

George Zheng • Athman Bouguettaya

Web Service Mining

Application to Discoveries of Biological
Pathways

Foreword by Boualem Benatallah

 Springer

George Zheng
Science Applications International Corporation
Systems & Technology Solutions
7990 Science Applications Ct.
Vienna, VA 22182
USA
GEORGE.ZHENG@saic.com

Athman Bouguettaya
CSIRO ICT Center
Computer Science & Information
Technology
Australian National University
Canberra, ACT 2601
Australia
athman.bouguettaya@csiro.au

ISBN 978-1-4419-6538-7 e-ISBN 978-1-4419-6539-4
DOI 10.1007/978-1-4419-6539-4
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010929768

© Springer Science+Business Media, LLC 2010

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To all my family and friends.

George Zheng

To Malika.

Athman Bouguettaya

Foreword

The new computing environment enabled by advances in service oriented architectures, mashups, and cloud computing will consist of service spaces comprising data, applications, infrastructure resources distributed over the Web. This environment embraces a holistic paradigm in which users, services, and resources establish on-demand interactions, possibly in real-time, to realise useful experiences. Such interactions obtain relevant services that are targeted to the time and place of the user requesting the service and to the device used to access it. The benefit of such environment originates from the added value generated by the possible interactions in a large scale rather than by the capabilities of its individual components separately. This offers tremendous automation opportunities in a variety of application domains including execution of forecasting, office tasks, travel support, intelligent information gathering and analysis, environment monitoring, healthcare, e-business, community based systems, e-science and e-government.

A key feature of this environment is the ability to dynamically compose services to realise user tasks. While recent advances in service discovery, composition and Semantic Web technologies contribute necessary first steps to facilitate this task, the benefits of composition are still limited to take advantages of large-scale ubiquitous environments. The main stream composition techniques and technologies rely on human understanding and manual programming to compose and aggregate services. Recent advances improve composition by leveraging search technologies and flow-based composition languages as in mashups and process-centric service composition.

The authors provide a very informative discussion on the state of the art in service mining. The motivation is simple: existing service composition techniques are only appropriate when a service composer knows how to search effectively for required services and have the programming skills to perform composition using flow based languages or general purpose programming language. The approach proposed by the authors rely on discovering useful and interesting services that can be used in situations where users lack precise knowledge to perform targeted search of services to be composed. The realization of the service mining approach poses a difficult challenge to address the critical issue of effort-less and flexible exploration

and manipulation of services in autonomous and highly evolving service spaces. The authors provide the fundamental concepts behind Web service mining and illustrate the proposed concepts and techniques in application related to biological pathway discovery. The book is timely, provides a thorough scientific investigation and also has practical relevance in the general area of composition. The book will be of particular interest to researchers and professionals wishing to learn about relevant concepts and techniques in service composition and mining.

Sydney, Australia

Boualem Benatallah

December 2009

Preface

The ubiquity of the Web is enabling innovations that were thought to be beyond reach just two decades ago. The Web has now in effect become the dominant medium for all human and economic activities. The advent of Web services almost a decade ago has accelerated this trend and included such activities as e-government, e-commerce, and e-science applications. The ultimate goal of this enabling technology is the use of Web services as independent components that are automatically (i.e., without human intervention) formed and that could dissolve or persist post demand-completion.

The fast increasing number of Web services is transforming the Web from a data-oriented repository to a service-oriented repository. Web services are anticipated to form the underlying technology that will realize the envisioned “sea of services”. Web services have originally been driven by standardization bodies, thus eliciting a wide acceptance in businesses and governments. The Service-Oriented Architecture (SOA) was conceived as the IT industry response to leveraging the Web as the center of all activities. This was achieved by adapting and evolving such technologies as CORBA to be Web congruous. As aforementioned, the development of Web services has so far mostly been the result of standardization bodies usually operating on a consensus basis and driven by market considerations. In this context, innovation and long-term market effects have not been the primary concerns. The standardization process has so far been very fragmented, leading to competing and potentially incompatible Web service infrastructures that lack a sound foundational framework. To maximize the benefits of this new technology, there is a need to provide a rigorous methodology for specifying, selecting, trusting, optimizing, composing, and mining Web services.

In this book, we focus on service mining. We describe a novel foundational framework that lays out a theoretical underpinning for the emerging field of service mining. We describe a disciplined and systematic framework for the efficient mining of Web services functionalities using non-functional properties, called Quality of Web Service (QoWS) parameters. The key components of this approach revolve around a novel service model that provides a formal abstraction of Web service modeling and organization. We draw inspiration from chemical processes for ser-

vice organization methodologies, and drug discovery for service mining techniques. We use this inspiration as the target application as a proof-of-concept of the proposed framework and algorithms.

We distinguish between Web service composition and Web service mining in terms of the stated *a priori* goals or lack thereof, respectively. For instance, service composition has traditionally taken a top-down approach. The top-down approach requires a user to provide a goal containing specific search criteria defining the exact service functionality the user expects. Often, the more specific the query and search criteria are, the smaller the search space and more relevant the composition results will be. The specificity of the search criteria would reflect the interest and often knowledge of the service composer about the potential composability of existing Web services. Since the composer is typically only aware of and consequently interested in some specific types of compositions, the scope of such a search is usually very narrow. The top-down approach thus works well only if the service composer clearly knows what to look for and the component Web services needed to compose such services are available. Another view, as taken by service mining, approaches service composition from the bottom-up. It aims at exploring the full potential of the service space without any *a priori* knowledge of what exactly is in it. Instead of starting the search with a specific goal, a service engineer may be interested in discovering interesting and useful service compositions as a result of the search process. For performance reasons, a general goal may be provided at the beginning to scope down the initial search space to a reasonable size. The interesting outcome of service mining is the ability to find useful and unexpected compositions. Thus, unlike the search process in the top-down approach that is strictly driven by the search criteria, the search process in the bottom-up approach is *serendipitous* in nature leading potentially to great innovations.

Virginia, USA
Canberra, Australia

George Zheng
Athman Bouguettaya

December 2009

Acknowledgments

I thank my wife Lin for her love and patience, and my daughter Sydney for the joy and hope she constantly brings. I thank my good friend, Traxon Rachell, for his continuing support, advice and inspiration; Folashade (Sharday) Adeyosoye, a good friend and former colleague, for his help in setting up my testbed environment for the application of this work; and Thien Le, a good friend and colleague, for his tips and interesting conversations. I thank SAIC for providing me with the environment and financial support for completing this work. Finally, I thank my mother's love and dedication to me and to my father, whose continuing battle against terrible diseases in the past few years has given me the courage to get involved with bioinformatics – an unfamiliar research field, and to make it a fertile ground for applying this work.

George Zheng

I would like to acknowledge the support of my family during the preparation of this book: my wife Malika, my children: Zakaria, Ayoub, and Mohamed-Islam. I would also like to thank my employer CSIRO (Australia) for providing me the environment to successfully finish this work.

Athman Bouguettaya

Contents

1	Introduction	1
1.1	Defining Web Service Mining	5
1.2	Major Challenges	6
1.3	Our Contributions	8
1.4	Book Organization	9
2	An Ontological Perspective of Web Services	11
3	Hints from A Molecular Analogy	15
3.1	A Simple Analogy	16
3.2	A More Complex Analogy	17
3.3	Molecular and Operational Recognition	18
3.4	Usefulness and Side Effects	22
3.5	Effects of External Conditions	22
3.6	A Eukaryotic Cell	24
3.7	A Process Perspective	25
3.8	Drug Discovery Process	26
3.9	Web Service Modeling of Biological Processes	27
4	Web Service Mining Framework	31
4.1	Pre-screening Planning	33
4.1.1	Scope Specification	33
4.1.2	Search Space Determination	34
4.2	Screening	35
4.2.1	Filtering	35
4.2.2	Static Verification	46
4.2.3	Linking	50
4.2.4	Effects of Operation Diversity and Complementarity	53
4.3	Post-screening Analysis and Evaluation	55
4.3.1	Objective Evaluation	55
4.3.2	Interactive Hypothesis Formulation	65

- 4.3.3 Verification and Predictive Analysis via Runtime Simulation 72
- 4.3.4 Subjective Evaluation 74
- 5 Application: Biological Pathway Discovery 77**
 - 5.1 Free Text Based Description 78
 - 5.2 Computer Models 78
 - 5.3 A Service Oriented Modeling and Deployment Approach 79
 - 5.3.1 Conceptual Service Oriented Models 80
 - 5.3.2 WSDL/WSML Description and Deployment 82
 - 5.3.3 WSDL Service Modeling of Biological Processes 83
 - 5.3.4 WSML Service Wrapping of WSDL Service 83
 - 5.3.5 Lowering and Lifting Adapters 87
 - 5.4 Experiment 88
- 6 Related Work 111**
 - 6.1 Web Service Composition 111
 - 6.1.1 Business Process-Oriented Web Service Composition 111
 - 6.1.2 Semantics-based Web Service Composition 115
 - 6.1.3 Contrasting Web Service Mining and traditional Web Service Composition Approaches 116
 - 6.2 Contrasting Web Service Mining and Data Mining 117
 - 6.3 Log-based Interaction and Process Mining 117
 - 6.4 Natural Computing 117
- 7 Conclusions 121**
- References 125**
- Glossary 131**
- Index 135**