Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom
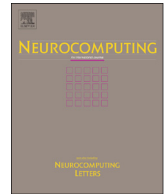
Original Software Publication

SoNeR: Social Network Ranker

Gajo Petrović, Hamido Fujita *

Iwate Prefectural University Takizawa, Iwate, Japan

ARTICLE INFO

ABSTRACT

Social Network Ranker (SoNeR) is software that retrieves, from the Semantic Web, documents describing people (users) and ranks them based on their popularity. It provides supportive tools for obtaining the related FOAF/RDF documents, parsing them, detecting identity synonyms and ranking people. The software is developed and implemented in the Java environment; and consists of the following modules: (1) semantic web crawler (collects FOAF/RDF documents), (2) document parser (extracts relevant information from FOAF/RDF documents), (3) synonym identity tagger (used to manually tag synonym pairs), (4) identity synonym merge and detection (5) ranking algorithms. While all these modules can be used separately, they are also unified in one Desktop GUI application.

---

* Corresponding author.
  E-mail addresses: gajopetrovic@gmail.com (G. Petrović), issam@iwate-pu.ac.jp (H. Fujita).

Software metadata

(Executable) Software metadata description	
Current software version	*1.1.1*
Permanent link to executables of this version	https://github.com/Neurocomputing/NEUCOM-D-14-02598
Legal Software License	*MIT license*
Computing platform / Operating System	*All systems capable of running Java 8.x* http://java.com/en/download/help/sysreq.xml
Installation requirements & dependencies	*Java 8.x*
If available Link to user manual - if formally published include a reference to the publication in the reference list	https://gajop.github.io/SoNeR
Support email for questions	gajopetrovic@gmail.com

Code metadata

Code metadata description	
Current Code version	*1d71751abc39f3a0c85b8dd037a0c890d0c453cb*
Permanent link to code / repository used of this code version	https://github.com/Neurocomputing/NEUCOM-D-14-02598
Legal Code License	*MIT license*
Code Versioning system used	*Git*
Software Code Language used	*Java, SQL*
Compilation requirements, Operating environments & dependencies	*Java, Maven Maven dependencies (which are obtained automatically by running 'maven') include: ant, ant-launcher, apache-jena-libs, libsvm, gex4j, log4j, post-gresql, mysql-connector-java, sqlite-jdbc*
If available Link to developer documentation / manual	https://gajop.github.io/SoNeR
Support email for questions	gajopetrovic@gmail.com

1. Introduction

In recent years we have seen an increase in the use of structured data such as RDFa [1], Dublin Core [2], schema.org [3] to supplement the classic web (written in HTML) and even attempts to fully expose the structured data in RDF/XML relying on standards such as the ones defined by the FOAF project [4,14]. Friend of a friend (FOAF) project defines a Semantic Web Ontology [5] for describing people and their relationships, often defined in RDF/XML, which is a format that uses a wide-spread and verbose representation that is XML with the simplicity of RDF triplets. The benefit of using structured data lies in providing a machine understandable format, which can be used as a data source by arbitrary programs, such as helping search engines better detect the content of a given page. That said, running algorithms against a dataset held in raw RDF/XML is computationally expensive (both in memory and speed) and therefore is recommended to transform it into a simpler and better optimized format for any demanding processing.

As motivation for this software, we have looked at the problem of analyzing datasets based on the FOAF ontology by running complex algorithms over a large social network. In particular we have focused on ranking users of social networks that support the FOAF format. Any algorithm whose result depends on the content of the entire social network, or even just large parts of it, is certainly not expected to be able to produce results in real-time, at the very least because of the time it takes to download and parse all documents. Therefore running ranking algorithms such as PageRank [6,7] or HITS [8], which take the entire social network (dataset) into account in order to compute the result require certain preparations. With that in mind, we have created a tool that obtains FOAF/RDF documents, parses them and puts them in a format (RDBMS) better optimized for reading speed and memory. To support this argument, we have conducted simple experiments by comparing computation time of operations such as reading entity fields, which have shown raw RDF/XML to be orders of magnitude slower than an RDBMS – extracting the required fields from raw RDF/XML took around 23 s whereas the same operation
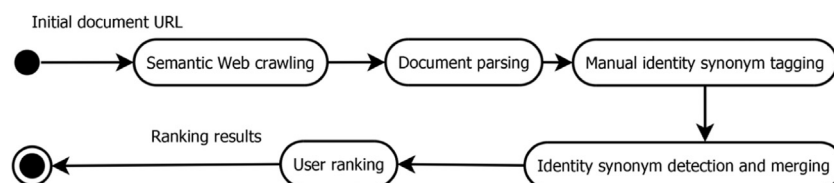

Fig. 1. Software architecture and usage.