Information Systems Department
Faculty of Computer & Information Sciences
Ain Shams University

Semantic-based Cloud Services Discovery and Recommendation

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of PhD in Computer and Information Sciences
to
Department of Information Systems
Faculty of Computer and Information Sciences
Ain Shams University

by
Yasmine Mohamed Abd ElMonem Afify
Assistant Lecturer
Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

Under Supervision of

Prof. Dr. Mohamed Fahmy Tolba
Scientific Computing Department
Faculty of Computer and Information Sciences
Ain Shams University

Prof. Dr. Nagwa Lotfy Badr
Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

Assoc. Prof. Ibrahim Fathy Moawad
Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

2016

Acknowledgement

Throughout my PhD study, I received support from amazing people whom I wish to acknowledge. First and foremost, I would like to express my sincere gratitude to my supervisor Professor Mohamed Fahmy Tolba for the continuous support and guidance throughout my study. He has built and directed an environment that granted me an opportunity to learn and practice my research skills.

I am very grateful to my supervisor Professor Nagwa Lotfy Badr for her encouragement, remarkable suggestions and insightful comments. Her expertise and experiences influenced and helped me doing research.

I would like to express my genuine appreciation to my supervisor Dr. Ibrahim Fathy Moawad whose valuable advice, constructive discussions and precious comments have transferred the long journey of PhD to a great and fruitful experience.

I would like to give deep thanks to my parents, sister and brother for their endless help, support and love. My heartfelt thanks to my husband for his encouragement, impressive patience and understanding. They have always been there for me whenever I needed them.

This thesis is dedicated to my son, Adham.

I would like to give him an advice to press on. Persistence is much of what separates the successful from the pure wishful thinkers. Have faith in yourself and always pursue your dreams.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AHP	Analytic Hierarchy Process
API	Application Programming Interface
AS	Autonomous System
AW	Average Width
BF	Business Function
BFO	Business Function Ontology
BPM	Business Process Management
CB	Content-based filtering
CC	Cloud Computing
CF	Collaborative Filtering
CRM	Customer Relationship Management
CS	Cloud Service
CSV	Cluster Signature Vector
DL	Description Logics
DM	Document Management
ECBR	Extended Case Based Reasoning
ERP	Enterprise Resource Planning
ESD	Enriched Service Description
GFS	Google File System
HAC	Hierarchical Agglomerative Clustering
IaaS	Infrastructure-as-a-Service
IC	Information Content
ICT	Information and Communication Technologies
IoS	Internet of Services
IR	Information Retrieval
LCS	Lowest Common Subsumer
MAE	Mean Absolute Error
MCDM	Multi-Criteria Decision Making
MD	Maximum Depth
MW	Maximum Width
NAICS	North American Industry Classification System
OWL	Web Ontology Language
PaaS	Platform-as-a-Service
QE	Query Expansion

QL	Query language
QoS	Quality of Service
RDF	Resource Definition Framework
RF	Relevance Feedback
RFS	Request for Service
RMSE	Root of Mean Square Error
RS	Recommender System
RT	Response Time
SaaS	Software-as-a-Service
SAW	Simple Additive Weighting
SLA	Service Level Agreement
SLO	Service Level Objectives
TF/IDF	Term Frequency/Inverse Document Frequency
TopK	Neighborhood Size
TP	Throughput
UDDI	Universal Description, Discovery and Integration
UI	User Interface
UML	Unified Modeling Language
UNSPSC	United Nations Standard Products and Services Code
URL	Uniform Resource Location
USDL	Unified Service Description Language
VA	Virtual Appliances
VIP	Virtual IP
VM	Virtual Machine
VU	Virtual Unit
W3C	World Wide Web Consortium
WS	Web Service
WSDL	Web Service Description Language
XaaS	Everything-as-a-Service

Abstract

Substantial number of vendors are offering their applications as Cloud Services (CS) to leverage the benefits of Cloud Computing (CC). Businesses and individuals have to select the CS that suits their requirements from a large pool of services. Consequently, there is an increasing demand for enhancing such a time-consuming and error-prone task.

This thesis proposes CSRDS, a semantic-based system that allows the registration, discovery and selection of cloud services. The main building blocks of the proposed system include: Hybrid matchmaking of CSs, metadata semantic similarity model, semantic-based business-oriented request-service matchmaking algorithm and CS selection based on non-functional requirements.

Despite its efficiency, the proposed CSRDS system provides the same set of services to all users, with no consideration of users interest or history. Motivated by the fact that the cloud users appreciate customized recommendations, we propose CSRec, a personalized reputation-based QoS-aware recommender system for CSs. Its main building blocks include: Feedback-based CS evaluation, hybrid collaborative filtering approach for personalized CS recommendation, new users similarity measure, dynamically built user profile, reputation calculation for CSs and 3-D justification for the recommended CSs.

The key enabler to the realization of the proposed work is the proposed CS ontology, a comprehensive full-fledged CS domain ontology with querying capabilities. It serves as a knowledge base for the proposed systems.

Extensive experimental evaluation is conducted in order to assess our proposed work. In order to carry out reliable evaluation, three real-world service datasets are used. The evaluation metrics include matchmaking, prediction and set recommendation metrics. The proposed work is compared to benchmark approaches. Case studies demonstrate that the proposed system is feasible, effective and relevant. Experimental results show that the proposed system has superior performance against the traditional approaches in real use-cases. The performance enhancement is investigated under different parameter settings.

Besides, the evaluation of the introduced CS ontology consists of three processes. First, its coverage, consistency and expressiveness were verified via semantic reasoner and semantic modeling. Second, its feasibility was confirmed via prototype implementation while its practicability was validated via case study. Finally, the ontology quality evaluation metrics prove that CS ontology captures a comprehensive, consistent, rich, accurate and relevant representation of the CS domain and that its concepts are interpretable.

List of Publications

1. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "A Semantic-based Software-as-a-Service (SaaS) Discovery and Selection System" Proceedings of the 8th IEEE International Conference on Computer Engineering & Systems (ICCES'2013), pp. 57-63, November, 2013. DOI: 10.1109/ICCES.2013.6707171
2. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "Cloud Services Discovery and Selection: Survey and New Semantic-Based System" A. E. Hassanien et al. (eds.), Bio-inspiring Cyber Security and Cloud Services: Trends and Innovations, Intelligent Systems Reference Library 70, Springer-Verlag Berlin Heidelberg, pp. 449-477, 2014. DOI: 10.1007/978-3-662-43616-5_17 (SJR: 0.19)
3. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "Concept recommendation system for cloud services advertisement" In International Conference of Advanced Machine Learning Technologies and Applications, Communications in Computer and Information Science (pp. 57-66), 2014. Springer International Publishing. DOI: 10.1007/978-3-319-13461-1_7 (SJR: 0.15)
4. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "Handbook of Research on Cloud Services Publication and Discovery" Accepted in Handbook of Research on Machine Learning Innovations and Trends, IGI Global, Feb 2016.
5. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "A Personalized Reputation-Aware Approach for Cloud SaaS Recommender Systems". Concurrency and Computation: Practice and Experience Journal, Wiley, April 2016. DOI: 10.1002/cpe.3877 (IF: 0.94)
6. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "A Comprehensive Business Domain Ontology for Cloud Services " Submitted to Applied Ontology Journal, IO Press, April 2016. (IF:0.6)
7. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "Ontology-based SaaS Catalogue for Cloud Services Publication and Discovery" Accepted in Asian Journal of Information Technology, May 2016. (SJR: 0.18)
8. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba " Enhanced Similarity Measure for Personalized Cloud Services Recommendation". Submitted to Concurrency and Computation: Practice and Experience Journal, Wiley, May 2016. (IF: 0.94)
9. Yasmine M. Afify, Ibrahim F. Moawad, Nagwa Badr and M. F. Tolba "An Enhanced Distance-based Similarity Measure for User-based Recommendations" Accepted in the 2nd International Conference on Advanced Intelligent Systems and Informatics (AISI16), October 2016. (SJR: 0.15)

1. Introduction

1.1 Research Motivation

Cloud Computing (CC) has become a new paradigm for the provision of computing Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). Huge numbers of Cloud Services (CS) have emerged and their adoption is accelerating due to interesting benefits such as cost savings, ease of use, operational efficiency and flexible business capabilities.

For businesses and individuals to benefit from CC, they have to select the CS that suits their requirements from a large pool of available services. Joining the into-cloud move guarantees major economic benefits to the organization/user if risks can be minimized.

1.2 Problem Statement

Cloud Service identification and discovery remains a hard problem due to different service descriptions, non-standardized naming conventions and diverse features of CSs. Users head for search engines and CS online directories in order to discover a CS that satisfies their requirements. However, the former returns many inaccurate results, which is a waste of time and the latter suffers from major limitations in their search capabilities and size. As a result, efficient CS discovery represents the first challenge to the widespread CS adoption.

In addition, CSs may provide the same functionality but differ in their characteristics or the Quality of Service (QoS) attributes they offer. Users find it tedious and time-consuming task to evaluate and compare the available cloud offerings from the cloud providers' websites and documentations. Consequently, informed CS selection represents the second challenge to the widespread CS adoption.

The explosive number of CSs has led to widespread use of service Recommender Systems (RSs). Therefore, evaluation and personalized CS recommendation represent the third challenge to the widespread CS adoption.

A strong key enabler to CS discovery and recommendation is the use of ontology in the CS domain. However, no broad cloud ontology for this purpose has dominated yet. Therefore, a comprehensive CS ontology represents the fourth challenge to the widespread CS adoption.

1.3 Research Objective

Driven by the aforementioned challenges, our research objective is to contribute to the efficient embracing of the CSs via addressing the challenges that hinder their widespread adoption. The roadmap to achieving this objective is characterized into several directions. First, enable higher competence of the CS life cycle activities, namely: Registration, discovery and selection/recommendation. Second, eliminating the conceptual and terminological confusion among the CSs. Third, provide cloud users with a unique experience that satisfies their needs. Fourth, support the into-cloud move by minimizing the risks involved in employing the CSs into business. Fifth, maximize the CSs reachability and subsequently the cloud provider profit. Our research key contributions are summarized in the following sub-section while presented in detail throughout the thesis.

1.4 Key Contributions

During the realization of our research objectives, we have achieved significant research contributions, which are shortly outlined in the following subsections.

1.4.1 Cloud Service Registration, Discovery and Selection System

A semantic-based CS registration, discovery and selection system (CSRDS) is proposed. CSRDS employs semantic approaches in order to standardize the advertisement process (uniform representation for CSs), to serve as a semantic-based repository for service offers and to assist the cloud user in finding a CS that satisfies his functional and non-functional requirements. Uniquely, the business perspective of CSs is utilized in order to, firstly, eliminate the lack of standardization problem, secondly, provide proficient search capabilities based on concrete service functionalities and thirdly, utilize both the CS characteristics and QoS metrics in the selection process.

A hybrid services matchmaking algorithm is proposed to cluster the CSs into functional groups, which integrates semantic-based services metadata and ontology-based hierarchical matching. We also propose an enhanced semantic similarity model for services metadata matchmaking. Moreover, we introduce a semantic-based business-oriented request-service matchmaking algorithm that maps keyword-based user requests to real CS offers. Finally, we address a major deficiency in existing works on CS discovery, which cannot help users refine their requirements when there is no feasible solution. Uniquely, we address the discovery partial matching case by introducing an ontology-based query expansion approach that increases the chance to find appropriate services to the user requirements.