

Dynamic Visual Analytics for Elicitation Meetings with ELICA

Zahra Shakeri Hossein Abad*, Munib Rahman*, Abdullah Cheema*, Vincenzo Gervasi†, Didar Zowghi‡, Ken Barker*

* Department of Computer Science, University of Calgary, Calgary, Canada

{zshakeri, murahman, cheemaab, kbarker}@ucalgary.ca

†Department of Computer Science, University of Pisa, Italy, gervasi@di.unipi.it

‡School of Software, University of Technology Sydney, Australia, didar.zowghi@uts.edu.au

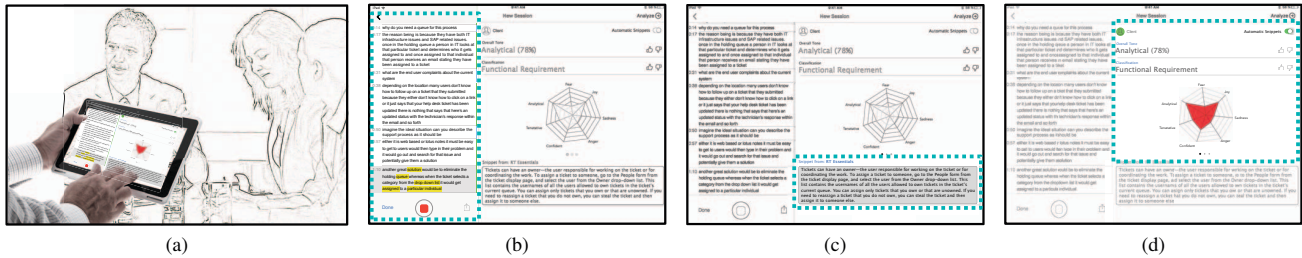


Fig. 1. (a) The application of ELICA during the collaboration with Stakeholders, (b) diarization feature included in the speech to text component, (c) extraction bar, which represents the extracted relevant information, (d) visualization and analytical component.

Abstract—Requirements elicitation can be very challenging in projects that require deep domain knowledge about the system at hand. As analysts have the full control over the elicitation process, their lack of knowledge about the system under study inhibits them from asking related questions and reduces the accuracy of requirements provided by stakeholders. We present *ELICA*, a generic interactive visual analytics tool to assist analysts during requirements elicitation process. *ELICA* uses a novel information extraction algorithm based on a combination of Weighted Finite State Transducers (WFSTs) (generative model) and SVMs (discriminative model). *ELICA* presents the extracted relevant information in an interactive GUI (including zooming, panning, and pinching) that allows analysts to explore which parts of the ongoing conversation (or specification document) match with the extracted information. In this demonstration, we show that *ELICA* is usable and effective in practice, and is able to extract the related information in real-time. We also demonstrate how carefully designed features in *ELICA* facilitate the interactive and dynamic process of information extraction.

Index Terms—Requirements elicitation, Natural language processing, Tool support, SVMs, Information extraction

I. INTRODUCTION AND MOTIVATION

Requirements elicitation is an information-intensive task, drawing on different sources of information to infer a consensus on the requirements of a system. There have been numerous attempts to improve the flow of information, knowledge, or decision in elicitation process (e.g. [1], [2]). However, two major obstacles still stand in the way of providing an effective communication between the systems analyst and clients: First, analysts and users often have different mental models and use different terminologies, filling the knowledge gap between the domain experts and the analysts during the requirements elicitation process is itself an open research problem. Second, even

if we had a perfect tool that could fully support analysts with requirements-related information, extracting and classifying the extracted information in real-time still remains challenging since this requires a dynamic modeling and analyzing the input text. For example, consider a situation in which an analyst is tasked to work on the requirements for a project from an unfamiliar domain. The misinformation during the requirements elicitation meeting reduces the accuracy of the requirements provided by stakeholders [3].

This paper describes *ELICA* (requirements ELICitation Aid), a tool for extracting and classifying requirements-relevant information during the elicitation process. Our focus is on both challenges: given an existing domain repository, how to extract requirements-relevant information in real-time. *ELICA* leverages the existing domain repositories as well as machine learning techniques to assist analysts to access relevant information about the application domain in real-time (Fig. 1a). In *ELICA*, we use an off-the-shelf API—IBM Watson [4]—to convert spoken utterances to text strings and to diarize meeting data (Fig. 1b). To extract requirements-relevant information (Fig. 1c) we design and implement a hierarchical language model to measure the *lexical association* [5] between the latest window of a conversation (or text snippet) and the existing domain repositories. To extract this information in real-time, we apply Weighted Finite State Transducers (WFSTs) [6], powerful technique which handle variable-length feature vectors and support dynamic modeling and analysis of textual data. We leveraged the closure properties of WFSTs and the well-defined and flexible algebraic operations for these tools to enhance the SVMs technique for dynamic classification of the extracted requirements-relevant information (Fig. 1d).

II. ELICA PROTOTYPE

The architecture of ELICA is shown in Fig. 2. The entire ELICA system consists of two parts, the first part is responsible for extracting and classifying requirements relevant information, including *dynamic generative model* (using WFSTs) and *real-time classifier* (using SVMs); the other provides support for recording data and interaction with the extracted information, including *speech to text*, *speech diarization*, and *visualization* components. A typical user interaction process in ELICA engages both analytical and interactive components.

Data Repository: ELICA stores the domain repository and the application data (e.g. users' information, diarized conversation, and extracted information) in MongoDB on Amazon EC2, which communicates in JSON. ELICA is able to keep the data up-to-date by running the diarization, tone analyzer, extraction, and classification components in real-time.

Extracting and Classifying requirements: Once the conversation between the client and the analyst has been initiated and transcribed, the most recent window of the conversation, along with the existing domain repository will be used as inputs to the information extraction module. After building the hierarchical language models, this module calculates the intersection between the two language models and returns the "relevant" terms which can be used to measure the lexical association between the on-going conversation and the existing repositories and fetch those parts of the domain repository that contain those relevant terms. The analyst can then interactively select among the returned text snippets those which best describe the on-going discussion. ELICA not only returns the extracted relevant snippets but also classifies these snippets to Functional/Non-Functional Requirements (NFRs) as well as sub-categories of NFRs (e.g. usability, availability, security).

Interactive Visualization: ELICA offers an interactive interface that combines intuitive visualization about speakers' intentions such as confidence level, analytical tone, and emotions with a range of visual-textual information such as diarized speech (using a different color for each speaker) and extracted relevant information (highlighting content-carrying terms). Fig. 2 outlines these components which are included in the *IBM Watson APIs* module. As the association between each recorded utterance and its relevant extracted snippet is stored in the App data storage, an analyst will be able to use this information at any time during/after the conversation and can easily access and review those parts of the conversation history (Fig. 1b) that are of interest at the moment. Facilities that allow the user to rate and analyze the results of the data analysis are also provided (Fig. 1d).

Application Scenarios: The main three potential scenarios of ELICA are: **[S1: Elicitation Meetings]:** An analyst is assigned to work on the requirements for a project from an unfamiliar domain. To support them during their elicitation meeting, ELICA selects inside each document of domain repository those textual snippets that are most relevant for the most recent part of the conversation; **[S2: Resumption/Information Recall]:** An analyst is working on a requirements

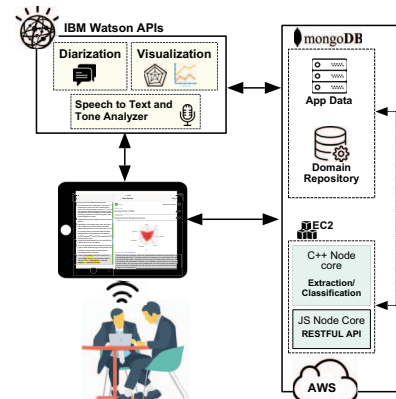


Fig. 2. The overall architecture of ELICA.

exploration task (e.g. elicitation meeting, discussion with other team members, ...) and needs to pause this task to address an incoming task. ELICA uses the recorded textual data during the exploration task as well as the existing problem domain documents to assist analysts better manage the issues of memory recall [7], [8] after resuming the elicitation task; and **[S3: Elicitation from Existing Documents]:** An analyst is assigned to an ongoing project and they need to become familiar with a possibly large amount of documentation that has already been produced. ELICA uses the existing problem domain documents to select textual snippets that are most relevant to the part of the system under investigation.

III. CONCLUSION AND FUTURE WORK

The ELICA tool was developed to achieve two main goals. The first was to propose an effective tool for extracting requirements-relevant information during the elicitation process. This tool records, analyzes, and synthesizes the existing repository documents and incoming textual data in real-time and enables easy exploration of relevant information. Second, once ELICA has extracted and classified the relevant information, it supports analysts to interact with textual and visual results to increase their engagement with the extracted information and enhance their exploration abilities. So far, both components of the ELICA tool has been evaluated with an industrial dataset, yielding promising results, and we plan to further evaluate the tool in real development environments.

REFERENCES

- [1] M. Noaen, Z. S. H. Abad, and B. H. Far, "Let's hear it from retta: A requirements elicitation tool for traffic management systems," in *2017 IEEE 25th International Requirements Engineering Conference (RE)*, 2017, pp. 450–451.
- [2] M. Gall and B. Berenbach, "Towards a Framework for Real Time Requirements Elicitation," in *1st International Workshop on MRE*, 2006.
- [3] R. Appan and G. J. Browne, "The Impact of Analyst-induced Misinformation on the Requirements Elicitation Process," *MIS quarterly*, pp. 85–106, 2012.
- [4] R. Akkiraju, "IBM Watson Tone Analyzer? new service now available," *JUL*, vol. 16, pp. 1–13, 2015.
- [5] P. Pecina, "Lexical Association Measures and Collocation Extraction," *Language Resources and Evaluation*, vol. 44, no. 1, pp. 137–158, 2010.
- [6] M. Mohri, F. Pereira, and M. Riley, "Weighted Finite-State Transducers in Speech Recognition," *Computer Speech & Language*, vol. 16, no. 1, pp. 69–88, 2002.
- [7] Z. S. H. Abad, G. Ruhe, and M. Bauer, "Task Interruptions in Requirements Engineering: Reality Versus Perceptions!" in *Requirements Engineering Conference (RE), 2017 IEEE 25th International*, IEEE, 2017, pp. 342–351.
- [8] Z. S. H. Abad, O. Karras, K. Schneider, K. Barker, and M. Bauer, "Task Interruption in Software Development Projects: What Makes Some Interruptions More Disruptive Than Others?" in *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*, ACM, 2018, pp. 122–132.