JOHANNES KEPLER
UNIVERSITÄT LINZ

Project for a PR/Bachelor's Thesis

Priv.-Doz. Mag. Dr.
Rick Rabiser
CDL MEVSS, ISSE

T +43 732 2468 4363
F +43 732 2468 4345
rick.rabiser@jku.at
http://mevss.jku.at/rabiser

Natural Language Data Querying

for the REMINDS Monitoring Framework

REMINDS (http://mevss.jku.at/reminds) [1] is a tool-supported framework for monitoring systems of systems at runtime. It comprises a flexible runtime monitoring infrastructure providing support for different roles and a requirements monitoring model covering the requirements to be monitored, the constraints checking adherence of a system's behavior to its requirements, the events and data produced by systems at runtime, and the probes instrumenting systems to intercept events and data at runtime.

In practice, besides monitoring the system at runtime, engineers are also interested in querying existing and archived events, data, as well as constraint violations, e.g., from databases or event traces recorded as files. Unfortunately, formulating queries on such databases and files, e.g., using SQL or other structured formats, can be quite challenging and creates a barrier preventing stakeholders from leveraging existing data.

The goal of this project is to explore if natural language queries could be used to retrieve information (events, data, constraint violations) collected and stored by the REMINDS framework:

- Investigate/find potential tools or libraries for natural language queries, e.g., see [2] and [3]
- Design and develop a query application to be used with REMINDS data bases
- Explore the capabilities of the query application by:
  - Using existing predefined queries
  - and writing new domain-specific queries,
  - to support different scenarios,
  - using test data obtained from REMINDS.
- Create a showcase and a brief (user as well as developer) documentation

To demonstrate what we expect, consider the following example: REMINDS recorded an event stream from monitoring an automation software system. Events could indicate states of the system, such as Initializing, Starting, Running, Completing, Finished, and Idle. A constraint checks that from Starting to Finished no more than 30 seconds pass. REMINDS recorded several violations of this constraint that occurred in a week-long recorded run of the system. A natural language query to find these violations among hundreds of thousands of events and several other violations could be: "Show me all violations regarding system performance" or, more specific, "Does the system always finish fast enough?". The application to be developed should enable such or similar queries.

[1] M. Vierhauser, R. Rabiser, P. Grünbacher, K. Seyerlehner, S. Wallner, and H. Zeisel, "ReMinds: A Flexible Runtime Monitoring Framework for Systems of Systems" Journal of Systems and Software, 2016 (http://dx.doi.org/10.1016/j.jss.2015.07.008).

[2] P. Pruski, S. Lohar, W. Goss, A. Rasin, and J. Cleland-Huang. "TiQi: answering unstructured natural language trace queries". Requirements Engineering, 20(3), 215-232, 2015. (http://dx.doi.org/10.1007/s00766-015-0224-4)

[3] https://opensource.com/business/15/7/five-open-source-nlp-tools

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenberger Straße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696