Defining Patterns for a Conversational Web

Emanuele Pucci
Politecnico di Milano
Milano, Italy
emanuele.pucci@polimi.it

Isabella Possaghi
Norwegian University of Science and Technology
Trondheim, Norway
isabella.possaghi@ntnu.no

Claudia Maria Cutrupi
Norwegian University of Science and Technology
Trondheim, Norway
claudia.m.cutrupi@ntnu.no

Marcos Baez
Fachhochschule Bielefeld
Bielefeld, Germany
mbaezpy@gmail.com

Cinzia Cappiello
Politecnico di Milano
Milano, Italy
cinzia.cappiello@polimi.it

Maristella Matera
Politecnico di Milano
Milano, Italy
maristella.matera@polimi.it

ABSTRACT

Conversational agents are emerging as channels for a natural and accessible interaction with digital services. Their benefits span across a wide range of usage scenarios and address visual impairments and any situational impairments that may take advantage of voice-based interactions. A few works highlighted the potential and the feasibility of adopting conversational agents for making the Web truly accessible for everyone. Yet, there is still a lack of concrete guidance in designing conversational experiences for browsing the Web. This paper illustrates a human-centered process that involved 26 blind and visually impaired people to investigate their difficulties when using assistive technology for accessing the Web, and their attitudes and preferences on adopting conversational agents. In response to the identified challenges, the paper introduces patterns for conversational Web browsing. It also discusses design implications that can promote Conversational AI as a technology to enhance Web accessibility.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **User studies**; **Accessibility technologies**.

KEYWORDS

Conversational UIs, Conversational Web Browsing, Conversational Patterns

ACM Reference Format:
Emanuele Pucci, Isabella Possaghi, Claudia Maria Cutrupi, Marcos Baez, Cinzia Cappiello, and Maristella Matera. 2023. Defining Patterns for a Conversational Web. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3544548.3581145

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581145

1 INTRODUCTION

The Web has evolved over time, but still, it remains a visual experience that might be inadequate for certain user categories or in certain usage situations. Accessing the Web with limited visual abilities can be granted by voice assistive technologies. Screen readers are the most adopted tools, especially by blind and visually impaired (BVI) people. However, their reading paradigm is not without problems [15, 58, 60]. On the one hand, despite the extended efforts dedicated to investigating how to improve Web accessibility [57], many websites are still designed regardless of accessibility guidelines, impeding screen readers to read what is coded in the HTML. On the other hand, the information published on the Web is conceived for visual consumption. Therefore, even with perfectly accessible websites, the transposition to the vocal paradigm might be critical.

Conversational AI is emerging as a technology that can support a more inclusive interaction with digital services [1, 10], offering benefits not only to BVI users but also to other populations that in different usage situations may take advantage of voice-based interactions [4, 18]. There are works focusing on the design of virtual assistants that complement visual pages to offer direct and quick access to the website content (e.g., [14]). Other works propose conversational agents (CAs) for searching the Web [16], or enable end users to customize their virtual assistants for searching the Web [24, 25]. This interest highlights the potential and the feasibility of adopting conversational AI for making the Web truly for everyone. Yet, there is still a lack of concrete guidance to inform designers and developers about how to deliver effective conversational experiences on the Web [26]. Emerging guidelines and heuristics [34, 36, 52] (i) tend to focus on general accessibility and not on the specific interaction abilities of BVI users, and (ii) provide general recommendations for CA design that do not capture Web browsing aspects. Our work wants to address this gap by analyzing practices and challenges for Web browsing by BVI users, to then contribute with reusable design patterns that capture solutions to conversational Web browsing challenges in informational websites.

Our goal is also to go beyond what current CAs offer for browsing the Web. As reported in the literature [16] and remarked by the users involved in our studies, *"one of the biggest limitations of CAs is that they help search and locate something interesting, but then they stop just after opening a website"*. Overcoming these limitations

implies not only addressing technological challenges for the integration of AI models and Web architectures. It also demands a deep understanding of the users' needs, to identify conversational flows that can help users browse the Web without the barriers posed by current voice-based technologies, both screen readers and CAs.

Motivated by this goal, we conducted a human-centered design (HCD) study that, following a "research-through design" approach [63], involved a sample of BVI users in several sessions, also asking them to co-design prototypes as vehicles for inquiring about foundational aspects of emerging design challenges. This paper illustrates how this process helped us identify a design space characterizing the notion of the *Conversational Web*, and progressively transpose it into a set of design patterns for conversational Web browsing. The main contributions of the paper are:

- The understanding of *challenges* for Web browsing by BVI people, thanks to an HCD process that extensively adopted rapid prototyping for ideating and validating with the users concrete conversational artifacts, gaining users' feedback directly in the form of conversational solutions embedding users' preferences and desiderata.
- The identification of *design dimensions* for conversational Web browsing, which also suggest how to turn the guidelines that for years have characterized the usability of the visual Web into opportunities for designing conversational user interfaces for the Web.
- The definition of *conversational patterns, related abstractions, and foundations* that embed, in a reusable format, the user experience and Web browsing strategies elicited through the HCD process.
- Reflections on the *lessons learned* from the collected insights, *limitations*, and emerging *design considerations* that can drive future research towards promoting a Conversational Web.

After discussing the related works (Section 2), the paper illustrates the HCD process and the identified challenges and design dimensions for conversational Web browsing (Section 3). It then illustrates the resulting conversational patterns (Section 4) and their preliminary validation (Section 5). Finally, it discusses limitations and further aspects that the studies highlighted (Section 6), which draw relevant future work for the design of CAs for the Web (Section 7).

2 RELATED WORK

Our work draws on technologies for integrating Conversational AI into the Web and explores methods to guide, within this technological framework, the design of CAs for the Web. With the aim of highlighting challenges that are still unsolved, this section illustrates related works in these two areas.

2.1 Bringing Conversational AI to the Web

The path to a conversational Web started with the early improvements to the linear navigation of screen readers. *To lower the complexity* of using screen readers, prominent approaches have proposed speech-based extensions that would accept spoken commands as shortcuts to screen reader functionality [6, 55]. Other approaches introduced strategies for *content organization and navigation*, such as machine-based segmentation of content to enable navigation through semantically related content [12, 13], summarisation of web content [28], and non-visual skimming [3]. *Speech optimizations*, through the exploration of faster text-to-speech [27] and multiple speech channels [27, 62], try to speed up and enrich navigation by catering to the abilities of BVI users. *Web automation* approaches also enable the user-defined [11] or automatic [5, 31] creation of macros for repetitive web-browsing actions.

Superimposing Conversational AI (CAI) on existing systems is now gaining momentum [9] and is suggesting new directions for the interaction with digital services. CAI is adopted to grant access to data and services at different levels, from extending GUIs of apps and websites [8, 33, 44], to adding natural-language front ends to Web services, processes [45, 47, 54, 61] and data repositories [17, 42]). On the Web, CAI is often exploited to build pop-up bots [9], i.e., assistants embedded within websites that offer services, such as conversational FAQ and help. However, these solutions do not focus on website navigation and content fruition. Tighter integration between websites and CAI is instead achieved by multi-experience websites, which offer both visual and conversational interfaces on the same content and functionality. For example, a recent approach proposed by Planas et al. [40] speeds up the development of multi-experience Web sites thanks to modeling abstractions and a related domain-specific language. However, the developer still defines the conversational experience by hand as a detached application.

Lately, researchers have proposed approaches that leverage CAI to offer alternative interaction paradigms for the Web. Cambre et al. [16] explore the use of open Internet technologies to build a virtual assistant for the Web implemented as a plugin for the Firefox browser. It enables interactions with browser functionality, and the natural-language expression of complex queries on Google Search for locating specific content items on the Web. Ripa et al. [44] then focus on facilitating the end-user generation of information bots out of website content. The approach is based on Web augmentation and relies on an annotation tool, to allow users to structure the content feeding the bot, and a flow editor to define the order and structure of responses and reading behavior. The end-user is in charge of the conversation design.

Related to all these approaches, some papers promote the idea of a Conversational Web [8] to enable users, especially those challenged by visual interaction paradigms, to express and fulfill their Web browsing goals by engaging in conversations mediated by a CA. One fundamental principle for this paradigm is enabling access to the web's wealth of services and information, even when websites are not equipped with ad-hoc conversational extensions, and provide (to the possible extent) a homogeneous conversational interface across websites. Progress towards this paradigm mainly refers to technical challenges and directions for tight integration between Web platforms and CAI [20, 39].

The above works bring valuable contributions, yet there is still limited guidance on how to design conversational interactions that can be generated by interpreting the structure of an existing website and transposing content and functionality of the visual Web into conversational experiences. Indeed, most of these approaches outsource the design of conversational interactions to developers or to the final users. Our work wants to take the lessons learned from formative user studies to propose reusable design patterns that can be natively offered by websites.

2.2 Design guidelines and patterns for Conversational AI

The literature highlights the potential benefits of conversational paradigms to enable a better user experience for people with disabilities, including BVI people [2]. At the same time, studies using a variety of methodologies have brought a better understanding of the design challenges and the unmet needs of this population when it comes to designing CAs [1, 2, 41]. Prominent challenges relate to the design of input mechanisms, control over the presented information, the interaction modalities, and even privacy when interacting through voice [1, 2, 41]. Branham and Roy [15] suggest that guidelines for the design of voice assistants might contribute to solving these challenges, but argue that current proposals do not properly meet the needs of BVI people. Here, contributions go from general Human-AI interactions guidelines [4], to industry and platform-specific guidelines [59], and recommendations for accessible CAI [34, 52].

Among the efforts toward accessible CAI, Leister et al. [34] analyzed 29 different sources of guidance under disability classifications and assessed the applicability of Web content accessibility guidelines (WCAG 2.1) for CA design. In the end, they derived 23 design considerations for accessible conversational interfaces. Stanley et al. [52] performed a review of 17 different sources from research and practice, and synthesized their findings in 157 recommendations derived from standards, empirical studies, and UX analysis. Similarly, Murad et al. [36] develop (high-level) heuristics for voice UIs inspired by GUI guidelines. While extremely valuable, these accessibility design considerations are general and not targeted to the needs and capabilities of BVI users. More specific guidance comes from a few empirical studies that do leverage the skills and capabilities of blind users [21, 30]. For example, Choi et al. [21] explored speech-rate configurations to meet the exceptional listening abilities of blind users. This study further supports the need for specific design considerations for this population but also highlights that solutions might depend on the type of task, content, and context of use. Still, these considerations do not address Web browsing tasks.

Design patterns for CAI are unexplored to a larger extent. Progress is being made but in specific solutions, such as VERSE [56], which explores the integration of screen reader functionality and voice assistance, with a focus on information search on the Web. VERSE incorporates interesting design ingredients, such as combining breath and depth in search, and supporting navigation commands through different input modalities. However, the actual approach to exploration (of a website) is limited to simple navigation commands specific to Wikipedia. Bouguelia et al. [14] propose patterns for task-oriented chatbots for Web services, and highlight the implications on dialog structure, required abstractions, and supporting infrastructure. Our work elaborates on such implications and adapts them to conversational patterns for Web browsing. The results illustrated in the following sections are grounded on the analysis of the needs and abilities of blind and visually impaired people; however, they still can contribute to a much-needed understanding of how to leverage CAI for accessing the Web.

3 FORMATIVE STUDIES

In the period from April 2021 to January 2022, we conducted a series of formative studies that guided the identification of patterns for the new conversational paradigm (see Figure 1). The studies were authorized by the research ethical committee of the Politecnico di Milano university (Opinion no.11/2021). In total, we involved 26 BVI participants[1] (6 self-identified as females while the rest as males), reached out through three Italian BVI associations (*Unione Italiana Ciechhi e Ipovedenti (UICI), Real Eyes Sport, Associazione Disabili Visivi (ADV)*). The involvement of BVI users allowed us to focus on the most stringent requirements for a Web browsing paradigm detached from the visual channel.

As illustrated in the following, at each step of the process the insights gained from the users progressively guided the identification of conversational patterns for accessing the Web. This was also possible thanks to the adoption of a "research-through design" approach [63] characterized by the co-design and validation of intermediate prototypes on which the users could act directly to express their opinions on conversational solutions. The study material and excerpts of the collected data are available at: https://tinyurl.com/Studies-data.

3.1 Remote preliminary interviews with experts of assistive technologies

We started our research in April 2021 with preliminary unstructured interviews carried out with 3 digital-technology experts (average age: 47), all but one blind, who educate and assist BVI people in learning and using assistive technologies. Each interview lasted about 2 hours and, due to COVID-19 restrictions, was conducted remotely through video-conferencing tools. To learn from the experts how to approach the design for BVI people, the whole research team (i.e., all the authors of this paper) attended the remote meeting, with two researchers acting as moderators. The involvement of the experts was significant not only for gathering their personal experience when they approach the Web but also as a proxy to understand the challenges faced by the BVI people they help with their educational activities. The discussion revolved around how screen readers support Web navigation, and to which extent BVI people could accept CAs as an alternative technology. The interviews were video-recorded. The transcripts were analyzed by the two moderators, to perform an inductive thematic analysis [35]. Independently, they double-checked the material; then, iteratively reduced a few variations on the emerging themes (10%), till reaching a complete agreement. The results were finally discussed with the whole team that in the end agreed on three main emerged aspects: (i) **the complexity of understanding the website structure and locating information**, even for users who are highly experts in screen readers' usage; (ii) **accessibility issues** related to the interpretation by screen readers of the page HTML code; (iii) **limitations of CAs**, first of all, the lack of support for browsing information on the Web, which was also in line with findings from the literature [16, 41]. To deepen and verify these aspects with a larger number of users, we defined and validated with the experts a mixed-method questionnaire including 25 questions, both closed and open-ended, covering

[1]Detailed data on participants are available at: https://tinyurl.com/Studies-data

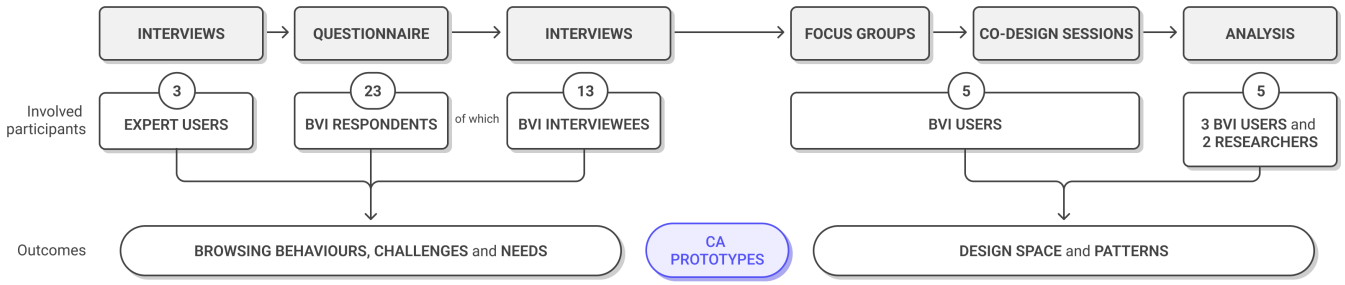

Figure 1: The Human-centered process for the identification of challenges and design dimensions for conversational Web browsing.

three main aspects: **experience with Web browsing**, **experience with screen readers**, and **experience with CAs**.

3.2 Online questionnaire and remote interviews

In May 2021, we advertised the questionnaire in the monthly newsletter of the UICI association and spread it also among BVI people belonging to the other two associations we were in contact with. Table 1 reports representative questions for each addressed aspect[2]. We received a total of 23 responses from users aged between 18 and 65 years;14 were totally blind, and the others had severe visual impairments. Among several aspects, the gathered answers revealed that the Web sites the participants access the most are information-intensive and that accessibility errors in a Web page layout are barriers that they can tolerate. The answers to the open questions, instead, highlighted that more server issues occur for content browsing, especially during the initial orientation within a website's content. At the beginning of July, we then conducted semi-structured interviews with 13 users who in the online questionnaire had declared to be frequently engaged with digital services and assistive technologies and had expressed their interest in follow-up activities. They were aged between 18 and 60 years (6 females, average age: 42), and 8 were totally blind. Each interview was moderated by two researchers, lasted about 1 hour, and was held remotely. The discussion was guided by the open-ended questions of the online questionnaire referring to the experience with Web navigation, screen readers, and CAs, but went deeper into the different challenges, probing participant responses, and encouraging them to provide details and clarifications. The questionnaire had highlighted a range of problems and opportunities, but it was the interviews that provided depth on the causes. After integrating the new data with the ones gathered through the questionnaire, the two researchers independently identified the emerging themes through an inductive thematic analysis [35]. The results were discussed in 2 sessions, after which they agreed upon the relevant findings and came to the definition of the **preliminary challenges** illustrated here below and summarized in Table 1.

Experience with Web browsing. The participants highlighted an intrinsic lack of accessibility for many websites. **P15** said: *"Even without visuals I would like to experience the Web page without being disappointed every time I need to interact with an unreadable visual element that I can't understand"*. They reported usability problems (e.g., confusing information architecture) that can impact any user, not only BVI people. **P24** said: *"I cannot detect how the content is structured. The real problem is not incorrectly coded links, but not being able to find the right information"*. They also commented on difficulties in identifying the semantics of links. **P5** said: *"Sometimes I don't clearly know what I am searching for, and I hope my navigation would be guided by meaningful links"*.

Experience with Screen Readers. When using screen readers, learning the website's structure is a necessary first step. However, the participants commented that this is also the most demanding activity for many websites. **P4** observed: *"Sometimes I open a site and it takes me 15 mins or more to understand what it is about, where I am, and where I can navigate to"*. **P22** said *"It is a mystery to me where they [links] can take me. I cannot predict my movements on the site"*. The participants also complained about the high variability of the content organization, which prevents them from identifying strategies that can work across different websites (**P8**: *"It is challenging to identify standard exploration strategies for different websites"*; **P21**: *"I keep going by trial and error. The effort is reduced over time thanks to my experience in browsing, but it remains significant"*). A frequent observation came up about the need of identifying the navigational context, i.e., where the users came from and where they could navigate to (**P12**: *"Sometimes I get lost because I cannot remember the page where I came from, and I cannot even return to the home page because I don't remember the path!"*). Almost all the participants reported a high effort to follow the serial reading of the page content (**P23**: *"I usually get distracted because I have to keep in mind everything that the screen reader tells me. Reading one paragraph per time would be more relaxing"*). They give up on browsing when they encounter page components with a strong visual connotation, such as online forms, calendars, and pop-ups, which the screen readers cannot interpret properly or even intercept (**P26**: *"Every time I close a pop-up window, I always ask myself what I could have accepted."*).

Experience with CAs. Frequently when talking about CAs, the participants complained about the limited understanding capabilities and the inability to serve many of their information needs (**P18**: *"This technology looks intuitive, but I frequently need to rephrase my request and often it is unable to give me an answer"*). The participants also mentioned the lack of a natural conversation flow (**P14**: *"Answers are often disconnected or repetitive, it can't remember what I*

[2]The whole set of questions (translated to English from the original language), and the answers gathered for the closed questions are available at: https://tinyurl.com/Studies-data

Aspects identified with the Experts	Representative questions in the questionnaire and interviews	Preliminary challenges
Experience with Web browsing	What are, in your opinion, the biggest challenges encountered while browsing the web?	a) Lack of accessibility b) Difficult orientation c) Unclear link semantics
Experience with screen readers	What are, in your opinion, the biggest limitations and difficulties encountered when using screen readers?	d) Complex memorization of website structure e) High variability of the content organization f) Forced exploration by "trial and error" g) Problems with navigational context identification h) Difficulties in serial reading i) Extreme reliance on visual order of page element
Experience with CAs	In what context has your interaction with a CA not been particularly useful?	j) Misleading or partial information acquisition k) Lack of natural conversation l) Unreliable management of misunderstandings m) Lack of feedback when performing tasks

Table 1: Evolution from the aspects identified with the help of the experts to the preliminary challenges gathered through the questionnaire and the interviews.

asked before and comes up with the same information."). Given these lacks, participants remarked they don't trust current CAs. They perceive the technology's potential but will engage with it only if it shows benefits for easing Web navigation. **P16** said: *"I would be happy with a CA retrieving Web content with a minimum number of steps, without being overwhelmed with useless embellishments and superfluous text"*.

3.3 In-presence focus groups

To refine the challenges identified in the previous step, we invited the interviewed users to participate in focus groups. 5 participants responded to our invitation. They were aged between 18 and 48 years (2 females, average: 24 years), two of them were totally blind and the remaining were affected by severe peripheral visual impairments. In the middle of July, we held two in-presence focus groups, the first with 3 participants, and the second with 2 participants. Two researchers moderated the two sessions, in which they observed the users while *i)* browsing the Web with screen readers and *ii)* interacting with a CA purposely designed to browse selected Web pages. Since the majority of the previously identified challenges referred to information-seeking and content-exploration issues, Web browsing was the main investigated dimension.

3.3.1 Web browsing with Screen Readers. In a 1-hour session, we observed the users and discussed with them how they navigated websites with screen readers to (i) access content (e.g., reading comments on YouTube or an article on a national news Web site[3]), and (ii) perform operations (e.g., leaving a comment for a YouTube video or buying a train ticket online). The websites on which they operated were chosen considering the browsing habits that users reported during the previous interviews, and the usability level[4].

[3]https://www.studenti.it/

[4]We referred to Google Chrome Lighthouse (https://developer.chrome.com/docs/lighthouse/overview/), an open-source, automated tool for assessing and improving the quality and correctness of websites.

The participants were asked to browse both accessible and non-accessible websites. They easily browsed YouTube, and they ascribed their success to the *"content hierarchy that helps orientation"* (**P23**). When navigating an informative website, they remarked that an essential factor for orientation is the consistency of the navigation strategies across different websites (**P25**: *"I find useful to spot in another website the same structure as Wikipedia, especially for the linear and clear distribution of content across different page sections."*). Several obstacles were instead encountered when searching for information on a railway's website, as the home page organization did not suggest a reading order, and the content was not adequately segmented (**P22**: *"Accessing this site is stressful and tedious, as it is full of unnecessary and confusing information. I keep using this site only because the screen reader enables me to skip some parts quickly"*).

3.3.2 Web browsing with CAs. We concluded the focus groups with a 30-minute session where we observed the participants interacting with a CA for browsing a few Wikipedia pages. The researchers designed the CA by considering the challenges identified after the interviews and injecting the characteristics summarized in the central column of Table 2 as possible solutions. Wikipedia was chosen for the familiarity of the participants with this site. In addition, as also highlighted by other works [56], its layered, well-organized structure would have facilitated the analysis of conversational mechanisms to move across different layers of a website — not just reading a webpage's content. The resulting prototype addressed conversational mechanisms for getting oriented, browsing, and reading the content available within pages. We thus asked the participants to use the prototype to elicit insights, triggered by concrete examples, on how they would envision browsing websites mediated by a CA. The participants left several positive comments about the way the initial orientation and then content browsing were supported by the CA (**P2**: *"I get a glimpse of the main information and understand the navigation options as soon as I enter the website. This way I can start exploring the content without being forced to browse the whole site"*), and about the overall experience (**P3**: *"It's nice to browse the*

Preliminary challenges	Focus-group CA characteristics	Co-Design (user-defined) CA characteristics
a) Lack of accessibility	Standardized conversational experience detached from the visual layout	Standard welcome message and consistent navigational commands across different Web sites
b) Difficult orientation	Layered content structure and guidance for progressive exploration	Fast-served requests for specific content
c) Unclear link semantic	Pertinent link labelling	Preview of link's target content
d) Complex memorization of website structure	Outlining the primary information and navigation options on each page	Landmarks for the main menu and the main navigational paths
e) High variability of content organization	Standardized dialog structure	Bookmarking specific elements with thematic categories
f) Forced exploration by "trial and error"	Content direct access through keyword-based search	Content access by Q&A
g) Problems with navigational context identification	Overview for any reached page	Navigational history
h) Difficulties in serial reading	Content segmentation	Summarization of text segments
i) Extreme reliance on visual order of page elements	Content reading independent of visual layout	Feedback on visual content not conveyed through conversation
j) Misleading or partial information acquisition	No filtering or alteration of the original content	Page segmentation and summarization, but with on-demand access to the original content
k) Lack of natural conversation	Rich intent library for CA training	Increased dialogue fluency but also conciseness
l) Unreliable management of misunderstandings	Interactive dialogue for narrowing down the user goal	Fallback and scaffolding intents
m) Lack of feedback when performing tasks	Informing the user on what the CA understands and does	Providing feedback and asking for confirmation before executing operations

Table 2: Solutions to the preliminary challenges as 1) injected in the CA adopted in the focus group (central column), and 2) suggested by the users in the co-design session (right-hand column).

site interactively, as in a real conversation: definitely less boring!"). The participants found the person-chatbot relationship trustworthy, as the CA conveyed transparency and authority in terms of content management. **P1** stated: *"In comparison with the original Web pages, the content you are presenting through the CA is not overly filtered or cut! The hierarchy of information dictated by visual features has been nicely transposed into the conversation. I appreciate it, as I don't want technology to make decisions for me about the content to be read"*. They however complained about the poor control of the navigation flow (**P1**: *"The flow of information is OK, but I should be able to stop it if I want and directly access only the information I ask for."*). They also signaled the lack of customization options for voice configuration (**P2**: *"I wish I could change the pitch, for example, to emphasize when an information item has been successfully found!"*). Our prototype did not cover this aspect, as we purposely wanted to focus on mechanisms for content organization and navigation.

3.4 Co-design

After one week, we organized a 2-hour in-presence co-design session with the same 5 participants, moderated by the same two researchers as in the previous activities. With the aim of identifying further preferences on how to organize the conversation for Web browsing, the participants were asked to define by themselves the conversation for browsing a website of their choice. They choose the website of an escape room that they, especially the youngest, knew very well, yet they claimed it was inaccessible via screen readers. As soon as the first ideas emerged, the researchers helped the participants implement the conversations with *DialogFlow*[5] and deployed them on Google Assistant. Participants could run the prototype on their mobile phones and Alexa (see Figure 2). They played with the prototype and iteratively modified it to refine their ideas. The analysis of the proposed solutions highlighted three design dimensions that we discuss in the following.

Navigation. Similar to the design of the CA adopted for the focus group, the participants organized the content hierarchically, to enable a layered exploration, but they also proposed new elements. They worked on providing feedback on the reached status when landing into a new browsing node, and on explaining the possible actions that could be invoked next (**P22**: *"First, I want to understand where I am and what I can do next; then I want the CA to proceed by reading the information on the page"*). They reflected on mechanisms to formulate fast-served requests for specific content (**P23**: *"When I open the website, I'd like to ask if it can offer what I am looking for, without necessarily having to scroll through everything"*). The users also discussed the benefits of tagging nodes and clustering them into meaningful categories (**P24**: *"It would be interesting to bookmark elements during my navigation, to easily retrieve them and get oriented even in successive browsing sessions!"*.

Content reading. Similar to the design of the focus-group CA, the participants segmented the page content to quickly locate items of interest (**P26**: *"A serial content reading is OK, but I would like to stop the CA if needed, or to proceed reading only under specific request, for example: "next please!""*). As a new element, for each segment they considered summarization and keyword-highlighting techniques (**P25**: *"It always takes me too long to figure out why this content is the result of my research. Highlighting the main concepts would help me a lot, similar to the preview in Google Search!"*).

Bot architecture. The participants confirmed the importance of being able to invoke, at any moment, commands for going back, asking for help, and navigating to landmark pages.

[5]https://dialogfow.cloud.google.com/

Figure 2: Co-design session. On the left: the researchers translate participants' ideas into a running prototype. On the right: one participant works on modifying the prototype.

3.5 Final analysis and validation of the design dimensions

The focus groups and the co-design activities were audio-video recorded. The two researchers moderating these activities individually took notes on significant participants' behaviors and aloud comments. They transcribed their notes and their post-experience considerations and extended them by integrating the video analyses. Separately, they performed a deductive thematic analysis, clustering codes from *in-vivo coding* [35] into the themes already identified in the previous phases. The two researchers independently double-checked 70% of this material. The initial reliability value was 80%. After discussion, the researchers agreed upon the differences and reduced the variations till they reached a full agreement.

As further validation, the participants in the previous co-design session were invited to revise the identified insights. 3 of them (1 female, average age: 28 years old) accepted to participate in a 2-hour focus group. To facilitate the discussion, the researchers built a new version of the CA previously defined for browsing Wikipedia pages, now integrating the additional conversational mechanisms that came out during the co-design session. The 3 participants were first asked to interact with the new prototype, then to express their opinions on the relevance of the identified aspects. In the end, with the agreement of the 3 participants, the researchers reached a consensus on the following design dimensions, which expressed in more detail all facets that emerged during the whole process:

- **Map of the navigable space.** It is fundamental to provide mechanisms for the users to understand the website structure. Mechanisms for link predictability and for keeping track of the navigational context can help users fluidly move along the main website areas. These mechanisms subsume a hierarchical content organization sustaining a systematic exploration through different layers.
- **Intelligible and quick navigation mechanisms.** The CA should support different navigation strategies. The participants discussed mechanisms for in-depth explorations to narrow down navigation options along the content hierarchy. Punctual, fast-served content requests were frequently discussed as a help to locate desired content, along with the capability of bookmarking information nodes for direct access to the content of interest.
- **Segmentation and summarization of page content.** Unneeded and tedious reading of content must be prevented. Segmenting the page content and summarizing segments through short descriptions and keywords could help users digest the content and identify in advance whether it is interesting.
- **Conversation-scaffolding intents.** It is fundamental to control the conversation through *general default commands*, i.e., those for invoking fallback and recovery paths and receiving help, and *conversation-oriented default commands*, i.e., those for accessing landmark pages and bookmarks.

4 PATTERNS FOR CONVERSATIONAL BROWSING

For each design dimension that emerged from the formative studies, we considered the solutions embedded in the last CA validated with the three participants, and translated the specific conversation instances into general *conversational patterns*. The aim was to achieve abstract tools to capture the knowledge gained from the users and transfer it as guidance on specific design aspects of CAs for Web browsing [53]. Interestingly, some patterns relate to traditional usability dimensions for the Web (e.g., orientation, navigation, content fruition [38]); however, still, their novelty lies in the provision of solutions for accessing the Web through CUIs. Since the definition of patterns has implications for abstractions, models, and supporting technical infrastructures [14], in the following we start from foundational concepts to later dive into solutions that respond to the challenges identified through the formative studies. Pattern presentation is organized thematically, with categories referring to the dimensions outlined in Section 3.5. Final reflections on the dialog structure suggest how to grant homogeneous browsing experiences across different websites.

4.1 Conversation-oriented Navigation Tree

Our patterns refer to a model, the *Conversation-oriented Navigation Tree (CNT)*, which suggests a hierarchical organization of the content to be conveyed through conversation - the study participants indeed frequently remarked on this aspect. As illustrated in Figure 3, for each page in a website the CNT represents a hierarchical nesting of both content elements and navigational structures, which recalls the HTML DOM organization but introduces new conversation-oriented elements. Each node in the tree, which we call *conversational node*, can be a content paragraph, a navigation menu, a link, or any other element in the webpage that can be presented independently of the others and has a role in the exploration of the website content. More specifically, the internal nodes represent aggregated content structures (e.g., "today's articles" on the Wikipedia Home Page) or navigational indices (e.g., the main navigation menu). The leaf nodes then represent the page's actual content. This node granularity serves the purpose of building an incremental exploration of webpage content through tree traversals.

The CNT enables conversational interactions with websites in two crucial ways. First, it organizes the website information architecture, independently of the visual layout organization and the presence of accessibility and semantic tags in the DOM definition, making it easier to navigate and consume content by CAs. Second, it contains a minimal set of attributes that are important for the generation of conversational interfaces. Each node is indeed associated with *descriptions*, which can be already available in the HTML code (e.g., alt-text for images), or purposely generated to enable the conversation (e.g., summaries of text content [32]) thus overcoming the lack of accessibility meta-data that affects many websites. A node might also store *keys indexing its content*, which are useful for locating and directly access to nodes. Specific nodes can be labeled as *conversational landmarks*, to have global visibility in the conversation. A *node type* indicates the role of the specific element and defines the type of operations that can be performed on it (e.g., content, link). Additional attributes can be derived from page content processing and attached to the nodes.

With the above approach, the problem space for enabling conversational interaction with websites is reduced to deriving CNT representations. This can be done by relying on automatic approaches, already explored in the technical literature [7], that build and handle the CNT during the website navigation by exploiting HTML annotations purposely added during the website authoring, or by employing techniques for the automatic segmentation [23] and summarization of webpage content [32].

4.2 Orientation - Shaping the map of the navigable space

The CNT organizes information so that, by applying tree-traversing strategies and providing content previews for the traversed nodes, a CA can help the user grasp the overall website organization.

View in the large. As soon as the user enters a website, the interaction with the Home Page must convey the main thematic areas and the main navigational components. Figure 4 illustrates an example of a conversation with the Wikipedia Home Page. On the left side, the CNT represents the hierarchy of the page components. The related conversation, on the right side, starts with a short description of the website content, as extracted from the CNT root node, plus a preview of the main thematic areas linked from the Home Page and the navigation header - these last are the child nodes of the CNT root. The same strategy can be adopted when the user enters the inner areas of the website, by recursively visiting any subtree representing content aggregations. Rollback actions and direct access to landmark nodes (see Section 4.5) further sustain the exploration of the content organization.

Navigational context. Every time the user enters a new navigation node, the CA must offer information on the navigational context. As reported in Figure 4, besides presenting a short description of the reached node, the children nodes can be introduced to help the users understand where they can move to (e.g.: *"Here you can read the introduction or follow one of the available links [...]"*). Each node can dynamically store a reference to the node the user came from, which can be presented to the users as an option to easily go back (e.g.: *"You came from the Home Page"*). To keep the conversation fluid, context information can be supplied on demand, i.e., only if the user asks for it (e.g.: *"Do you need more information for localizing the node?"*).

Link predictability. A preview of the content reachable through a link can help understand a-priori what the target content is about and avoid useless navigation steps. Considering the CNT model, each node representing a link can store a short description of the target content. It can be derived from the HTML link-title tag, when available; the CNT can also provide more contextualized descriptions, extracted from the target CNT node the link points to. The last part of the conversation in Figure 4 illustrates an example.

4.3 Navigation – Intelligible and quick navigation mechanisms

The hierarchical content organization suggested by the CNT lets the user browse the navigable space by moving down in the hierarchy (*in-depth exploration*) and exploring the content at each layer (*in-breadth exploration*). The indexing keys associated with each node also allow the users to locate content by direct queries.

Exploration of thematic areas. As represented in Figure 5, the CA assists the exploration of each thematic area by allowing the users to *i)* move horizontally across the pages at a level of the content hierarchy (e.g.: *"[...] you can find "Name and Symbol", "Physical Characteristics", [...]"*), and *ii)* move vertically (e.g.: *"Go down to "Composition"*) to reach inner layers offering further details on a given topic. Considering the CNT representation of a website, this pattern requires strategies for traversing subtrees representing thematic clusters of conversational nodes.

Q&A. The users must be able to formulate punctual, fast-served requests for specific content. If we consider the CNT representation of a website, this requirement implies that each node in the tree is labeled with keys characterizing its content and serving node localization and direct access (e.g.: *"Is there anything about "Jupiter's rings"*, in Figure 5).

Bookmarking. The users must be able to bookmark nodes for later access to important pieces of information; this can be beneficial for personalizing the navigation experience and have a positive impact on orientation. The CNT addresses this requirement by storing in each node possible user-generated labels. As illustrated

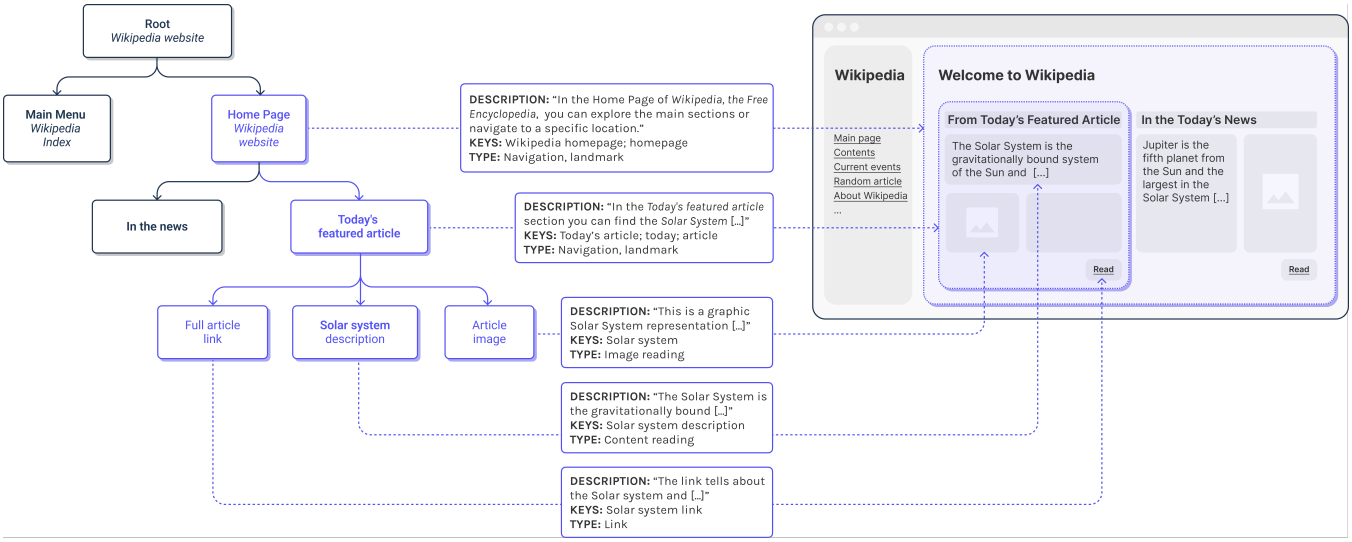

Figure 3: Simplified CNT model for the Wikipedia Home Page. The tree nodes represent page elements, following the hierarchical organization of the visual content. The dotted rectangles highlight the node metadata that feed the conversation.

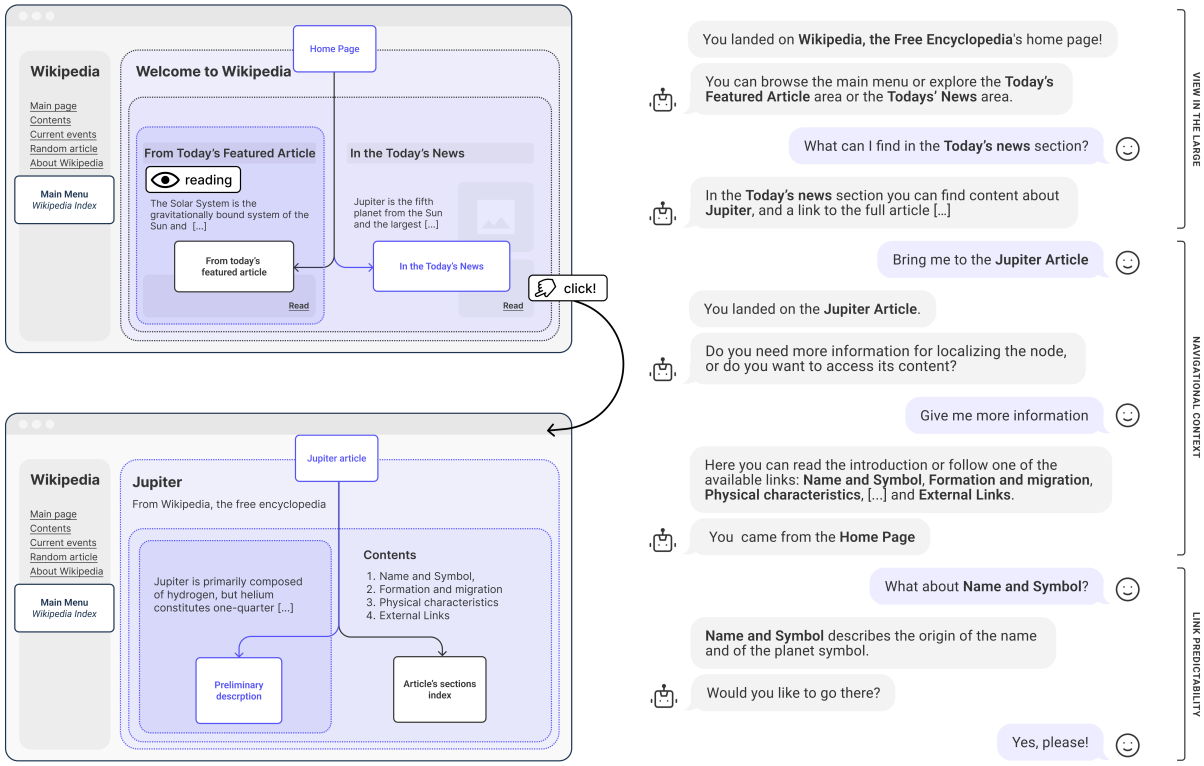

Figure 4: Orientation Patterns. On the left: visual organization of two Wikipedia pages and related CNT. On the right: example of conversation supporting the exploration of the two pages. Natural-language requests trigger reading and navigation actions.

in Figure 5, the CA must then be able to understand and serve user intents to *i)* label the nodes, and *ii)* retrieve those nodes in successive browsing sessions (see Section 4.5).

User-defined node clustering. This is a specialization of the bookmarking pattern, which allows the users to group bookmarked nodes in clusters representing thematic areas that are meaningful

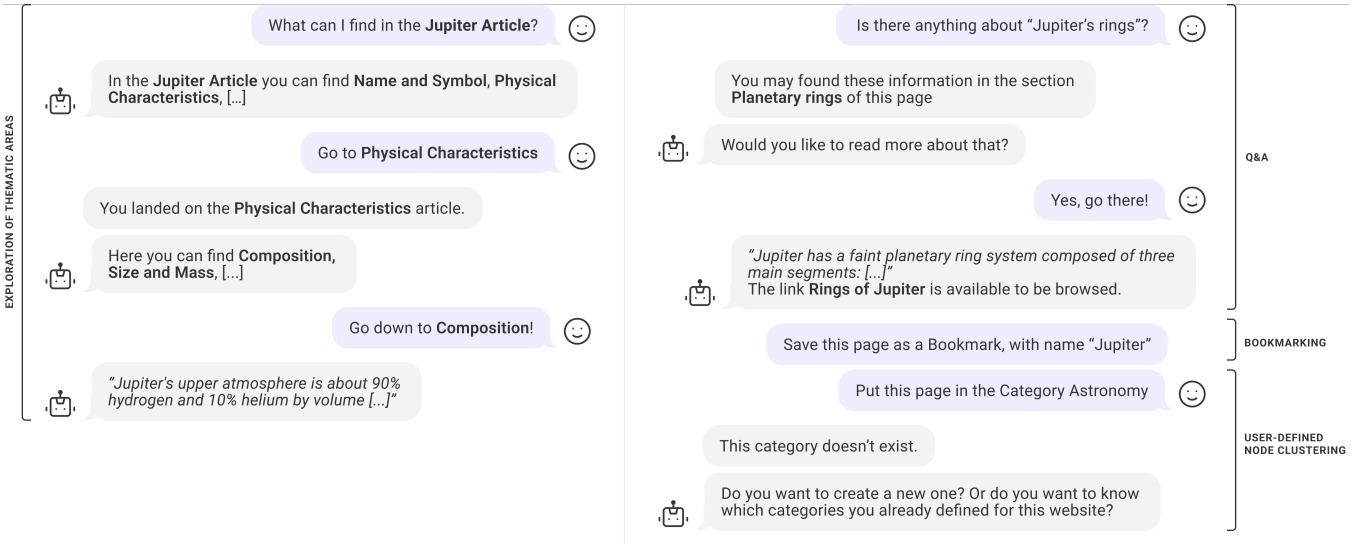

Figure 5: Navigation patterns for Jupiter's Wikipedia article.

to them and could help them recall a navigation context (e.g., the category "Astronomy" in Figure 5).

4.4 Content reading – Summarizing and segmenting the page content

Content reading patterns suggest how to segment, summarize, and index the website content, and how to let the user move through the resulting conversational nodes thanks to dedicated content-reading commands. The original page content must be preserved and should be entirely accessible if the users ask for this option.

Content Segmentation. Page content can be divided into segments that the users can quickly scroll through appropriate commands. Figure 6 reports an example of content reading based on segment scrolling. The CNT can support this pattern if an adequate node granularity is defined. Besides considering the DOM structure of a webpage, vision algorithms can be applied to parse the visual appearance of a webpage and identify segments [22, 23]. A vision-based segmentation preserves the consistency between the visual and the conversational Web - a desire users frequently expressed during formative studies.

Skimming mechanisms. Content summarization can give a preview of a node, for example when a new node is entered or when a link must be traversed and the user wants to know a-priori what can be found in the target node. It can prevent unwanted, or unneeded, content readings and navigations. An example of this pattern is reported in Figure 6, where the CA presents a summary of what can be found in the Jupiter article before reading the entire page. Summaries can be manually defined by the CA designer or automatically generated by summarization techniques [32].

Conversational tag cloud. Inspired by the visual Web, conversational tag clouds can convey the key concepts characterizing a node's content. Together with summaries, tags can help users assess the relevance of a node before fully reading it. In the example in Figure 6, the CA lists tags that represent the most relevant information within the reached node. In a CNT, each node can store these tags, which can be extracted through summarization techniques [32].

4.5 Conversation control – Providing access to conversation-scaffolding intents

The tree-based organization of conversational nodes and the metadata stored in each node help manage intents for controlling the conversation and the navigational context, such as:

- **Default actions:** for invoking help, moving back and forth along the conversation steps, listening again to a CA utterance, and accessing bookmarked nodes.
- **Current status:** for grasping what the CA is doing and where the reached content is located, i.e., answering questions such as "Where I came from?", and "Where can I go from here?"
- **Landmark nodes:** to quickly move to the Home Page and to other nodes that are representative of the main thematic areas. All the nodes labeled as landmarks are globally reachable in the CNT; the CA must be able to present them at any step of the conversation, upon users' request.

Figure 7 illustrates examples of conversations where the user invokes control commands.

4.6 Dialog structure

An important factor for orientation, extensively highlighted by our formative studies, is the homogeneity of browsing strategies across different websites, which is very difficult to achieve given the variability of website designs. Variability in the visual Web might not constitute a problem as visual cues can guide the users to understand how the website is organized. With conversational experiences, it is instead paramount to prioritize consistency across different websites or webpages, to speed up the familiarization with the information structure [46].

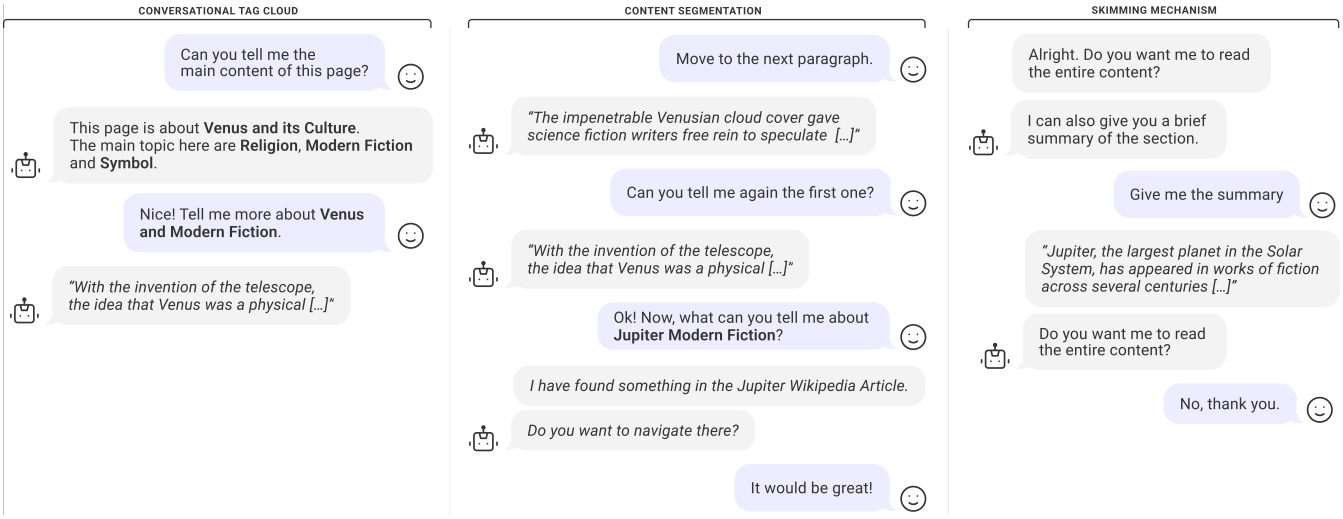

Figure 6: Content reading patterns for Venus and Jupiter Wikipedia article.

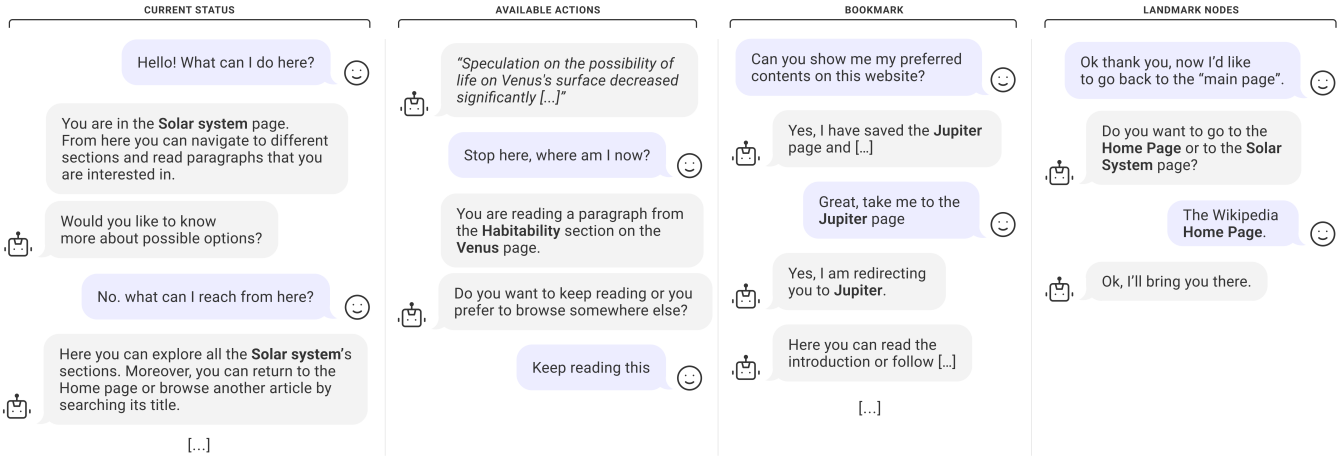

Figure 7: Conversation-control patterns for Solar System and Jupiter Wikipedia articles.

Independently of the visual organization of webpages, the CA should provide consistent mechanisms that can help the users retrieve familiar conversational strategies. This can be achieved by adopting a consistent dialogue flow across different websites, with the same mechanisms for exploring and navigating into the available areas, accessing node content, and invoking general commands for conversation control. Figure 8 schematically shows a typical dialog organization based on the patterns illustrated above. The vertical arrow highlights a possible "conversational journey", with a flow of user intents that go from orientation to navigation and then content reading. This flow is not rigid, as the user's adoption of the available commands might depend on the specific use and navigational context. However, the CA must support those Web browsing intents, together with control intents (represented by the horizontal arrow) that can be invoked at any conversational step.

4.7 CNT and pattern implementation

The identified patterns were implemented within a platform, ConWeb, that sustains the conversational Web browsing paradigm. As described in [7], ConWeb is a middleware that acts in between the Web client issuing webpage requests and Web servers providing the pages. Every time a webpage is accessed, its HTML code is parsed and the CNT is built automatically by instantiating its nodes with the detected page elements. For those websites natively designed for being accessed through ConWeb, the page parsing can rely on purposely added meta-data. When meta-data are not available, techniques for page visual segmentation and text summarization are used, as described above for each pattern.

The platform then automatically generates the dialogue for serving the user requests. On the client side, a Web browser extension handles the voice-based interaction using libraries for speech-to-text and text-to-speech conversion. Thanks to the integration with an NLP pipeline [43], server-side components: *i)* interpret the user natural-language request to extract intents and entities, and maps these elements onto one of the available patterns, each pattern being implemented as an *intent handler*; *ii)* using a headless browser [49], transform the interpreted requests into navigation actions to keep the navigation status updated, and *iii)* automatically generate

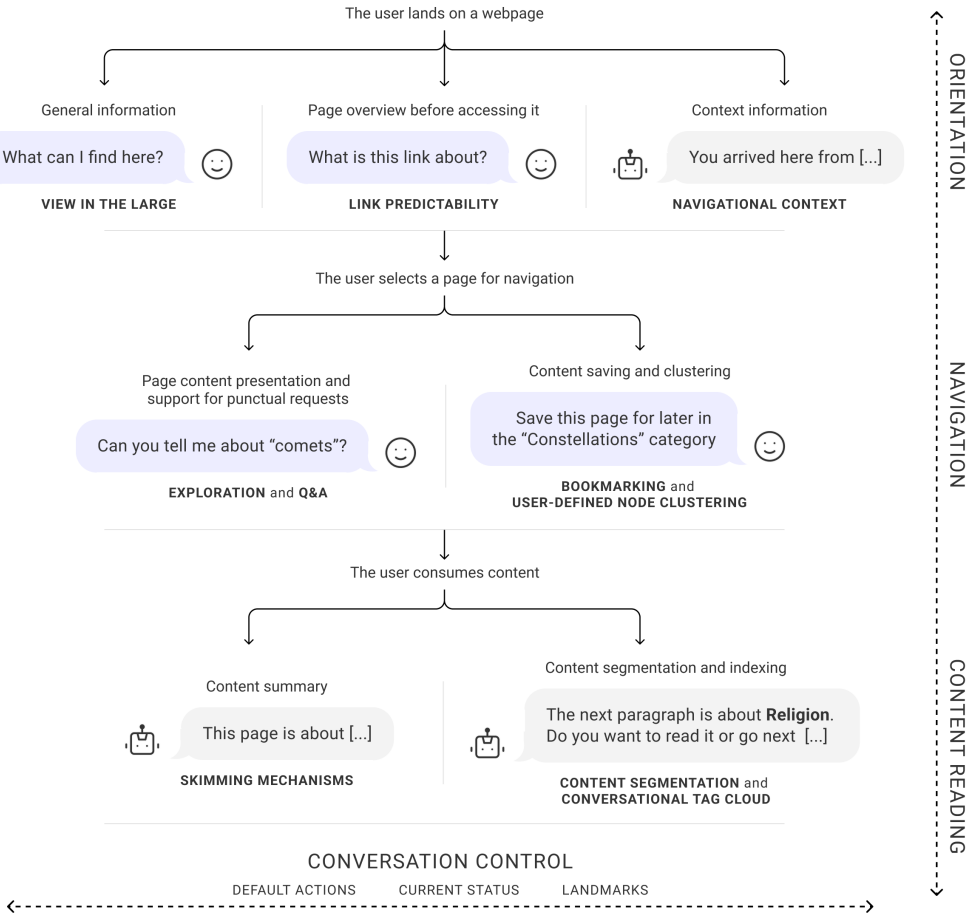

Figure 8: In-the-large organization of the dialog flow. The vertical arrow highlights the typical sequence of intent invocation. Control intent represented by the horizontal arrow can be invoked at any conversational step.

natural-language responses based on the activated intent handlers. The technical feasibility, the performance of the NLP-based generative techniques for handling the dialog, and the optimization strategies to achieve low response times are discussed in [7].

5 VALIDATION

Thanks to the availability of the ConWeb prototype, in January 2022 we conducted a preliminary evaluation of the identified patterns. We configured a CA to browse three Wikipedia pages: the Home Page and two other pages presenting content on the Solar System. The examples reported in Figures 4-7 are excerpts of the supported conversations. We then tested the prototype with 4 users who had previously participated in our user research activities (1 identified herself as a female; the others as males; average age: 30). Three of them already knew the idea behind the new conversational paradigm, having been involved in the previous formative activities. One of them had participated only in an initial interview and his involvement was interesting for validating the paradigm's ease of learning. Due to COVID restrictions, the sessions were organized remotely. ConWeb was installed on a server at the Politecnico di Milano. To avoid asking each participant to install the required Web browser extension, we deployed the CA on a purposely configured Web site. Thus, the participants were asked to connect to the website URL and from there they could navigate the Wikipedia pages through conversation[6].

The participants were individually asked to use the CA. After a brief introduction about the goal of the validation, the users were provided with general instructions about how to run the CA in their browser. Then they were assigned 9 browsing tasks, purposely defined to trigger the adoption of the defined patterns: 3 tasks on the Wikipedia Home Page focused on orientation aspects; 3 tasks on the Solar System article were conceived to assess content browsing aspects; 3 final tasks on the Jupiter and Venus articles focused on content reading. The CA fully integrated scaffolding commands to control navigation. The users were encouraged to navigate the site without further guidance. The researchers took note of the strategies adopted by the users.

Overall Performance and Experience. The participants successfully performed the tasks, except for **P22** who reported difficulties while performing Task 1. He formulated a request that

[6]A video demonstrating the scenario used for pattern validation is available at https://tinyurl.com/demo-validation

did not specify explicitly what he was looking for (*"I want to read again the other paragraph"*), and the CA responded with a generic error alert. **P22** commented that the answer should have been more specific about the error: *"I would like to figure out what to do or how to ask for help. ConWeb could ask me: What information do you want? Can you tell me what you would like to read?"*. This request suggests that the context tracking and the feedback on the system status could be improved [48]. However, in line with the findings of previous research investigating human-chatbot relationships [50, 51], it also lets us think that the participant considered the paradigm natural, til the point that he expected a very accurate language comprehension.

In general, the users manifested a good attitude and engagement with the interaction paradigm (**P23**: *"Controlling the website with my voice It's amazing"*; **P24**: *"Truly, it appears as a handy explorational method"*). They found it consistent with other CAs (**P23**: *"I find it natural to ask these things to ConWeb because with Alexa I'm used to doing this way"*). They liked the dialog-based interaction even if personalization was not allowed (**P22**: *"The conversation is very clear and pleasant, even if I was not able to configure the speech speed"*). The interaction paradigm was considered easy to learn (**P21**: *"After trying it two or three times, it comes naturally to me asking these things as I easily learned that ConWeb can do that"*).

Orientation. With the first proposed scenario on the Wikipedia Home Page, we wanted to assess orientation aspects. When performing these tasks, the users appreciated the ease of understanding the website structure and the navigational context. **P22** said: *"Having an initial page preview, for deciding later whether to explore the page in detail or not, helps me create my model of the site organization without accessing in detail any piece of content"*. One unexpected result is that the users extensively used the Q&A intent since the very first interactions with the site, although in our design hypotheses this mechanism was not conceived to facilitate orientation (**P23**: *"I like it because it's like finding a location on a map, I don't have to remember too many things"*).

Navigation. The 3 successive tasks were designed to understand how the users would perceive the patterns for content browsing. The users tried all the commands to explore the available areas. **P24** commented: *"There is a significant reduction in mental effort: it is possible to memorize smaller portions of the website, as the CA reminds the available navigation options, and it is easier to move along the different levels of the content hierarchy, also from the bottom to go up and reach the Home Page"*. As for the previous tasks, the most appreciated feature was the Q&A command (**P21**: *"Asking directly for a topic is an impressive option and represents one of the best features for me"*).

Content reading. For the last groups of tasks, the webpage content was chunked and organized according to the Skimming and Content Segmentation patterns. These mechanisms were not only spotted, yet also appreciated (**P22**:*"[. . .] I like that ConWeb is not "choosing for me", but rather makes it clear what I am selecting and which paragraphs I can listen to later."*). The participants positively commented on the intuitiveness of content reading mechanisms and the ease of moving along the different segments (**P21**: *"It anticipates many of the reading actions I'm thinking about, especially having pauses between different paragraphs, and being enabled to return to a previous intermediate segment"*).

Conversation control. In all three navigation scenarios, participants found the interaction with the CA reliable. Even if they remarked the need for customization options (see next section), which were not considered in the defined patterns, the users perceived the CA as reliable thanks to the direct mapping of its information structure with the webpage content (**P22**: *"There seems to be a high adherence with and a faithful interpretation of the Wikipedia content: the content is not redundant, and I understand the logic of the ConWeb proposal"*). Scaffolding patterns, such as knowing the actions that can be invoked at a given step, proved highly effective. More in general, users were not suspicious of the technology. Instead, they highlighted the potential of such a CA, also devising its use for screen readers' augmentation (**P23**: *"I've always wanted a hybrid between a screen reader and a voice assistant!"*).

6 DISCUSSION

The HCD process adopted in this work allowed us to progressively move from practices and challenges faced by BVI users to design dimensions and ultimately conversational patterns for Web browsing. In this section, we further reflect on the insights gained during the process and discuss the implications of our findings for the design of accessible conversational experiences on the Web.

6.1 Specificity for Web browsing

Current literature offers a valuable general framework for designing accessible conversational experiences [34, 37, 52] but the focus on Web browsing is limited. Our HCD process specifically reflected on Web browsing practices and challenges faced by BVI, and involved the observation, analysis, and co-design of conversational experiences across different informational websites. We run the evaluation on Wikipedia for having a quality representative baseline (i.e., a well-structured and accessible website) with which the users had prior experience, so that to gather user feedback on Web browsing and navigation through conversation, not on other factors that might derive from an unnecessary complexity in the website structure and accessibility problems. Thanks to the specific attention posed on content browsing mechanisms, also in comparison with other prominent solutions addressing Wikipedia, such as VERSE [56], our proposal goes beyond basic navigation commands and proposes articulated patterns that seem to respond to the Web browsing challenges identified during the HCD process.

6.2 Web browsing by BVI users

The preliminary validation suggests that the devised patterns can effectively support Web browsing while addressing some of the most prominent challenges faced by BVI users today. Participants perceived that the conversational paradigm enabled a more *natural interactions*, allowing them to more directly express their Web browsing requests without having to enact complex user workflows operating a screen reader. They perceived a *reduced cognitive effort*, offloading the need for memorizing the structure of websites. They acknowledged being *facilitated in the creation of a mental map*, without having to explore the entire content of websites, and that the dialog flow *lowered the barriers to learn* the navigational

Figure 9: Two participants interacting with the ConWeb during the final pattern validation.

approach and capabilities. While these results are preliminary, the response was positive and provides a foundation for larger-scale studies on the impact of the conversational paradigm and patterns in Web browsing scenarios.

6.3 Conversational Web for all

The patterns for conversational Web browsing were developed through close engagement with BVI people as users who can benefit from the technological development behind the idea of a conversational Web. However, reflecting on extreme cases is a powerful design approach and can also inform the design of a more general solution [2]. Our intuition is that the identified patterns can be beneficial for other user categories, including sighted users. However, further large-scale studies are needed to verify to what extent these patterns apply to a larger population and can thus contribute, in general, to the design of conversational agents.

6.4 Customization for specific needs and preferences

Besides aiming at an approach that can be beneficial for a large user population, it should be noted that going beyond the conversation flow and addressing the actual information presentation and consumption would require more tailored solutions. In this case, bringing voice-based interactions to a common denominator could hinder the accessibility for BVI users [15]. At the same time, other specific needs for inclusivity could also be hampered. The feedback from our own evaluation, as well as prior research on screen readers [27, 62] and numerous studies on voice-based interactions for BVI users [1, 21] suggest a need for a wider spectrum of voice-based interaction solutions, and high-level of control and configurability over the different nuances of the user experience (e.g., pitch, style, speed, multiple channels) that would not fit a single persona. Our studies have not investigated these aspects, as we purposely wanted to focus on content browsing and navigation. However, the literature offers multiple findings (e.g.,[1]) that can be integrated into our conversational paradigm.

Likewise, users expressed diversified needs for content reading, especially concerning in-text link detection: from communicating explicitly the presence of a link (**P2**: *"Personally, I would like to hear the word "link" every time, but I am aware that in Wikipedia this could be irritating for someone"*), to providing recognizable earcons (**P1**: *"I prefer a nice background sound instead of listening to an interrupted dialogue"*), or even to disabling any mechanism to prevent annoying repetitions (**P4**: *"I first read the entire text and then proceed by searching for links I could be interested in"*).

This variability in the participants' preferences for voice-based content presentation lets us think that the configurability of conversational elements is the key to improving the conversational experience for any user, and will therefore be the object of further studies and design activities.

6.5 Further insights

The HCD process shed light on several aspects. For the definition of patterns, also considering the recurrence of the emerged themes, we focused on content browsing dimensions. However, several complementary perspectives, relevant to conversational Web browsing, were also unfolded. We summarize here the most prominent considerations as suggestions for future research directions.

Direct queries vs exploration. Recent studies have identified undergoing changes in mental structures for file browsing and storing among young people [19]. Years of googling and social network browsing have changed how information is perceived and retrieved. During the validation phase, participants tried indeed to access information with direct queries, both for getting oriented and exploring the navigation options. Even while consuming the content of single pages, direct requests were the most used (*"What about x?"*; *"Where can I find y?"*). Contrary to our expectations, the layered exploration from the home page to the inner nodes, which emerged as a need during the formative studies, was then mostly abandoned. Future investigations could examine whether this trend is confirmed with a more extensive and diversified set of users.

Multi-experience paradigms. Participants with visual capabilities, who use the visual segmentation of the text and the length of paragraphs to orient themselves within a webpage, pointed to a

deeper synergy among visual and conversation paradigms (**P23**: *"I can also orient myself "by sight" while using ConWeb! I cannot read the text, but I identify the correspondence between text reading and the "visual shape" of the paragraphs."*). In line with previous works [40], one interesting direction to meet the diversity of needs could therefore be a mixed paradigm that, besides leveraging the conversation, could also offer an integration of visual and conversational access. As reported in the literature this would be beneficial for other classes of users, for example, older adults [29].

Integration with screen readers. For most of the participants, the opportunity of naturally conversing with a webpage was thrilling, yet some participants expressed doubts. **P22**, who in the final phases of the study expressed great enthusiasm for the new paradigm, in the initial focus groups had observed: *"Screen readers will always be my preference as assistive technology for their "passive" nature as I would not be forced to have a verbal interaction"*. Also in line with previous work [56], this suggests that the next studies should put effort into discovering new synergies among existing assistive technologies (such as screen readers) and the conversational paradigm, to give the users a chance to choose the most adequate one depending on their tasks and the context of use. Since the identified patterns and the logic governing our technological platform do not prescribe any specific input and output mode, the integration is feasible and can be achieved through adequate communication and synchronization mechanisms, with the conversational paradigm becoming an orchestrator.

6.6 Limitations

Together with the previous considerations, the limitations of our work will be worth addressing as future work to consolidate the notion of the Conversational Web.

Patterns coverage. Our studies focused on conversational experiences often inspired by informational websites with a regular and consistent organization, e.g., Wikipedia. If, on the one hand, this allowed us to addressmany aspects of conversational Web browsing, on the other hand, it could constitute a limitation for the generalization of the approach to other classes of websites. The need to address the presentation of dynamic components already emerged during the conducted studies: when using the screen reader to interact with a train-reservation website, the participants were not able to select the departure date on a calendar component (**P25**: *"These letters spoken by the screen reader I believe are the days' initials, but it took me a bit to figure out"*; **P26**: *"We had a lot of troubles inserting the departure time since we had to select it from a calendar visualization"*). So far, we have given priority to the design of patterns focusing on textual content within Web pages but our current work is already devising solutions for intercepting and presenting dynamic page components. We will also extend the evaluation to the navigation of more complex websites. However, we are confident that the defined patterns can still provide a solid basis for organizing conversational access.

CNT management. On the technical side, our approach relies on the construction of the CNT model, which is built automatically every time a webpage is accessed. The dialog is also automatically built depending on the created CNT instance. In our current prototype, these mechanisms work perfectly if the website HTML is augmented with proper tags (i.e., for websites natively instrumented to be accessed through ConWeb). They are highly accurate, even in the absence of *ad-hoc* tagging, when websites have a regular structure, while the page interpretation may fail with highly dynamic websites and dynamic components. This might hinder the adoption of ConWeb, and that is why further work will focus on making page interpretation more robust.

Diversity of study participants and sample size. One limitation of the studies was the age of the participants who were almost all young adults. The recruitment for these phases was done on a volunteering basis and the young adults were the most enthusiastic. While on one side this can hinder the representativeness of the insights for the regular user population, on the other side user comments pinpointed the youngest users as the adequate target. Older participants might be too accustomed to screen readers to appreciate or even show an attitude toward new technology. For example, **P17** said: *"Learning screen reader technology was a slow and difficult process. At my age [64 years old] I do not want to spend even more time trying something new because what I use already works enough for me!"*. Also, participants were particularly hard to find for the study, resulting in a limited sample size. Future studies can specifically focus on assessing the attitude toward the new proposed technology among users of different ages, thus enhancing diversity and improving the robustness of the results. Gender-related issues could also be investigated, since out of the 26 participants in our studies only 6 were female. Finally, the study participants were very interested and motivated to adopt the assessed technology. Diversity in relation to the participants' experience with screen readers and digital technology should be pursued.

7 CONCLUSIONS

This paper has discussed a new paradigm for conversational Web browsing, as emerged from a human-centered process conducted with a sample of BVI users. The illustrated results aim to fill the current gap in the literature about concrete guidance on how to design conversational agents for the Web. Our belief is that effective guidelines are those defined by directly involving the users to co-design *with them* possible solutions. To consolidate this strategy, we are planning new and large-scale studies, with a balanced involvement of participants based on their diverse characteristics, to further validate the identified patterns and investigate in detail how the nuances emerged during the studies, yet not covered by our patterns, can be captured by meaningful conversational solutions.

More in general, in line with recent standardization initiatives[7], our goal is to promote the notion of *Conversational Web* by means of innovative Web technologies that can natively support conversational access. Our current efforts are therefore devoted to consolidating the ConWeb prototype, to understand how Conversational AI and NLP techniques can efficiently sustain the generation of a dialog system like the one discussed in Section 4. In our current prototype, the CA capability of enabling Web browsing relies on the availability of ad-hoc HTML tags within the webpage code that enable the CNT construction. We are now consolidating techniques for the lightweight integration of NLP and AI techniques with Web technologies, to build "domain knowledge" by automatically

[7] An example is the *conversation tag* introduced by schema.org

extracting from the HTML code relevant features of the website content and functionality. The dream is to have technologies "inclusive by design", that seamlessly grant access to the Web through voice interfaces, without any extra effort for the development of ad-hoc conversational agents.

8 ACKNOWLEDGEMENTS

This research is partially supported by the PNRR-PE-AI FAIR project funded by the NextGeneration EU program.

We are grateful to the associations *Unione Italiana Cieche e Ipovedenti* (UICI), *Associazione Disabili Visivi* (ADV), *Real Eyes Sport*, and to the users who took part in our studies, for the help given in the definition of the ConWeb paradigm.

REFERENCES

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. 2018. " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proc. of the 20th Int. ACM SIGACCESS Conference on Computers and Accessibility*. 249–258.

[2] Ali Abdolrahmani, Kevin M Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M Branham. 2020. Blind leading the sighted: drawing design insights from blind users towards more productivity-oriented voice interfaces. *ACM Trans. on Accessible Computing (TACCESS)* 12, 4 (2020), 1–35.

[3] Faisal Ahmed, Yevgen Borodin, Andrii Soviak, Muhammad Islam, IV Ramakrishnan, and Terri Hedgpeth. 2012. Accessible skimming: faster screen reading of web pages. In *UIST*. ACM, 367–378.

[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[5] Vikas Ashok, Syed Masum Billah, Yevgen Borodin, and IV Ramakrishnan. 2019. Auto-suggesting browsing actions for personalized web screen reading. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. 252–260.

[6] Vikas Ashok, Yevgen Borodin, Yury Puzis, and IV Ramakrishnan. 2015. Capti-speak: a speech-enabled web screen reader. In *W4A*. ACM, 22.

[7] Marcos Baez, Cinzia Cappiello, Claudia M Cutrupi, Maristella Matera, Isabella Possaghi, Emanuele Pucci, Gianluca Spadone, and Antonella Pasquale. 2022. Supporting Natural Language Interaction with the Web. In *International Conference on Web Engineering*. Springer, 383–390.

[8] Marcos Baez, Florian Daniel, and Fabio Casati. 2019. Conversational web interaction: proposal of a dialog-based natural language interaction paradigm for the web. In *International Workshop on Chatbot Research and Design*. Springer, 94–110.

[9] Marcos Baez, Florian Daniel, Fabio Casati, and Boualem Benatallah. 2020. Chatbot integration in few patterns. *IEEE Internet Computing* 25, 3 (2020), 52–59.

[10] Moshe Chai Barukh, Shayan Zamanirad, Marcos Baez, Amin Beheshti, Boualem Benatallah, Fabio Casati, Lina Yao, Quan Z Sheng, and Francesco Schiliro. 2021. Cognitive augmentation in processes. In *Next-Gen Digital Services. A Retrospective and Roadmap for Service Computing of the Future*. Springer, 123–137.

[11] Jeffrey P Bigham, Tessa Lau, and Jeffrey Nichols. 2009. Trailblazer: enabling blind users to blaze trails through the web. In *IUI*. ACM, 177–186.

[12] Yevgen Borodin, Faisal Ahmed, Muhammad Asiful Islam, Yury Puzis, Valentyn Melnyk, Song Feng, IV Ramakrishnan, and Glenn Dausch. 2010. Hearsay: a new generation context-driven multi-modal assistive web browser. In *WWW*. ACM, 1233–1236.

[13] Yevgen Borodin, Jalal Mahmud, IV Ramakrishnan, and Amanda Stent. 2007. The HearSay non-visual web browser. In *Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A)*. ACM, 128–129.

[14] Sara Bouguelia, Hayet Brabra, Shayan Zamanirad, Boualem Benatallah, Marcos Baez, and Hamamache Kheddouci. 2021. Reusable Abstractions and Patterns for Recognising compositional conversational flows. In *Int. Conf. on Advanced Information Systems Engin.* Springer, 161–176.

[15] Stacy M Branham and Antony Rishin Mukkath Roy. 2019. Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 446–458.

[16] Julia Cambre, Alex C Williams, Afsaneh Razi, Ian Bicking, Abraham Wallin, Janice Tsai, Chinmay Kulkarni, and Jofish Kaye. 2021. Firefox Voice: An Open and Extensible Voice Assistant Built Upon the Web. In *Proc. of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[17] Nicola Castaldo, Florian Daniel, Maristella Matera, and Vittorio Zaccaria. 2019. Conversational data exploration. In *international conference on Web Engineering*. Springer, 490–497.

[18] Yung-Ju Chang, Chu-Yuan Yang, Ying-Hsuan Kuo, Wen-Hao Cheng, Chun-Liang Yang, Fang-Yu Lin, I-Hui Yeh, Chih-Kuan Hsieh, Ching-Yu Hsieh, and Yu-Shuen Wang. 2019. Tourgether: Exploring Tourists' Real-time Sharing of Experiences as a Means of Encouraging Point-of-Interest Exploration. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4 (2019), 128:1–128:25. https://doi.org/10.1145/3369832

[19] Monica Chin. 2022. Students who grew up with search engines might change STEM education. https://www.theverge.com/22684730/students-file-folder-directory-structure-education-gen-z.

[20] Pietro Chittò, Marcos Baez, Florian Daniel, and Boualem Benatallah. 2020. Automatic generation of chatbots for conversational web browsing. In *International Conference on Conceptual Modeling*. Springer, 239–249.

[21] Dasom Choi, Daehyun Kwak, Minji Cho, and Sangsu Lee. 2020. " Nobody speaks that fast!" An empirical study of speech rate in conversational agents for people with vision impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[22] Michael Cormer, Richard Mann, Karyn Moffatt, and Robin Cohen. 2017. Towards an improved vision-based web page segmentation algorithm. In *2017 14th Conference on Computer and Robot Vision (CRV)*. IEEE, 345–352.

[23] Michael Cormier, Karyn Moffatt, Robin Cohen, and Richard Mann. 2016. Purely vision-based segmentation of web pages for assistive technology. *Comput. Vis. Image Underst.* 148 (2016), 46–66. https://doi.org/10.1016/j.cviu.2016.02.007

[24] Florian Daniel, Maristella Matera, Vittorio Zaccaria, and Alessandro Dell'Orto. 2018. Toward truly personal chatbots: on the development of custom conversational assistants. In *Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services, SE4COG@ICSE 2018, Gothenburg, Sweden, May 28-2, 2018*, Hamid R. Motahari Nezhad, Rao Mikkilineni, Boualem Benatallah, Fabio Casati, Schahram Dustdar, Gordana Dodig-Crnkovic, and Adrian Mos (Eds.). ACM, 31–36. https://doi.org/10.1145/3195555.3195563

[25] Michael H Fischer, Giovanni Campagna, Euirim Choi, and Monica S Lam. 2021. DIY assistant: a multi-modal end-user programmable virtual assistant. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*. 312–327.

[26] Asbjørn Følstad, Theo Araujo, Effie Lai-Chong Law, Petter Bae Brandtzaeg, Symeon Papadopoulos, Lea Reis, Marcos Baez, Guy Laban, Patrick McAllister, Carolin Ischen, et al. 2021. Future directions for chatbot research: an interdisciplinary research agenda. *Computing* (2021), 1–28.

[27] João Guerreiro and Daniel Gonçalves. 2015. Faster Text-to-Speeches: Enhancing Blind People's Information Scanning with Faster Concurrent Speech. In *ACM SIGACCESS*. ACM, 3–11.

[28] Simon Harper and Neha Patel. 2005. Gist summaries for visually impaired surfers. In *ACM SIGACCESS Conf. on Computers and Accessibility*. ACM, 90–97.

[29] Michael Heron, Vicki L. Hanson, and Ian W. Ricketts. 2013. Accessibility Support for Older Adults with the ACCESS Framework. *International Journal of Human–Computer Interaction* 29, 11 (2013), 702–716. https://doi.org/10.1080/10447318.2013.768139 arXiv:https://doi.org/10.1080/10447318.2013.768139

[30] Jonggi Hong, Christine Vaing, Hernisa Kacorri, and Leah Findlater. 2020. Reviewing Speech Input with Audio: Differences between Blind and Sighted Users. *ACM Transactions on Accessible Computing (TACCESS)* 13, 1 (2020), 1–28.

[31] Tessa Lau, Julian Cerruti, Guillermo Manzato, Mateo Bengualid, Jeffrey P Bigham, and Jeffrey Nichols. 2010. A conversational interface to web automation. In *UIST*. ACM, 229–238.

[32] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019).

[33] Toby Jia-Jun Li and Oriana Riva. 2018. KITE: Building conversational bots from mobile apps. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 96–109.

[34] Kate Lister, Tim Coughlan, Francisco Iniesto, Nick Freear, and Peter Devine. 2020. Accessible conversational user interfaces: considerations for design. In *Proceedings of the 17th International Web for All Conference*. 1–11.

[35] Jimmie Manning. 2017. In vivo coding. *The international encyclopedia of communication research methods* 24 (2017), 1–2.

[36] Christine Murad, Cosmin Munteanu, Benjamin R Cowan, and Leigh Clark. 2019. Revolution or evolution? Speech interaction and HCI design guidelines. *IEEE Pervasive Computing* 18, 2 (2019), 33–45.

[37] Theresa Neil. 2014. *Mobile design pattern gallery: UI patterns for smartphone apps*. " O'Reilly Media, Inc.".

[38] Jakob Nielsen. 2000. Designing web usability. (2000).

[39] Alessandro Pina, Marcos Baez, and Florian Daniel. 2020. Bringing Cognitive Augmentation to Web Browsing Accessibility. In *International Conference on Service-Oriented Computing*. Springer, 395–407.

[40] Elena Planas, Gwendal Daniel, Marco Brambilla, and Jordi Cabot. 2021. Towards a model-driven approach for multiexperience AI-based user interfaces. *Software and Systems Modeling* 20, 4 (2021), 997–1009.

[41] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. " Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on human factors in computing systems*. 1–13.

[42] Mickael Rajosoa, Rim Hantach, Sarra Ben Abbes, and Philippe Calvez. 2019. Hybrid question answering system based on natural language processing and SPARQL query. In *Proc. 3rd Int. Workshop Appl. Knowl. Represent. Semantic Technol. Robotics*. 94–102.

[43] Rasa. 2022. *Rasa Platform Docs*. https://rasa.com/docs/

[44] Gonzalo Ripa, Manuel Torre, Sergio Firmenich, and Gustavo Rossi. 2019. End-User Development of Voice User Interfaces Based on Web Content. In *IS-EUD 2019*. Springer, 34–50.

[45] Yara Rizk, Abhishek Bhandwalder, Scott Boag, Tathagata Chakraborti, Vatche Isahagian, Yasaman Khazaeni, Falk Pollock, and Merve Unuvar. 2020. A unified conversational assistant framework for business process automation. *arXiv preprint arXiv:2001.03543* (2020).

[46] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for Building an Open-Domain Chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, Paola Merlo, Jörg Tiedemann, and Reut Tsarfaty (Eds.). Association for Computational Linguistics, 300–325. https://doi.org/10.18653/v1/2021.eacl-main.24

[47] Donya Rooein, Devis Bianchini, Francesco Leotta, Massimo Mecella, Paolo Paolini, and Barbara Pernici. 2022. aCHAT-WF: Generating conversational agents for teaching business process models. *Software and Systems Modeling* 21, 3 (2022), 891–914.

[48] Abigail See, Stephen Roller, Douwe Kiela, and Jason Weston. 2019. What makes a good conversation? How controllable attributes affect human judgments. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 1702–1723. https://doi.org/10.18653/v1/n19-1170

[49] Selenium. 2022. *The Selenium Browser Automation Project*. https://www.selenium.dev/documentation/

[50] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My Chatbot Companion - a Study of Human-Chatbot Relationships. *Int. J. Hum. Comput. Stud.* 149 (2021), 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

[51] Marita Skjuve, Ida Haugstveit, Asbjørn Følstad, and Petter Brandtzaeg. 2019. Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology* 15 (02 2019), 30–54. https://doi.org/10.17011/ht/urn.201902201607

[52] Jeff Stanley, Ronna ten Brink, Alexandra Valiton, Trevor Bostic, and Becca Scollan. 2022. Chatbot Accessibility Guidance: A Review and Way Forward. In *Proceedings of Sixth International Congress on Information and Communication Technology*. Springer, 919–942.

[53] Jenifer Tidwell. 2010. *Designing interfaces: Patterns for effective interaction design*. " O'Reilly Media, Inc.".

[54] Mandana Vaziri, Louis Mandel, Avraham Shinnar, Jérôme Siméon, and Martin Hirzel. 2017. Generating chat bots from web API specifications. In *Proc. of the 2017 ACM SIGPLAN Int. symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*. 44–57.

[55] Bostjan Vesnicer, Janez Zibert, Simon Dobrisek, Nikola Pavesic, and France Mihelic. 2003. A voice-driven web browser for blind people. In *Eighth European Conference on Speech Communication and Technology*.

[56] Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W White. 2019. Bridging screen readers and voice assistants for enhanced eyes-free web search. In *The world wide web conference*. 3590–3594.

[57] W3C. 2017. *W3C Accessibility Guidelines (WCAG) 3.0*. https://www.w3.org/TR/wcag-3.0

[58] WebAIM. 2017. *The WebAIM Million: An accessibility analysis of the top 1,000,000 home pages*. https://webaim.org/projects/million/

[59] WebAIM. 2022. *Alexa Brand Guidelines for Amazon Developers*. https://developer.amazon.com/en-US/alexa/branding/alexa-guidelines

[60] Shunguo Yan and PG Ramachandran. 2019. The current status of accessibility in mobile apps. *ACM Trans. on Accessible Computing* 12, 1 (2019), 1–31.

[61] Shayan Zamanirad, Boualem Benatallah, Moshe Chai Barukh, Fabio Casati, and Carlos Rodriguez. 2017. Programming bots by synthesizing natural language expressions into API invocations. In *32nd IEEE/ACM Int. Conf. on Automated Software Engineering (ASE)*. IEEE, 832–837.

[62] Shaojian Zhu, Daisuke Sato, Hironobu Takagi, and Chieko Asakawa. 2010. Sasayaki: an augmented voice-based web browsing experience. In *Proc. of SIGACCESS 2010*. ACM.

[63] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 493–502.