LANCE
A Generic Benchmark Generator for Linked Data

T. Saveta\(^1\), E. Daskalaki\(^1\), G. Flouris\(^1\), I. Fundulaki\(^1\), A.-C. Ngonga Ngomo\(^2\)
\(^1\) FORTH-ICS, \(^2\) University of Leipzig
(jsaveta, eva, fgeo, funduli)@ics.forth.gr, ngonga@informatik.uni-leipzig.de

Motivation
The widespread adoption of Semantic Web Technologies and the publication of large interconnected RDF datasets and ontologies in the Web has made the integration of data a crucial task. Data linking in this context is essential in order to provide an integrated view of the underlying information; this is achieved by instance and schema matching techniques. To aid the users to choose among the systems that perform such tasks, a number of benchmarks have been developed.

Demonstration
LANCE is a novel instance matching benchmark generator for assessing instance matching techniques for RDF data with an associated schema.

LANCE Approach
LANCE is a flexible, generic and domain-independent benchmark generator which takes into consideration RDFS and OWL constructs in order to evaluate instance matching systems. LANCE supports:
- Semantics-aware transformations
- Standard value and structure based transformations\(^{[2,4]}\)
- Weighted gold standard based on tensor factorization
- Varying degrees of difficulty and fine-grained evaluation metrics

Transformations-based Test Cases

<table>
<thead>
<tr>
<th>Value-based</th>
<th>Semantics-aware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blank Character Addition/Deletion</td>
<td>RDF/OWL</td>
</tr>
<tr>
<td>Random Character Addition/Deletion/Modification</td>
<td>SD</td>
</tr>
<tr>
<td>Token Addition/Deletion/Shuffle</td>
<td>SQ</td>
</tr>
<tr>
<td>Country &amp; Simple Abbreviation</td>
<td>SC</td>
</tr>
<tr>
<td>Date Format</td>
<td>f-measure</td>
</tr>
<tr>
<td>Synonym/Antonym</td>
<td>Value</td>
</tr>
<tr>
<td>Stem of a Word</td>
<td>Value</td>
</tr>
<tr>
<td>Multilingualism</td>
<td>Value</td>
</tr>
<tr>
<td>Structure-based</td>
<td>Schema Triples</td>
</tr>
<tr>
<td>Property Addition/Deletion</td>
<td>SC</td>
</tr>
<tr>
<td>Property Aggregation</td>
<td>SC</td>
</tr>
<tr>
<td>Property Extraction</td>
<td>SC</td>
</tr>
</tbody>
</table>

Combination of transformations:

More than one transformation types per instance.

Simple (SC): One transformation per triple.

Complex (CC): Combination of two transformations per triple (value-based and structure-based or value-based and semantics-aware).

Applicability

- Evaluated LogMap\(^2\), LIMES\(^7\) running the EAGLE\(^1\) algorithm and OIO\(^9\)
- Entire source dataset transformed

Comments:
LogMap responds well to the value-based test cases but fails to find matches when the instance is involved in semantics-aware test cases. EAGLE and OIO: give very good precision results for the value-based test cases but are not able to find matches in the remaining categories.

References


This work was partially supported by the projects LDBC (FP7-ICT-2011-1 #317548), DIACHRON (ICT-2011-4.3, #601043) and PARTHENOS (#654119).

This poster is a companion to the accepted ISWC demo \(^{[1]}\) and research paper \(^{[2]}\).