Linked Data Lifecycle

- Storage/Querying
- Extraction
- Search/Explore
- Revision/Authoring
- Linking/Fusion
- Enrichment
- Quality Analysis
- Evolution/Repair
Section 1

Topics
Benchmarking? What is that?

Very briefly: Evaluate a system in a controlled environment and measure its *effectiveness* and *efficiency*. 
Topics
Benchmarking – GERBIL

Open Datasets → Benchmark Core → GERBIL

Systems
- Web service calls
- Your Annotator
  - Interface
  - Configuration
  - Interface

GERBIL

© DICE Group (Data Science) 2020

Data Science Suite V

July 13, 2020 6 / 20
**Topics**

**Benchmarking – GERBIL**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Extraction</td>
<td>Text</td>
<td>Structured data</td>
</tr>
<tr>
<td>Question Answering</td>
<td>Questions</td>
<td>Answers</td>
</tr>
<tr>
<td>Fact Checking</td>
<td>Facts</td>
<td>Veracity values</td>
</tr>
<tr>
<td>Machine Translation</td>
<td>Text</td>
<td>Text</td>
</tr>
</tbody>
</table>

**Summary**

- **Problem:** Development and maintenance of new GERBIL instances for different domains
- **Solution:** Extend GERBIL to become a light-weight benchmarking platform
- **Goal:** Develop GERBIL 2.0
Further information:

- M. Röder, R. Usbeck, and A. Ngomo: *GERBIL – Benchmarking Named Entity Recognition and Linking consistently*. 2018

Github projects:

- https://github.com/dice-group/gerbil

Technologies:

- Java / Maven
- UML
- RDF
Summary

- **Problem:** We need to improve crawlers for gathering data from the Data Web.
- **Solution:** Create a benchmark for such crawlers.
- **Goal:** Improve our Data Web crawler benchmark.
**Summary**

- **Problem:** We need to improve crawlers for gathering data from the Data Web.
- **Solution:** Create a benchmark for such crawlers.
- **Goal:** Improve our Data Web crawler benchmark.

---

**Diagram**

- **Seed URIs**
- **Frontier**
  - Normalizer
  - Filter
  - Queue
- **Worker**
  - Fetcher
  - Analyzer
  - Collector
  - Sink
- **URIs**

---

**Notes**

- DICE Group (Data Science)
- Data Science Suite V
- July 13, 2020
Topics
Benchmarking – ORCA

RDF Data Generator

Node Data Generator

Benchmark Controller

Evaluation Module

Sink

Linked Data Crawler

seed URIs
Further information:
- Paper on arxiv.org

Github projects:
- https://github.com/dice-group/orca
- https://github.com/dice-group/Squirrel

Technologies:
- Graph theory
- RDF, RDFa, Microformat, microdata, JSON-LD
- Java / Maven
- Docker
**Summary**

- **Problem:** We would like to be able to scale knowledge graphs.
- **Solution:** Create an algorithm that is able to mimic real-world graphs.
- **Goal:** Improve this library with respect to runtime and functionalities.
Further information:

- Master thesis
  https://hobbitdata.informatik.uni-leipzig.de/teaching/projectgroups/Thesis-Final-lemming.pdf

Github projects:

- https://github.com/dice-group/Lemming

Technologies:

- Graph theory
- RDF
- Java / Maven
Summary

- **Problem**: Application server is needed to transform data from RDF to JSON
- **Solution**: Let the UI work directly with RDF
- **Goal**: Implement a prototype based on templates
Summary

- **Problem**: Application server is needed to transform data from RDF to JSON
- **Solution**: Let the UI work directly with RDF
- **Goal**: Implement a prototype based on templates
Further information:

- Our DICE group website already relies on the template-based translation of RDF to static pages.
  https://github.com/dice-group/dice-website

Technologies:

- RDF, SPARQL
- JavaScript
Section 2

Summary
We want you

- **Create new software**: Develop new software and research prototypes.
- **Enhance code**: Improve existing solutions.
- **Participate**: Bring your own ideas in.
We Offer

- **Running software**: Open-source, industry-grade solutions
- **Real data**: Billions of facts from Wikipedia, bio-medicine, etc.
- **Expert tutors**, who developed the core software
- **Follow-up**: Topics can be extended to MSc thesis
- **Publications** at top conferences (ISWC, ESWC, WWW)
That’s all Folks!

Thank you!

Topics:
- Benchmarking
  - GERBIL
  - ORCA
  - Lemming
- Generic RDF UI templates

The topics are subject to change.
More information at
https://dice-research.org