

A Graphical User Interface for SILK Data Link Discovery Framework

Rina Singh¹, Jan Hidders², Feng Xia¹, Jialiang Kang¹

¹School of Software, Dalian University of Technology, China

²WIS Group, Delft University of Technology, The Netherlands

rinasinghp@gmail.com; a.j.h.hidders@tudelft.nl; f.xia@acm.org; jialiangkang@126.com

ABSTRACT

In the field of linked data, interlinking previously unlinked datasets that are available on the linked open cloud is still a big challenge. Silk is one of the tools that allow one to do interlinking between data items within different linked data sources. The main goal of our work is to simplify the process of specifying linking conditions using Silk. Specifying the correct conditions in the Silk-LSL is a complex task where a helpful interface can make a large difference. In this work, we propose Silk Magic as a useful tool that is capable of guiding the users through the process of specifying linking conditions during the creation of the Silk LSL program and hence simplifies the task of writing Silk-LSL programs. The tool for example allows to display the conditions as an interactive tree, and offers suggestions for class selection conditions and step expressions in path expressions.

Keywords

Linked Data, Link Specification Language, Query Builder, Path Expressions, Similarity Function

1. INTRODUCTION

The Semantic Web was invented by Tim Berners-Lee with the aim of constructing a machine-understandable web of data. To achieve this task, it is necessary to have huge amounts of data available on the web in a standard, readable, interconnected and manageable format [1]. Very few of the datasets published on the Linked Open Data cloud are interlinked with each other. Various tools and techniques have been developed for making interlinking between linked datasets. Silk-A Link Discovery Framework is one of the tools which allow one to do interlinking between data items within different Linked Data sources through the declarative language called Silk-LSL. The Silk Linking Specification Language (Silk-LSL) allows the user to write a script for specifying conditions that need to be met in order for resources to be interlinked [2]. The Silk-LSL file can be either written by manually or generated with the help of tools. The manual approach requires skilled users and hard work for manually discovering the classes and classes' properties of the datasets that need to be interlinked. The manual approach is too labor intensive, time consuming and works fine if there are small numbers of datasets but does not scale to larger datasets.

To simplify the process of specifying linking conditions during the creation of the Silk-LSL program, we create an interactive user friendly Graphical User Interface (GUI) for SILK called

SILK Magic tool that allows writing Silk LSL program for linking and discovering relationships between data items within different Linked Data sources in an intuitive way. Furthermore, the Silk Magic tool provides several features for editing linking conditions. These features provide an interactive tree view for specifying the linking conditions that are required in order to generate the LSL file. The main goal of our tool is to make it easy and as simple as possible for the users to specify the classes, its properties, retrieving data from datasets, specifying path expression that leads to generation of the final Silk-LSL program for interlinking datasets.

Silk version 3.4 has been launched on July 2011 with an interface called Silk Workbench to specify the linking conditions in order to perform the interlinking task between different datasets. The main goal of the Silk Workbench is to provide users an interface through which they can specify the linking conditions in order to perform the interlinking task between different datasets. The Silk Workbench [3] provides the three components: Workspace Browser, Linking Rule editor and Evaluation editor. In contrast, our SILK Magic tool will considerably simplify the process and be easier to use.

In the following, we will give an overview of the Silk Magic tool, and present some preliminary results on its evaluation.

2. SILK MAGIC TOOL

Silk Magic is a tool that guides users through the process of creating a Silk LSL file for interlinking data items between or within the same datasets. The tool is designed to be as intuitive and easy to use as possible. The process of specifying linking conditions that is supported by the application is organized in an explicitly structured way, which helps users to use it in a productive way without having to be an expert with the knowledge required for it. The main features of Silk Magic are:

- It allows users to choose a particular class of datasets from a list of classes
- It provides a query builder editor which allows users to make SPARQL queries
- It provides a tree-like structure for editing the similarity function. This editor helps to specify possible link conditions that need to be applied in order to interlink different datasets.
- It helps users with making path expressions in an easy way.

The screen shot in Figure 1 shows the main interface of the Silk Magic tool through which users can proceed for specifying linking conditions. Users can choose a dataset from the list provided and select a specific class of that dataset which they want to match. The Source/Target restriction button allows user to build the query. The similarity function button allows specifying the linking condition and making the path expression. After specifying all the condition and once all the conditions are met, tool will allow generating Silk LSL file.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym '13, August 05 - 07 2013, Hong Kong, China

Copyright 2013 ACM 978-1-4503-1852-5/13/08 ...\$15.00.

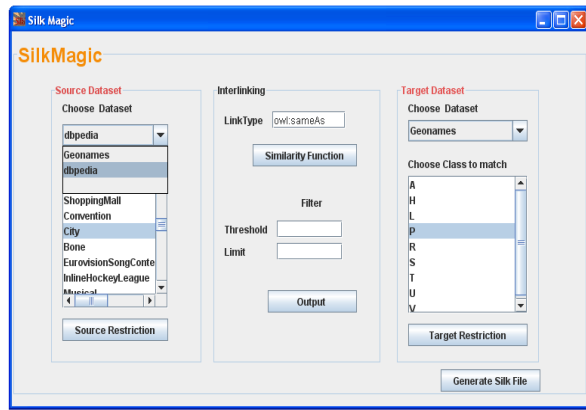


Figure 1: Silk Magic tool.

The general process of creating a new Link Specification file using Silk Magic application is shown in Figure 2.

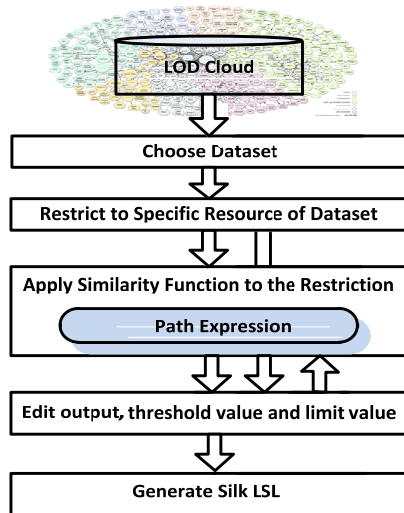


Figure 2: Steps for creating Silk-LSL file using Silk Magic.

The Silk Magic approach consists of five main steps to generate a Silk LSL file for discovering relationships between data items within different datasets. Each step is discussed below:

Choosing the Dataset: There have been lots of datasets published in the Linked Open Data (LOD) cloud; choose the datasets which need to be interlinked from the LOD.

Restriction to specific resource of the dataset: After selecting the particular datasets, users have to specify the particular resources (data items) of the dataset that they want to link with other datasets.

Apply Similarity Function to the Restriction: Link conditions need to be applied to the properties of the class in order for them to be interlinked. Properties of the selected class can be extracted by executing SPARQL on the datasets. Various condition operators are provided by Silk that can be used on these extracted properties for interlinking such as: aggregation, similarity metrics, path expressions and transformation.

Output, Link Type and Filter: This phase allows users to specify the output file where users want to store the generated links. The output file can be either in Ntriple or OAEI alignment format. Link Type allows users to specify which type of

relationship they want to establish, e.g. owl:sameAs, and Filter is used to only generate those links that have a similarity score above a threshold value.

Generate Silk-LSL: After specifying each of the previous phases, users can now generate a Silk-LSL file. This generated file by Silk Magic tool then feed to Silk tool which generates all the similar entity specify in the program.

3. EVALUATION

The current version of the Silk Magic application successfully generates the LSL file that can be used to interlink cities between datasets DBpedia and Geonames. To do interlinking between other datasets such as drugbank, dailymed, sider and so on, the corresponding datasets need to be queried through their SPARQL endpoints or their ontology files need to be loaded into Sesame. Afterwards, users can generate LSL files that do interlinking between those datasets too.

First a Silk LSL file was written manually that matches all the cities of the Netherlands between datasets DBpedia and Geonames. While writing the file manually, it is required to know the details about the datasets ontology to get familiar with its classes and properties of the classes' in order to query and build the path expression. Users need to know SPARQL knowledge in more detail to query particular datasets. This process is very time-consuming which require hard work and makes the linking process slow. However, users want to get their job done without knowing much; in other words, we can say they want their job to be done in an easy way and fast.

With the help of Silk Magic tool users can choose the dataset from the list, once the users have selected particular dataset, all of its classes and classes properties are extracted automatically by the tool which saves time. These extracted classes are used for making triples while building query to restrict to specific class of the dataset with the help of query builder editor. Query Builder editor of Silk Magic tool simplify the process of making not only simple queries but also supports nested queries. It is found that with the help of this feature users can successfully generate any kind of query in correct format. The extracted properties of the classes are used for making path expression for selecting sets of resources or literal in the RDF graph that is needed for discovering semantic relationships. Path expression editor of the tool provides a list of classes, classes' properties, properties of properties in order to make the path expression. The editor support users to make any level of path expression.

ACKNOWLEDGMENTS

Part of this work was done when the first author was a Master student at Delft University of Technology. This work was partially supported by the Innovation Fund of School of Software, Dalian University of Technology.

REFERENCES

- [1] Campbell, L. M., & MacNeill, S.: The Semantic Web, linked and open data. *Briefing paper, JISC CETIS, 2010.*
- [2] Volz, J., Bizer, C., Gaedke, M., & Kobilarov, G.: Silk—a link discovery framework for the web of data. In *Proceedings of the 2nd Linked Data on the Web Workshop (LDOW2009)*, Madrid, Spain, 2009.
- [3] R. Isele, A. Jentzsch, C. Bizer, and J. Volz. Silk Workbench. 2012. https://www.assembla.com/spaces/silk/wiki/Silk_Workbench.