

Review of Master's Thesis

Student: Kozák David, Bc.
Title: Indexing of Big Text Data and Searching in the Indexed Data (id 22961)
Reviewer: Smrž Pavel, doc. RNDr., Ph.D., DCGM FIT BUT

- 1. Assignment complexity** **more demanding assignment**
The topic required studying and understanding a broad range of relevant fields - semantic enrichment of texts, distributed processing, advanced querying languages, etc., and using the knowledge in the realization of the search system.
- 2. Completeness of assignment requirements** **assignment fulfilled with enhancements**
The goals were fully achieved and the resulting solution even extends the expected results towards practical applicability in distributed environments - for example, the system regularly checks individual nodes running the index servers and takes care of the graceful degradability of the system as a whole.
- 3. Length of technical report** **in usual extent**
The Master's thesis has 60 pages, including the list of references and appendices.
- 4. Presentation level of technical report** **95 p. (A)**
The thesis has a clear and logical structure, chapters 2 and 3 cover the state of the art in relevant areas of text processing and indexing, the core of the text deals with the design, implementation, and evaluation of the new indexing system.
- 5. Formal aspects of technical report** **95 p. (A)**
The text is written in very good English and its typographic quality is very high too.
- 6. Literature usage** **85 p. (B)**
The author references relevant scientific resources and presents results of a clear compilation of the current state of the art in the field. On the other hand, I would appreciate a broader survey of existing approaches to semantic querying, including methods combining the search in a combination of structured and textual data.
- 7. Implementation results** **100 p. (A)**
The technical quality of the realized system is excellent, it is fully functional and provides a base for further research work in the semantic search.
- 8. Utilizability of results**
The system is already used for searching in semantically enriched data within the Knowledge Technology Research Group at FIT BUT.
The student presented his work at the student conference Excel@FIT and was awarded in one of its categories.
- 9. Questions for defence**
What steps would be needed to replace the MG4J system by another indexing server, e.g., Manatee?
- 10. Total assessment** **95 p. excellent (A)**
The overall contribution of the presented Master's thesis is significant. The realized system is fully functional and has been already employed as the primary search solution in the Knowledge Technology Research Group at FIT BUT. The text has very good quality and the work was awarded at the Excel@FIT conference.

In Brno 29 June 2020

Smrž Pavel, doc. RNDr., Ph.D.
reviewer