RDFa crawler
Webcrawler including RDFa parsing

STEFAN ACHMÜLLER, ROLAND GRITZER, MATHIAS
GSCHWANDTNER
Universität Innsbruck
February 11, 2017

Abstract

Enabling to scrap the world wide web for RDFa data. Give a URL to start from, furthermore black- and whitelisting is possible. DRFa data, integrated within html documents, is generated by applying rdfa.info sequence. RDFa data is annotated by the RDF n-triple format.

Keywords: webcrawler, parser, rdfa

Contents

1 Problem Definition

Need to get information from web - clientside info building with rdfa, webcrawler to automate info gathering.

2 Methodology

Do research: javascript, node, schema.org, etc. Split work into parser and crawler. Set up front end (testing and presentation). Start implement. Put together and write documentation.

3 Outcome

1 package to download via node npm. 2 separate functionalities within package: crawler (without parser), and parser (without crawler).