≋ NYU   **Junior Garcia <jfg388@nyu.edu>**

Pre-rebuttal reviews for your UIST 2023 submission #7937

1 message

Nathalie Riche and Juergen Steimle <uist23a@precisionconference.com>   Thu, May 25, 2023 at 8:05 PM
Reply-To: program2023@uist.org
To: Junior Garcia Ayala <jfg388@nyu.edu>

Dear Junior Garcia Ayala,

Your UIST 2023 submission

   7937 - CREST: Mediating Collaborative Search and Agreement for Group Property Bookings

has undergone peer review and the reviews are now available online (or should be shortly) at
https://urldefense.proofpoint.com/v2/url?u=http-3A__https-3A__new.precisionconference.com_uist23a&d=DwIDaQ&c=
slrrB7dE8n7gBJbeO0g-IQ&r=LFgT5eHJjh739rLb4hZAVQ&m=421Zx8Cgdd0uqOdm38fA-Lv-
qSjOtRmazk4d64SbtVLojI2EOadP6Z1WqprfC76x&s=HOFWIoVM5WK0opPtYGY--5QkLyd8CmayzdJOCEx5vCw&e= .
They are also enclosed at the end of this email.

For your reference, here are relevant statistics regarding overall scores across all papers that are still under
consideration: The mean overall score is 2.77 (stdev = 0.64) with a median of 2.67. Overall scores ranged from 1.3 to
4.3. The 75th percentile score is 3.2. While there is no explicit bar, UIST has traditionally had an acceptance rate below
25%.

All authors this year are given a rebuttal opportunity. For low-scoring papers, you should examine the primary review
("1AC") to see if a rebuttal is encouraged and what to address. The primary reviewer may indicate whether a rebuttal is
likely to make a difference or not based on their reading of the reviews and discussions. A decision not to submit a
rebuttal--even for a high-scoring paper--may be interpreted as an unwillingness to consider the reviewers' comments.

Our rebuttal period is two weeks, and the rebuttal text is limited to 5000 characters. They should address factual issues
and not bring in new data unless explicitly requested by the 1AC.

The deadline for submitting your rebuttal is Thursday, June 8, 2023, 5 pm PT.

More details are below (your reviews are included at the end):

The Primary ("1AC") reviewer should have provided a summary of the issues raised by external reviewers and may have
specified things to address. Please ensure that you cover these in your rebuttal and any other important concerns raised
in the reviews. You should use the rebuttal to correct any substantive errors or misconceptions or answer specific
questions posed by the reviewers.

The rebuttal is for addressing factual errors in the reviews, responding to specific questions for clarification raised by
reviewers, or including novel material as long as it fits within the character limit. Again, note that the rebuttal is limited to
5000 characters in length and must be self-contained. External links, e.g., to video material, are not allowed unless
explicitly requested by the 1AC. Since the external reviewers do not know the authors' names, the authors should
maintain their anonymity in their rebuttals.

The rebuttal is an important part of the UIST review process. The content of your rebuttal may be discussed among
reviewers when arriving at a final recommendation, and among the Program Committee during the PC meeting.

For more information about the UIST review process, please see:
https://urldefense.proofpoint.com/v2/url?u=https-3A__uist.acm.org_2023_author-2Dguide_&d=DwIDaQ&c=
slrrB7dE8n7gBJbeO0g-IQ&r=LFgT5eHJjh739rLb4hZAVQ&m=421Zx8Cgdd0uqOdm38fA-Lv-
qSjOtRmazk4d64SbtVLojI2EOadP6Z1WqprfC76x&s=DI9I2sVamn-ZGZpVAJPpC7U0dNx4brPIzt4K_vq1veU&e=

Thank you,

Jürgen Steimle and Nathalie Henry Riche
UIST 2023 Program Chairs

=========================

1AC review (reviewer 4)

Meta Review Overall Rating

2.0 Probably Reject

The Meta Review

The paper presents Crest, a collaborative search and agreement tool to find group property bookings. A formative study is used to determine key challenges and tasks, which are implemented in a functioning tool that is evaluated with 22 participants.

Overall, this paper has received mixed to negative reviews and reviewers seem to agree that there are shortcomings and concerns with the current presentation of the work. While the individual reviews contain detailed comments which they authors should consider for the rebuttal and potential revision, I summarize the key issues below:

1) Formative study (R2 , R1) – While the formative study is an important step in identifying the key tasks and challenges, the very small N and choice of participants make these results not very reliable, strong or representative. It is unclear how and if these findings would be different with different groups or setup. It lacks rigor and reliability.

2) Too focused domain (2AC, AC) – This tool addresses quite a specific problem and it is not clear what underlying or more generalizable research problem this work addresses.

3) Implementation and design (2AC, R1, R2) – reviewers all point to issues around rigor, lack of details and problems in the presentation of the system. (e.g., R2s comment on the mismatch between challenges and implementations and features).

4) User study (R2, R1, 2AC) – the study is limited as it focused mainly on user dynamics and not the UI/UX evaluation (R2). Further, the demographics lack representation, meaning the results may skew the research findings and limit their generalizability (AC). It is unclear whether IRB was optained and what compensation (if any) participants received. Finally, the study lacks rigor and should be more detailed in how the analysis is performed (2AC, R2).

While these reviews point to quite significant issues, the authors can submit a rebuttal to address some of these concerns. But as it stance, I follow the review recommendations and think this paper in its current state cannot be accepted to UIST.

Meta Review Systems Contribution

(blank)

Personal Rating

(blank)

Expertise

Summary and Contribution

(blank)

The Personal Review

(blank)

2AC review (reviewer 3)

score 2.5/5

Personal Rating

2.5 Weakly Reject

Expertise

Knowledgeable

Summary and Contribution

The paper's main contribution is CREST, a tool designed to aid group decision-making in rental property booking. It introduces features like a mediating bot and contract creation component, and the specific design solution is interesting and potentially valuable. However, the study has several limitations:

Narrow Scope: The paper focuses primarily on rental property group booking without addressing broader problems and solutions in collaborative decision-making tools.

Limited Participant Diversity: The user study is skewed towards university academic and residential communities, potentially limiting the findings' generalizability.

Absence of Broader Research Focus: The research focuses on CREST's design rather than on generalizable research questions, limiting its broader contribution to the field.

The Secondary Personal Review

With the quality of submissions for this year's UIST conference exceptionally high, this review considers the relative contribution and rigor of each paper within this competitive context. The paper under review here presents CREST, a tool designed for collaborative real estate search, aiming to address group decision-making challenges in the rental property group booking process. While the work is commendable, the review shall highlight key areas that could be improved to better align the research with broader and more abstract problems, increasing its potential impact on the field.

The paper effectively recognizes a gap in existing property search tools like Airbnb, which largely fail to offer collaborative search features or ways to mediate agreement within their systems. The authors present CREST as a tool enabling asynchronous search and agreement on a rental property among remote users. The tool incorporates several features designed to aid search, group communication, and agreement, including a mediating bot, contract creation, and signing component, and templates specifying house rules to negotiate group contributions.

However, the authors appear to approach the problem from a somewhat limited scope, targeting a relatively narrow domain. While rental property group booking presents unique challenges, it could be more beneficial to explore this as a case study within a broader context of collaborative decision-making tools. Extracting generalized problems and solutions from the group property search scenario could elevate the research's contribution and relevance to the wider field.

The user study conducted to evaluate CREST's effectiveness sourced participants from a university's academic and residential communities. While this approach may seem practical, it fails to represent the diversity of consumers a real estate search tool like CREST would typically serve. The lack of representation from different social groups may skew the research findings and limit their generalizability. Future studies could benefit from a more diverse participant pool to ensure more robust and inclusive results.

Despite presenting some interesting design elements, the paper ultimately frames its solution in a somewhat narrow scope. The work could benefit from a shift towards a more generalized research focus, which might offer more meaningful contributions to the field. This would entail identifying broader research questions and challenges in the realm of collaborative decision-making tools, with the CREST tool serving as a compelling case study within this larger framework.

In conclusion, while the paper presents a novel tool with potential practical value, it falls short in terms of broader research contribution. For this paper to stand out in the competitive UIST conference submissions, the authors would need to expand their research focus to address more generalized issues, enrich their participant pool for user studies, and provide a more compelling justification for their chosen design elements.

----

reviewer review (reviewer 1)
score 3/5

Overall Rating

3.0 Borderline Paper

Expertise

Passing Knowledge

The Review

This paper introduces CREST which is an integrated property booking system that support collaborative search and agreement within a group and seeks to cover what existing systems currently lack. The paper is well motivated and written. The methods are suitable but lacking in rigor. The discussions are appropriate. The topic of the paper is suitable for UIST. The remainder of my review are suggestions for improvement.

Introduction:

For "However, these somewhat upscale, packaged offerings do not address the needs of a group with a unique set of housing preferences and co-living arrangements [22].", instead of a reference that the reader needs to look for, it would be better to have a line providing an example.

Design:

In general, all sections and sub-sections are well substantiated. However, the need-finding falls on a very small group of users which lacks rigor.

For section 2.3, regarding "c5. Lack of Autonomy". This reads more like a claim by the authors (especially as no concrete references are provided). If there are no immediately suitable papers, perhaps the authors can look into relevant papers that may not necessarily be on bookings but that come to the same conclusion. Otherwise, this leaves me rather skeptical.

For section 2.4, the authors mention that "we carefully craft the language used in the different UI elements to create a clear understanding that the goal is to agree on a specific property". Was there a method that authors used to ensure the rigor of this?

Again for section 2.4, the reference for "Hong et al.'s Collaborative Dynamic Queries system" should be put in the quoted line.

Evaluation:

For section 3.1, regarding demographics, if the authors also asked whether the participants have used booking systems before, it would be good to report this as prior experience could influence the way they use either the CREST or baseline booking system.

Again for section 3.1, some comments: Was IRB approval obtained? And what was the minimal compensation participants received?

For section 3.2, the authors could do a statistical test of the average ratings to assess significance. Otherwise, the results were well presented and discussed. It was overall easy to read except for Figure 6 that was not as intuitive. Also,

while the author chose not to present results based on time, I wonder how quickly it took users to reach agreement between the baseline and CREST.

Remaining sections:

The remaining sections are well discussed. I wonder however if users reported disliking any features or interactions of CREST as it could also feed into future work.

----

reviewer review (reviewer 2)
score 2/5

Overall Rating

2.0 Probably Reject

Expertise

Knowledgeable

The Review

The paper proposes an asynchronous collaborative tool CREST for groups booking rental properties. A formative/need-finding study with 2 groups (N=5) is done to identify the tasks and the challenges. Design principles from related literature are adopted to support the user needs. A treatment-control study is performed to evaluate CREST against a currently available booking platform as the baseline.

The research direction toward a collaborative tool for facilitating group booking is interesting and needed. The motivation for the work is well-established in the introduction.

A major criticism of the work is that the formative study, which identifies the needs, tasks, and challenges as well as shapes the design and implementation of the tool, is limited. First of all, depending on the needs of only 2 groups (N=5; a couple, a friend group) can vastly deviate from the generalized user needs. Secondly, each group performed only one booking, which bears the possibility that some of the needs were specific to the preference for that particular booking and did not capture all the needs associated with people's general booking experience. Even though the paper mentions prior work supporting the need-finding results, (a) the specific references are not directly mentioned in Section-2, and (b) it does not guarantee that the formative user study reached a convergence or achieved generalizability.

The formative user study does not explain what kind of qualitative analysis was performed on the unstructured conversations to retrieve the results (section-2.2, 2.3).

The identified tasks and challenges are explained well. The need for asynchronous communication is well-established through the findings (e.g., disconnected tooling, discontinuity). The paper successfully finds the technological gaps in fulfilling the needs for the research goals (e.g., no platform supporting group search, etc.). One suggestion would be to reframe the need-finding study to focus on prior work and current technological solutions to explain the gaps, as currently the inputs from 2 user groups are not well-tied with each gap.

The proposed prototype of the tool is interesting and has the potential to be useful. Especially, the facilitation of conflicting preferences (page-3, c4) is a useful addition. Table-1 ties the design principles with the features of the tool in a useful way. The collaborative query panel, the collabo-ratio, etc. concepts appear useful and add value to the tool.

The implementations of different features of the tools seem questionable. Users can have "soft" and "strict" preferences (e.g. staying within budget can be a strict preference vs having a sound system can be a soft preference). They may also have a rank for their preferences (e.g., having a kitchen might be more important than having a sound system). The current implementation does not capture

such needs. The reasons why the proposed implementations follow the proposed manner are not explained clearly. These are possibly stemming from the limitation of the need-finding study. The implementation also does not record other features available in other platforms such as time spent in the system. The time needed to finish a task can be an indicator of the level of efficiency. Specific features to measure the usability/efficacy of the system are not correctly incorporated.

The user study is more focused on exploring group dynamics rather than system/design/UI validation. For example, the results discuss more on various group dynamics (e.g., someone preferred X over Y) leading to unsuccessful booking rather than evaluating the success/limitation of the tool. The groups are also synthetically formed and do not fully align with the research goals (i.e., the members do not know each other and have no real-world motivation to collaborate; rather they have the motivation to focus on negotiation to fulfill personal goals as they receive monetary benefits for that.) It is also not explained why a between-subject study is chosen over a within-subject study. The paper should provide separate demographic distributions for the control and the treatment groups (e.g., what is the male:female distribution in CREAT/baseline assessment groups, etc.). The analysis should also be more rigorous. For example, t-tests should be performed to report the significance of the results (e.g., t-test on the satisfaction levels of the baseline-CREST users). The normalized event sequence to discuss the user's decision-making process is done properly. The average number of events needed by treatment-control groups is unclear. The quotes from the participants show the potential of the tool (e.g., page-11 interaction with novel features). The evaluation of the chatbot indicates that it can keep the users engaged even though the group interactions are asynchronous.

Overall, the proposal of the tool targets a very promising research direction, and the paper has the potential to add value to the UI/UX community. However, the system design, the user study, and the evaluation need to be more thorough, rigorous, and aligned with the research goals.