Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	David Košťál		
Název práce	Complex experiment support through the IVIS framework		
Rok odevzdání	2025		
Studijní program	Informatika	**Studijní obor**	Softwarové a datové inženýrství

Autor posudku	Tomáš Petříček	**Role**	Oponent
Pracoviště	Department of Distributed and Dependable Systems		

Text posudku:

The thesis presents a primarily engineering contribution to the international and collaborative academic project ExtremeXP. The aim of ExtremeXP is to allow users to systematically conduct experiments in order to explore and understand data. The thesis develops and presents a sub-system for storing information about experiments that will serve as one of the main components of the overall ExtremeXP system.

The technical aspect of the thesis consists of multiple interconnected aspects. First, the author designed the model of experiments, he implemented an API for storage and retrieval of data on experiments and he integrated the API into an existing IVIS platform. As part of the work, he also updated the IVIS system itself. The author also managed the deployment of the system to make it available to project partners.

EVALUATION

On the engineering side, the project involves a fair amount of technical work. It required becoming familiar with a large existing system (IVIS), designing an extension based on requirements collected from project partners, implementing the server-side and extending the client-side part of the system. The technical realization of the system shows good engineering practices - ranging from the use of Docker for deployment to the use of Postman for API development. The thesis provides a clear description of the final system design.

That said, there are likely interesting aspects of the work that are not covered by the thesis - and so remain hidden. The collaborative nature of the ExtremeXP project (and the fact that the author participated in the ExtremeXP meetings directly) means that there was likely significant amount of work done on requirements gathering, revisions to the initial model etc. It would be interesting to document and reflect on this process and the methodologies used.

The brief "Evaluation" section in the thesis also notes that the resulting system is used by four project partners, which shows the clear value of the work. The section, however, also briefly notes that the API was then extended - the thesis does not reflect on this. If it did so, it might present interesting observations on software engineering.

In summary, the thesis presents a well-executed primarily engineering project. It would be even better if it attempted to critically reflect on the project - and perhaps tried to draw more general lessons about software engineering from the project.

MINOR POINTS

The phrasing in the opening of chapter 3 is a bit confusing ("The thesis contributed with the division of data...") - I see what you are trying to say here, but it would be better to say "The author contributed...".

The link in footnote on page 63 is a link to an academic paper on a repository. A citation would be better here.

QUESTIONS

The thesis clearly explains that traceability does not imply reproducibility. I would be curious what exactly the objectives of traceability here are. In particular, there is a nice existing framing of the difference between reproducibility and replicability [2]. Where would traceability fit?

On the technical side - the system provides a relatively lightweight layer on top of the underlying data storage. I was wondering if you considered using some kind of a system that automatically generates an API based on a data model? Two possibilities would be something like PostgREST [3] or GraphQL [4]? Is this something that could work, or are there reasons why not? And could GraphQL make the system more flexible so that project partners do not later ask you for more changes in the system?

I believe the general framing of the ExtremeXP would benefit from some broader context. For example, can you explain how does the project relate to existing scientific workflow systems such as Apache Taverna [1]?

[1] https://esciencelab.org.uk/products/taverna/
[2] https://book.the-turing-way.org/reproducible-research/overview/overview-definitions.html
[3] https://docs.postgrest.org/en/v12/
[4] https://graphql.org/

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 3. February 2025 **Podpis**