

A Process-Oriented View of Website Mediated Functionalities

Martin Junghans

Institutes AIFB and KSRI,
Karlsruhe Institute of Technology, Germany
junghans@kit.edu

Abstract. Functionalities that are offered in form of interactive websites are omnipresent. It requires a substantial manual and tedious effort to use the functionalities within increasingly sophisticated use cases. Structured descriptions enable the development of methods that support users in dealing with everyday tasks. We present how a process-oriented view of the Web is gained and helps end users, for instance, in finding information scattered across multiple websites.

The Web is not only an interlinked collection of documents. Functionalities provided in form of interactive Web pages and Web applications are offered and consumed on a regular basis by most of the end users connected to the Internet. Web pages serve as a front-end to access services and information of the Deep Web. In contrast to the Semantic Web with its aim to allow providers to annotate their services such that automatic discovery and composition are enabled, website mediated functionalities target primarily at human use.

By observing end users and their **browsing behavior**, the Web is perceived as a pool of functionalities solving simple tasks. Users select functionalities and use them in a certain sequence in order to achieve a goal. For example, in order to arrange a travel or to buy a high rated and cheap product, several individual functionalities are composed in the users' minds. Logic dependencies between inputs and outputs have to be managed manually and often the same inputs are provided at multiple websites repeatedly.

A large portion of the Web can be seen as a set of distributed and networked processes. They can provide access to information and cause effects during their execution. Also, they can require multiple interactions (such as form submissions and link selections) with the user. Unfortunately, these processes are currently not explicit. Users have to compose them every time again on their own. However, a **process-oriented view** on the Web requires an explicit and structured description of the functionalities and processes. In our approach, we model end user browsing processes that describe how users have to interact with which website at which time in order to consume functionalities and reach the desired outcomes. The descriptions allow to support users in combining, sharing, and reusing the processes, which capture previous efforts to achieve a goal [1].

We learned from the lack of public Semantic Web Services that we cannot rely on providers to create semantic annotations and to adopt a top-down formal semantics of Web services. So, we let users capture their browsing processes by existing **Web automation scripting** tools, which monitor, describe, and partially automate the process

execution. In a bottom-up approach [2], semantic annotation can be added by users when needed, e.g., to describe elements of interactions and Web pages. Semantic annotations of websites can also be derived from the scripts [3].

Information Search Based on a Process-Oriented View For many practical purposes, end users need information that is scattered across multiple websites. Consider for example an end user who is interested in knowing the names of the chairs of a particular track at the previous WWW conferences. As of today, Web search engines do not deliver satisfactory results for queries similar to “track chairs of all WWW conferences”. In order to obtain the required information the end user has to pose multiple queries to a search engine, browse through the hits, and aggregate the required information fragments outside of the found Web pages.

End users need help in selecting the pages that are relevant for obtaining the scattered information. Such a help must contain at least the set of the pages that the end user should visit, and support for invoking all the pages of the set easily. More advanced help could contain the complete end user browsing process including support for data flow between the user and the pages as well as among the pages, and control flow if there are data dependencies among inputs and outputs of Web pages in the set. We aim at providing the end users with a list of browsing processes that are relevant for a given information need instead of a list of links to Web pages. Each browsing process in the list of hits will lead the end user to the required information.

In order to search for existing browsing processes, e.g., from a repository shared with other users, we developed efficient **discovery** techniques. We proposed an offline classification of processes [4], which is based on formally defined classes, and the use of offline and online index structures [5] to efficiently locate desired browsing processes from large repositories. The discovery allows to reuse existing browsing efforts. The composition of browsing processes promises to create solutions to the information need that have not been executed before.

Acknowledgments

This paper presents results of joint efforts of a group of colleagues including Sudhir Agarwal (Stanford University), Charles J. Petrie (Stanford University), and the author.

References

1. Agarwal, S., Junghans, M.: Swapping out coordination of web processes to the web browser. In Brogi, A., Pautasso, C., Papadopoulos, G., eds.: ECOWS, IEEE (2010) 115–122
2. Agarwal, S., Petrie, C.J.: An alternative to the top-down semantic web of services. IEEE Internet Computing **16**(5) (2012) 94–97
3. Agarwal, S., Junghans, M.: Towards simulation-based similarity of end user browsing processes. In Daniel, F., Dolog, P., Li, Q., eds.: ICWE, Springer (2013) 216–223
4. Junghans, M., Agarwal, S., Studer, R.: Behavior classes for specification and search of complex services and processes. In Goble, C., Chen, P., Zhang, J., eds.: ICWS, IEEE (2012) 343–350
5. Junghans, M., Agarwal, S.: Efficient search for web browsing recipes. In: ICWS, IEEE (2013) (to appear)