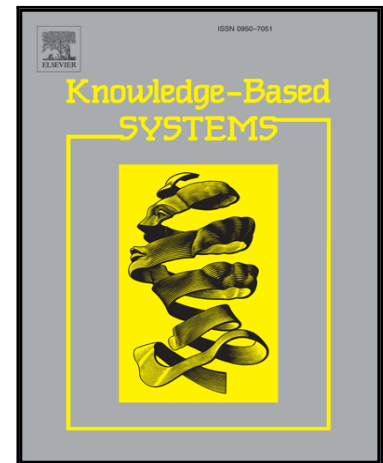


Accepted Manuscript

CrumbTrail: an Efficient Methodology to Reduce Multiple Inheritance in Knowledge Graphs

Stefano Faralli, Irene Finocchi, Simone Paolo Ponzetto, Paola Velardi

PII: S0950-7051(18)30162-X
DOI: [10.1016/j.knosys.2018.03.030](https://doi.org/10.1016/j.knosys.2018.03.030)
Reference: KNOSYS 4279



To appear in: *Knowledge-Based Systems*

Received date: 19 January 2018
Revised date: 20 March 2018
Accepted date: 22 March 2018

Please cite this article as: Stefano Faralli, Irene Finocchi, Simone Paolo Ponzetto, Paola Velardi, CrumbTrail: an Efficient Methodology to Reduce Multiple Inheritance in Knowledge Graphs, *Knowledge-Based Systems* (2018), doi: [10.1016/j.knosys.2018.03.030](https://doi.org/10.1016/j.knosys.2018.03.030)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

CrumbTrail: an Efficient Methodology to Reduce Multiple Inheritance in Knowledge Graphs

Stefano Faralli¹, Irene Finocchi², Simone Paolo Ponzetto¹, Paola Velardi²

¹ *University of Mannheim Germany*

{stefano,simone}@informatik.uni-mannheim.de

² *University of Rome Sapienza*

{finocchi,velardi}@di.uniroma1.it

Abstract

In this paper we present CRUMBTRAIL, an algorithm to clean large and dense knowledge graphs. CRUMBTRAIL removes cycles, out-of-domain nodes and non-essential nodes, i.e., those that can be safely removed without breaking the knowledge graph's connectivity. It achieves this through a bottom-up topological pruning on the basis of a set of input concepts that, for instance, a user can select in order to identify a domain of interest. Our technique can be applied to both noisy hypernymy graphs – typically generated by ontology learning algorithms as intermediate representations – as well as crowdsourced resources like Wikipedia, in order to obtain clean, domain-focused concept hierarchies. CRUMBTRAIL overcomes the time and space complexity limitations of current state-of-art algorithms. In addition, we show in a variety of experiments that it also outperforms them in tasks such as pruning automatically acquired taxonomy graphs, and domain adaptation of the Wikipedia category graph.

Keywords: Semantic Networks, Ontologies, Knowledge Graph Pruning

1. Introduction

In the recent years, research in information extraction and knowledge acquisition produced knowledge resources on a scale that was arguably hard to imagine a few years ago. Web-scale open information extraction systems like NELL
 5 [1] or ReVerb [2] have been successful in acquiring massive amounts of machine-

Download English Version:

<https://daneshyari.com/en/article/6861439>

Download Persian Version:

<https://daneshyari.com/article/6861439>

[Daneshyari.com](https://daneshyari.com)