A Tool to Remotely Collect and Visualize Users' Interactions with Web-Based Content

Robert B. Watson
University of Washington
Campus Box 352315,
Seattle, WA 98195
01-206-685-1557
rbwatson@uw.edu

Jan H. Spyridakis
University of Washington
Campus Box 352315,
Seattle, WA 98195
01-206-685-1557
jansp@uw.edu

ABSTRACT
In a remote, unmoderated experiment, we tested readers' performance and perceptions as they used web-based documents to perform information-seeking tasks. The high value of specific participant experiences motivated us to develop a low-overhead tool to collect specific information remotely about readers' experiences with web-based documents. The tool we developed overlays a web document and prompts readers with a specific request to which they respond by highlighting a location in the document. In our experiment, we found the interaction to be very clear and easy for readers of a web site to use. The tool required very low communication bandwidth and processing power to collect and analyze the data and produced easy to interpret response visualizations and statistical analyses. This paper describes our experience with the tool in the course of the experiment. Further, we propose broader applications of the tool to support research activities in other document genres and, to adapt it for professional use in content management systems and web publishing platforms.

CCS Concepts
• Human-centered computing~Empirical studies in interaction design • Human-centered computing~User studies • Human-centered computing~Information visualization

Keywords
Remote user experience assessment; Web analytics; User experience testing; Documentation feedback; API documentation.

1. INTRODUCTION
While designing a remote, unmoderated, online experiment to measure readers' task performance and perceptions of application programming interface (API) reference topics, we sought a method to learn more about participants' experiences with the topics. Our experiment measured how participants' performance and perceptions were affected by variations in the information concept count and visual design element count of the topics [1].

SIGDOC '16, September 21-23, 2016, Arlington, VA, USA
© 2016 ACM. ISBN 978-1-4503-4495-1/16/09...$15.00
DOI: http://dx.doi.org/10.1145/2987592.2987601

While designing the experiment, we found no shortage of instrumentation and analysis methods to measure task-completion time and participant perceptions in remote unmoderated studies. However, we wanted to collect specific information about the participants' experience with the tasks and the topics they saw in in the context of those tasks. In a moderated experiment, this would be a simple action—the researcher would ask participants questions about their experience and observe their behavior. We wanted to collect similar information in our unmoderated study; however, we could not find a suitable tool.

In our experiment, participants used API reference topics to perform information-seeking tasks that are common in software development. After participants completed a task, we wanted to know where in the topic they found the information that most influenced their task response. Common methods for collecting such information in an unmoderated study apply tools such as free-text or multiple-choice responses to questions in a questionnaire and rating and ranking responses. Passive methods such as mouse and click tracking were also considered; however, the task did not require participants to click on the topic during the task—it was simply an information-seeking task—and we did not want to complicate the task or confound the results by adding an unnecessary interaction. In earlier unmoderated, remote studies, we experienced better completion rates when the cognitive load of the responses was low [2], so we felt that the more common methods might seem unnecessarily intrusive to participants and complicate the analysis for the researchers. To fill this gap, we developed Spot2Know, a low-overhead method to collect this information. Spot2Know collects the desired responses in a way that is very easy for participants to use and produces data that researchers can analyze statistically and display graphically.

After applying Spot2Know in the context of our experiment and reviewing the results, we realized that practitioners and researchers could also apply Spot2Know to collect information about reader experiences with other types of documents and contexts. The methods we developed to collect, analyze, and present the data from Spot2Know are simple, focused on a specific research question, and adaptable to presenting data in compelling and easy-to-understand formats. The analysis methods we used in our experiment could provide meaningful and useful data to authors and stakeholders without requiring complex statistical analysis, while not forfeiting the ability to perform more sophisticated analyses on the same data. This paper describes the tool that we developed, the analysis methods we applied in the experiment, and how Spot2Know and its related methods can be applied to answer research questions in a variety of document contexts and genres.

2. BACKGROUND

Mouse-click and mouse-motion tracking have been used for many years to track reader interest, reader focus, and even to serve as a proxy for eye tracking without the need for dedicated eye-tracking hardware [3,4,5,6]. While these technologies provide a wealth of information about how readers interact with content in a web browser, the amount of information they provide is considerable—much more than was needed to answer a simple research question. We had several concerns with collecting mouse-motion tracking in the context of a remote study. We were concerned that the bandwidth required to record the data might introduce delays in the study, which would affect participants' experiences and the precision of the timed measurements. We were also concerned that the additional data would complicate the analysis, especially when the research question did not require it.

The principles of the mouse-click and mouse-motion recording techniques, however, could be distilled to collect only the information required to answer the research question. In doing so, the data collected, stored, and processed would require much less bandwidth to send to the server and much less processing power to study. While providing a prompt to direct readers' responses is more invasive than passive click and motion tracking, it makes it possible to draw specific conclusions from the data.

The API topics used in the experiment were based on samples from the open-source documentation reviewed by Watson et al. [7]. The task orientation of the experiment was based on earlier studies of software developers interacting with documentation [8].

The experiment that included Spot2Know was designed and conducted using established remote experiment design practices [9,10,11]. Because we were looking for a specific response to a specific research question in the experiment, we elected to use Spot2Know instead of passive measurement instruments such as click, link, and mouse-motion tracking.

3. METHOD

After reviewing the available tools and methods, we decided that the ideal instrument for our research question should:

- Incur a cognitive load on the part of the participant similar to existing response methods, such as a Likert scale.
- Contain the directness and specificity of a questionnaire.
- Be as easy to analyze statistically as Likert-scale data.

The result was Spot2Know, the document feedback tool we developed for the experiment. Spot2Know was implemented in JavaScript and stored as an external file that was included in the web pages evaluated in the experiment.

The information-seeking task of the experiment presented participants with a scenario and asked them to decide whether the topic they would see in the next step contained the information described in the scenario. The experiment timed how long participants took to make that decision. To collect information about the participants' experiences with the document, Spot2Know was programmed to appear immediately after participants completed the timed, information-seeking task, but before they were shown the perception questionnaire. This placement collected the data we sought while the information-seeking experience with the documents was still fresh.

After the timed task, Spot2Know did the following.

1. Prompted participants for the desired feedback. Figure 1 [1] shows the prompt used in our experiment, which asked participants to click on the location in the topic that most influenced the decision they made in response to the preceding timed information-seeking task.
2. Highlighted the location that participants specified by displaying a shaded circle where they clicked. Figure 2 [1] shows how a selected location appeared to participants.
   a. Participants could change their selection if they wanted.
   b. Participants had the option to select **Not applicable** instead of clicking on the document—for example, it might be difficult and possibly frustrating for participants to select a location in the topic if the topic did not contain the information requested by the information-seeking task.
   c. Participants clicked the **Save** button to record their selection and continue.
3. Collected and transmitted the following data to the server:
   a. The horizontal *(x)* and vertical *(y)* location of where the participant clicked in the topic.
   b. Whether the participant selected a spot in the topic or the *not applicable* option.
   c. How many times participants changed their selection before saving.
   d. The amount of time participants spent in the Spot2Know interaction.
4. Returned control to the web page after participants saved their selection.

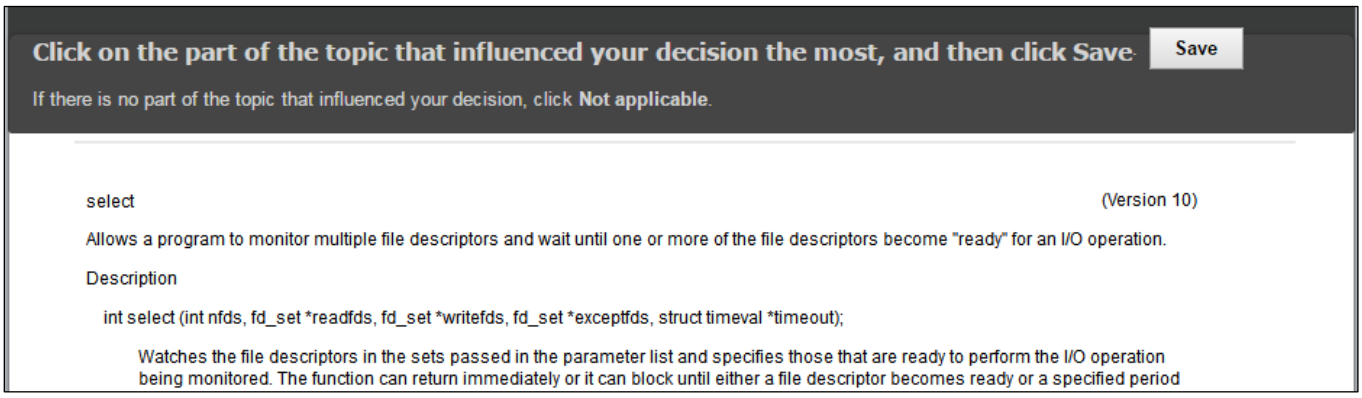

Figure 1. Spot2Know prompt at the top of a sample topic

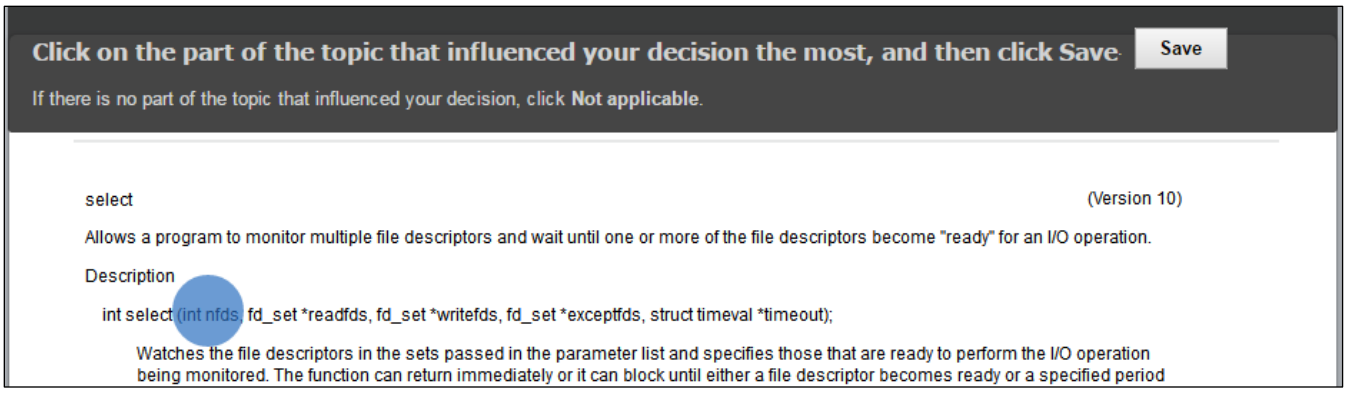

Figure 2. Selected location highlighted in Spot2Know

3.1 Participant Experience

We measured how long participants took to respond to each screen in the experiment, including Spot2Know and the perception questionnaires. Measuring participant response time gave us the ability to compare the times that participants spent to indicate the most influential spot in the topic they reviewed and the time they spent responding to the more traditional scales used to collect perception data.

The participants who self-reported as being proficient in English spent an average of 17.65 seconds (N = 833, St. Dev. = 19.09) to indicate the most influential spot in the topic. Only 5% of the participants changed their selected spot and those who did changed it only once. Those who changed their response took an average of 8.57 seconds longer to respond. While participants could select the *not applicable* option, only 10% of them did. There was no significant difference between the interaction times of the participants who selected a spot in the topic and those who selected the *not applicable* option.

3.2 Data Collection

The data collected by Spot2Know provided information about each participant's experience with the web-based document and aggregated graphical summaries and statistical analyses of the responses. The data collected, transmitted, and stored by Spot2Know were designed to be as compact as practical so as to not adversely affect participants' experiences. Spot2Know reported the data to the study database after each participant response by sending 120 characters or less to the server and using less than 260 bytes of space in the database on the server. As a part of the study, the locations of the sections of each topic were entered into the database so that the *(x,y)* coordinates of the response locations could be related to topic sections during the analysis. The sections in every topic variation were the Introduction, Description, Parameters, Return Values, and Notes. Some topic variations also included Related Topics and Examples sections. The section location data enabled categorical data analyses and comparisons of results between topic variations with different layouts and content placements.

In our implementation, a third-party survey vendor provided the survey software that implemented the experiment's study protocol and a private web server hosted Spot2Know's software and data collection.

3.3 Data Analysis

The first data analysis we performed was to overlay the participant spots on the document image as the example in Figure 2 shows. Aggregating the responses in this manner with the prompt produces an informative and easy-to-read visualization. For prompts such as the one used in the experiment—"Click on the part of the topic that influenced your response the most"—presenting the data in this format reflects what participants saw when they made their selection, making it easy for researchers to visualize participants' experiences at a glance.

Figure 3 [1] shows an example of the visualization we used. This display provided us with a collection of participant experiences aggregated into a single image. In our visualization, we plotted the spots of participants who responded correctly to the preceding task in a different color than those who did not, making any difference in response easily visible. The number of participants who selected *not applicable* is indicated in the grey banner below the task scenario.

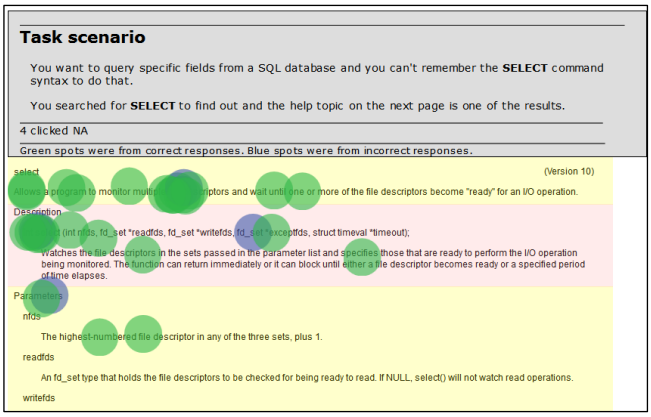

Figure 3. Spot interaction visualization

The second analysis we performed was to compare the responses across different topics. While the visualization provided an easily viewed answer to a research question on a single page, it is difficult to see patterns that might exist in sets or groups of different pages. To study this type of data, we applied several statistical analyses. We evaluated the standard deviation of the vertical *(y)* spot locations to provide a statistical description of the variation that we observed in the graphical representations.

For example, pages with a narrow vertical spread of responses, such as the example in Figure 4 [1], also had smaller standard deviations of the vertical *(y)* spot locations. Pages with broader response spreads, such as the example in Figure 5 [1], had larger standard deviations of the vertical *(y)* spot locations.

Vertical and 2-dimension clusters illustrate the degree to which participants agreed, but when compared across topic variations with different content layouts, the analysis that best informed our multivariate experiment was the categorical comparisons. Evaluating the frequency of responses by category enabled comparisons across experimental conditions.

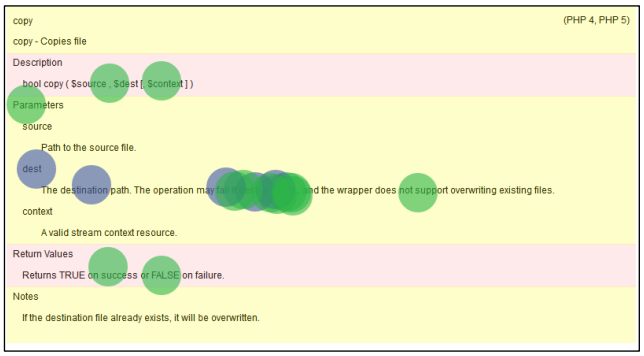

Figure 4. Visualization example with small vertical Std. Dev. (62.19)

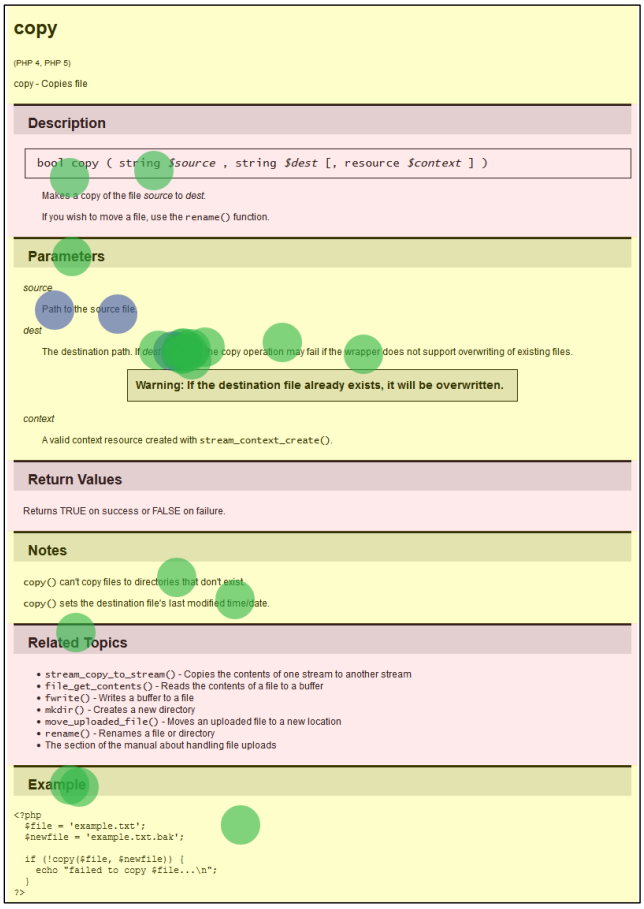

Figure 5. Visualization example with large vertical Std. Dev. (230.98)

Because the topic sections had different physical locations in the different variations of a topic, this categorization enabled comparisons of response frequencies within each topic section across experimental conditions to identify statistically significant differences. Figure 6 is an example from the study's analysis website that shows how we visually compared the responses to different experimental conditions by topic section. In Figure 6, the large number in the shaded column to the left of the bar graph is the total number of responses in that condition and the smaller number below it is the number of correct task responses (those in which the task response matched the experimental condition). The color of the column with these numbers reflects the percentage of correct responses. The five numbers in the column to the right of the colored columns describe the values of the experiment's timed task that preceded participants' interaction with Spot2Know. From top to bottom these values represent:

- The longest task response time in the data set.
- The upper limit of the 95% confidence interval.
- The mean response time.
- The lower limit of the 95% confidence interval.
- The shortest task response time in the data set.

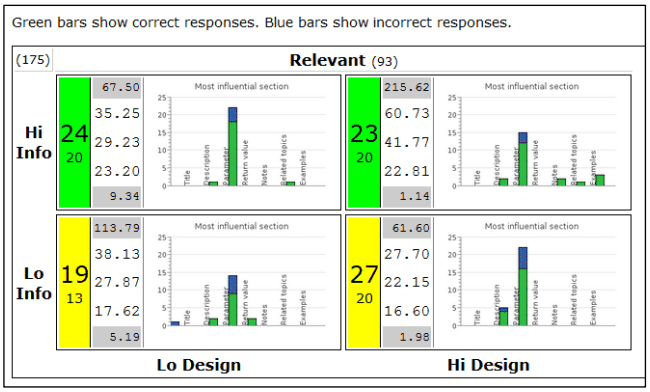

Figure 6. Bar graph visualization example

Identifying topic sections enabled Chi-square tests on the resulting categorical data to identify statistically significant differences between the different experimental conditions—tests that could not be performed with only the raw location data.

4. DISCUSSION

In our experiment, Spot2Know provided us with specific data in response to a specific prompt, which was easily evaluated by both graphical and statistical reports. By matching the prompt to the research question and the topic, Spot2Know makes it easy for readers to respond and researchers to analyze their responses. When the reader can respond to the prompt by indicating a single location in a document, Spot2Know provides a low-overhead, easy-to-use method to collect this data from readers and study participants. While mouse motion and click tracking provide similar data, the method applied by Spot2Know involves a more focused and lower-overhead approach, which makes it easier to implement, evaluate, and apply to large populations and documentation sets. Research conducted using Spot2Know, however, has some limitations that the experimental design must not exceed. The next sections review the benefits and limitations we found in our application of the tool to our experiment.

4.1 Benefits Experienced

The ease of use and ease of analysis make Spot2Know suitable to more applications than just that of our experiment. The average response time of 17.65 seconds suggests that it is easy for readers to use. In our experiment, we found that Spot2Know made it easy to offer both response visualizations and statistical analyses that enabled us to group sets of documents for further analysis. We also found that Spot2Know could be integrated with other remote measurement instruments to gain an even more complete picture of participants' motivations and response patterns.

4.1.1 Individual Document Study

Spot2Know offers visual and statistical feedback on individual web documents. In our experiment, we used the data from Spot2Know to evaluate readers' experiences with a specific document; however, other questions could be asked in other studies and contexts. An example of a prompt for another context might be, "Click on the part of the document you found most helpful" or "Click on the part of the document you found most interesting." The feedback could inform improvements in the specific topic and in other, similar or related documents.

4.1.2 Document Group Study

Sets of documents could be organized into groups for analysis, while retaining the ability to study the individual documents in the set. API reference documentation, such as our experiment tested, is an example of such a group in that it consists of many different documents or topics that all share a common format. Questions about the topics could be analyzed by topic section to understand how readers perceive the overall structure and format of the documentation set. Spot2Know provides a way for readers to provide specific feedback about, for example, the section of the documents in the group that helps them the most at a very low cost to the user and the researcher. The information that Spot2Know provides about readers can feed into the content strategy for the next iteration of similar documentation.

4.1.3 Integration with Other Instruments

Spot2Know can be employed with other measurement and response instruments. For example, after readers interact with Spot2Know, a free-text comment box could be used to ask for additional comments. Likewise, Spot2Know could be used in response to other feedback tools, such as having readers interact with Spot2Know after they respond to another measurement instrument.

4.2 Limitations Observed

The data that Spot2Know provides is only as good as the prompt. Prompts need to be considered in the context of readers' goals and the nature of the document being evaluated. As with survey questions in questionnaires, the prompt must be focused [8] and, in the case of Spot2Know, able to be answered by locating a spot in the topic being reviewed. Further, the focused aspect of Spot2Know emphasizes the need for researchers to know how they plan to use their findings to make the most out of Spot2Know interaction data. This section reviews some of the limitations we observed while developing and using Spot2Know.

4.2.1 Provide Effective Prompt Text

The utility of Spot2Know and the value of the data depend greatly on the prompt. The prompt must elicit a response that can be provided easily by clicking on the document or on a negative response. Trapping the reader such that they cannot exit the interaction without responding on the document could frustrate the reader and produce invalid data for the researcher.

The benefit of Spot2Know is the virtually transparent interruption it presents to a reader's task flow. Aspects that increase the cognitive load of the interaction, such as complex prompts or failing to provide a negative or opt-out response, make the interruption by the Spot2Know interaction difficult to ignore and reduce the tool's utility and value. Longer prompts, and more complex and intrusive interactions reduce the likelihood of responses. In content interactions that are typically brief, Spot2Know provides a way to collect data about readers' experiences that might otherwise be missed by lengthier or more intrusive interactions. The specific prompt and response format employed by Spot2Know makes it easier to associate the data to a specific research question and less likely to misunderstand the data when compared to responses provided to less specific feedback instruments, such as the "thumbs-up/thumbs-down" interaction.

4.2.2 Provide an Opt-Out Option

In pilot testing, before we deployed the experiment that used Spot2Know, we found that the spot feedback interaction was not sufficient without a non-response option. Dillman [8] suggests that a non-response option be provided when the question might not apply to the participant. For example, a prompt such as, "Click on the part of the topic that helped you" presumes that the topic actually helped the reader, when, in fact, that might not have been the case. The result of our pilot-test feedback was to add a *not applicable* option to provide the participant with a non-response option.

4.2.3 Understand How You Will Use Your Findings

As with any survey instrument, researchers should understand how they plan to use the data they collect before they start collecting it. With Spot2Know, this is an important consideration because some analysis methods require additional preparation before they can be used Spot2Know. While the raw data collected by Spot2Know can be used to generate heat-map-like displays, categories must be identified and mapped to each document and document variation to use the categorical comparisons between documents.

Fortunately, choosing one analysis or presentation option does not preclude using another. Encoding the document categories could be done in the document's HTML and automated, or entered manually before or after the experiment is conducted, as long as it is encoded before attempting a categorical analysis.

4.3 Research Applications

Our initial application of Spot2Know was to visually analyze individual responses and perform statistical comparisons between the topics and topic variations of the experiment. In our experiment, we used the heat-map-like display and the standard deviations of click locations to evaluate our scenarios, tasks, and documents and we used the categorical comparison to identify differences between topics and topic variations.

We used the heat-map-like display to review where participants found the information when they decided whether a topic variation was relevant to the timed task they had just performed. These reviews helped us know whether they found the information where we would have expected them to find it or if they found it somewhere else. Using the graphical presentations of the responses, we could get a general sense of the quality of the relationship between the scenario and the document. For example, if participants' responses clustered tightly around the information that we would expect it to, we probably had a good match between the scenario and the document. If participants' responses, however, did not cluster tightly, we might look into the quality or clarity of that task, the scenario, or the topic.

Spot2Know provided our unmoderated, remote user experience study with the extra insight into participants' experiences that, while taken for granted in moderated studies, is not typically available in unmoderated studies. Researchers could, for example, use this method to validate study assumptions and compare participant actions to expectations at an individual participant level and aggregate this information into summaries and statistics.

4.4 Practitioner Applications

The analytical and decision-support value that Spot2Know provided to our experiment and its minimal intrusion into the user's experience with the content suggest that Spot2Know could support practitioners in broader contexts as it supported our research. Reviewing the Spot2Know results with the results of our experiment revealed other possible applications for the data and visualizations that Spot2Know produced—applications that could benefit practitioners as it did our research, especially when used with data from other measurement instruments.

In our experiment, we used only a few analysis methods and additional possibilities remain to be explored. For example, the data from Spot2Know could be used to select cases to examine by filtering data to include only the cases where readers selected a specific section of the document. Another application that might be valuable to authors would be to use the data from Spot2Know to identify the areas of a document that are most helpful to the people who spent the shortest time on the page. At the same time, another opportunity might be to study how long readers who found a specific document section to be most interesting spent on the page. For practitioners, Spot2Know provides deeper knowledge about the parts of a document that readers find most helpful or most interesting at a very low data-collection and analysis cost. This information could influence the content strategy for future revisions of a document or future documents.

5. FUTURE WORK

Before the analysis possibilities described in the previous section can become widespread, some work remains. The Spot2Know code that is inserted into the web documents being studied needs to be productized and made suitable to a wider range of content, web sites, and web-site managers. While the data that Spot2Know produced in the experiment are interesting, collecting and analyzing them is currently too labor-intensive to be practical for practitioners. The analysis tools need to be simplified and analysis of the most common use cases should be automated so that site managers and authors can incorporate the tool and the data analysis into their authoring routine.

The easiest way to incorporate the Spot2Know tool and its analysis tools into a routine practice for web-content authors is to incorporate it into content management systems (CMSs) that they use to author and publish web content. For example, as a part of the authoring process, the CMS could incorporate settings for feedback frequency and the prompt text for each topic or topic group. The CMS could present reports and dashboards that facilitate reviewing how the content is performing and provide the tools to quickly evaluate and update pages that the CMS identifies as needing revision. Spot2Know prompts could be simplified to support common examples and key performance metrics by default to simplify or automate prompt selection and analysis. Prompt frequency (how many page views between prompting readers for a response) could also be configured.

Developing Spot2Know to support the most valuable use-cases and incorporating it into the content management and publishing process will give authors detailed and useful information about their content—information that is currently difficult to obtain and apply to authoring processes. The information provided by Spot2Know can be studied at the individual topic level and the topic group level to inform content decisions at all levels. By making Spot2Know an organic part of web-content authoring, authors and site managers will be able to provide the most value to their readers and stakeholders with as few iterations as possible.

6. REFERENCES

[1] Watson, R. B.. 2015. *The Effect of Visual Design and Information Content on Readers' Assessments of API Reference Topics*. Doctoral dissertation, University of Washington, Seattle, WA.

[2] Watson, R. (2012). Developing best practices for API reference documentation: Creating a platform to study how programmers learn new APIs. In *Proceedings of the 2012 IEEE International Professional Communication Conference (IPCC)* (pp. 1–9). IEEE.

[3] Huang, J., White, R. W., & Dumais, S. 2011. No clicks, no problem: Using cursor movements to understand and improve search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1225–1234). New York, NY, USA: ACM.

[4] Rodden, K., & Fu, X. 2007. Exploring how mouse movements relate to eye movements on web search results pages. In *Web Information Seeking and Interaction* (pp. 29–32).

[5] Atterer, R., Wnuk, M., & Schmidt, A. 2006. Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th International Conference on World Wide Web* (pp. 203–212). New York, NY, USA: ACM.

[6] Navalpakkam, V., & Churchill, E. 2012. Mouse tracking: Measuring and predicting users' experience of web-based content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2963–2972). New York, NY, USA: ACM.

[7] Watson, R., Stamnes, M., Jeannot-Schroeder, J., & Spyridakis, J. H. 2013. API documentation and software community values: A survey of open-source API documentation. In *Proceedings of the 31st ACM International Conference on Design of Communication* (pp. 165–174). New York, NY, USA: ACM.

[8] Brandt, J., Guo, P. J., Lewenstein, J., Dontcheva, M., & Klemmer, S. R. 2009. Two studies of opportunistic programming: Interleaving web foraging, learning, and writing code. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1589–1598). New York, NY, USA: ACM.

[9] Dillman, D. A., Smyth, J. D., & Christian, L. M. 2008. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (3rd ed.). Wiley.

[10] Bartell, A. L., & Spyridakis, J. H. 2012. Managing risk in internet-based survey research. In *Proceedings of the 2012 IEEE International Professional Communication Conference (IPCC)* (pp. 1–6). IEEE.

[11] Spyridakis, J. H., Wei, C., Barrick, J., Cuddihy, E., & Maust, B. 2005. Internet-based research: providing a foundation for web-design guidelines. *IEEE Transactions on Professional Communication*, 48(3), (pp. 242–260). IEEE.