Designing Metrics to Evaluate the Help Center of Baidu Cloud

Zhijun Gao
gaozhijun@pku.edu.cn
Peking University
University of Twente

Yuxin Gao
gaoyuxin@pku.edu.cn
Peking University

Jingjing Xu
xujingjing03@baidu.com
Baidu AI Cloud Group

ABSTRACT

Help centers are mainly designed to assist users with their product uses. The question as to how we measure the quality of a help center remains unanswered. As the first step of a joint research initiated by Peking University and Baidu Cloud that aims to develop a set of computable metrics to evaluate the quality of help centers, this experience report shares the results of data analysis on correlation between user behavioral data and technical documentation quality. The documents and data we use are a suite of cloud computing services provided by Baidu Cloud. The report begins with an introduction of the research goal; following reviews on the related work, it then lays out the design of the experiments with user data collected from Baidu Cloud. In our experiments, we categorize all documents into three groups and try to identify which metrics would affect documentation quality most. The result shows that the key index that contributes most to the model is PV/UV. At last, the report concludes with our current experimental efforts and future work in our plan.

CCS CONCEPTS

• **Human-centered computing → User centered design**.

KEYWORDS

help center evaluation, web metrics, technical information, quality evaluation

ACM Reference Format:
Zhijun Gao, Yuxin Gao, and Jingjing Xu. 2019. Designing Metrics to Evaluate the Help Center of Baidu Cloud. In *Portland '19: Broadening the Boundaries of Communication Design, October 04–06, 2019, Portland, OR.* ACM, New York, NY, USA, 7 pages. https://doi.org/https://doi.org/10.1145/3328020.3353936

1 INTRODUCTION

Online help centers provide users with hypertext information in an organized fashion. Users may resort to help centers when they encounter problems with the product, and quality online self-services save labor costs and improve service efficiency. It then brings up the question of how we can evaluate the quality of a help center with a quantitative approach such that we can predict a quality score given any help center. The help center first provides navigation and search functions, enabling users to quickly find the information they need. This information should then help solve users' problems. If it is not resolved, there should be a way for user feedback, so that the service provider can further improve the corresponding content of the help center or the related products.

Help centers are separated from other websites for discussion here because of their specific use of help. We further divide a help center into two parts: the website as an information container and the information inside it. As the information container, a help center needs to place the information properly, which means it should provide users with a fine guide to the target information through navigation or searching function. The information inside, which refers to technical information in this paper, helps users to actually solve their problems. As the technical information we mention here is presented as a form of online technical documentation, using technical documentation quality models can help with designing metrics and assessing them.

This paper is structured as follows. We first discuss how the term "quality" was defined as in previous works in Section 2, which also presents reviews on documentation and help center qualities. In Section 3, we investigate related works of quality assessments on the two parts of a help center, i.e., the container and the content. We present our current efforts in Section 4, and corresponding experiment results in Section 5 followed by conclusions and future plans in Section 6.

2 DEFINE QUALITY

The term "quality" is often vague and intangible as it is subjectively felt or judged, but not exactly measured by some standard. The whole concept of quality can be confusing so in order to evaluate the quality of help centers, so it should first be defined.

Crosby [8] defined quality as "conformance to requirements," which suggests that there must be a set of requirements that cannot be understood. The requirements are not necessarily universal, but set by an entity or for a single product. Juran [16] defines quality as "fitness for use", mainly considering the customers, their requirements and expectations on the product, and their particular use. It then requires the product to possess multiple elements of fitness for use. These characteristics can be further divided into parameters for quality assessment. The two definitions of quality may seem unrelated, but complement each other in practice. Customers' requirements and expectations can guide the product requirements, and a product that conforms to requirements is usually fit for use.

SIGDOC '19, October 04–06, 2019, Portland, OR

ACM ISBN 978-1-4503-6790-5/19/10…$15.00
https://doi.org/10.1145/3328020.3353936

2.1 Documentation Quality

Early definitions of documentation quality are mostly set for printed ones, due to the reason that computers and the Internet were not popularized as they are today. Bandes [2] selected eight characteristics of technical documentation quality, including psychological quality, time-oriented quality, contractual quality, ethical quality, physical appropriateness, accuracy, completeness, and usefulness. Detienne et al. [10] adopted a two-dimensional grid model from Seawright and Young [24] to visually plot the definition of technical communication quality. Their six major categories of quality definitions are transcendent, design-based, product-based, customer based, value-based, and strategic, plotted into the two-dimensional model as in Figure 1.

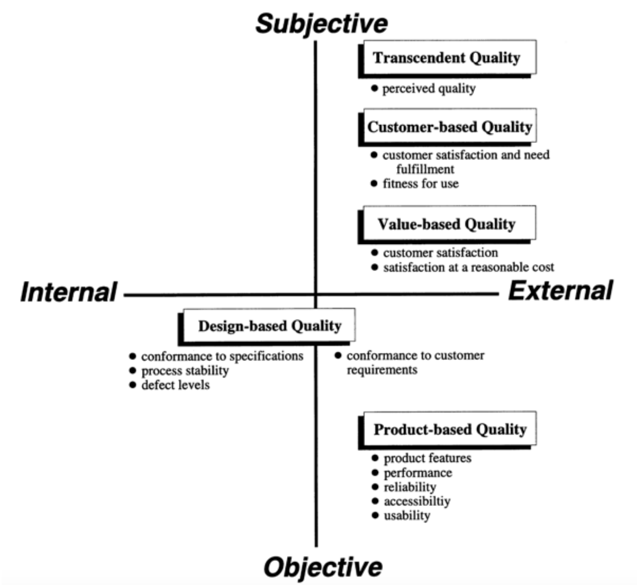

Figure 1: A 2-D grid model on definition of technical communication quality[10]

Waller [26] proposed a set of sixteen criteria for documentation benchmark from four aspects: language, design, relationship, and content. Carey, Lanyi, Lango et al. [4] presented three well-known quality characteristics for technical information based on comments from users and on the authors' experience in writing and editing technical information. "Easy to use" implies that good technical information should possess such quality characteristics as task orientation, accuracy, and completeness. "Easy to understand" means that the information needs to be of clarity, correctness, and style. Quality technical information should be organized properly with high retrievability and visually effective to be "easy to find". Zhi et al. [32] modeled the quality of documentation content using Content Quality, which has several attributes as its subclass: accessibility, accuracy, author-related, completeness, consistency, correctness, information organization, format, readability, similarity, spelling and grammar, traceability, trustworthiness, up-to-date-ness.

2.2 Help Center Quality

We decide to separate help centers from other kinds of websites, because they have more specific use, which is often about guiding users to the information they need to solve their problems. To evaluate the quality of a help center, it is necessary to be clear about what kind of quality we are looking for. German company K15t [6] found a way to improve their help center quality by asking the following questions about navigation, findability, solution and problems: 'Is our self-service portal easy to navigate?', 'Are visitors finding the information they are looking for?', 'Are visitors actually solving their problems?', and 'which problems are readers trying to solve?'. Erin Cochran [5] proposed several metrics to measure the success of a help center. Such metrics include the number of self-service users, self-service success rate, content engagement, and findability. And Zendesk [30] added more aspects such as content relevance, navigation success, and mobile device optimization.

We define the quality of a help center from two perspectives – findability and problem solving. Findability has two attributes as its subclass: navigability and search efficiency. Navigability describes the effectiveness of a help center's navigation function, which means users can click from one page to another until they find the one containing the information they need. Search efficiency describes the correctness of search engine results.

Problem-solving efficiency of the help center is measured from two perspectives; one is whether the user's problem is solved, and the other is the remaining problems. Whether the user's problem is solved, on the one hand, can be inferred from user's engagement with the document. On the other hand, we can also base the document quality on the user-rated score and the number of submitted tickets. The latter is used by Zendesk [30] in its self-service score– ratio between the total number of unique visitors that interacted with help content and the total number of unique users with tickets.

3 RELATED WORK

Current studies on general web quality testing often gather server log data for user behavior analysis. Clickstream data [3], mouse movement activity [19], eye tracking data [12], and AJAX application data [1] can all be used to evaluate user experience on a website or to pinpoint problems they may encounter. When assessing help center quality, we can borrow heavily from those approaches used in web quality evaluation. In this section, we categorize related works into navigability and documentation assessment. As the information container, a help center functions mainly in navigation and search. With search function improvement relating more to search engine optimization, we specifically take navigability and documentation assessment into consideration.

3.1 Navigability

Website designs incorporate multiple disciplines that affect web quality and user experience. As in the case of a help center, we consider navigation a crucial part since users always visit help centers seeking information they need, mostly following a path guided by the help center's navigation system. That being said, navigation systems provide an important means for supporting people's browsing and path selections to locate target information on a web site [27]. There has not been much work that targeted at evaluating the navigability of help centers specifically. There are, however, a large number of research conducted on measuring the navigability of other kinds of websites. Naturally, we believe that the navigability evaluation for any arbitrary website should not differ much, since navigation is the general process of determining a path to be travelled through a chosen environment [9]. It is claimed that the navigation design of a website should help users answer three fundamental questions when browsing the site, and they are: 'Where am I?', 'Where have I been?', and 'Where can I go?' [20].

To answer these questions is by no means an easy task and recent works concerning navigability evaluation mainly fall into three categories. The first is to analyze usage data of websites, mostly the log files containing user click-stream data. The second is to consider the web navigation structure, i.e., the outline to achieve an easy maneuver for the users accessing the site [11]. The third is to assess navigability against a set of criteria or check list [31]. When considering the website usage data or its navigation structure, graph theory is often adopted. Typically, a graph G is defined as a set of vertices, or points V and a set of edges E, where V denotes all webpages in a given website, and E represents all hyperlinks on those webpages; any vertex A in V is connected to another vertex B by an edge in E when there is a hyperlink from webpage A to B. In this case, when a user travels through pages in a web session, all those pages visited and hyperlinks clicked could form a graph G for assessment. Kalczynski et al. [17] employed such a graph-based approach to capture the navigational complexity. They analyzed 485 individual goal-oriented sessions on different websites and studied ten complexity metrics to select the ones that can be used to predict the result, i.e, whether the task is completed or not. Zhang et al [31], also following the graph theory, proposed a set of metrics for website navigability measurement, such as the total number of links on a website. Kaur and Dani [18] utilized similar evaluation metrics, and investigated the structural properties collected by HTML parsers deployed on banking websites. These measurements are mainly based on static hyperlink structure, and Winoto et al. [33] attempted to abstract dynamic user behavior as a Markov model, to serve as the basis for navigability measure. Markov model, borrowed from probability theory, is a stochastics model used to model randomly changing systems [14]. Fang et al. [13] proposed a data-driven approach to measure the website navigability guided by information foraging and information-processing theories.

3.2 Documentation assessment

The next part we are interested in is the assessment of help center's technical documentation. Supposing the user has trudged through the navigational structure of a help center, to the webpage containing the technical documentation, he still needs to locate the target information on the page, and quality documents should help users finish this step rather effectively. Then there is a question we need to answer, i.e., - how do we measure the documentation quality?

The work of Wingkvist et al. [29] reported an approach for using metrics to quantitatively measure documentation quality, based on the Goal-Question-Metric paradigm: predefined quality goals are continuously assessed and visualized by the use of metrics. They carried out two experiments with 'clone detection' and 'test convergence' analysis borrowed from software testing domain, and the experiments showed that quality issues can be identified. Garousi et al. [15] came up with a joint method comprised of two parts - first gathering document access logs and second asking for experts' opinions by a questionnaire-based survey. And this approach identified the most relevant factors affecting documentation quality. Shpak el al. [25] implemented a web-based information testing tool called "QAnalytics" to access information quality of documentation provided by a website. This tool offers an HTML proxy to the web-site and allows the tester to trace all interaction events between a user and the website in a web session, and then processes the data for analysis. The case study suggested that the precision and recall can help determine how understandable the webpages are; furthermore, some other statistical results such as the deviation in the number of clicks and time spent support their findings.

3.3 Conclusion

We investigated current studies which are primarily on websites' navigability evaluation and documentation quality assessment. While these studies did manage to capture some quantitative mea-surements on navigational complexity and criteria on assessing document quality, few has tried to define what a quality documen-tation is or to design a set of metrics specifically for help canter quality. Although previous quantitative approaches were verified to be capable of evaluating documentation quality in some way, they have failed to form systematic document categorizations and to fur-ther define assessment rules for documents of certain categories. In light of drawbacks mentioned here, we hope to build upon existing studies and aim to develop a set of computable metrics to evaluate the quality of help centers.

4 METHODOLOGY

In this section, we present the methodology used, which is struc-tured as three parts. Part one concerns with the key observations during the technical writing process and the assumptions we based on for the experiments. In part two we introduce how we pre-process the data collected before analysis. Finally, in the last part we briefly talk about the experiment design.

4.1 Observation and Assumption

The quality improvement for technical documentation will experience a 3-stage cycle of "development – measurement – cognition". In the development stage, new documentation are created or old ones get modified. The next would be measuring the quality of the documentation acquired from stage one. The last step then is to assess the documentation given the quality reflected by the measurement results. The documentation could be of high quality– such that it enables users to find information and solve problems effectively, or it may be problematic in some ways. If it is, then we will be back to the stage one and have it revised. After iterative efforts, documentation quality should be improved.

The most critical step in this cycle is the measurement of quality and there are three ways by which Baidu Cloud commonly evaluates documentation quality, including comparison between company standards and industry standards, company's internal review and analysis, and results from user evaluations on the documentation. There is, however, no single standard that could quantify and predict

Table 1: Variables Defined in the Experiment

Dependent Variable	Independent Variables
User Ratings	PV/UV downstream contribution rate exit rate average time on site landing page rate evaluation rate

the quality of an arbitrary documentation. We aim at the proposal of such standard and as the first step of our research, we explore in this report the correlation between user behavioral data and technical documentation quality.

Commonly speaking, users would give rather positive ratings on quality documentations since the latter tend to help solve their problems. Thus we carry out our experiments on the following three assumptions: 1) a correlation exists between user behavioral data and document evaluation scores given by the user, 2) the user behavioral data can be used to partially predict the evaluation scores and 3) user evaluation results of the documentation can reflect documentation quality. That is to say, there is a positive correlation between the two. We could then use the user behavioral data to predict the document quality, or user ratings in the sense that these two are positively correlated.

4.2 Data Processing

We collected user ratings and user behavioral data on Baidu Cloud's product documentations, all from Baidu Cloud Statistics Q2 2017 through Q4 2018. They were formulated below and listed in table 1.

With the intention of predicting quality, or ratings of a document given user behavioral data, the user ratings function as dependent variable in the setting. For any documentation, its user rating is a score from number 1 to 5, a higher score indicating a better quality. The independent variable, therefore, includes six types of user behavioral data we collected. The PV/UV [22] is calculated as the number of page views divided by unique visitor; the downstream contribution rate is the number of downstream browsings divided by page views, where the downstream browsings of a certain page refers to the number of browsings it brings for other pages; the exit rate represents the percentage of visitors to a page on the website from which they exit the website to a different website [28]; the average time on site calculates the amount of time a visitor spends on the site; the landing page rate is the ratio between the number of conversions in a given time frame and the total amount of visitors; lastly, the evaluation rate marks the percentage of visitors who give evaluation on the document.

As an illustration, supposing that the product description in the page for document A comprises five level one headings, then when a user browses the page, it will generate one landing page, four downstream browsings, one exit, five PV's and UV's, coupled with an 80% downstream contribution rate, a 20% exit rate and a 10% landing page rate. By the same token, for a page of document B with ten level one headings, it creates one landing page, nine downstream browsings, one exit, ten PV's and UV', with a 90% downstream contribution rate, a 10% exit rate and a 10% landing page rate.

4.3 Documentation Categorization

The total number of documentations with user ratings is huge and many of them are excluded in this experiment, for the reason that many are only rated a few times and this may cause the evaluation to be biased. In effect, we first choose all 142 documents that have 30 or more user evaluations for a general analysis, and then 265 documents with 20 or more user evaluations were selected. The next step is to categorize these 265 documents into multiple groups based on their contents, since documentations for different types of products and services might achieve contrasting results. Specifically, there are three groups of documentations in this experiment and they are dealt with separately.

Documentations of group one are 117 development documents on product development; users checking on them are assumed to possess at least some programming basics. An example of such documents is the Application Programming Interface (API) document for Baidu Cloud's cloud services (https://cloud.baidu.com/doc/BCC/API.html#.E5.88.9B.E5.BB.BA.E5.AE.9E.E4.BE.8B). Then there is the second group of 61 documents that are just operation procedures any user can simply follow along; to clarify, we take the document of a creation process on how to set up an instance of Baidu Cloud's cloud server as an example (https://cloud.baidu.com/doc/BCC/GettingStarted.html#.E6.93.8D.E4.BD.9C.E6.AD.A5.E9.AA.A4). Last but not the least, the 87 informational documentations are those that only cover product and service information, such as product prices or functionalities; for instance, the introduction document on what Baidu Cloud server is, its features and advantages is a documentation of such type (https://cloud.baidu.com/doc/BCC/ProductDescription.html#.E4.BB.8B.E7.BB.8D).

4.4 Experiment Design

Under the three assumptions mentioned earlier, we conduct correlation and regression analyses on dependent and independent variables formulated above. In other words, for every one of the six independent variables, we carry out the two analyses on it with the dependent variable, i.e., the user ratings. The two types of analysis are both based on multivariate distribution, which simply refers to the distribution of multiple variables. The correlation analysis notifies association, or the presence of relationship between two variables [23], while the regression analysis serves to predict the value of dependent variable on the basis of the independent variable [7]. Note that "predict" should not be interpreted as the ability to predict events in the future beyond the limits of data analysis [21]. To put it differently, the regression indicates to what extent does the change in dependent variable explain the dependent variable.

5 RESULTS

At a first attempt, we conduct the experiments on all 142 documents with 30 or more user evaluations. Presented here (Figure 2) is the sketch on the analyses between user ratings, which is the dependent variable and PV/UV, being one of the six independent variables. The solid line reveals the linear correlation between the two, while the dashed line demonstrates the quadratic result from regression analysis. We find that the key indices that contribute most to the model are PV/UV and the evaluation rate; the overall degree of interpretation is 43%, meaning there is a relationship between two variables but the correlation and prediction is not strong. This is a reasonable rate since all data observed were scattered around on the coordinates plane, and on top of that, we see a few outliers that are detached from the main group. As we suggested above, different document types might achieve contrasting results so our next attempt is to conduct our experiments on three types of documentations as categorized in Section 2.

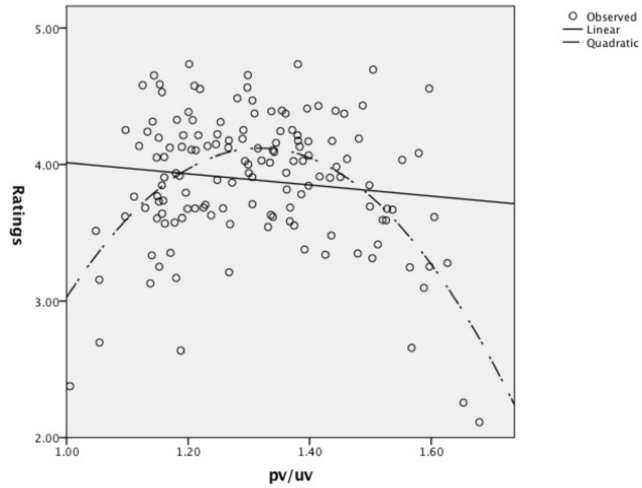

Figure 2: Correlation and regression analysis on all 142 doc-uments with 30 or more user ratings, with pv/uv being the independent variable.

In the final analysis, we check on the experimental results on the three groups of documents and furthermore, we all choose to present here the analysis on the PV/UV (Figure 3-5). Since as it seems that from our data, the PV/UV explains the ratings most for all three types of documentations. The analysis on development documentations reveals a negative correlation between two variables (Figure 2); in this case, a bigger PV/UV implies a smaller rating score. The logic underlying such correlation is that development documentations contain some parts that need programming, which is more time-consuming and difficult to operate on; more pageviews will result in a lot more time spent on those pages compared to the time needed to go over other types of documentations; consequently, a bigger PV, provided that UV, the denominator stays unchanged, may cause users to give a lower rating on the documents, as it expects tons of work from them. The degree of interpretation on development documents is 54.3%, indicating some predictive value of the model. The key indices are PV/UV, exit rate and downstream contribution rate. As for the informational documents (Figure 4) and documentations for operation procedures (Figure 5), there are no convincing correlations. Hence the analyses also justify our statement that different types of documents produce various trends and results.

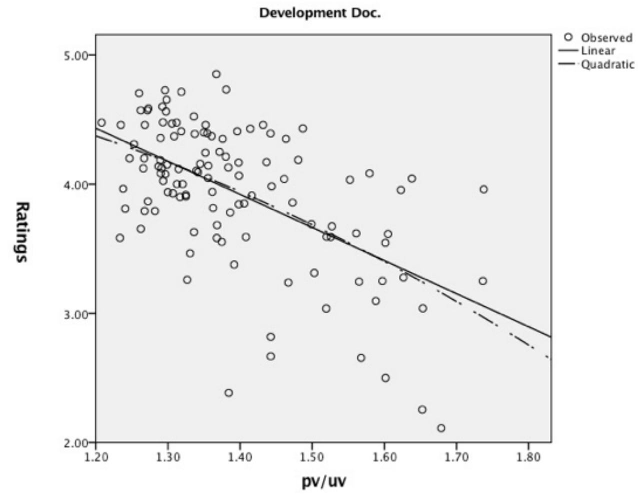

Figure 3: Correlation and regression analysis on 117 development documentations, with pv/uv being the independent variable.

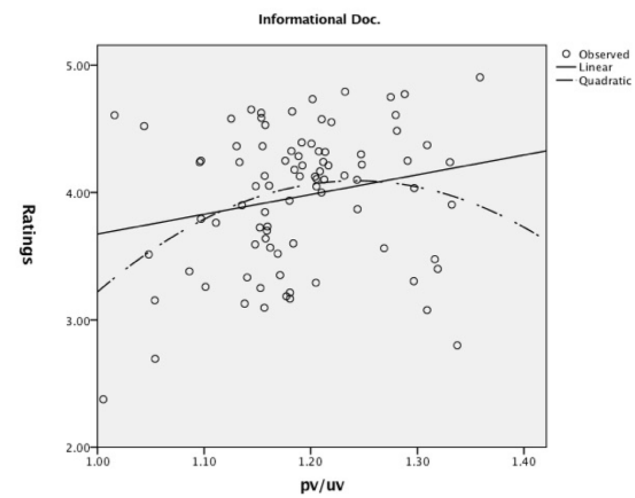

Figure 4: Correlation and regression analysis on 87 informa-tional documentations, with pv/uv being the independent variable.

6 CONCLUSION AND FUTURE WORK

This study is the first attempt toward our research goal, which is to propose a set of computable metrics that could implement a standard to quantify and evaluate the quality of help centers. We share the correlation and regression experiments conducted on Baidu Cloud's data, specifically the analyses on relationships between user behavioral data and documentation quality, where the latter in this case is the user ratings. However, the scope of this experiment is limited and certainly there are still many flaws to it. The data analysis part of the research, based on the three assumptions mentioned in Section 2.1, lacks qualitative analysis

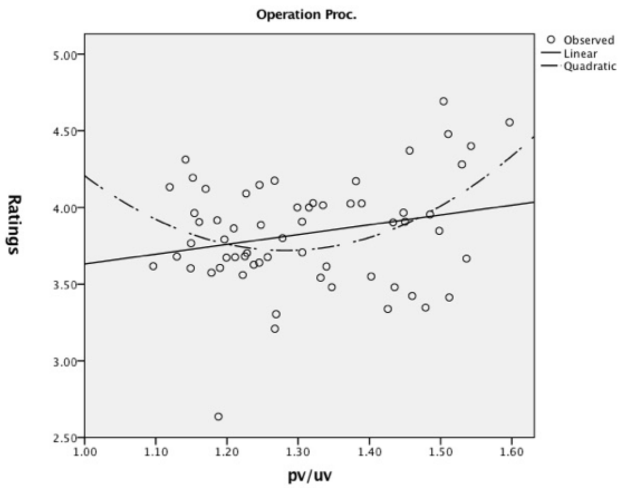

Figure 5: Correlation and regression analysis on 61 docu-ments of operation procedures, with pv/uv being the inde-pendent variable.

of technical documents. Since there are no standard definition on what a good technical document is, or qualitative approaches to evaluate the document quality, we take user ratings as a substitute. That is, however, not necessarily true since there could be many factors related to what score an arbitrary user would give. Rating a technical document is a subjective event; for example, the users' moods could also affect on how they rate things; some users just tend to give high ratings while it does not hold for some others. Moreover, the document classification is rough and there are potentially more accurate categorizations. Due to the difference in the number of users of Baidu Cloud products, the available documents scored are more concentrated on certain products, so the model may have different prediction strength on other types of documents as it is the case for three types of documents in this experiment. What is equally important should be the document contents. For example, such as the lengths of a document and of its subsections at the same page. It is also possible to construct a multi-feature user score for a more reliable predictive model.

In the follow-up study, we will first conduct user research and refine the elements of high-quality technical documentation from the user's perspective. Combining with results from the previous research, we plan to further the qualitative analysis of document quality and find out the general indicators used to measure the quality of documents; thereupon we select the parts that can be collected directly from back-end database for testing. For the purpose of evaluating the findability of a help center, we plan to research into the path through which the users try to find solutions to their problems. In essence, combining the help center findability score with the predicted document quality, we will iteratively establishing, testing and modifying models for help center quality evaluation, aiming to propose a standard architecture for the quantification of help centers.

REFERENCES

[1] Richard Atterer and Albrecht Schmidt. 2007. Tracking the interaction of users with AJAX applications for usability testing. (2007), 1347. https://doi.org/10.1145/1240624.1240828

[2] H. Bandes. 1986. Defining and controlling documentation quality : part I. *Technical Communication: Journal of the Society for Technical Communication* 33, 1 (1986), 6–9. https://www.jstor.org/stable/4189229

[3] Randolph E. Bucklin and Catarina Sismeiro. 2003. A model of web site browsing behavior estimated on clickstram data. *Journal of Marketing Research* 40, 3 (2003), 249–267. https://doi.org/10.1509/jmkr.40.3.249.19241

[4] Michelle Carey, Mc Fadden Lanyi, Deirdre Longo, Eric Radzinski, Shannon Rouiller, and Elizabeth Wilde. 2014. *Developing Quality Technical Information: A Handbook for Writers and Editors.* IBM Press.

[5] Erin Cochran. 2016. The Six-Step Guide to Robust Help Center Metrics. Retrieved 2019 from https://blog.rjmetrics.com/2016/03/09/the-five-step-guide-to-robust-help-center-metrics/

[6] K15T Company. 2018. How metrics help us measure help center effectiveness. Retrieved 2019 from https://www.k15t.com/blog/2018/01/how-metrics-help-us-measure-help-center-effectiveness#HowMetricsHelpUsMeasureHelpCenterEffectiveness-problems

[7] R. Dennis Cook and Sanford Weisberg. 1982. *Criticism and Influence Analysis in Regression.* Vol. 13. Sociological Methodology. 313–361 pages.

[8] P. B. Crosby. 1979. *Quality is free: The art of making quality certain.* McGrawHill.

[9] Rudolph P. Darken and John L. Sibert. 1993. A Toolset for Navigation in Virtual Environments.

[10] Kristen Bell Detienne, Karl L Smart, and Kristen Seawright. 1995. Defining Quality in Technical Communication - A holistic Approach. *Technical communication* (1995).

[11] Glen Doss. 2014. Designing Effective Web Navigation. *Towns University, Center for Applied Information Technology* (2014).

[12] C. Ehmke and S. Wilson. 2007. Identifying Web Usability Problems from Eye-tracking Data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers.* HCI.

[13] Xiao Fang, Paul Jen-Hwa Hu, Michael Char, Han-Fen Hu, Zhuo Yang, and Olivia R. Liu Sheng. 2012. A Data-Driven Approach to Measure Web Site Navigability. *Journal of Management Information Systems* 29, 2 (oct 2012), 173–212. https://doi.org/10.2753/MIS0742-1222290207

[14] Paul A. Gagniuc. 2017. Markov Chains: From Theory to Implementation and Experimentation. (2017), 1–256.

[15] G. Garousi, V. Garousi, M. Moussavi, G. Ruhe, and B. Smith. 2013. Evaluating usage and quality of technical software documentation: An empirical study. In *International Conference on Evaluation Assessment in Software Engineering.* https://doi.org/10.1145/2460999.24610038

[16] J. M. Juran and A. Blanton Godfrey. 1998. *Juran's Quality Control Handbook (5th Edition).* McGrawHill.

[17] Pawel J. Kalczynski, Sylvain Senecal, and Jacques Nantel. 2006. Predicting On-Line Task Completion with Clickstream Complexity Measures: A Graph-Based Approach. *International Journal of Electronic Commerce* 10, 3 (2006), 121–141.

[18] Arvinder Kaur and Diksha Dani. 2013. The Web Navigability Structure of E-Banking in India. *International Journal of Information Technology and Computer Science* 5, 5 (apr 2013), 29–37. https://doi.org/10.5815/ijitcs.2013.05.04

[19] F. Mueller and A. Lockerd. 2001. Cheese: tracking mouse movement activity on websites, a tool for user modeling. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2001), 279–280. https://doi.org/10.1145/634067.634233

[20] Jakob Nielson. 2000. *Designing Web Usability: The Practice of Simplicity.* New Riders.

[21] Boston University School of Public Health. 2001. Introduction to Correlation and Regression Analysis. Retrieved January 17, 2019 from http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_multivariable/bs704_multivariable5.html

[22] Jonathan W. Palmer. 1988. *Web Site Usability, Design, and Performance Metrics.* Vol. 13. INFORMS. 151–167 pages. https://www.jstor.org/stable/23011053

[23] J. L. Rodgers and W. A. Nicewander. 1988. *Thirteen ways to look at the correlation coefficient.* Vol. 42. The American Statistician. 59–66 pages.

[24] Kristie W. Seawright and Scott T. Young. 1996. A Quality Definition Continuum. *Interfaces* 26, 3 (1996), 107–113. www.jstor.org/stable/25062136

[25] O. Shpak, W. Lowe, A. Wingkvist, and M. Ericsson. 2014. A method to test the information quality of technical documentation on websites. In *International Conference on Quality Software.* 296–304. https://doi.org/10.1109/QSIC.2014.48

[26] Rob Waller. 2011. What makes a good documentâĂŕ? The criteria we use. (2011).

[27] Jane Webster and Jaspreet S. Ahuja. 2006. Enhancing the Design of Web Navigation Systems: The Influence of User Disorientation on Engagement and Performance. *Mis Quarterly* 30, 3 (2006), 661–678.

[28] WEELYTICS. 2016. Bounce rate vs Exit rate Use Cases. Retrieved January 20, 2019 from https://weelytics.com/category/digital-analytics/

[29] A. Wingkvist, M. Ericsson, R. Lincke, and W. Lowe. 2010. A metrics-based approach to technical documentation quality. In *International Conference on*

the Quality of Information and Communications Technology. 476–481. https://doi.org/10.1109/QUATIC.2010.88
[30] Zendesk. 2016. 6 steps for measuring self-service success. Retrieved April 20, 2018 from https://www.zendesk.com/blog/measuring-self-service-success/
[31] Yanlong Zhang, H. Zhu, and S Greenwood. 2004. Website Complexity Metrics for Measuring Navigability. In *International Conference on Quality Software*. IEEE.
[32] Jun Zhi, Golara Garousi, Shawn Shahnewaz, and Guenther Ruhe. 2015. Cost, benefits and quality of software development documentation: A systematic mapping. *Journal of Systems and Software* 99, C (2015), 175–198. https://doi.org/10.1016/j.jss.2014.09.042
[33] Yuming Zhou, Hareton Leung, and Pinata Winoto. 2007. MNav: A Markov Model-Based Web Site Navigability Measure. *IEEE Transactions on Software Engineering* 33, 12 (2007), 869–890.