A Leader-Driven Open Collaboration Platform for Exploring New Domains

Michael Weiss
Carleton University
TIM Program
Ottawa, Canada
michael_weiss@carleton.ca

Ibrahim AbuAlhaol
Carleton University
TIM Program
Ottawa, Canada
ibrahimee@ieee.org

Mohamed Amin
Carleton University
TIM Program
Ottawa, Canada
mohamed.amin@carleton.ca

ABSTRACT

This paper describes the design and initial evaluation of a leader-driven open collaboration platform for exploring new domains. The goal of this platform is to enable the collaboration of subject matter experts across knowledge boundaries. Traditionally, new domains are explored from within a single specialist or a focused group perspective. However, this often introduces bias. Collaboration helps reduce such bias by providing access to a broader range of information sources, increasing the chances for producing new insights in a new domain. However, it also introduces a new problem: variance between the contributions made. Variance makes it difficult to produce a coherent document. In this paper, we derive propositions about how leader-driven open collaboration is expected to help reduce bias while containing variance. We also offer an initial evaluation of these propositions based on our observations from developing an initial prototype of the open collaboration platform.

CCS Concepts

•Information systems → Crowdsourcing; •Human-centered computing → Collaborative and social computing systems and tools;

Keywords

Leader-driven open collaboration; new domains; bias; coherence; crowdsourcing; co-creation; literature reviews

1. INTRODUCTION

Today's big challenges (whether climate change, searching for alternative energy sources, or protecting ourselves from cyber attacks) require us to become good at exploring new domains, so that we can find feasible and practical solutions to those challenges. For example, network operators face threats from a wave of new malware releases, as many as 80,000 per day. Their analysts are looking for practical ways to mitigate this massive threat. As part of this task, they may need to learn about code reuse attacks, since most malware is created by modifying existing malware or by reusing existing techniques to bypass cyber defences. Information about code reuse attacks can be found in a variety of information sources, including papers, websites, and videos. How can analysts stay on top of the latest types of attacks and prevention techniques?

Traditionally, new domains are explored from within a single specialist perspective (i.e., an organization, a department in an organization, an academic discipline). This perspective influences what kind of information sources are consulted and how the information is synthesized. In other words, the perspective can introduce a bias. Collaboration helps reduce such bias by providing access to a broader range of information sources, thus increasing the chances for producing new insights and opening a new perspectives on the domain. However, collaboration also introduces a new problem: variance between the contributions made. This makes it difficult to produce a coherent document.

In this paper, we describe the design and initial evaluation of a leader-driven open collaboration platform for exploring new domains. The goal of the platform is to enable the collaboration of subject matter experts across knowledge boundaries. A preliminary design was described in previous work [15]. Adopting a design science approach, we first articulate a model for collaborative writing based on the cognitive science of writing and prior research on crowd-based collaborative writing and crowd synthesis. We then derive propositions on the expected benefits of the approach, and provide an initial evaluation of those propositions based on an early prototype of the collaboration platform.

The article first provides more background on exploring new domains, the structure of literature reviews, and prior work on collaborative writing and leadership in collaboration. Next, it presents the design of a leader-driven open collaboration platform for exploring new domains. This section is followed by propositions about how the proposed platform can help reduce bias while containing variance, and by observations on an initial prototype of the collaboration platform. The final section presents our conclusions.

2. BACKGROUND

In this section, we provide additional background on exploring new domains, the structure of literature reviews, collaborative writing, and leadership in collaboration.

2.1 Exploring new domains

A standard approach to exploring a domain is to conduct

OpenSym '16 Berlin, Germany

ISBN 978-1-4503-4451-7/16/08…$15.00

DOI: http://dx.doi.org/10.1145/2957792.2957797

Table 1: Elements of a literature review

Focus	The authors should state the purpose or focus of the literature review.
Relevance	The authors also need to make a case for the relevance of the review.
Glossary	The literature review should define any unusual terminology.
Sources of information	The authors of the literature review need to report on the databases searched.
Search terms	To limit the number of papers to review, the authors should turn the main concepts of the domain into search terms.
Selection criteria	The literature review should describe on what grounds papers were included or excluded from the review. Such criteria help avoid bias in paper selection.
Synthesis	The information obtained from the literature should be organized into common streams. A goal is to identify agreements, disagreements, and gaps.
Limitations	The authors should identify weak points of the review and areas for future work.
Conclusion	The conclusion should relate back to the focus and summarize the major findings of the literature review and identify its contributions to knowledge.

a literature review [7]. However, conducting a literature review in a new domain presents unique challenges. Whereas in an existing domain, researchers can use established classifications of knowledge to guide their search for and interpretation of the literature, this is not the case for a new domain. The task of the researcher is to make sense of evidence when it does not fit existing models and classifications and to extend existing knowledge accordingly.

Exploring a new domain can be thought of as looking for anomalies in the evidence that cannot be explained by what is already known [15]. A particular challenge in exploring a new domain is that the very criteria for searching the domain are co-evolving with our understanding of the domain. At the outset of a literature search, there are few established criteria for what the researchers should be looking for. The authors in [10] refer to as a "needle in a haystack problem where the appearance of the needle is unknown".

2.2 Structure of a literature review

The goal of a literature review is to synthesize the current knowledge on a given topic based on previously published research. Creating a literature review involves searching through the literature, retrieving sources of information, and synthesizing the findings of those sources [7]. We distinguish three broad categories of literature reviews: narrative, qualitative systematic, and quantitative systematic literature reviews [7]. Given the fragmented and evolving nature of the literature in a new domain, the type of literature review most suitable for exploring a new domain is a hybrid between a narrative and qualitative systematic literature review. Its elements are summarized in Table 1.

2.3 Collaborative writing

Several recently proposed collaborative writing systems use a crowdsourcing approach to complete the writing task [8, 14]. Crowdsourcing is a technique for leveraging a group of collaborators to solve complex problems [3]. In crowdsourcing, there are two types of users: requesters and crowd members [3]. Requesters are people or organizations who define a problem and aggregate partial solutions produced by the crowd. Crowd members are users who contribute.

The Ensemble system [8] is based on a cognitive model of writing in [6]. This model views the writing process as a series of rhetorical problems that frame the writing task. For each writing task, there is a top-level rhetorical problem that includes the constraints given to the writers and the goals the writers create for themselves. This top-level problem can be further decomposed into sub-problems.

In the MicroWriter system, the task of writing is decomposed into three types of subtasks: generation of ideas, labeling ideas to identify groups of related ideas, and writing paragraphs from related ideas [14]. A key insight is that each subtask should be completed with limited awareness of what has been done already and what others are doing. To this end, the context required to complete a writing subtask is embedded within the subtask itself.

2.4 Collaboration and leadership

The authors in [11] identified several challenges for leaders in collaboration platforms. They grouped these challenges into four areas: completion, originality, subjectivity, and ownership. The platform proposed in this paper with its four-phase workflow (i.e. collection, staging, synthesis, and refining) assists leaders in overcoming such challenges. Originality is achieved in the staging phase by selecting winning drafts. Subjectivity is avoided in the synthesis phase, and completion is reached in the refinement phase. Ownership is accomplished by giving users access to all comments and drafts and enabling leaders and contributors to refine the final article with clear records of contribution.

3. DESIGN OF A LEADER-DRIVEN COLLABORATION PLATFORM

The design of our leader-driven collaboration platform is modeled on previous work on leader-driven collaborative writing in [8] and crowd-based categorization of documents in [1]. In a leader-driven approach to collaborative writing, there are two types of participants: leaders, who constrain and specify the nature of the contributions – the lead author of a literature review sets the scope of the literature review and guides the synthesis process, and contributors, who are recruited to focus on specific writing tasks.

Following the cognitive model in [6], we conceptualize creating a literature review as a series of writing tasks. These typically include tasks for each of the standard elements to build a literature review. However, a more fine-grained decomposition is usually required, e.g. separate sections for each literature stream need to be added. These sections are defined by the lead author of a literature review.

For each section, the lead author motivates the need for the section and specifies a prompt or question (such as "define key features of topic X", or "identify examples of X") that helps focus the contributors' work. Contributors, as well as the lead author, provide alternative drafts in reply to the question. Finally, contributors or lead authors can comment on and categorize those drafts. It is up to the

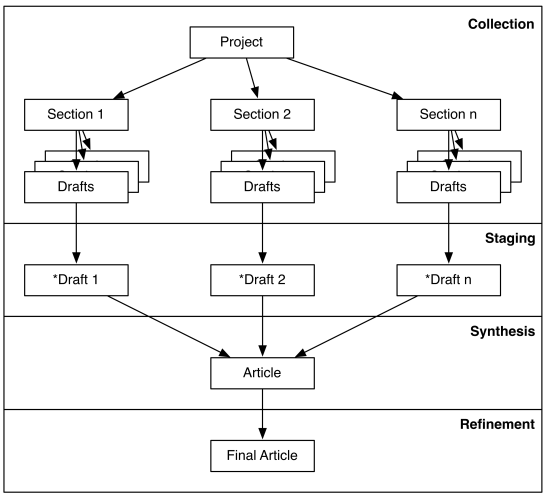

Figure 1: Workflow for leader-driven collaboration.

leader to choose the best draft for each section.

Figure 1 shows the workflow supported by the collaboration platform. At the collection stage, the leader solicits contributions from contributors in the form of drafts that address specific questions. At the staging stage, the leader selects the drafts to include into the article. At the synthesis stage, the leader composes an article from the selected drafts. This article is revised at the refinement stage. While the figure suggests a strictly linear flow, in actual use this will be an iterative process, where drafts can be staged as soon as drafts for some of the sections have been collected, a tentative article can be composed by synthesizing those drafts, and the creating of the final article involves the leader editing drafts, asking for additional information in the form of new drafts as necessary, and organizing the content.

4. PROPOSITIONS AND OBSERVATIONS

Based on the above design, we developed propositions about how leader-driven collaboration can help reduce bias while containing variance. Below, we describe these propositions and their rationale. We also offer an initial evaluation of the propositions based on our observations from developing an initial prototype of the collaboration platform. This prototype was implemented using the Drupal content management system, which greatly simplified building the prototype and allowed for quick iterations. The observations are based on an experiment with two groups of researchers conducting literature reviews in new cybersecurity-related domains using the collaboration platform. Members of the team were asked for their impressions on using the prototype and how it helped them carry out a literature review.

4.1 Reducing bias

Bias can blindside an organization to miss important evidence. Existing knowledge in one area can prevent an organization from accessing knowledge in another area. An organization's established predictive model can cause it to ignore evidence that contradicts the predictive model [2]. As noted by [7], literature reviews may be biased by the researcher's perspective, both in terms of the literature included and how it is synthesized. One way of reducing this bias is to conduct a more rigorous, systematic literature review. Bias can also be reduced by increasing the diversity of perspectives employed during the literature review.

Open collaboration allows collaborators to share information sources and to build on each other's findings. The diversity (in terms of area of expertise, culture, etc.) of the collaborators helps reduce bias that may exist when the research is conducted by a single person. The literature on team diversity predicts that increasing knowledge diversity in a team positively affects the range of information accessible to it [5]. In the experiment, contributors read sources contributed by others and challenged their interpretations by commenting and replying to comments.

H1a: *Open collaboration helps collaborators find and interpret information sources in the new domain.*

A significant task during the exploration of a new domain is to group the knowledge into research streams, and to identify research gaps. In [1], crowd members iteratively categorize text fragments. When a crowd member is asked to categorize a fragment, they see how other fragments have been categorized. They can then decide to put the new fragment into an existing category or create a new one. In our experiment, one leader asked for examples of code reuse attacks. As contributors collected and categorized the examples, they produced a taxonomy of code reuse attacks, including one novel type of code reuse attack.

H1b: *Open collaboration helps collaborators organize the knowledge in a new domain.*

Researchers provide support for the relevance of a research problem as well as the validity of its solution in the form of evidence [7]. According to [12], evidence has two attributes: type (e.g. prior literature, presentations, examples), and validity (i.e., following a research methodology, authorization by experts, and support by a large community). The authors in [4] describe a crowdsourced process for evaluating literature which involves decomposing a research question into subtasks that can be distributed to the crowd. Participants in the experiment reported that the platform helped them explore different types of evidence that included papers, related news websites, videos, and online lectures. It also helped them identify the key researchers in the problem domain, lending credibility to the evidence.

H1c: *Open collaboration helps collaborators find and organize evidence, in particular non-traditional evidence.*

Creating a taxonomy is an important starting point for conducting research in a new domain. It provides context and helps direct the attention of researchers to relevant problems. In the experiment, the taxonomy helped participants identify different definitions of the "anticipation" concept. It also helped them understand the relationship between anticipation and prediction and articulate the differences between those concepts. In the collaboration platform, comments were used to categorize concepts and link them.

H1d: *Open collaboration helps collaborators create a taxonomy of concepts and their relationships in a new domain.*

4.2 Containing variance

While helping contain bias, collaboration introduces a new problem: variance between the contributions is the result of

stylistic differences and inconsistencies in content, and is due, in part, to different levels of motivation to contribute. Variance makes it difficult to produce a coherent document. Misaligned contributions from different authors can impair the consistency of a document [13].

In leader-driven open collaboration, there are two types of participants: leaders and contributors. Leaders are responsible for the overall vision and flow of the document, while contributors provide input in their specific areas of expertise [8]. Leaders and contributors thus have complementary motivations. In our experiment, leaders were generally quite opinionated about the structure and direction of the articles, while contributors were looking to leaders to define goals.

H2a: *Collaborators adapt to the asymmetric collaboration structure involving a leader and multiple contributors.*

Coherence is an important quality of a document that reflects how consistent its narrative is. A document is coherent, when there is a "smooth and natural progression of ideas between them" [13]. In the experiment, the leaders reported that having a list of section templates to choose from helped them maintain a clear structure. We also observed that leaders synthesized contributed drafts into final drafts. However, to firmly conclude that a leader-driven approach increases coherence, we need to operationalize the coherence construct and evaluate a larger sample of articles.

H2b: *Leader-driven collaboration allows collaborators to create more coherent documents.*

We expect that leaders seek and incorporate feedback (in the form of drafts and comments) from contributors. In the Ensemble system in [8], leaders reported that they found the perspectives of other participants beneficial. In our experiment, leaders often gave specific directives to contributors on what kind of feedback they sought, for example, "identify the reasons why anticipation is critical in cybersecurity referring to examples in the popular press".

H2c: *Feedback from contributors can improve the content and quality of the article created, but it can also lead to a wide range of dispersed ideas and disagreement.*

User engagement is important for the adoption of a collaboration platform [9]. Contributors need to perceive that their contributions have impact. Similarly, leaders need to feel that they succeeded in orchestrating the production of quality and content-rich articles. In the experiment, contributors felt that their contributions were valued. Leaders and contributors effectively collaborated on reducing variance.

H2d: *Contributors perceive that their expertise is valued and leaders are able to harness their contributions.*

5. CONCLUSIONS

In this paper, we described a leader-driven approach to open collaboration for the exploration of new domains. We derived a set of propositions, and presented observations on an initial prototype of the platform. Our initial observations lend tentative support for our hypothesis that a leader-driven approach can help reduce single perspective bias while containing the variance between contributions. In the next step of our research, we will conduct a larger-scale, formal experiment to test the propositions.

6. ADDITIONAL AUTHORS

We would like to recognize the contributions of Chris Budiman, Raed Iskander, Ali Abu Alhawa, Mahmoud Gad, Ahmed Shah by adding them as co-authors of this paper.

7. REFERENCES

[1] P. André, A. Kittur, and S. Dow. Crowd synthesis: Extracting categories and clusters from complex data. *ACM Conference on Computer-Supported Cooperative Work and Social Computing*, pages 989–998, 2014.

[2] J. Attenberg, P. G. Ipeirotis, and F. J. Provost. Beat the machine: challenging workers to find the unknown unknowns. *Human Computation: Papers from the 2011 AAAI Workshop*, pages 2–7, 2011.

[3] J. P. Bigham, M. S. Bernstein, and E. Adar. Human-computer interaction and collective intelligence. In T. Malone and M. S. Bernstein, editors, *Handbook of Collective Intelligence*, pages 57–83. MIT Press, 2015.

[4] A. W. Brown and D. B. Allison. Using crowdsourcing to evaluate published scientific literature: methods and example. *PloS ONE*, 9:e100647, 2014.

[5] K. Dahlin, L. Weingart, and P. Hinds. Team diversity and information use. *Academy of Management Journal*, 48:1107–1123, 2005.

[6] L. Flower and J. R. Hayes. A cognitive process theory of writing. *College Composition and Communication*, 32:365–387, 1981.

[7] B. N. Green, C. D. Johnson, and A. Adams. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of Chiropractic Medicine*, 5:101–117, 2006.

[8] J. Kim, J. Cheng, and M. Bernstein. Ensemble: Exploring complementary strengths of leaders and crowds in creative collaboration. *ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 745–755, 2014.

[9] R. Kraut and P. Resnick. *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press, 2012.

[10] A. Lin, A. Huynh, G. Lanckriet, and L. Barrington. Crowdsourcing the unknown: The satellite search for genghis khan. *PloS ONE*, 9:e114046, 2014.

[11] K. Luther, C. Fiesler, and A. Bruckman. Redistributing leadership in online creative collaboration. *ACM Conference on Computer Supported Cooperative Work*, pages 1007–1022, 2013.

[12] A. J. Onwuegbuzie, N. Leech, and K. M. Collins. Qualitative analysis techniques for the review of the literature. *The Qualitative Report*, 17:1–28, 2012.

[13] N. H. D. Silva and P. Henderson. Narrative-based writing for coherent technical documents. *ACM International Conference on Design of Communication*, pages 208–215, 2007.

[14] J. Teevan, S. Iqbal, and C. von Veh. Supporting collaborative writing with microtasks. *Conference on Human Factors in Computing Systems*, pages 2657–2688, 2016.

[15] M. Weiss. Crowdsourcing literature reviews in new domains. *Technology Innovation Management Review*, 6:5–14, 2016.