LadderBot: A requirements self-elicitation system

Tim Rietz
Institute of Information Systems and Marketing (IISM) Karlsruhe Institute of Technology (KIT) Karlsruhe, Germany tim.rietz@kit.edu

Alexander Maedche
Institute of Information Systems and Marketing (IISM) Karlsruhe Institute of Technology (KIT) Karlsruhe, Germany alexander.maedche@kit.edu

Abstract—[Context] Digital transformation impacts an ever-increasing amount of everyone’s business and private life. It is imperative to incorporate user requirements in the development process to design successful information systems (IS). Hence, requirements elicitation (RE) is increasingly performed by users that are novices at contributing requirements to IS development projects. [Objective] We need to develop RE systems that are capable of assisting a wide audience of users in communicating their needs and requirements. Prominent methods, such as elicitation interviews, are challenging to apply in such a context, as time and location constraints limit potential audiences. [Research Method] We present the prototypical self-elicitation system “LadderBot”. A conversational agent (CA) enables end-users to articulate needs and requirements on the grounds of the laddering method. The CA mimics a human (expert) interviewer’s capability to rephrase questions and provide assistance in the process. An experimental study is proposed to evaluate LadderBot against an established questionnaire-based laddering approach. [Contribution] This work-in-progress introduces the chatbot LadderBot as a tool to guide novice users during requirements self-elicitation using the laddering technique. Furthermore, we present the design of an experimental study and outline the next steps and a vision for the future.

Index Terms—User, Requirements Elicitation, Wide Audience, Conversational Agent, Self-Elicitation, Laddering

I. INTRODUCTION

Digital transformation has brought a variety of information systems into everyone’s business and private life with a substantial impact on business and society [1]. We observe a transformation towards a digital society, stressing the influence of the Internet on many traditional services, which advocates a power shift towards the user [2]. In the face of persistently high failure rates of IS development projects, it is imperative that an increasing number of users is involved in RE processes, with a varying degree of technological and methodological expertise [3]. The scalable elicitation of user requirements is crucial for developing software that meets needs and demands and to reduce project failure [4]. Consequently, RE needs to be performed with a wide range of users that are novices at contributing requirements to development projects [1].

For requirements elicitation, interviews have been used most widely [5]. Especially the laddering interview is considered a very effective technique for eliciting relevant information for articulating requirements [5]. Laddering produces comprehensive and structured insights due to the method’s hierarchical nature. In laddering, an interviewer identifies a seed attribute, an initial topic, and asks a series of “why…” questions to uncover and clarify needs and related attitudes [6]. While having its roots in personality psychology, laddering has already seen usage for requirements elicitation [4] (e.g., to elicit Customer Attribute Hierarchies [7]). Essentially, requirements are elicited as attribute-consequence-value (ACV) chains [6]. Since laddering interviews require highly trained and experienced interviewers, the availability of suitable interviewers imposes a bottleneck onto elicitation interviews [6]. Tool support is necessary to enable requirements elicitation with a wide range and number of users [8].

Several tools to aid with wide audience elicitation have been proposed over the years. AnnotatePro allows users to submit requirements that can be drawn on their screens [9]. Given the common problems with requirements quality, such as completeness and ambiguity, exploring natural-language (NL) based elicitation systems gained traction. Pérez and Valderas (2009) combine visualization-based RE with NL to reduce ambiguity and inconsistency in end-user RE. Derrick et al. (2013) evaluated an embodied conversational agent to facilitate a group workshop that used prompts to guide and assist during user story formulation [10]. However, these tools do not suffice in providing a solution to both challenges introduced: Annotation-based tools primarily enable RE for iteratively improving existing systems; NL tools commonly require a requirements engineer to facilitate the process, hence retaining a bottleneck for wide audience integration [11]; additionally, existing research rarely considers (methodological) guidance for novice end-users. Tools such as FAME [12] and ASSERT [13] cater to novices, but only on the side of a novice analyst, not novice users, hence not enabling self-elicitation. A literature gap remains in extending RE techniques to wide audiences. Guidance and assistance are necessary to facilitate the elicitation of high-quality requirements from novice users [14], [15]. We utilize a conversational agent (CA) in the form of a chatbot to mimic a human interviewer’s capability to guide an interview [10]. Chatbots allow us to include a wide audience of users, independent of personal, time, or location restrictions and may guide novice users through laddering interviews. Therein, we extend our previous research on (semi-automated RE be explicitly focusing on the collection of unstructured data on the basis of self-elicitation interviews [16], [17].
II. Conceptual Foundations

A. Common issues in user elicitation interviews

To understand the implications for a novice-centric self-elicitation system, we need to understand the characteristics of the requirements (self-)elicitation behavior of novices. In this article, we refer to self-elicitation of requirements rather than self-service RE system. As the user is guided in uncovering their requirements, rather than being enabled to create a service with a direct benefit for themselves, we argue that self-elicitation serves as a better term to describe the process.

So far, RE literature rarely focuses on characteristics of novice users to be supported in elicitation processes [1]. Commonly, novice RE analysts are the focus of supporting activities [18]. However, insights from analyzing the behavior of novice analysts in elicitation processes may serve as a guideline for how to provide appropriate support for requirements self-elicitation.

Notably, one of the most frequently observed downsfalls in elicitation performed with novice users or by novice analysts is a lack of structure [19]. A lack of structure results in interviewers not digging deep enough when conducting interviews, impacting requirements correctness [15]. Since especially novice users are not familiar with communicating requirements, which may be rooted in an incomplete understanding of their own needs, the task of uncovering the cause of a need or requirement falls to the interviewer. Otherwise, interviews lead to ambiguous interview statements at the wrong level of abstraction [13]. Without uncovering the cause of, or foundation for user needs, the development of disruptive solutions stagnates. We can avoid common mistakes of novice analysts that happen during interviews, such as question formulation, ordering, and question omission through a predefined interview structure [18]. Furthermore, the analyst’s behavior, such as lack of confidence or unprofessionalism, or inadequate time management, has a substantial impact on the results of an interview [21]. Hence, both structural and behavioral interview guidelines are necessary for eliciting high-quality requirements.

Analysts should be educated in thinking in relations, hence applying model-based reasoning rather than object-attributes to increase the performance of requirements analysis [20]. We propose that by using an elicitation structure following the laddering technique, we can enable users to generate requirements in a relation-focused fashion, contributing to the quality of requirements specification. Fig. 1 provides an overview of how the conceptual foundations feed into the development of LadderBot.

B. The laddering interview technique for RE

Laddering is a cognitive interview technique with its roots in personality psychology that utilizes a structured approach for data-gathering [6]. For RE, cognitive techniques, in comparison to traditional, collaborative, or contextual techniques, are commonly used to acquire knowledge. As such, requirements are not directly communicated but extracted from the structure and content of user knowledge based on rich enough information [21]. Herein, cognitive techniques provide the most natural interaction with end-users [21].

![Fig. 1. Overview of the conceptual foundations of LadderBot](image)

Laddering was introduced as a method to elicit superordinate items from subordinate ones, to clarify the relations between items obtained using the repertory grid method, with its origin in personal construct theory. However, the laddering technique has primarily been used for knowledge-elicitation in marketing and advertising [22]. As such, the technique has become a tool for the means-end theory in marketing. The means-end theory distinguishes three levels of abstraction of meaning that users obtain from a purchase or consume [6]. These three levels are described as ACV chains: attributes – consequences – values [23]. Attributes as the least abstract level describe “concrete, physical, or observable characteristics” of products. Despite the notion initially describing physical products, we may use the idea for digital products like software, too [24]. Consequences constitute the second level of abstraction. They describe what a product provides a user with, either on the positive (benefits) or negative side (costs). A product can have functional or non-functional, e.g., psychosocial, consequences. Values are the most abstract...
level. They represent a user’s wishes, goals, and needs and are the end state a customer is trying to achieve through a purchase. An exemplary ACV chain in a software development context has the following form: Providing default values (A) – No need to fill out data repeatedly (C) – Happiness (V) [24].

The laddering technique usually comprises three steps: elicitation of attributes, a ladder of representation, and representing and analyzing the results. Attributes serve as the seed for the interview, in the form of lower order characteristics with implications for higher-order cognitive processes and determine the direction of the interview. As such, multiple methods of generating attributes have been used, depending on the purpose of related study. The laddering interview itself follows a straightforward structure. Participants are asked why a particular attribute is important to them, using a series of “why…?” questions while navigating through the ACV chains. E.g., an interviewer might ask, “why is starting process X from the landing page is important?”.

A content-coding procedure initializes the analysis process of laddering interviews. These codes are then used to build a summary matrix, visualizing each chain from each participant, showing the included codes per chain. Subsequently, an aggregate implication matrix is formed, showing the aggregated information across interviews. This matrix contains all direct and indirect relations between attributes, consequences, and values. Finally, we can visualize the aggregate implication matrix as a hierarchy value map, a tree diagram showing either only direct or both direct and indirect relations at a specified cut-off value (for examples, see [24]–[26]).

C. Form and Function of Chatbots

The goal of CAs, as McTear (2002) puts it, is the “[…] effortless, spontaneous communication with a computer”. Klopfenstein et al. (2017) conducted a systematic analysis of one of the instantiations of CAs, chatbots, categorizing advantages for users and developers. They find instant availability, a gentle learning curve, and platform independence to be among the most prominent benefits. Hence, we argue that chatbots serve as a promising form of CAs for approaching a large number of users. Instant availability and platform independence enable barrier-free interaction with the system. A gentle learning curve, resulting from an interaction mode that is familiar to novice users, writing, creates an effortless experience. Multiple variants of chatbots have seen use over the years, which can be differentiated according to form and function. The form of a chatbot describes the arrangement of aspects that do not primarily contribute to the utility of the bot (similar to non-functional requirements). For example, anthropomorphism comprises methods for making the appearance and behavior of a bot more human-like. Function describes aspects related to general performance, such as the bot’s dialogue control strategy. A frame-based bot uses question templates to provide information back to a user. These systems do not have pre-determined dialogue flows but adapt to user input, e.g., a software problem reparation tool.

Despite a renewed research interest in chatbots, due to advances in artificial intelligence, the integration of CAs into RE remains spare. Derrick et al. (2013) investigated the effect of a simple scripted agent in facilitating group elicitation sessions with users while other studies developed prototypes for frame-based agents in interview scenarios. While these studies evaluated the general applicability of CAs as facilitators of elicitation processes, to the best of our knowledge, no evaluation of chatbot-based requirements elicitation with a wide audience of end-users has been conducted, comparing the performance of a system with established processes on the basis of measures such as performance and perception.

III. LadderBot

LadderBot uses a two-column visualization, with a graphical representation of ACV chains on the left and a frame-based chatbot on the right, as shown in fig. 2. Initially, LadderBot welcomes users and provides a short explanation of the interface and the interview process. We adapted the subsequent laddering interview structure from Jung (2014). To begin the interview, LadderBot asks the user to state the three most frequently used features of a system as seed attributes for each chain. The following process is then repeated until participants constructed three chains. At the beginning of each chain, LadderBot asks an initial question to elicit the first consequence for the current attribute:

LadderBot: “As 2. example, you said Email. Why do you use Email? What do you obtain by using the function?”
User: “I need to know if someone needs something from me, and see if I got any updates from the services I signed up for.”

Rather than asking an initial default question, LadderBot integrates the specific attribute that users selected into question formulation. The line of questioning for consequences and values is repeated until a value is identified, or the user is unable to provide a more precise answer. When asking why-questions repeatedly, the chatbot will rely on four techniques for rephrasing questions to help and guide the user. We adapted these techniques from suggestions for human interviewers on how to conduct laddering interviews, as described in table 1. Fig. 3 depicts a visual overview of how the solution works in a laddering interview. For now, the four techniques are applied by LadderBot randomly. The rephrasing techniques primarily incorporate the seed attribute of the current ladder into the question formulation.

![LadderBot interface](image-url)
User replies are used for rephrasing only in the form of quotes, for ensuring that the resulting question makes sense. The visualization of the current status of the interview on the left side updates for each elicited consequence. The graphical representation of ACV chains may assist users in structuring their thoughts and uncovering new relations [20]. When asking a series of questions, a human interviewer would need to identify if the user has described the value that they satisfy through an attribute to end the elicitation for a specific attribute or to end the interview in general (e.g. [25]). As the current iteration of LadderBot is not capable of recognizing whether a user has already described a final value on its own, the bot requires the user to indicate if they want to continue the laddering process for the current attribute, or switch to the next chain. The user can make this indication with a predefined command ("stop"). The questioning process for each of the three ladders is continued until the stop command is given and LadderBot concludes the session.

The current implementation of LadderBot does not impose restrictions on the length of an answer of a user, to keep the interaction with the chatbot as natural as possible. Long replies impose a challenge for LadderBot in formulating an appropriate question as a response. As such, user replies are incorporated in questions only as complete references. Furthermore, LadderBot uses the three features provided by the users at the start of the interview to formulate more direct questions, as we identified these replies to be rather short. Users are currently not capable of making changes to previous answers. However, we plan to include this functionality in future iterations. As the technological foundation of LadderBot, we use the Microsoft Bot Framework on node.js. To visualize elicited ACV chains, we integrate the bot into a web application build on the frameworks d3.js and bootstrap. This architecture allows for a straightforward reconfiguration of the artifact to change the laddering use case or the interview structure.

IV. EXPERIMENTAL STUDY DESIGN

To evaluate LadderBot, we will conduct an experimental study. The experiment procedure and the applied measurements will partially build on previous studies that evaluated elicitation techniques [33], [34] or used the laddering technique as part of their experiment design [26].

We will conduct the study with students from a large university in Germany in an experimental lab designed for conducting scientific studies. As laddering case, we recreate the laddering structure applied by Jung (2014) to elicit the users’ goals for smartphone use. Such results may be used to uncover requirements to develop or improve an IS for smartphones. Similar to the original study, we will invite students as participants, while controlling for the participants’ experience with development projects and laddering interviews. Around 200 students will be invited, randomly selected from a pool of potential participants.

TABLE I. QUESTION REPHRASING TECHNIQUES

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative laddering</td>
<td>Ask the user why they do not do something or do not want to feel a certain way</td>
<td>What problems could be caused by Email? How would Email have to change to mitigate these problems?</td>
</tr>
<tr>
<td>Exclusion</td>
<td>Ask the user to imagine a situation where an attribute or consequence does not exist</td>
<td>Imagine you could not use Instagram. What alternatives to Instagram would you use and why?</td>
</tr>
<tr>
<td>Retrospective</td>
<td>Ask the user to imagine their behavior in the past and compare it to now</td>
<td>Has your perception of this changed compared to a couple of years ago? If so, why is that and what changed?</td>
</tr>
<tr>
<td>Clarification</td>
<td>Repeat a reply back to the user and ask for clarification</td>
<td>Okay, you just said “I want a real-time newsfeed”, right? In the context of Instagram, could you explain that to me in more detail?</td>
</tr>
</tbody>
</table>
The experimental study will use a between-subject design with three treatments. Across treatments, participants will be asked to conduct a self-elicitation of their goals in smartphone use. Treatments will be characterized by the available interview tool and the interview visualization. In treatment (1), participants will use an established version of a “pencil-and-paper” laddering questionnaire [26]. However, a digital questionnaire will be used to increase comparability with other treatments. In treatment (2), participants will use the same questionnaire as in treatment (1) but augmented with the visualization used in LadderBot to keep track of already elicited ladders. In treatment (3), participants will use LadderBot to complete the laddering interview. As such, only one of either the visualization or the interview tool presented to participants is changed between treatments. Thereby, we aim to increase the comparability of results between treatments while being able to evaluate the visualization and chatbot interface features of LadderBot separately.

We will evaluate the treatments using a combination of quantitative measurements. Herein, we rely on the established procedure for analyzing the results of the laddering interviews [6]. We will calculate abstractness and centrality based on an aggregate implication matrix, which represents direct and indirect linkages between attributes, consequences, and values. Abstractness indicates whether constructs are predominantly at the beginning (attributes) or ends (values) of a chain. Constructs become increasingly abstract from means to ends. As such, it is a measure of importance in the means-ends structure [6]. Centrality measures the extent to which a concept is connected to all other concepts in the matrix and is used to evaluate the importance of a concept. Additionally, we will use the amount of direct/indirect links, the number of elicited consequences and values, and the time taken for comparing treatments [33]. Furthermore, after the treatments, we will apply a self-reporting questionnaire to collect the participants’ perceptions regarding the following constructs on a 7-point Likert scale: Understandability, Learnability, Efficiency, Effectiveness, and Enjoyment [34], as well as multiple constructs from the Big Five personality test. The self-reported measurements allow us to compare the perception of LadderBot against the established computer-based laddering questionnaire. Finally, we will incorporate multiple control questions, to evaluate the influence of experience, age, or gender, amongst others, on the experiment results.

V. ROADMAP AND CONCLUSION

This paper presents our work-in-progress for building LadderBot, a requirements self-elicitation system capable of guiding a novice user through a laddering interview to generate attribute-consequence-value chains as follows: Elicitation guidance & assistance - the user is supported through randomized rephrasing of questions based on an established guideline for interviewers; Dynamic visualization - Elicited attributes, consequences, and values are visualized for the user and continuously updated throughout the interview process.

We propose an experimental study design to evaluate LadderBot against the traditional approach of pencil-and-paper laddering using a digital questionnaire. As we will use the proposed structure for the evaluation of LadderBot and its subsequent iterations, the scenario and the generated dataset might be helpful for other researchers for comparing CA-driven tool support for RE. Several comparisons of elicitation techniques have identified laddering as a very potent technique. However, only a limited amount of research describes approaches to creating tool support for laddering, especially for tool-supported self-elicitation of user requirements. A similar approach to our work-in-progress comes from Kassel and Malloy (2003), who attempt to automate requirements elicitation through combining domain knowledge, a software requirements specification (SRS) template and user needs as XML in a tool-based approach [35]. However, their focus lies on closed-ended questions, while the laddering tool proposed in our article relies on the detail introduced by open-ended questions.

Overall, we expect LadderBot to allow the elicitation of requirements from users without the need for highly qualified interviewers. Furthermore, enabling users to self-elicit requirements creates the potential to come in contact with a broad range of users, hopefully improving software development projects through detailed insights. In the spirit of “RE for everyone” [1], tool support for users enables developers to get an idea of the expectations of society and supports the end-to-end value co-creation between an outer- and an inner circle of systems development teams: between users and system engineers, analysts and developers. Additionally, with LadderBot, we also wish to show a proof-of-concept for using chatbots for RE, which may inspire the usage of the technology with elicitation techniques other than laddering in the future (e.g., 5W2H).

We are currently working on finalizing the LadderBot artifact and setting up a pre-test for the initial evaluation of the tool. Moving forward, we envision multiple adjustments to LadderBot, which will be evaluated in future studies: Enable the tool to use an interviewing technique (retrospective, ...) not randomly but based on measurements from the interview process, such as time since asking a question or based on user characteristics, e.g. cognitive styles [36]. For example, should a user diverge a specified amount from the average response time, the bot may provide additional assistance through question reformulation. Furthermore, future iterations of LadderBot will explore ways of generating content codes for the analysis of laddering interviews automatically. When dealing with a large number of self-elicitation interviews, it becomes necessary to provide requirements analysts with support in generating aggregate implication matrices and hierarchy value maps, ideally through an automated aggregation of results as well as an interactive visualization.

REFERENCES


