Crowdsourcing HCI for the institutional repository

Stephanie Taylor and Emma Tonkin

Abstract

It is said that one must not judge a book by its cover, but does that extend to researchers and cover sheets? Cover sheets excite enough discussion for technical and policy reasons – impact on metadata, necessity of use, branding , impact on publishers and so forth – to relegate the questions of their usability and efficacity to the bottom of the pile. In this era of cutting costs and trimming budgets, who has the money to spend on detailed investigation of anything that does not immediately impact on the core functions of institutional repositories: encouraging deposits, repository upkeep and so forth? In this paper we demonstrate the use of a crowdsourcing platform to run an extensive between-subjects HCI experiment designed to explore the impact of cover sheets upon common user tasks such as identification of document elements such as publication date and location of publication, and evaluate user perceptions of document layout.

Introduction

The bottleneck for HCI experiments is often the need to seek suitable experimental subjects, defined in general as kind individuals who share the time and inclination to further the cause of science for very little reward. Usability researchers the world over are familiar with the problem: bribe them with book tokens, enter them into prize draws, tempt them with small rewards – whatever works (and does not prejudice the outcome of the experiment). Experimental subjects are rare and kind creatures and on the whole those inclined towards experimental psychology are grateful to them and fearful of making too many demands on their time.

Consequentially, in the last few years, crowdsourcing has been exhaustively explored by usability researchers for suitability as a source for, essentially, cheap labour (Kittur et al, 2008; Paolacci et al, 2010; Buhrmester et al, 2011) Crowdsourcing through Amazon's Mechanical Turk, for example, is a fast way to collect responses to questions destined to non-specialists.

At times crowdsourcing is not appropriate: some problems are specialist and require specialist responses. In the repository environment, understanding the perspectives of repository managers can only be achieved directly, through asking repository managers, or indirectly, through literature review and exploration of available resources.

In this paper we explore the use of crowdsourcing to support a usability-focused investigation of the design and use of cover sheets.

The cover sheet question

Cover sheets, in the context of institutional repositories, are additional pages concatenated onto documents, typically through the use of PDF manipulation libraries such as pdftk. These are usually prepended to the document and are therefore the first thing that users see upon downloading and opening the PDF.

A discussion of views on and purposes of cover sheets can be found in Tonkin, Taylor & Tourte (2013). The authors conducted a survey of institutional repository managers in the United Kingdom with the intention of determining repository managers' viewpoints on the cover sheet. The survey identified the proportion of repository managers who currently employ cover sheets on some or all documents, approximately two thirds of respondents, and the proportion of repository managers who do not, but would like to do so, which constitute approximately half of the remainder. The barriers that this population identify to cover sheet adoption primarily referred to lack of demand, followed by lack of support from the technical platform used and by scalability issues.

We base this investigation on findings from Tonkin et al (2013), in particular, the identification of the purposes of cover sheets from the perspective of repository managers.

Respondents identified the following uses for cover sheets during that survey:

> Institution or departmental branding (>60%)
> Document provenance (>60%)
> Citation information (>50%)
> Document provenance (~40%)

Respondents also identified presentation of copyright information as a key reason for making use of cover sheets. Around 1/3 of respondents answered in this way, although the true number may be much higher for methodological reasons.

The purpose of a cover sheet from a repository manager's perspective, then, is the clear presentation of a number of relevant pieces of information about the document it encapsulates: document provenance, departmental/institutional branding, the means by which the document may be cited, the origin of the document and copyright information. From this we extract the following hypotheses:

1. The use of a cover sheet should increase the extent of user responses to simple indexing tasks (for example, as they are presented with more information, readers should be able to provide more information about the paper, such as author, publication venue, citation information)
2. The use of a cover sheet should increase the accuracy of user responses to simple indexing tasks – as this information is clearly presented on a prepended informational sheet, readers should be able to more accurately identify salient information such as author, publication venue, citation information...)

Additionally, we consider the following hypotheses:

3. Cover sheets improve user perceptions of layout
4. For users engaged in simple information extraction/indexing tasks, cover sheets improve user perceptions of readability of the document
5. Cover sheets may be viewed by users as redundant, as they may duplicate information.

We then design an experiment permitting us to explore these hypotheses using crowdsourced labour.

Experimental design

As the experiment will be crowdsourced, we cannot rely on observational methods. Instead, we depend on quantitative evaluation of user input and on user ratings.

We identify a number of cover sheet designs and prepend them to simply-styled documents. Extracting the first three pages as images, we then design the experiment using an A-B between subjects design. Group A receive pages 1 and 2 as images and are asked to complete a simple data extraction test, then asked for their opinions regarding the document layout, readability and level of redundancy inherent in the document design, as well as some basic demographic information such as level of education. Group B complete the same task and questionnaire, making use of pages 2 and 3 of the document. The groups are balanced in size.

The data extraction tests are intended to serve several purposes. Firstly, they represent a basic test of participant engagement, permitting us to check whether participants have attended to the article(Kittur et al, 2008) rather than answering questions randomly. Secondly, they can also be used as a gold standard against which to compare user input, in order to detect and invalidate unacceptably poor-quality input (e.g. the Mechanical Turk participant who chooses to fill in all text fields with 'asdf' in order to reduce the time taken to complete the task). Additionally, they permit us to establish the visibility of the descriptive information contained in the cover sheet – whether participants are reliably able to identify points such as publisher and publication date.

Results

Demographics show that all respondents are high school graduates or above: 38% have some experience of higher education; 44% are university graduates, 8% are postgraduates. Respondents are primarily in the 18-24 bracket (42%); 31% are 25-33, 15% are 34-44, 9% are 45-54, 3% older.

Hypothesis 1: Extent of data extracted.

The extent of data provided by users was higher in the case in which cover sheets were used. This is predictable, since it is often the case that preprints do not specify the venue of publication (journal/conference proceedings, etc). As cover sheets typically contain this information participants were significantly less likely to answer these questions with 'N/A' (not applicable), but instead provide an answer. Interestingly, very few participants correctly identified the citations provided on the cover sheet.

Hypothesis 2: Accuracy of user-generated metadata

Although the extent of data extracted is higher in the cover sheeted case, the accuracy of the metadata proves to be lower on average (see Fig. 1). This excludes both the citation identification task, for which performance was extremely poor in both cases, and minor typographical/ transcription errors (e.g. misidentification of 'rn' as 'm'), which do not differ significantly between the cases.
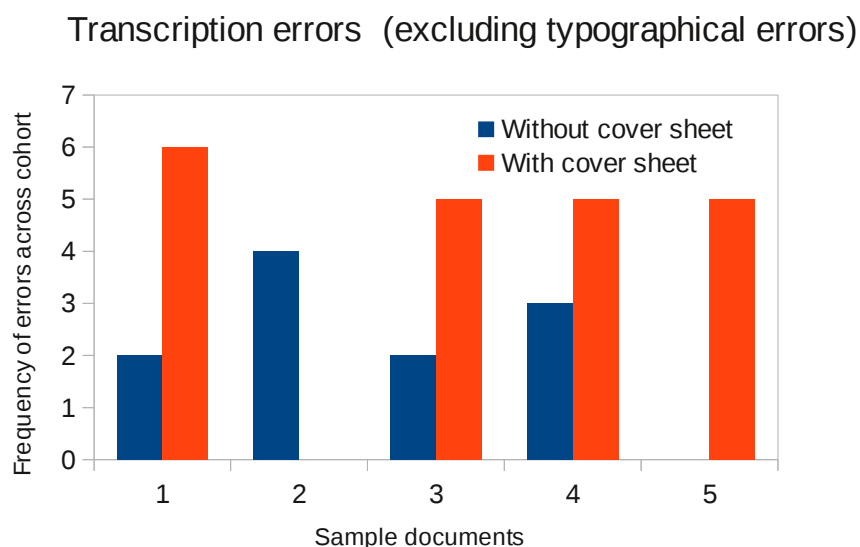

Figure 1: Accuracy of user-extracted data.

The source of the majority of these errors is participant misidentification of the place of publication. One-third of all participants in the cover-sheeted case misidentified the repository itself as the publisher, even though several cover sheet templates tested explicitly identify the publisher. Additional confusion is caused by the frequent practice of qualifying preprints as 'Unpublished', often alongside publisher data. The nuanced meaning behind this (e.g. the specific version placed online as a preprint is not the published version) is not well-understood by participants.

User perception of readability

Hypothesis 4 is not borne out, with $F_{(11,60)}=0.845$, $p=0.597$. Whilst there is a weak correlation (mean readability values for non-cover sheeted documents 3.2, mean for cover sheeted documents 3.5), no statistically strong correlation was identified between user perception of readability and the prepending of a cover sheet.

User perceptions of layout

Hypothesis 3 may be decomposed into two component parts. Statistically strong differences were identified in user perceptions of the various documents presented to them, with $F_{(11,60)}=2.77$, $p=0.057$. That is, users consistently scored certain documents highly and others poorly. To what extent was this variation led by cover sheet presence, or variations in cover sheet design? A MANOVA shows no clear group variation – it would appear that the cover sheet is not a strong

governing factor in perception of document layout.

User perceptions of cover sheet value

Finally, the participants were asked to score the first page of the document on a scale of 'redundant' to 'useful'. The results viewed across the cohort do not clearly separate the two cases - a MANOVA shows F- approximation of 0.56, approximate p of 0.58. Reviewing the data more closely there are examples in which a statistically significant difference is seen in user scores between cases A and B, and examples in which it is not; user perception of relevance appears to correlate with the quantity of information in the cover sheet and with its format, but this is a subject for further research.

Conclusions and further work

Relating the findings shown here to cover sheet use in general it is clear that a positive result found here is that it is useful to provide metadata not evident from the preprint itself, notably place of publication. Less positively, the addition of branding information leads to confusion on the part of the user relating to the publisher. There is no evidence from this data set that users correctly identify citations as citations or make much use of them.

It is suggested that some of the difficulties identified here relate to unfamiliarity with the concepts of institutional repositories and institutional repository branding, leading to misidentification/misclassification of entities. It is further suggested that cover sheet layouts are a source of some confusion due to the subversion of everyday understanding of document layouts. We are used to documents obeying certain presentational rules in most cases, which cover sheets do not typically obey. Not only does this confuse automated systems but it can also confuse the user.

Future work will include controlled testing of templates built with the findings from this test in mind, including both the better cover sheet templates identified here and alternatives such as a marginal 'stamp' on an existing document.

Finally, as an experiment in the use of Mechanical Turk/crowdsourcing approaches for low-cost HCI, we find that this experiment has been a success. We were able to complete an extensive between-subjects experiment involving almost a hundred participants at a cost of under £50 in total. Although this approach would not work for all types of problems it can work well where, as in this case, the problem can be expressed as a large number of short activities to be carried out by non-expert users.

References

Paolacci, Gabriele, Chandler, Jesse and Ipeirotis, Panagiotis G. (2010). Running Experiments on Amazon Mechanical Turk. Judgment and Decision Making, Vol. 5, No. 5, 411-419

Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08). ACM, New York, NY, USA, 453-456. DOI=10.1145/1357054.1357127 http://doi.acm.org/10.1145/1357054.1357127

Buhrmester, Michael, Kwang, Tracy and Gosling, Samuel D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science, Jan 2011 Vol 6 no.1 3-5. DOI=10.1177/1745691610393980

Tonkin, E., Taylor, S. and Tourte, G. (2013). Cover Sheets Considered Harmful? *ELPUB-2013*, June 2013. In press.