Panel on Cognitive Service Engineering

Boualem Benatallah
University of New South Wales (UNSW)
Sydney, Australia
b.benatallah@unsw.edu.au

Fabio Casati
University of Trento, Italy and
Tomsk Polytechnic University, Russia
first.last@unitn.it

ABSTRACT

Cognitive services and conversational digital assistants are emerging as the engine that powers natural interactions between humans, software services, devices and "things" - supported by advances in AI and human computations. Not surprisingly, many large and small tech companies are rushing to occupy this space by providing platforms for building cognitive services and conversational bots. Digital assistants interact in a natural way (through text or voice) with both software and humans to get information and perform actions, from checking the weather to booking a restaurants and a cab ride, managing cloud resources, answering simple scientific questions, and preparing a decaf latte using IoT enabled coffee machines. User requests or tasks are often expressed in natural language, an interaction ensues to clarify the intent and the details, and the answer is sought - or the appropriate service or device is invoked - based on the cognitive service understanding.

While the potential of this new wave of services is exciting, it also brings significant challenges: we are far away from the comfort of developing deterministic software that responds to API calls by invoking other APIs. Now we have to understand, guess, explore options, take decisions based on probabilistic models over a large set of possible intents and services, all while engaging with users. Doing so brings a large set of engineering challenges related to the development, training, tuning and evolution of such services. This panel will discuss such challenges and identify interesting opportunities for research as well as promising trends.

KEYWORDS

cognitive services, chatbots, natural language interfaces

1 PROBLEM AND SCOPE

Cognitive computing brings a paradigm shift in computing by opening up new frontiers for the design, programming, engineering and management of software and software-defined environments. Natural language is emerging as the interface of choice to support, often via chatbots, interactions among humans, software, devices and things. Delivering a satisfactory user experience in this context is very different from both traditional service invocation (where a client invokes an API with precise parameters) but also from Web search, where we have the luxury of returning a set of options to the user, something we rarely can afford in chatbots.

To a large extent, chatbots engineering, and in general cognitive service engineering, is still an art. While there are a number of techniques and tools (e.g., natural language processing, entity and intent recognition, intent-code mapping rule languages) that facilitate the basic understanding of how to map user expressions to intents and even the extraction of key entities from user expressions, chatbot developers still need to map "by hand" this inferred information to API invocations (to then deliver actual services to the user). Furthermore, activities such as bot training can be expensive and error prone, and to a large extent we are still understanding the guidelines and best practices for how to do so.

When services are conversational and require anything but simple and natural interactions with users, we are faced with even bigger challenges as we still lack the understanding of how to model and manage conversational contexts, besides very simple situations. Another source of complexity surfaces whenever the service required by the user needs invocation of multiple APIs, possibly in a certain order. This problem has been widely studied, though not really solved effectively yet, in the area of Web service composition and it is even more complex here because our understanding of users' intents is probabilistic rather than deterministic. This is particularly problematic in an environment where the kind of capabilities services can perform keeps growing, as more "things" can be controlled by IT and have an API.

To a large extent this is uncharted territory for software engineering: we are shifting from a world of deterministic services that receive precise instruction and execute them by selecting among a small and relatively static set of services, to probabilistic reasoning where we recognize user intents and identify - with short latency - how to constructively interact with the user, clarify such intents, and ultimately fulfill them by leveraging a mapping of intents to a large and evolving set of APIs, or even humans through crowdsourcing. There are significant gaps and risks in the cognitive service-enabled endeavour, with many unsolved theoretical and technical challenges stemming from fuzzy and ambiguous intent abstractions, rapid evolution of digital technologies, and growing concerns about the unintended consequences of the digital age including security, privacy and quality control concerns.

2 STRUCTURE AND OBJECTIVES

The panel will discuss and explore the main challenges in the sound and effective engineering of cognitive services, which are the untapped opportunities and promising research directions, how different computing disciplines can effectively contribute to solve these challenges. The panel will be organized as a round of statements followed by a round of discussions, with the goal of actively engaging expert panelists and attendees. At the end the moderators will summarize and share the main ideas, observations and findings from the panel.