Why I don't care for your F1 score

Evaluating Knowledge Extraction Tools for Industry Use

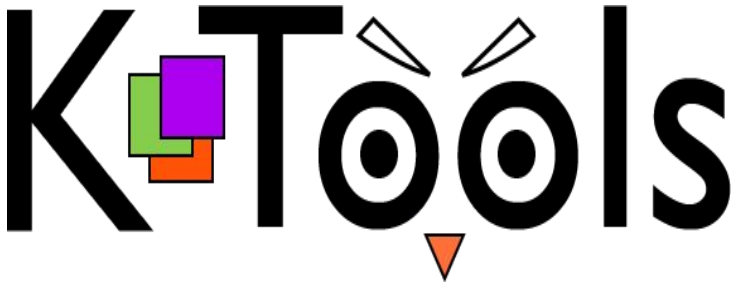
Panos Alexopoulos

Head of Ontology

textkernel

Machine Intelligence for People and Jobs

Knowledge Extraction Tools



Questions to ask before you "buy"

- **Q1:** *What extraction tasks and semantics does the tool exactly support?*
- **Q2:** *What performance range can we expect from the system and under what conditions?*
- **Q3:** *What are the ingredients and recipe(s) needed to use the tool?*
- **Q4:** *How can we troubleshoot/optimize the tool's performance?*
- **Q5:** *How do we operationalize and maintain the tool?*

Q1: Exact supported extraction tasks

- We want to determine whether and to what extent the tool can support the exact tasks we want to perform.
- Generic taglines and descriptions like “knowledge extraction from textual resources” or “machine reading for the Semantic Web” are not enough,
- Instead, it is important that both the input/output and the extracted semantics the tool supports are clearly and completely defined.

Q2: Performance range and conditions

- We are looking to learn something more than the precision and recall scores that the tool has achieved in a couple of experiments.

- The reason is that often performance varies significantly across different scenarios:

  - DBPedia Spotlight
    - 81% on a set of 155,000 wikilink samples
    - 56% on a set of 35 paragraphs from New York
    - 34% on the AIDA/CO-NLL-TestB dataset

  - AGDISTIS
    - 76% on the AQUAINT dataset
    - 60% on the AIDA/CO-NLLTestB dataset
    - 31% on the IITB dataset

Q2: Performance range and conditions

- What we need is a clearly expressed, educated generalization of the tool's performance, including the conditions under which the tool will perform best and worst.

- Possible conditions include:

    - Type and size of input text
    - Availability, quantity and quality of background knowledge
    - Target semantics
    - Domains

Q3: Ingredients and recipes

- Not the technical details of how to install or run the tool!
- What we are looking for methodological guidelines and best practices on how to use the tool so as to get the best possible results
- This is especially important for middleware tools that are highly versatile and configurable and require specific procedures and expertise to adapt them for a given use case.

Q4: Troubleshooting

- How can we troubleshoot and improve the tool's performance if the latter proves not to be satisfying for our case or data.
- Some tools may be black boxes, only specifying input and output, but others often have user-exposed parameters that can be configured to affect performance.
- In both cases, it can be hard for engineers that are no experts in Knowledge Extraction to diagnose and improve a problematic performance.

Q5: Operationalization and Maintenance

- We want to understand what it would take to put the particular tool in a production environment.
- Also how easy would be its maintenance.
- For example, if the tool is provided as a service, we need to know how much we can trust it to be available and up to date.
- Or, if the tool uses particular data or knowledge bases, it's important to know how often and in what ways these should be updated so that performance does not decay over time.

Thank you!

Panos Alexopoulos
Head of Ontology

E-mail: alexopoulos@textkernel.nl

Web: http://www.panosalexopoulos.com

LinkedIn: www.linkedin.com/in/panosalexopoulos

Twitter: @PAlexop