
# A similarity-based framework for service repository integration

Fedelucio Narducci, Marco Comerio, Carlo Batini, Marco Castelli

PII: S0169-023X(16)30131-8
DOI: doi: 10.1016/j.datak.2016.08.001
Reference: DATAK 1571

To appear in: *Data & Knowledge Engineering*

Received date: 8 April 2014
Revised date: 29 July 2016
Accepted date: 4 August 2016

Please cite this article as: Fedelucio Narducci, Marco Comerio, Carlo Batini, Marco Castelli, A similarity-based framework for service repository integration, *Data & Knowledge Engineering* (2016), doi: 10.1016/j.datak.2016.08.001

Available online at www.sciencedirect.com

ScienceDirect

Data & Knowledge Engineering 00 (2016) 1–20

Journal Logo

# A Similarity-based Framework for Service Repository Integration

Fedelucio Narducci[a], Marco Comerio[b], Carlo Batini[b], Marco Castelli[c]

[a]*Department of Computer Science*
*University of Bari Aldo Moro*
[b]*Department of Informatics, Systems and Communication (DISCo)*
*University of Milano-Bicocca*
[c]*BTO Research*

**Abstract**

Nowadays, repositories of services are becoming increasingly useful in the management of many public and private service provider organizations. In order to make a repository an integrated representation of all services delivered in an organization, a unified representation is desirable. Since several repositories of services, each potentially characterized by heterogeneous and conflicting representations, may coexist in the same organization or in cooperating organizations, the need for service repository integration techniques is emerging. In this paper, we investigate the problem of integrating heterogeneous service repositories. We first provide a conceptual model for describing services and semantic relationships among them. Then, we define a multi-level similarity function that is able to discover similarities between services belonging to different repositories, and to suggest candidate relationships among services. The proposed function combines a simple keyword-based matching with a more complex semantic matching that exploits the Explicit Semantic Analysis technique for generating a representation of services based on Wikipedia concepts. These combined techniques are implemented in the SCAn (Service Correspondence Analyzer) framework that supports the human expert during the repository integration process. The framework has been evaluated in a real-life scenario and the results demonstrate the effectiveness of the proposed approach.



*Keywords:* service science, service repository, repository integration, similarity function, semantic service representations, methodologies and tools, knowledge management applications.

## 1. Introduction

The service development pipeline generally includes five stages: exploration, planning, design, production, and management. A service repository helps the standardization, the service discovery and composition, the management of stakeholder expectations, the cost transparency, and the pricing mechanisms [1]. Notwithstanding the multidisciplinary efforts carried out in service science research [2], the design and planning of services in digital service ecosystems [3] still sees a focus on the technological perspective as the prevailing one. Nevertheless, the situation is evolving and it is worth noting that some of the most recent methodologies for the so-called service-oriented software engineering lifecycle compared in [4] consider two levels of services: *abstract* (or conceptual) and *concrete* (or physical) services. Basically, abstract services provide the conceptual template for the manual, semi-automated, or else

*Email addresses:* {fedelucio.narducci@uniba.it} (Fedelucio Narducci), {comerio@disco.unimib.it} (Marco Comerio), {batini@unimib.it} (Carlo Batini), {marco.castelli@btoresearch.com} (Marco Castelli)