

Review report of a final thesis

Czech Technical University in Prague

Faculty of Information Technology

Student: Viacheslav Kroilov
Reviewer: Ing. Daniel Langr, Ph.D.
Thesis title: Numerical database system
Branch of the study: Computer Science

Date: 8. 6. 2017

<p><i>Evaluation criterion:</i></p> <p>1. Difficulty and other comments on the assignment</p>	<p><i>The evaluation scale: 1 to 5.</i></p> <p><i>1 = extremely challenging assignment, 2 = rather difficult assignment, 3 = assignment of average difficulty, 4 = easier, but still sufficient assignment, 5 = insufficient assignment</i></p>
<p><i>Criteria description:</i> Characterize this final thesis in detail and its relationships to previous or current projects. Comment what is difficult about this thesis (in case of a more difficult thesis, you may overlook some shortcomings that you would not in case of an easy assignment, and on the contrary, with an easy assignment those shortcomings should be evaluated more strictly.)</p> <p><i>Comments:</i> The thesis analyses memoization systems, designs and implements new concurrent versions intended for multi-threaded environments, and conducts their experimental evaluation. This is generally a challenging problem, since memoization systems need to be updated not only while inserting new data, but also while searching for the current ones. Simple locking of whole system would not provide scalable performance and efficient implementation need to use advanced concurrent access strategies.</p>	
<p><i>Evaluation criterion:</i></p> <p>2. Fulfilment of the assignment</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p><i>1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled</i></p>
<p><i>Criteria description:</i> Assess whether the thesis meets the assignment statement. In Comments indicate parts of the assignment that have not been fulfilled, completely or partially, or extensions of the thesis beyond the original assignment. If the assignment was not completely fulfilled, try to assess the importance, impact, and possibly also the reason of the insufficiencies.</p> <p><i>Comments:</i> The thesis goals were accomplished</p>	
<p><i>Evaluation criterion:</i></p> <p>3. Size of the main written part</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p><i>1 = meets the criteria, 2 = meets the criteria with minor objections, 3 = meets the criteria with major objections, 4 = does not meet the criteria</i></p>
<p><i>Criteria description:</i> Evaluate the adequacy of the extent of the final thesis, considering its content and the size of the written part, i.e. that all parts of the thesis are rich on information and the text does not contain unnecessary parts.</p> <p><i>Comments:</i> The thesis text is 40 pages long, additional detailed results of the experiments can be found in Appendix B. Chapter 1 includes necessary theory concerning involved data structures and thread synchronization. Most of the text discusses designed data structures and algorithms, their implementation and experimental results.</p>	
<p><i>Evaluation criterion:</i></p> <p>4. Factual and logical level of the thesis</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>95 (A)</p>
<p><i>Criteria description:</i> Assess whether the thesis is correct as to the facts or if there are factual errors and inaccuracies. Evaluate further the logical structure of the thesis, links among the chapters, and the comprehensibility of the text for a reader.</p>	

Comments:

The whole thesis is very-well written and despite of the complex topic, the text can be easily understood. It contains all the necessary parts and the experimental section shows and discusses the measured results. I have only very few comments here:

— Pp. 13: "It is a huge drawback since it is natural to run scientific applications on many-core systems." It may not be as huge drawback as it seems. Many large-scale HPC applications uses pure MPI without shared-memory parallelization (threading). Only applications intended for "single-node" many-core systems and applications build upon hybrid MPI+threading (MPI+OpenMP) need multi-threaded algorithms / data structures.

— Pp. 25. "...std::hash for an integer is defined as the number itself (at least on some compilers..." A compiler does not define C++ Standard Library functions. These depend on the implementation of C++ Standard Library, e.g, GNU Libstdc++. Generally, different compilers can use the same implementation of C++SL and the same compiler can work with different implementations of C++SL.

— I would like to see a short sample code how the library is used. Even though it is generally recommended not to put source code into text of theses, when the text describes a developed library, some code might be useful (such as API and minimal working examples of library usage).

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

5. Formal level of the thesis

98 (A)

Criteria description:

Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 14/2015, Article 3.

Comments:

The typographic and language level of the thesis is excellent. The results are clearly presented by using graphs and tables. I haven't found any typos or grammar errors in the text. My only comment is as follows:

— Text sometimes overruns the right margin, see, e.g., pp. 7, 11, 12, etc. Similarly, sometime left margin is overrun as well (see, e.g., pp. 24). This should not happen.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

6. Bibliography

100 (A)

Criteria description:

Evaluate the student's activity in acquisition and use of studying materials in his thesis. Characterize the choice of the sources. Discuss whether the student used all relevant sources, or whether he tried to solve problems that were already solved. Verify that all elements taken from other sources are properly differentiated from his own results and contributions. Comment if there was a possible violation of the citation ethics and if the bibliographical references are complete and in compliance with citation standards.

Comments:

I haven't found any problems here. Just one comment/advice: It is generally recommended that a sentence should remain valid even if references are removed. For instance, it is better to write:

"The exact AVL tree implementation was described by Knuth [5, p. 458]."

than

"The exact AVL tree implementation is described in [5, p. 458]." (as at page 5).

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

7. Evaluation of results, publication outputs and awards

95 (A)

Criteria description:

Comment on the achieved level of major results of the thesis and indicate whether the main results of the thesis extend published state-of-the-art results and/or bring completely new findings. Assess the quality and functionality of hardware or software solutions. Alternatively, evaluate whether the software or source code that was not created by the student himself was used in accordance with the license terms and copyright. Comment on possible publication output or awards related to the thesis.

Comments:

Main outcomes of the thesis consists of a design of concurrent memoization systems, their implementation in the form of standalone library, and also its experimental evaluation. All these outcomes are of high quality. Only few minor comments:

— I do not like the idea of making the code C++14 dependent. C++14 might not be even nowadays supported by all mainstream C++ compilers used in HPC environments (such as compilers from Intel, PGI, Cray, IBM, etc.). Or, I would like to see the compiler compatibility test for the code and different compilers.

— I would not call the folder with source files "lib". "lib" is a standard name for a folder with compiled library files.

Evaluation criterion:

No evaluation scale.

8. Applicability of the results

Criteria description:

Indicate the potential of using the results of the thesis in practice.

Comments:

The developed memoization library has a high potential for usage in HPC codes. Though, generally, in multi-threaded programs one wants to avoid memoization and prefer fixed-sized databases, it is not always possible. Moreover, improving memory subsystems is very expensive in comparison with increasing computational power of computer systems in general. The amount of memory per core (or per GFLOP) is continuously decreasing and this will become a significant challenge for programmers especially with future HPC systems based on accelerators/coprocessors with fast but small memories (e.g., GPGPUs, Intel MICs, etc.). Memoization might therefore become a necessary part of many HPC codes.

Evaluation criterion:

No evaluation scale.

9. Questions for the defence

Criteria description:

Formulate any question(s) that the student should answer to the committee during the defence (use a bullet list).

Questions:

- What features of C++14 are used in the code? Do they include language features only, or C++ Standard Library features as well?
- Can the library be used without Boost.math? As written, Boost.math is required only by the benchmark program. It would be bad practice to introduce Boost-dependency into the whole library just because of benchmarking.
- I can't see the reason for implementing hash tables by class inheritance. Why not use simply a single templated class:
template <class EvictionPolicy> FixedHashTable { ... };
where EvictionPolicy would determine the eviction policy? See "policy-based design" idioms / design patterns. It would be much simpler and readable. Inheritance and CRTP might be a powerful tool, but it should not be used if it is not necessary. Is it really necessary here?

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

10. The overall evaluation

98 (A)

Criteria description:

Summarize the parts of the thesis that had major impact on your evaluation. The overall evaluation **does not** have to be the arithmetic mean or any other formula with the values from the previous evaluation criteria 1 to 9.

Comments:

This is an excellent thesis with practically no errors and I had only very few minor comments to its content and form. Moreover, the concurrent memoization is a viable research topic and the outcomes of the thesis might turn out to be very useful for the HPC community in the future.

Signature of the reviewer: