextualised Markov processes of user sessions and propose a probabilistic clustering using expectation maximisation to detect groups of similar (i) sessions and (ii) users.

## 1. INTRODUCTION

Electronic text books may offer a multitude of benefits to both teachers and students. By representing learning content in various ways and enabling alternative trajectories of accessing learning objects, electronic text books offer great potentials for individualised teaching and learning. Although technological progress passed by schools for a long time, inexpensive electronic devices and handhelds have found their way into schools and are now deployed to complement traditional (paper-based) learning materials.

Particularly text books may benefit from cheap electronic devices. Electronic versions of text books may revolutionise rigour presentations of learning content by linking maps, animations, movies, and other multimedia content. However, these new degrees of freedom in presenting and combining learning materials may bring about new challenges for teachers and learners. For instance, learners need to regulate and direct their learning process to a greater extent if there are many more options they can choose from. Thus, the ultimate goal is not only an enriched and more flexible presentation of the content but to effectively support teachers in preparing lessons and children in learning. To this end, not only the linkage encourages users to quickly jump through different chapters but intelligent components such as recommender systems [4] may highlight alternative pages of interest to the user. Unfortunately, little is known on the impact of these methods on learning as such and even little is known on how such electronic text books are used by students.

In this article, we present insights on the usage of an electronic text book for history called the *mBook* [5]. Among others, the book has been successfully deployed in the German-speaking Community of Belgium. We show how data-driven analyses may support history teachers in preparing their lessons and showcase possibilities for recommending resources to children. Our approach is twofold: Firstly, we analyse user sessions to find common behavioural patterns across children and their sessions. Secondly, we aggregate sessions belonging to the same user to identify similar types of users. This step could help to detect deviating learners requiring additional attention and instructional support.

## 2. THE MBOOK

The *mBook* is guided by a constructivist and instructional-driven design. Predominantly, the procedural model of historical thinking is implemented by a structural competence model that consists of four competence areas that are deduced from processes of historical thinking: (i) the competency of posing and answering historical questions, (ii) the competency of working with historical methodologies, and (iii) the competency of capturing history's potential for human orientation and identity. The fourth competency includes to acquire and apply historical terminologies, categories, and scripts and is best summarised as (iv) declarative, conceptual and procedural knowledge.

Imparting knowledge in this understanding is therefore not about swotting historic facts but aims at fostering a reflected and (self-)reflexive way of dealing with our past. The underlying concept of the multimedia history schoolbook implements well-known postulations about self-directed learning process in practice. The use of the *mBook* allows an open-minded approach to history and fosters contextualised and detached views of our past (cf. [3]). To this end, it is crucial that a purely text-based narration is augmented with multimedia elements such as historic maps, pictures, audio and video tracks, etc. Additionally, the elements of the main narration are transparent to the learners. Learners quickly realise that the narration of the author of the *mBook* is also constructed, as the author reveals his or her construction principle.

## 3. METHODOLOGY

For lack of space, we only sketch the technical contribution. We devise a parameterised mixture model with  $K$  components to compute the probability of a user session. The

browsing process through chapters is modelled by a first-order Markov chain so that pages are addressed only by their chapter. The category model depends on the chapters as we aim to observe correlations between different types of pages. This may show for example whether galleries of some of the chapters are more often visited (and thus more attractive) than others and thus generate feedback for the teachers (e.g., to draw students attention to some neglected resources) and developers (e.g., to re-think the accessibility or even usefulness of resources). The model for the connection times is inspired by the approach described in [2] to capture repetitive behaviour across weeks. The final model is optimised by an EM-like algorithm.

## 4. EMPIRICAL RESULTS

In our empirical analysis, we focus on about 330.000 sessions collected in Belgium between March and November 2014 containing approximately 5 million events.

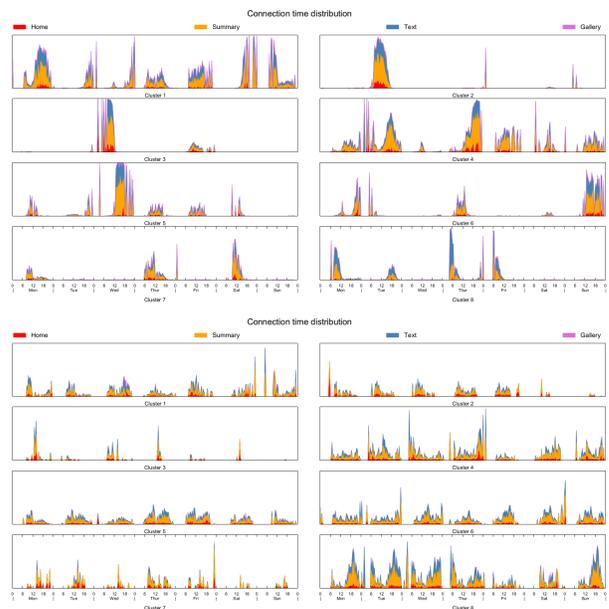
**Session-based View:** Figure 1 (top) shows the results of a session-based clustering. User sessions are distributed across the clustering according to the expressed behaviour. Clusters can therefore be interpreted as similar user behaviours at similar times. The visualisation shows that all categories are clearly visible for all clusters, indicating a frequent usage of all possible types of resources by the users. Cluster *C6* possesses half of the mass on the weekend of category *text*. This indicates more experienced users who like to form their opinion themselves instead of going to summary pages. The same holds for cluster *C8* that possesses in addition only a vanishing proportion of the *home* category. Small probabilities of category *home* as well as large quantities of category *text* indicate that users continuously read pages and do not rely on the top-level menu for navigation.

**User-based View:** Our approach can also be used to group similar users. To this end, we change the expectation step of the algorithm so that sessions by the same user are processed together. That is, there is only a single expectation for the sessions being in one of the clusters. Clusters therefore encode similar users rather than similar behaviour as in the previous section.

Figure 1 shows the results. Apparently, the main difference of the clusters is the intensity of usage during working days and weekends. Cluster *C2* for instance clearly focuses on working day users who hardly work on weekends compared to Cluster *C1* whose users place a high emphasise on Saturdays and Sundays. Cluster *C3* contains low frequency users who rarely use the mBook and exhibit the smallest amount of sessions and page views per session. Cluster *C8* contains heavy (at night) users with high proportions of category *text*. In general, we note that transition matrices are consistent between chapters in contrast to the session-based clustering, that is, test takers interact with most of the chapters.

## 5. DISCUSSION

Our results illustrate potential benefits from clustering learners for instructional purposes. In the first place, the probabilistic clustering approach shows a way how to condense a huge amount of logfile information to meaningful patterns of learner interaction. Classifying a student into one of several clusters reveals whether, when, and how the learner used



**Figure 1: Resulting clusters for the session- (top) and user-based (bottom) clustering.**

the materials offered by the electronic text book. Thus, the teacher can get information about the learners' navigation speed, whether part of the content was used in self-directed learning processes as expected, whether learners came up with alternative learning trajectories, and so on and so forth. This information can be used by the teacher in a formative way (cf. the concept of formative assessment, e.g., [1]), that is, it is directly used to further shape the learning process of students. For instance, in a follow-up lesson the teacher could simply draw the students attention to some parts of the book that have not or only rarely been visited. Moreover, history and learning about history could be reflected in a group discussion of learners who used the mBook resources of a particular chapter in different ways.

## 6. REFERENCES

- [1] P. Black and D. Wiliam. Assessment and classroom learning. *Assessment in Education*, 5(1):7–74, 1998.
- [2] P. Haider, L. Chiarandini, U. Brefeld, and A. Jaimes. Contextual models for user interaction on the web. In *Proc. of the Workshop on Mining and Exploiting Interpretable Local Patterns*, 2012.
- [3] Y. Karagiorgi and L. Symeou. Translating constructivism into instructional design: Potential and limitations. *IFETS*, 8 (1):17–27, 2005.
- [4] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors. *Recommender Systems Handbook*. Springer, 2011.
- [5] W. Schreiber, F. Sochatzy, and M. Ventzke. Das multimediale schulbuch - kompetenzorientiert, individualisierbar und konstruktionstransparent. In W. Schreiber, A. Schöner, and F. Sochatzy, editors, *Analyse von Schulbüchern als Grundlage empirischer Geschichtsdidaktik*, pages 212–232. Kohlhammer, 2013.