Evaluating E-books
Monica Landoni
University of Lugano (USI)
via G. Buffi 13, CH-6904 Lugano, CH
+41 58.666.4300
monica.landoni@unisi.ch

ABSTRACT
In this paper, we discuss the importance of evaluation and in particular the impact of user studies on the progress and improvement of e-books and their usability, with particular attention to their design and functionalities aimed at increasing overall user experience. While there is a consensus on the importance of good design and few authors have already been publishing guidelines and provided advice on good practice, still no much attention has been paid to evaluation and its impact on e-book quality. It would indeed be extremely useful for designers to have a common platform in terms of benchmarks, agreed procedures, criteria and measures to evaluate the impact of their products on users but this is far from happening. The reality is that quite a few researchers are setting up evaluation experiments and collecting evidence of good practice but mostly in isolation from other previous or current similar initiatives. Under the INEX Book Track a specific task called Active Reading Task (ART) has been designed to gather efforts across the e-book community. Over the last two years we have tried to raise awareness in related communities but we are still facing serious problems in terms of participation. Indeed, we understand how demanding taking part in ART maybe, as working with users, real or surrogates, always has a high cost in terms of time and resources, this is why we are working to make the overall experience as rewarding, efficient and painless as possible. We strongly believe that such initiative would bring lots of benefits to the e-book community at large, in terms of improved design and quality of e-books. This paper starts with a description of ART, as part of the INEX initiative. We will then point out known problems with participation and discuss how to increase visibility, awareness and participation to ART by involving as many members of relevant communities as possible and by providing them with support and examples of good practice.

Categories and Subject Descriptors
H.5.2 [User Interfaces]: Evaluation/methodology.

General Terms

Keywords

1. INTRODUCTION
In order to evolve any research community needs an arena where individual and groups can share and compare their results but comparison is not a trivial issue, it can only happen on agreed common grounds and takes quite a lot of hard work to set these up.

The Information Retrieval community is a good example of how beneficial the constitution of a common ground can be, with the establishment of the TElx Retrieval Conference, TREC initiative (trec.nist.gov), research is flourishing as researchers from both academia and industry have a place and a way to discuss and compare, in a very constructive fashion, their achievements. Of course, as for any entity of this sort, there are plenty of controversy and criticisms to the initiative per se but this, again, is promoting discussion and has been very beneficial to the community. It would be ausplicable for the e-book community to establish a similar initiative. We believe that the Active Reading Task (ART) could be the answer to this quest. Built on previous experiences in evaluating electronic resources and framed in a TREC-like context, it has been designed to facilitate running user studies while focusing on e-books and bringing together researchers in this multidisciplinary area. In the next session we introduce briefly INEX and give more details on ART.

2. INEX and The Active Reading Task (ART)
The main goal of INEX is to promote the evaluation of focused retrieval by providing large test collections of structured documents, common evaluation measures, and a forum for organizations to compare their results. It is important to note that INEX facilitates the process by allowing research groups interested in Focused Retrieval to identify themselves, and provides a centralized facility to make collaboration possible. Each year INEX acquires a number of specific tracks to fulfill changing research needs. In particular, the Book Track promotes inter-disciplinary research by looking into techniques for supporting users in reading, searching, and navigating the full texts of digitized books while providing a forum for the exchange of research ideas and contributions. Focusing on topics of interest in the fields of information retrieval (IR), human computer interaction (HCI), digital libraries (DL), and eBooks.

The track builds on a collection of digitized books, provided by Microsoft Live Book Search and the Internet Archive (for non-commercial purposes only). The corpus consists of over 50,000 digitized out-of-copyright books. The OCR content of the books is stored in an XML format, referred to as BookML. Most books also have an associated metadata file (*.mrc), which contains publication (author, title, etc.) and classification information in MArchine-Readable Cataloging (MARC) record format.

The Book Track has in turn a number of tasks for participants to choose and enrol in order to explore issues further. One of these is the Active Reading Task, ART.
ART aims at exploring, with user studies, how and why readers use eBooks in specific scenarios with a focus on eBook usability. The main aim of ART is to study how hardware and/or software tools for reading eBooks can provide support to users engaged with a variety of reading related activities, such as fact finding, memory tasks, or learning. The goal of the investigation is to derive user requirements and consequently design recommendations for more usable tools to support active reading practices for eBooks. The task is motivated by the lack of common practices when it comes to conducting usability studies of e-reader tools. Current user studies focus on specific content and user groups and follow a variety of different procedures that make comparison, reflection, and better understanding of related problems difficult. ART is hoped to turn into an ideal arena for researchers involved in such efforts with the crucial opportunity to access a large selection of titles, representing different genres and appealing to a variety of potential users, as well as benefiting from established methodology and guidelines for organising effective evaluation experiments. The methodology proposed is based on that devised in the EBONI project and is built upon essential HCI principles. It includes a range of well established methods to support usability studies and is in line with recent e-book evaluations [3,4,5,6]. In EBONI we run a number of different comparative experiments across disciplines, formats and devices and our methodology has proved both flexible and robust enough in all these scenarios [7,8]. We also investigated different contexts of reading from education in high schools and universities to research and finally reading for pleasure. Overall, we plan to share our experience with participants both formally, by providing a well structured framework for the design of their experiments and informally, by sharing the advice derived from our experience in this arena. By adopting a commonly well accepted paradigm like in EBONI we would ensure our user studies and their findings will not become obsolete but form, over time, a corpus of evidence for research community to learn and grow.

The ART task is to run one or more user studies in order to test the usability of established products (e.g., Amazon’s Kindle, iRex’s Ilaid Reader and Sony’s Readers) or novel e-readers by following the provided EBONI based procedure and focusing on INEX content. Participants may then gather and analyse results according to the EBONI approach and submit these for overall comparison and evaluation. The evaluation is task-oriented in nature. Participants are able to tailor their own evaluation experiments, inside the EBONI framework, according to resources available to them. In order to gather user feedback, participants can choose from a variety of methods, from low-effort online questionnaires to more time consuming one to one interviews, and think aloud sessions, all with the support and guidance from the ART research team.

2.1 TASK SETUP

Participation requires access to one or more software/hardware e-readers (already on the market or in prototype version) that can be fed with a subset of the INEX book corpus (maximum 100 books), selected according to participants’ needs and objectives. Participants are asked to involve a minimum sample of 15/20 users these will be asked to complete 3-5 growing complexity tasks and fill in a customized version of the EBONI subjective questionnaire, usually taking no longer than half an hour in total, allowing to gather meaningful and comparable evidence.

Additional user tasks and different methods for gathering feedback (e.g., video capture) may be added optionally. A crib sheet (see below) is provided to participants as a tool to define the user tasks to evaluate, providing a narrative describing the scenario(s) of use for the books in context, including factors affecting user performance, e.g., motivation, type of content, styles of reading, accessibility, location and personal preferences.

2.2 ART CRIB SHEET

A task crib sheet is a rich description of a user task that forms the basis of a given user study based on a particular scenario in a given context.

Thus, it aims to provide a detailed explanation of the context and motivation of the task, and all details that form the scenario of use:

- Objectives: A summary of the aims and objectives of the task from the users’ point of view, i.e., what is it that users are trying to achieve in this task.
- Task: Description of the task.
- Motivation: Description of the reasons behind running the task.
- Context: Description of the context of the task in terms of time and resources available, emphasis and any other additional factors that are going to influence task performance.
- Background: Description of any background knowledge required to accomplish the task.
- Completion: Description of how to assess whether the task has been completed or not.
- Success: Description of whether the task has been completed successfully.

Participants are encouraged to integrate questionnaires with interviews and think aloud sessions when possible, and adapt questionnaires to fit into their own research objectives whilst keeping in the remit of the active reading task.

We also encourage direct collaboration with participants to help shape the tasks according to existing research needs. In fact one of the participants explained how English written material was not much use for their experiments as they were targeting Korean speaking users, so it was agreed that they would use their own book collection while still adopting the ART evaluation framework to ensure results were comparable at the end.

Our aim is to run a comparable but individualized set of studies, all contributing to elicit user and usability issues related to eBooks and e-reading.

3. PARTICIPATION

We have run ART for two years now and have encountered major problems in recruiting and keeping participants from beginning to end. We have identified a number of issues affecting participation starting from:

- cross-disciplinary nature of ART and its implications on the target community to be invited and involved;
- need to raise awareness and increase visibility as much of raising the profile of such a new initiative;
Indeed ART suffers from a number of known problems, common to user studies on top of not having yet fully established itself with academics and industry even at a time when this kind of study would indeed be so necessary and crucial in order to provide essential indications for further developments in e-books and e-readers.

4. NEXT STEPS

For this third round of ART we are going to address some of the issues just highlighted in order to improve:

- Visibility and awareness of the initiative across research communities. ART is part of INEX, an initiative mainly addressed and dedicated to information retrieval scientists. We have already made some attempts to reach out at the Digital Libraries community but we need to go further and advertise ART across other related communities. For instance the e-learning community seems a promising option and will be our next target. We also aim at involving as many crucial stakeholders as possible, including publishers, librarians and even authors and readers. The industry and in particular e-reader producers would be a great addition and contribute to complete our e-book scenario from producer to consumer.

- Motivation is crucial in making sure researchers feel the importance of the initiative and sign for it. There has to be some kind of implicit reward or, even better status to be acquired by taking part in this experience in order to justify its cost; and this will have to come from inside the community itself as a recognition of the importance of running and discussing comparable evaluations. Times are more than ripe and projects such as the e-book observatory, funded by JISC, http://www.jiscebooksproject.org/, with the aim of monitoring impacts of e-books and developing new models for their market in academia, are sure signs of the attention this area is attracting from practitioners. As researchers, we should react timely and take on this challenge to drive e-book research in terms of usability. It would be a shame if after all the talks and papers on the advantages of e-books on their paper counterparts, in terms of flexibility and adaptation, readers would be left with book surrogates on fancy devices. It should be clear the role ART can have and the support it can give in testing new models for e-books, creative interfaces and innovative market strategies.

- Do-ability is what would keep participants working on the task from beginning to end. Already with the past two editions we offered help and support along the way both in designing evaluation tasks and building questionnaires and interview schemes. For this new edition we will provide extra services and tools with the aim of making the overall experience of setting up a user study as straightforward as possible while maintaining a level of flexibility to accommodate for individual study needs. We would also provide flexible deadlines in order to encourage participants in taking part in ART by respecting their schedules.

- Quality of content and its relevance to users. This is crucial to motivate and support user studies. We will provide tools for matching users to relevant content in the original collection as this was highlighted as a concern in previous edition.

We have also started an internal process of reflection in order to identify and be inspired by examples of good practice in similar initiatives, while better understanding the shortcomings and possible ways to improve the current version of ART.

Finally, we have already two promising research groups taking part in ART this year and will use their experiences as examples when advertising ART in future.

5. CONCLUSIONS

Evaluation of e-books in terms of their usability and general user experience is a necessity. It is not only the academics but also the laymen and women who are now investing in e-readers and buying e-books. A number of interested stakeholders are monitoring the situation with a keen interest on its financial implications. It is crucial that our research community plays a central and leading role in keeping users and usability central to the whole process. ART is a way to get researchers together and work in this direction. We hope that our renewed effort would increase participation and give our initiative the visibility and popularity it deserves.

6. REFERENCES


