

# Assessment of digital library design guidelines to support blind and visually impaired users: a study of key stakeholders' perspectives

Iris Xie

*School of Information Studies, University of Wisconsin-Milwaukee,  
Milwaukee, Wisconsin, USA*

Rakesh Babu

*Envision, Wichita, Kansas, USA, and*

Shengang Wang, Hyun Seung Lee and Tae Hee Lee

*School of Information Studies, University of Wisconsin Milwaukee,  
Milwaukee, Wisconsin, USA*

Received 31 May 2022  
Revised 28 June 2022  
Accepted 13 July 2022

## Abstract

**Purpose** – This study aims to investigate the perceptual differences of key stakeholders in assessing the Digital Library Accessibility and Usability Guidelines (DLAUG), in which design information is created and organized by types of help-seeking situations, to support blind and visually impaired (BVI) users. The stakeholders consist of BVI users, digital library (DL) developers and scholars/experts. The focus is on the identification of types of situations in which BVI users and developers show significant perception differences of DLAUG's relevance, clarity and usefulness than the other two groups, respectively, and the associated reasons.

**Design/methodology/approach** – An in-depth survey was conducted to examine the perceptions of 150 participants representing three groups of key DL stakeholders: BVI users, DL developers and scholars/experts. Both qualitative and quantitative analyses were applied.

**Findings** – The results show that BVI users and developers had significant perception differences of the relevance, clarity and usefulness of the DLAUG than the other two groups held on five situations, mainly because they played distinct roles in the development of DLs with differing goals and expectations for the DL design guidelines.

**Originality/value** – This is the first study that considers different DL stakeholders to assess DL guidelines to support BVI users.

**Keywords** Assessment, Guidelines, Blind and visually impaired users, Digital library design, Stakeholders

**Paper type** Research paper



## 1. Introduction

The complex and sight-centered design of digital libraries (DLs) hampers their accessibility and usability for blind and visually impaired (BVI) users. A DL is “an online collection of digitized or born-digital items, equipped with an interactive interface to enable diverse user

communities to find desired information effectively” (Xie *et al.*, 2021a, p. 995). A BVI user relies on screen readers to interact with desktop computers and mobile devices non-visually. Making DLs accessible and usable is critical for BVI users (Mátrai, 2018). Accessibility and usability of Web applications are mainly guided by the *Web Content Accessibility Guidelines* (WCAG) from the World Wide Web Consortium (W3C). However, studies indicate a gap between the accessibility and usability issues addressed by WCAG and the issues faced by BVI users on the Web (Kreps and Goff, 2015; Power *et al.*, 2012). Recent research shows that BVI users face many help-seeking situations when interacting with a DL (Gaona-García *et al.*, 2017; Xie *et al.*, 2018a, 2021a), which have not been fully addressed in the WCAG. A help-seeking situation refers to a problem that BVI users identify, indicating that some help is needed to facilitate their interactions and accomplish their goals/tasks (Xie *et al.*, 2021b). Additionally, research shows that developers and accessibility professionals have difficulty implementing WCAG because of its insufficient coverage of accessibility problems, confusing structure and vague vocabulary, as well as arduous implementation suggestions (Power *et al.*, 2012; Trewin *et al.*, 2010).

Relevance, clarity and usefulness are the frequently applied criteria for evaluating accessibility and usability guidelines (Alajarmeh, 2022; Calvo *et al.*, 2016). The *relevance* of guidelines is determined by checking if existing accessibility problems are addressed in the guidelines. *Clarity* is associated with clear and easily understandable guidelines in addressing instructions for potential readers. *Usefulness* is related to applicability and is associated with whether guidelines can be implemented successfully in designing a website or a DL.

For guidelines to be relevant, clear and useful, different perspectives from key DL stakeholders, such as BVI users, scholars/experts and developers, need to be investigated. Although previous research recognizes the importance of incorporating different stakeholders’ views in developing and evaluating a system (Selvadurai *et al.*, 2019; Tu *et al.*, 2021), there is a lack of research done in the DL context in examining the perception gap among users, developers and scholars. This research investigated diverse stakeholder perceptions when examining the *Digital Library Accessibility and Usability Guidelines* (DLAUG <https://sites.uwm.edu/guidelines/>) in terms of their relevance, clarity and usefulness. DLAUG is a set of guidelines covering diverse recommendations for designing an accessible and usable DL to support BVI users. DLAUG was developed based on a series of user studies that identified 37 unique help-seeking situations BVI users face in their interactions with existing DLs. Because of space limitations, in the present paper, the authors only propose and report on the research questions related to the perceptual differences between users and the other groups and between developers and the other groups, as the perspectives of users and developers are especially important (Cheoh *et al.*, 2020; Patel *et al.*, 2020).

The related research questions are as follows:

- RQ1. In which help-seeking situations do users show significantly different perceptions of DL design guidelines’ relevance, clarity and usefulness than scholars/experts and developers? What are the associated reasons for the differences?
- RQ2. In which help-seeking situations do developers show significantly different perceptions of DL design guidelines’ relevance, clarity and usefulness than scholars/experts and users? What are the associated reasons for the differences?

## 2. Literature review

### 2.1 *The importance of considering different stakeholders' perceptions*

Studies have highlighted the need to consider different stakeholders' views in system development and evaluation (Jia and Capretz, 2018; Selvadurai *et al.*, 2019; Velleman *et al.*, 2017). The same applies to developing DL design guidelines. Different stakeholders have differing views and understandings of accessibility and usability guidelines. A holistic evaluation of DLs should incorporate diverse stakeholders' views to ensure evaluation validity (Xie *et al.*, 2018a, 2021c; Zhang, 2010). In the DL context, users, scholars/experts and developers are recognized as the key stakeholders. Studies have highlighted the importance of incorporating users' insight and needs in designing a DL (Li and Liu, 2019; Ngimwa and Adams, 2013; Xie *et al.*, 2018b). While users' opinions are based on their previous interactions with DLs, developers' perspectives result from their work experience (Bai *et al.*, 2019). As scholars share some of the same expertise and viewpoints as users and DL developers (Xie *et al.*, 2018b), they consider both users' and developers' perspectives, bridging the perception gap between users and developers.

The perception differences during design, development and assessment are recognized to be influenced by stakeholders' backgrounds, expertise and work roles (Yesilada *et al.*, 2012, 2015). The users' role focuses on using a system and finding relevant information (Cooper *et al.*, 2012; Xu and Du, 2019). Colusso *et al.*'s (2017) study illustrated how different work roles influence developers' perceptions. As the developers' work role is to implement accessible information in practice, they criticize scholarly works for their insufficient actionable resources, misleading recommendations and differences in vocabulary use. Accessibility experts stress the advantages of a user-centered approach and training on WCAG more strongly than non-experts (Yesilada *et al.*, 2015). The key stakeholders' insights are needed for the guidelines to be relevant to the situations faced by BVI users, understandable to the key stakeholders and useful when designing a DL.

### 2.2 *Accessibility guidelines assessment*

**2.2.1 Relevance.** Research on the relevance of accessibility guidelines has concentrated on whether they reflect existing problematic situations that people with disabilities face (Calvo *et al.*, 2016; Rømen and Svanæs, 2012). Guidelines are considered relevant when WCAG 2.0 success criteria address problematic situations for users (Clegg-Vinell *et al.*, 2014; Power *et al.*, 2012).

Studies have revealed different stakeholders' views on the relevance of accessibility guidelines. Users with disabilities are the main participants in assessing the relevance of guidelines (Alajarmeh, 2022; Cheoh *et al.*, 2020; Clegg-Vinell *et al.*, 2014). Research involving users emphasized identifying accessibility problems which were used to evaluate the relevance of WCAG. Specifically, blind participants identified accessibility problems and associated severity levels, but only 50.4% of these situations were addressed in WCAG 2.0 success criteria (Power *et al.*, 2012). Similarly, Rømen and Svanæs (2012) noted that only 32% of situations identified by BVI users were acknowledged in WCAG 2.0. Alternatively, accessibility experts used other approaches to assessing accessibility (e.g. semi-automatic tools, code inspections). They identified several categories of design problems that were not addressed in WCAG 2.0 AA success criteria (Calvo *et al.*, 2016). Recently, Alajarmeh (2022) pointed out that there were still some problems that are not addressed in existing criteria in WCAG 2.1 in a mobile environment for visually impaired users.

**2.2.2 Clarity.** The clarity of guidelines is one of the topics examined in the assessment of accessibility guidelines and resources. Farrelly (2011) revealed the problems of WCAG 2.0, including the lack of clarity, obtuse language and convoluted organization. Previous studies

pointed out that WCAG should address the communication challenges different audiences face to improve its clarity (Abuaddous *et al.*, 2016; Watanabe *et al.*, 2017). They stressed the importance of using clear terminology and language to the targeted audiences (e.g. web designers) and providing relevant content suiting their background knowledge.

Some studies considered a lack of language clarity as one of the reasons for confusion in understanding guidelines (Swallow *et al.*, 2016). Snider *et al.* (2020) identified different usages and understandings of terminologies by diverse stakeholders and associated problems using the guidelines. Furthermore, developers have difficulty interpreting and understanding accessibility guidelines to implement during Web development (Abuaddous *et al.*, 2016). In Petrie *et al.*'s (2011) study, developers complained that the language used in accessibility guidelines was too foreign and vague to be applied. Moreover, developers were confused about instructions with vague language and unfamiliar categorization; these included categorizing guidelines as perceivable, operable, understandable and robust (Swallow *et al.*, 2016). Concerning clarity to the targeted audience, Brajnik *et al.* (2011, 2012) researched the impact of the expertise of scholars/experts when using the WCAG. They noted that the clarity of the guideline is vital to alleviate discrepancies between different evaluators (Brajnik *et al.*, 2011).

*2.2.3 Usefulness.* Prior research has investigated the usefulness of accessibility guidelines and tools in relation to their applicability in implementing guidelines (Gaggi and Pederiva, 2021). Previous research stressed the need to incorporate web developers' input in creating useful accessibility guidelines (Trewin *et al.*, 2010). Even with accessibility tools and guidelines, accessibility issues are prevalent on the Web (Alajarmeh, 2022; Haider and Yesilada, 2020).

Challenges faced by developers in creating an accessible Web are associated with the large volume of accessibility guidelines (Abuaddous *et al.*, 2016; Swallow *et al.*, 2016), the lack of tools encapsulating all accessibility criteria (Frazão and Duarte, 2020), difficulty understanding and prioritizing accessibility requirements (Trewin *et al.*, 2010) and insufficient details and explanatory examples (Farrelly, 2011). Patel *et al.* (2020) interviewed developers regarding the challenges in implementing an accessible website. The participants valued digestible-sized guidelines that could be implemented in a short time. They also preferred detailed instructions, such as how to build accessible components to resolve accessibility issues. Some studies developed accessibility resources to accommodate developers' challenges and encourage the development of an accessible Web (Gaggi and Pederiva, 2021). The tools and resources were not only designed to reduce the volume of information but also to provide the rationale for accessibility recommendations (Swallow *et al.*, 2016) and a comprehensive list of accessibility tools (Gaggi and Pederiva, 2021) to help developers understand and implement accessibility requirements.

### 3. Methodology

Survey design, using an in-depth survey instrument, was used for this study.

#### 3.1 Sampling

Users, scholars/experts and DL developers were the three key types of stakeholders in this study. In total, 150 participants were recruited, with 50 participants from each group. BVI users were mainly recruited with the help of national BVI organizations. The authors conducted literature searches in academic databases to collect contact information of scholars who have publications related to DLs and/or accessibility and usability of systems. Accessibility and usability experts were recruited via related listservs. An invitation flyer was sent to DL developers in different academic libraries across the USA to recruit them.

Participants with various demographic characteristics were recruited to represent these three groups.

Table 1 shows the basic demographic information of the participants. The main research areas of the 50 scholar/expert participants consist of accessibility, usability, DLs, BVI users and design guidelines and principles. DL developers involved in the study had an average of 9.31 years of experience in DL services. Their work titles include Web application developer, digital project developer, digital project librarian, head of digital collections and so forth.

3.2 Data collection

An in-depth survey was administered to 150 participants representing three types of stakeholders to obtain their assessment. The DLAUG provided associated guidelines for 37 situations that BVI users encounter when interacting with DLs. When assessing the guidelines for a situation using a seven-point Likert scale, participants rated the associated guidelines based on perceived clarity, relevance and usefulness, and the definitions of these terms were provided in the general instruction to the participants as specified in the Introduction. Additionally, participants were also instructed to provide reasons for their ratings. Figure 1 presents a sample of the survey.

3.3 Data analysis

First, one-way ANOVA was applied to analyze the numerical ratings to reveal the similarities and differences among the three groups of stakeholders in their assessment of the guidelines. Second, when any statistically significant difference was observed, a *post hoc* test using Tukey’s honestly significant difference (HSD) method was conducted to compare all possible pairs of group means to identify specific situations where a user or developer group rated the relevance, clarity or usefulness of the related guidelines significantly differently from the others. Moreover, to better understand the reasons behind the different ratings from

Demographic characteristics	User (n = 50) (%)	Scholar/expert (n = 50) (%)	Developer (n = 50) (%)	Total (N = 150) (%)
<i>Age</i>				
18–29	16	14	2	11
30–39	26	38	44	36
40–49	30	22	42	31
50–59	14	16	10	13
60–69	10	10	2	7
70+	4	0	0	1
<i>Gender</i>				
Female	66	44	60	57
Male	32	56	38	42
Other	2	0	2	1
<i>Ethnicity</i>				
Asian/Pacific islander	4	24	8	12
Black or African American	10	10	0	7
Hispanic or Latino	12	10	10	11
White	72	52	82	69
Other	2	4	0	2

**Table 1.** Demographic information of three groups of participants

**1) Difficulty accessing alternative text for an image**

Review the information presented on this page carefully, and share your feedback about the Guidelines. Specifically, rate the importance, clarity, relevance, and usefulness of the Guidelines that address BVI users' difficulty accessing alternative text for an image using a scale of 1 (Not at all) to 7 (Extremely), and provide your rationale for these ratings respectively. The following link will take you to a page that explains design guideline to address BVI users' difficulty accessing alternative text for an image:

<URL> <http://people.uwm.edu/guidelines/accessing/difficulty-accessing-alternative-text-for-an-image/>

- Please rate the **clarity** of the guidelines for this situation using a scale of 1 (Not at all) to 7 (Extremely):

Briefly justify your rating:

Please offer any suggestions, examples or resources to improve the efficacy of the Guidelines in addressing the difficulty accessing alternative text for an image below:

- Please provide your suggestions for the **guidelines**;
- Please provide your suggestions for **Associated techniques and methods**;
- Please provide your suggestions for **Desired features**; and
- Please provide your suggestions for **Examples and related resources**.

**Figure 1.**  
Example of the  
survey

stakeholders, the authors performed a qualitative analysis of participant comments regarding the relevance, clarity and usefulness of guidelines pertaining to the above quantitative analysis.

**4. Results**

This section presents the results of the two types of comparisons corresponding to *RQ1* and *RQ2*. First, the results show significantly different perceptions of the relevance, clarity and usefulness of DL design guidelines between users versus scholars/experts and developers. Second, the results show significantly different perceptions of the clarity and usefulness of DL design guidelines between developers versus users and scholars/experts. Additionally, associated reasons are offered for the quantitative results. [Table 2](#) presents the ANOVA results based on *RQ1* and *RQ2*.

Criterion	RQ	Situation	Users	Scholars	Developers	ANOVA
Relevance	<i>RQ1</i>	DLSW	6.64	5.36	5.60	$F(2,72) = 5.649, p < 0.05$
Clarity	<i>RQ1</i>	DAAT	6.08	4.84	4.88	$F(2, 72) = 7.585, p < 0.05$
		DURS	6.84	6.20	6.20	$F(2, 72) = 4.439, p < 0.05$
	<i>RQ2</i>	DLNA	6.44	6.56	5.56	$F(2, 72) = 6.205, p < 0.05$
		CDLS	6.68	6.28	5.40	$F(2, 72) = 6.986, p < 0.05$
Usefulness	<i>RQ1</i>	DAAT	6.68	5.60	5.80	$F(2, 72) = 6.796, p < 0.05$
	<i>RQ2</i>	DLNA	6.68	6.44	5.56	$F(2, 72) = 6.973, p < 0.05$

**Notes:** DLSW = difficulty locating a specific word/phrase; DAAT = difficulty assessing alt text; DURS = difficulty understanding results structure/layout; DLNA = difficulty locating a navigational aid; and CDLS = confusion about digital library structure

**Table 2.**  
Combined ANOVA  
results for  
comparison among  
stakeholders

#### 4.1 Users vs scholars/experts and developers

4.1.1 *Relevance.* There were significant differences among the three groups in rating of the relevance of the guidelines for difficulty locating a specific word/phrase (DLSW) (users:  $M = 6.64$ ; scholars/experts:  $M = 5.36$ ; developers:  $M = 5.60$ ,  $F(2, 72) = 5.649$ ,  $p < 0.05$ ). According to the *post hoc* test results, users perceived the guidelines for DLSW ( $M = 6.64$ ,  $SD = 0.86$ ) significantly more relevant than those of developers ( $M = 5.60$ ,  $SD = 1.50$ ,  $p < 0.05$ ) and those of scholars/experts ( $M = 5.36$ ,  $SD = 1.78$ ,  $p < 0.05$ ).

When assessing the relevance of DLAUG, BVI users considered whether the situation was relatable to their past experiences. One user commented, "I have often experienced difficulty locating a specific word or phrase in DL pages, so these guidelines are very important to provide potential remedies to BVI users" (UB19). The reason behind users' relevance perception might be that, as information seekers, they encountered this type of difficult situation during their interactions with DLs. Also, the associated guidelines seemed capable of addressing problems they had experienced, which fulfilled their expectations of the guidelines. Scholars/experts paid attention to whether the guidelines targeted the entire search process. For example, one scholar was concerned that the guidelines for DLSW were incomplete because "It is relevant only at the first search at the beginning of a search" (SB13). Scholars/experts' relevance perception was related to their research experience of conducting user studies, which enabled them to consider potential issues at different phases of users' search processes. Developers did not directly address relevance but mainly focused on issues related to clarity and usefulness. For example, DB9 requested more useful guidelines related to a recommended feature, stating that "It is good to think about how different kinds of 'search within the document' features might be made accessible, but this needs more detail" (DB9). It seems that DL developers focused on two things: identification of individual features that should be included in the guidelines and how to implement those features in practice. DL developers usually have less knowledge of BVI users' needs and behaviors than scholars/experts, and they focused more on practical issues of DL development. What DL developers expected from relevant guidelines were primarily detailed and easy-to-implement design examples and instructions.

4.1.2 *Clarity.* There were significant differences among the three groups in rating of the clarity of the guidelines for difficulty assessing alt text (DAAT) for an image (users:  $M = 6.08$ ; scholars/experts:  $M = 4.84$ ; developers:  $M = 4.88$ ,  $F(2, 72) = 7.585$ ,  $p < 0.05$ ) and difficulty understanding results structure/layout (DURS) (users:  $M = 6.48$ ; scholars/experts:  $M = 6.20$ ; developers:  $M = 6.20$ ,  $F(2, 72) = 4.439$ ,  $p < 0.05$ ). According to the *post hoc* test results, users ( $M = 6.08$ ,  $SD = 1.04$ ) perceived the guidelines for DAAT significantly clearer than developers ( $M = 4.88$ ,  $SD = 1.17$ ,  $p < 0.05$ ) and scholars/experts ( $M = 4.84$ ,  $SD = 1.57$ ,  $p < 0.05$ ). Users ( $M = 6.84$ ,  $SD = 0.47$ ) perceived the guidelines for DURS significantly clearer than developers ( $M = 6.20$ ,  $SD = 0.91$ ,  $p < 0.05$ ) and scholars/experts ( $M = 6.20$ ,  $SD = 1.12$ ,  $p < 0.05$ ).

When assessing the clarity of the guidelines for DAAT, BVI user participants paid attention to whether the guidelines concerning specific elements (e.g. alt text) were described in a concise manner. For example, UA2 valued the clarity of the guidelines for DAAT, commenting that "The guideline is quite clear to me. I love the focus on conciseness." The likely reason behind this clarity perception is that users have less system design knowledge of evaluating or implementing the guidelines, and they see conciseness as a dimension of clarity. From a different perspective, scholars/experts considered whether guidelines were described using theoretical concepts and whether design recommendations were easy to understand for DL developers. For example, one scholar mentioned that the:

Description of this guideline seems oriented to abstractions, and which require the creator of alt text to interpret from his or her own current knowledge. It is not realistic to expect individuals without specialized knowledge of the needs of BVI readers to know what constitutes "clear and concise" descriptions, or "context-sensitive" descriptions [...] (SA7).

Scholars/experts realized that developers might not have first-hand knowledge of BVI users' unique needs, and they expected that guidelines could link recommended features to illustrative examples of implementation to make it easier for DL developers to follow in practice. From a practical perspective, DL developers cared whether the provided examples accurately demonstrated the recommended techniques. For instance, DA3 pointed out that "The language of the guidelines is a bit confusing to follow in a practical manner. The examples weren't clearly demonstrating the technique as far as I could tell." In addition to language, developers considered the guideline structure to be an important factor related to clarity. DA10 focused on the structure of the writing for rationales and objectives, mentioning that "The specification of rationale and objective may need to change a little for the order to issue, phenomena, solution, and significance." The reason behind the developers' clarity perception is related to their role in implementing design guidelines because they care about how clearly the guidelines are written and how logically they are organized, which would impact the ease of understanding and applying the guidelines in practice.

*4.1.3 Usefulness.* There were significant differences among the three groups in rating of the usefulness of the guidelines for DAAT (users:  $M = 6.68$ ; scholars/experts:  $M = 5.60$ ; developers:  $M = 5.80$ ,  $F(2, 72) = 6.796$ ,  $p < 0.05$ ). According to the *post hoc* test results, users ( $M = 6.68$ ,  $SD = 0.80$ ) perceived the guidelines for DAAT significantly more useful than developers ( $M = 5.80$ ,  $SD = 1.26$ ,  $p < 0.05$ ) and scholars/experts ( $M = 5.60$ ,  $SD = 1.19$ ,  $p < 0.05$ ).

When assessing the usefulness of DLAUG, BVI users considered whether guidelines recommended multiple ways to help overcome the situation concerned. For example, UA19 mentioned that "I love that there are several different methods being considered, including the audio descriptions and the detailed descriptions." The reason behind this usefulness perception is that different methods could help address the difficult situation. For scholars/experts, they emphasized whether recommendations were illustrated using relevant examples. SA9 commented that "This provides some useful information. However, it may be even more so if the recommendations for image accessibility are hyperlinked to a notable example." The reason behind this usefulness perception is that some scholars/experts had research experience in assessing information retrieval systems, and they expected the guidelines to link recommended features to illustrative examples of implementation, which will help DL developers apply guidelines. When assessing the usefulness of DLAUG, DL developers tended to appreciate multiple best practice examples in the guidelines that they can follow in practice. To enhance the usefulness of the guidelines for DAAT, DL developers suggested that more examples of good alt text be included. For example, DA13 emphasized that "[...] more examples of good alternative text would make the guidelines more useful." DL developers are those who implement design guidelines in practice, and they expect that each recommended method/technique/feature is described along with multiple examples to facilitate its implementation.

## 4.2 Developers vs users and scholars/experts

*4.2.1 Clarity.* There were significant differences among the three groups in rating of the clarity of the guidelines for difficulty locating a navigational aid (DLNA) (users:  $M = 6.44$ ; scholars/experts:  $M = 6.56$ ; developers:  $M = 5.56$ ,  $F(2, 72) = 6.205$ ,  $p < 0.05$ ) and confusion about DL structure (CDLS) (users:  $M = 6.68$ ; scholars/experts:  $M = 6.28$ ; developers:  $M = 5.40$ ,  $F(2, 72) = 6.986$ ,  $p < 0.05$ ). According to the *post hoc* test results, developers perceived the guidelines for DLNA ( $M = 5.56$ ,  $SD = 1.19$ ) were significantly less clear than those of users ( $M = 6.44$ ,  $SD = 1.23$ ,  $p < 0.05$ ) and those of scholars/experts ( $M = 6.56$ ,  $SD = 0.82$ ,  $p < 0.05$ ). Developers perceived the guidelines for CDLS ( $M = 5.40$ ,  $SD = 1.63$ ) were significantly



less clear than those of users ( $M = 6.68$ ,  $SD = 0.85$ ,  $p < 0.05$ ) and those of scholars/experts ( $M = 6.28$ ,  $SD = 1.10$ ,  $p < 0.05$ ).

When assessing the clarity of DLAUG, DL developers considered whether the guidelines clearly differentiated good and bad DL structures and demonstrated associated methods. For example, when discussing the guidelines for CDLS, DB9 mentioned that “I don’t feel that this guideline was very clear about conveying what made a good or bad DL structure, and what methods were effective for conveying that structure to users.” The reason behind this clarity perception is that DL developers expected clear explanations of good and bad DL structures and associated methods to create ideal DL structures that BVI users desire. Unlike DL developers, BVI user participants considered things like how simple and understandable the writing was and whether technical jargon was used. For example, UB19 said, “the guidelines are expressed in clear, simple, and understandable language, devoid of technical jargon.” The reason behind this clarity perception is BVI users’ inadequate IT background related to DL development and interface design, making them prefer understandable language to technical or specialized jargon. For scholars, their assessment of clarity was based on whether the guidelines will clearly enhance accessibility and usability. For example, SB20 commented that “The text clearly stated how a page was to be put together so a screen reader user could locate navigational elements.” The reason behind this clarity perception is that scholars/experts expected these guidelines to explain how the intended accessibility could be clearly conveyed to DL developers. Also, scholars considered the structure of the guidelines. For example, SB19 mentioned that “Guidelines clearly describe the issues and potential resolution.” The reason behind this perception is that scholars considered the logical structure to be an important facet of guideline clarity and expected the guidelines to describe accessibility and usability issues in a logical way.

*4.2.2 Usefulness.* There were significant differences among the three groups in rating of the usefulness of the guidelines for DLNA (users:  $M = 6.68$ ; scholars/experts:  $M = 6.44$ ; developers:  $M = 5.56$ ,  $F(2, 72) = 6.973$ ,  $p < 0.05$ ). According to the *post hoc* test results, developers perceived the guidelines for DLNA ( $M = 5.56$ ,  $SD = 1.29$ ) were significantly less useful than those of users ( $M = 6.68$ ,  $SD = 0.75$ ,  $p < 0.05$ ) and those of scholars/experts ( $M = 6.44$ ,  $SD = 1.23$ ,  $p < 0.05$ ).

When assessing the usefulness of DLAUG, developers paid attention to whether there was a clear and prioritized recommendation list. For example, DB4 expressed her concern that “This is another kitchen-sink help page that lists a variety of recommendations of different sizes without a sense of priority.” To developers, they expected the guidelines to inform them of the relative importance or priority of specific guidelines, such as which guidelines/recommendations are required or optional. Unlike developers, users considered how the guidelines would help them in their daily interactions with DLs. For example, UB3 said, “Following these guidelines will make navigating pages much more efficient.” The reason behind users’ usefulness perception is that the recommended guidelines would help them efficiently interact with DLs. For scholars, they cared about whether the guidelines could help developers design a DL. For example, SB19 commented positively on the usefulness of the guidelines to help create easy-to-navigate DLs, saying that “Guidelines provide highly useful information for designers to provide an easily navigable site.” The reason behind this usefulness perception is that scholars assessed the guidelines from the perspective of DL developers and focused on how well users’ requirements can be translated into practice.

## 5. Discussion

This is the first study that investigates the perceptions of three types of stakeholders related to the relevance, clarity and usefulness of DL design guidelines. The findings of this study

echoed the results of some previous research. Most importantly, this study generates significant theoretical and practical implications. DLAUG was updated based on feedback from the three groups (BVI users, scholars/experts and DL developers). Figure 2 presents the perceptions of DL stakeholders regarding the relevance, clarity and usefulness of DLAUG and their standpoints.

### 5.1 Theoretical implications

The results of the study demonstrate that it is critical to incorporate perspectives from diverse stakeholders into DL design guidelines, confirming the results of prior research (Colusso *et al.*, 2017; Xie *et al.*, 2018b; Yesilada *et al.*, 2015; Zhang, 2010). Nevertheless, this research further explores the standpoints of stakeholders and their associations with their rating rationales on relevance, clarity and usefulness. Three groups of stakeholders had different perceptions of the relevance, clarity and usefulness of the DLAUG because of the following primary reasons. First, they played differing roles in the development of DLs. Previous research shows that stakeholders' roles influence their assessment of guidelines (Colusso *et al.*, 2017; Cooper *et al.*, 2012; Xu and Du, 2019; Yesilada *et al.*, 2015). This study specifies the roles of the three stakeholder groups. In their role, BVI users mainly use DLs for information-seeking purposes. Scholars/experts have a role in conducting research on issues related to BVI users' interactions with systems and conducting accessibility and usability testing for systems. The role of DL developers is to create accessible and usable DLs. Second, not only do they play different roles, but they also possess varying levels of experience and knowledge. Prior research points out that each stakeholder's opinion is affected by their experience and knowledge (Bai *et al.*, 2019; Xie *et al.*, 2018b; Yesilada *et al.*, 2015). This study further reveals that users experience help-seeking situations in their interactions with DLs, but they have inadequate knowledge of the accessibility and usability of DLs. Conversely, DL developers lack knowledge about BVI users, their behaviors and

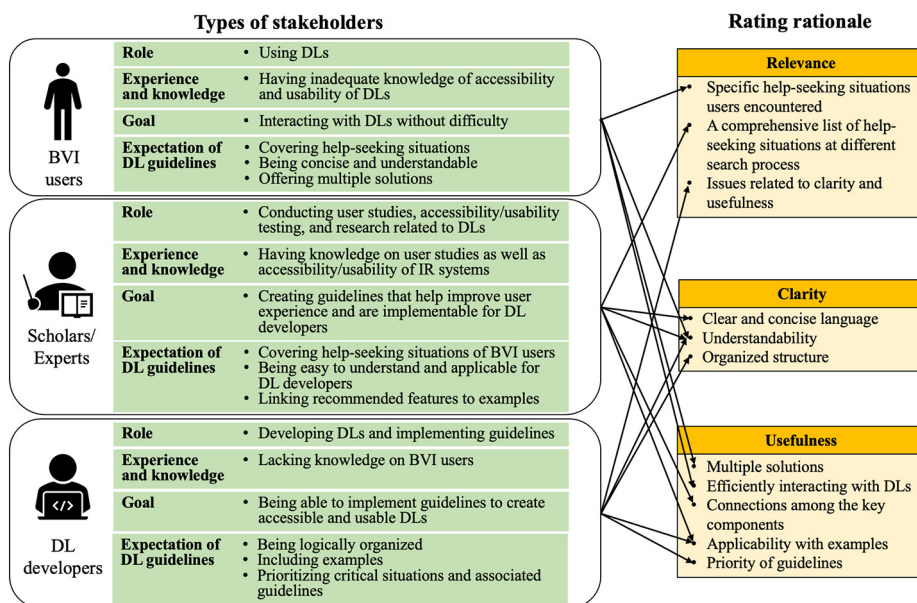


Figure 2. Perceptions of DL stakeholders on relevance, clarity and usefulness of DLAUG

---

their needs. Scholars/experts are the intermediaries who know about both users and the accessibility and usability of DLs. Third, this study identifies that these three groups have their own goals and associated expectations that previous studies failed to explore. The primary goal of users is to interact with DLs efficiently. Their expectations for the guidelines are that the problems they encounter in interacting with DLs are addressed in multiple ways, and that the guidelines are written clearly and understandably for laymen. The goal of developers is to realistically implement the guidelines in the development of DLs. They expect the guidelines to be logically organized, include sufficient examples and be prioritized, allowing them to successfully build accessible and usable DLs that support BVI users. The foremost goal of scholars/experts is to develop guidelines that not only address user problems but also facilitate implementation by the developers. They expect the guidelines to cover DL accessibility and usability problems in the search process and to be easy to understand as well as applicable. Additionally, the study shows how scholars/experts serve as the “go-between” group that helps reconcile the diverse needs of the BVI user and developer groups.

For relevance, it is imperative that DL design guidelines consider a comprehensive list of help-seeking situations that BVI users encounter in using DLs. Previous research has not systematically identified situations that BVI users face (Alajarmeh, 2022; Power *et al.*, 2012; Rømen and Svanæs, 2012). At the same time, most existing research only involved users; very few studied experts and/or developers. Scholars/experts relied on their knowledge of research and evaluation for guideline assessment (Calvo *et al.*, 2016; Power *et al.*, 2012). For relevance, users made their judgments based on their prior experience and the help-seeking situations they encounter. In rating the relevance, scholars/experts considered the problems faced by BVI users holistically. Their relevance judgment emphasized whether the recommended guidelines can cover all of the situations at different phases of a search process. DL developers’ evaluation concentrated on whether detailed and specific information was provided to demonstrate the relevance of associated guidelines. They did not have experience or knowledge of BVI users’ situations, so they mainly emphasized issues related to clarity and usefulness.

For clarity, language and structure are essential for DL guideline assessment. In previous research, scholars/experts emphasized clarity of the language to alleviate confusion in understanding guidelines (Brajnik *et al.*, 2011, 2012), while developers concentrated on the linguistic aspect of guidelines for implementation (Abuaddous *et al.*, 2016; Petrie *et al.*, 2011; Swallow *et al.*, 2016). No user input about clarity has been previously published. In this study, on the one hand, users assigned significantly higher scores for clarity of DLAUG in users versus scholars/experts and developers comparison. On the other hand, developers assigned the lowest scores for clarity of DLAUG in developers versus users and scholars/experts comparison. Users’ ratings were based on the clarity of the statements concerning situations and the conciseness of the guidelines because they did not have a design background. Scholars/experts’ judgments were based on both users and developers. They cared not only to what extent the guidelines were written in non-technical language but also whether developers could understand all guideline components. Developers’ assessments were determined by whether the guidelines were easy to follow, mainly whether there were examples offered to illustrate the recommended techniques. In addition, some participants were concerned about the structure of the guidelines, just as Swallow *et al.* (2016) found in their study.

For usefulness, in this study, users expected that the usefulness of the guidelines could be assessed based on whether the guidelines would help them efficiently interact with DLs. For scholars/experts, both previous research and this study emphasized the applicability of design guidelines. Petrie *et al.* (2011) stressed the importance of taking guidelines’ applicability into consideration. In this study, scholars/experts’ evaluations also checked

whether connections exist between key components, particularly whether DLAUG links recommended features to illustrative examples of their implementation. For developers, both previous works and this study suggested providing explanatory and illustrative examples (Farrelly, 2011; Patel *et al.*, 2020). Furthermore, developers went a step further to examine the provided design recommendations for adoption when creating DLs. In addition, they also cared about the priority of guidelines – i.e. which situations and associated guidelines need to be considered first. This is helpful in solving the issue of the large volume of accessibility guidelines (Abuaddous *et al.*, 2016; Swallow *et al.*, 2016). Of the three groups, developers were the most demanding and critical in their assessment of the usefulness of DLAUG.

### 5.2 Practical implications

Based on the survey results for the three groups of stakeholders, this study enhanced DLAUG. For relevance, developers requested useful guidelines related to recommended features, scholars/experts asked to provide comprehensive guidelines and users expected guidelines targeted at their encountered problems when using DLs. Therefore, the enhancement of DLAUG considered how to cover all critical help-seeking situations. For clarity, the guidelines were modified by replacing their technical terms with more generally understandable terms because users and scholars recommended not using technical language. In this way, all stakeholders, especially users and developers, could read and understand clearly. Where technical terms persisted, a glossary was provided. Also, developers suggested a more organized structure, so an introduction to the guidelines was added. Improving the guidelines' usefulness was mainly achieved by adding more how-to examples with suggested steps for implementation. Providing a prioritized recommendation list addressed developers' requests for a list of guidelines with proposed priorities. In addition, to address scholars/experts' concerns, a numbering system was applied to the entire set of guidelines to show the connections between components.

## 6. Conclusion

Despite the best intentions and sustained efforts of the scholarly and practitioner communities, the accessibility and usability of systems continue to be problematic for BVI users. Although existing design guidelines and standards (e.g. WCAG 2.2) are capable of enhancing BVI users' access to interface elements, research shows that these guidelines fall short of ensuring effective access and usability of systems for BVI users. The findings of the study highlight the importance of adequately accommodating the needs of key stakeholders – BVI users, developers and scholars/experts – to ensure the success of the design guidelines. It further characterizes each group according to their viewpoints, namely, roles, knowledge/experience, goals and expectations. These shape each group's accessibility and usability needs *vis-à-vis* the design guidelines, which, in turn, shape their perceptions of each guideline's quality. Taken together, the differing needs and perspectives of the three stakeholder groups make the design guidelines more holistic in nature. Hence, the success of a design guideline depends on how well it meets the needs of these three groups of stakeholders.

The findings of this study also imply that the success of a design guideline in delivering DL accessibility and usability hinges broadly on whether it meets three quality standards: relevance, clarity and usefulness. Relevance depends on how directly a design guideline speaks to a help-seeking situation encountered by target users, how comprehensive its coverage of the concerned search task is and how adequately it illustrates the recommended help features. Clarity depends on how well technical jargon is simplified for a non-technical reader and how logically it organizes information. Usefulness depends on how detailed and

illustrative the provided directions are and how well the priorities of implementation are set. Each of these quality standards is further explained with illustrations in this paper, which may serve as guidance for scholars and practitioners.

This study also has its limitations. The first limitation is its modest sample size. Each stakeholder group was represented by 50 research participants, which is insufficient for establishing statistical generalizability of the findings. The second limitation is related to the diversity of the developers recruited. All 50 developers who participated in the study were affiliated with academic libraries, which is only a subset of the DL developer community. Future research must address these limitations by expanding the sample size and participant diversity when applying the design guideline development approach. Future research could also involve converting the design guidelines into a set of success measures (akin to WCAG success criteria), which could form the basis for automated accessibility and usability testing for DLs and other information retrieval systems.

### References

- Abuaddous, H.Y., Jali, M.Z. and Basir, N. (2016), "Web accessibility challenges", *International Journal of Advanced Computer Science and Applications*, Vol. 7 No. 10, pp. 172-181.
- Alajarmeh, N. (2022), "The extent of mobile accessibility coverage in WCAG 2.1: sufficiency of success criteria and appropriateness of relevant conformance levels pertaining to accessibility problems encountered by users who are visually impaired", *Universal Access in the Information Society*, Vol. 21 No. 2, pp. 507-532.
- Bai, A., Stray, V. and Mork, H. (2019), "What methods software teams prefer when testing web accessibility", *Advances in Human-Computer Interaction*, Vol. 2019.
- Brajnik, G., Yesilada, Y. and Harper, S. (2011), "The expertise effect on web accessibility evaluation methods", *Human-Computer Interaction*, Vol. 26 No. 3, pp. 246-283.
- Brajnik, G., Yesilada, Y. and Harper, S. (2012), "Is accessibility conformance an elusive property? A study of validity and reliability of WCAG 2.0", *ACM Transactions on Accessible Computing*, Vol. 4 No. 2, pp. 1-28.
- Calvo, R., Seyedarabi, F. and Savva, A. (2016), "Beyond web content accessibility guidelines: expert accessibility reviews", *Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI '16)*, pp. 77-84.
- Cheoh, J.L., Beigpourian, B., Wei, S., Ferguson, D. and Ohland, M. (2020), "Examining the perceptions of people with disabilities on the use of accessibility standards in web interface design", *IEEE Frontiers in Education Conference (FIE '20)*, pp. 1-4.
- Clegg-Vinell, R., Bailey, C. and Gkatzidou, V. (2014), "Investigating the appropriateness and relevance of mobile web accessibility guidelines", *Proceedings of the 11th Web for All Conference (W4A '14)*, pp. 1-4.
- Colusso, L., Bennett, C.L., Hsieh, G. and Munson, S.A. (2017), "Translational resources: reducing the gap between academic research and HCI practice", *Proceedings of the Conference on Designing Interactive Systems*, pp. 957-968.
- Cooper, M., Sloan, D., Kelly, B. and Lewthwaite, S. (2012), "A challenge to web accessibility metrics and guidelines: putting people and processes first", *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, pp. 1-4.
- Farrelly, G. (2011), "Practitioner barriers to diffusion and implementation of web accessibility", *Technology and Disability*, Vol. 23 No. 4, pp. 223-232.
- Frazão, T. and Duarte, C. (2020), "Comparing accessibility evaluation plug-ins", *Proceedings of the 17th International Web for All Conference*, pp. 1-11.

- 
- Gaggi, O. and Pederiva, V. (2021), "WCAG4All, a tool for making web accessibility rules accessible", *IEEE 18th Annual Consumer Communications and Networking Conference (CCNC '21)*, pp. 1-6.
- Gaona-García, P.A., Martín-Moncunill, D. and Montenegro-Marin, C.E. (2017), "Trends and challenges of visual search interfaces in digital libraries and repositories", *The Electronic Library*, Vol. 35 No. 1, pp. 69-98.
- Haider, W. and Yesilada, Y. (2020), "Tables on the web accessible? Unfortunately not!", *Proceedings of the 17th International Web for All Conference*, No. 7, pp. 1-5.
- Jia, J. and Capretz, L.F. (2018), "Direct and mediating influences of user-developer perception gaps in requirements understanding on user participation", *Requirements Engineering*, Vol. 23 No. 2, pp. 277-290.
- Kreps, D. and Goff, M. (2015), "Code in action: closing the black box of WCAG 2.0, a Latourian reading of web accessibility", *First Monday*, Vol. 20 No. 9, available at <https://firstmonday.org/ojs/index.php/fm/article/view/6166/4900> (accessed 9 June 2021).
- Li, Y. and Liu, C. (2019), "Information resource, interface, and tasks as user interaction components for digital library evaluation", *Information Processing and Management*, Vol. 56 No. 3, pp. 704-720.
- Mátrai, K.R. (2018), "How to make an electronic library accessible", *The Electronic Library*, Vol. 36 No. 4, pp. 620-632.
- Ngimwa, P. and Adams, A. (2013), "The different roles of 'design process champions' for digital libraries in African higher education", *International Journal on Digital Libraries*, Vol. 13 Nos 3/4, pp. 119-133.
- Patel, R., Breton, P., Baker, C.M., El-Glaly, Y.N. and Shinohara, K. (2020), "Why software is not accessible: technology professionals' perspectives and challenges", *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '20)*, pp. 1-9.
- Petrie, H., Power, C., Swallow, D., Velasco, C.A., Gallagher, B., Magennis, M., Murphy, E., Collin, S. and Down, K. (2011), "The value chain for web accessibility: challenges and opportunities", *Proceedings of Accessible Design in the Digital World*, pp. 131-138.
- Power, C., Freire, A., Petrie, H. and Swallow, D. (2012), "Guidelines are only half of the story: accessibility problems encountered by blind users on the web", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, pp. 433-442.
- Rømen, D. and Svanæs, D. (2012), "Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users", *Universal Access in the Information Society*, Vol. 11 No. 4, pp. 375-385.
- Selvadurai, V., Vistisen, P. and Rosenstand, C.A.F. (2019), "Fruitful gaps in digital literacy: interpreting gaps in digital literacy among stakeholders in collaborative design research projects as an evolving innovative capacity", *The Design Journal*, Vol. 22 No. sup1, pp. 2045-2059.
- Snider, S., Scott, W.L. II. and Trewin, S. (2020), "Accessibility information needs in the enterprise", *ACM Transactions on Accessible Computing*, Vol. 12 No. 4, pp. 1-23.
- Swallow, D., Petrie, H. and Power, C. (2016), "Understanding and supporting web developers: design and evaluation of a web accessibility information resource (WebAIR)", *Proceedings of the 3rd International Conference on Universal Design (UD '16)*, pp. 482-491.
- Trewin, S., Cragun, B., Swart, C., Brezin, J. and Richards, J. (2010), "Accessibility challenges and tool features: an IBM web developer perspective", *Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A '10)*, pp. 1-10.
- Tu, Y.F., Hwang, G.J. and Lai, C.L. (2021), "Facilitating learning by the visually impaired: development and usability evaluation of a specially designed ubiquitous library", *The Electronic Library*, Vol. 39 No. 1, pp. 169-185.
- Velleman, E.M., Nahuis, I. and van der Geest, T. (2017), "Factors explaining adoption and implementation processes for web accessibility standards within eGovernment systems and organizations", *Universal Access in the Information Society*, Vol. 16 No. 1, pp. 173-190.

- Watanabe, W.M., Fortes, R.P. and Dias, A.L. (2017), "Acceptance tests for validating ARIA requirements in widgets", *Universal Access in the Information Society*, Vol. 16 No. 1, pp. 3-27.
- Xie, I., Babu, R., Castillo, M. and Han, H. (2018a), "Identification of factors associated with blind users' help-seeking situations in interacting with digital libraries", *Journal of the Association for Information Science and Technology*, Vol. 69 No. 4, pp. 514-527.
- Xie, I., Babu, R., Lee, H.S., Wang, S. and Lee, T.H. (2021a), "Orientation tactics and associated factors in the digital library environment: comparison between blind and sighted users", *Journal of the Association for Information Science and Technology*, Vol. 72 No. 8, pp. 995-1010.
- Xie, I., Babu, R., Lee, T.H., Wang, S. and Lee, H.S. (2021b), "Coping tactics of blind and visually impaired users: responding to help-seeking situations in the digital library environment", *Information Processing and Management*, Vol. 58 No. 5, available at: [www.sciencedirect.com/science/article/pii/S0306457321001084](http://www.sciencedirect.com/science/article/pii/S0306457321001084) (accessed 8 June 2021).
- Xie, I., Joo, S. and Matusiak, K. (2018b), "Multifaceted evaluation criteria of digital libraries in academic settings: similarities and differences from different stakeholders", *The Journal of Academic Librarianship*, Vol. 44 No. 6, pp. 854-863.
- Xie, I., Joo, S. and Matusiak, K. (2021c), "Digital library evaluation measures in academic settings: perspectives from scholars and practitioners", *Journal of Librarianship and Information Science*, Vol. 53 No. 1, pp. 130-152.
- Xu, F. and Du, J.T. (2019), "Examining differences and similarities between graduate and undergraduate students' user satisfaction with digital libraries", *The Journal of Academic Librarianship*, Vol. 45 No. 6.
- Yesilada, Y., Brajnik, G., Vigo, M. and Harper, S. (2012), "Understanding web accessibility and its drivers", *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '12)*, Vol. 19, pp. 1-9.
- Yesilada, Y., Brajnik, G., Vigo, M. and Harper, S. (2015), "Exploring perceptions of web accessibility: a survey approach", *Behaviour and Information Technology*, Vol. 34 No. 2, pp. 119-134.
- Zhang, Y. (2010), "Developing a holistic model for digital library evaluation", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 1, pp. 88-110.

### About the authors

Iris Xie is a Professor in the School of Information Studies at the University of Wisconsin, Milwaukee. She received her PhD from the School of Communication and Information at Rutgers University. Her research interests focus on the design and evaluation of digital libraries, interactive information retrieval, human-computer interaction, user studies and research methods. In addition to numerous journal papers, she has published two books: *Interactive Information Retrieval in Digital Environments* (IGI Publishing, 2008) and *Discover Digital Libraries: Theory and Practice* (Elsevier, 2016), co-authored with Krystyna Matusiak. Iris Xie is the corresponding author and can be contacted at: [hiris@uwm.edu](mailto:hiris@uwm.edu)

Dr Rakesh Babu is a Lead Accessibility Scientist in Envision, Inc. His research interests include accessibility and usability, digital libraries, mobile assistive technology, mental models and technology affordances. His research attempts to empower blind people in the information society. His interdisciplinary, socially relevant research has been funded by the Institute of Museum and Library Services, Online Computing Library Center, National Science Foundation and Research Council of Norway. He has published journal articles and refereed conference papers on the under-researched topic of information technology (IT) accessibility and users with vision impairment. He serves as a Reviewer for multiple journals and international conferences.

Shengang Wang is a Doctoral Student at the School of Information Studies, the University of Wisconsin-Milwaukee. His research interests include information behaviors of vulnerable populations, health informatics and qualitative research methods. His research has been published in venues like the *Journal of the Association for Information Science and Technology*, *Journal of the American Medical Informatics Association* and *Information Processing and Management*.

Hyun Seung Lee is a Doctoral Student at the School of Information Studies, the University of Wisconsin-Milwaukee. She received her Master of Library and Information Science from the School of Information Studies at Syracuse University. Her research interest lies in information behavior and health information behavior, specifically in social media context and information serendipity. Her research has been published in the *Journal of Association for Information Science*, *Information Processing and Management*, and *Technology* and *International Journal of Medical Informatics*.

Tae Hee Lee is a PhD student in the School of Information Studies at the University of Wisconsin Milwaukee. He has over eight years of working experience in the information technology (IT) fields, including an IT project and operation manager in the department of supporting information technology at the Samsung Fire and Marine Insurance, before studying at the University of Wisconsin Milwaukee. His current research relates to professionals or marginalized groups' user experience and information behavior. He has published articles in *Information Processing and Management* and the *Journal of Association for Information Science and Technology*.