



fcfm

Ciencias de la Computación

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

July 30th 2014

Evaluation of doctoral thesis – RnDr. Tomas Bartos

Dear Prof. Kratochvil,

Please find enclosed my evaluation for the aforementioned doctoral thesis. Should you have any questions, please don't hesitate to contact me.

Yours sincerely,

Benjamin Bustos
Associate Professor
Department of Computer Science
University of Chile

Examiner's Report on Ph.D. Thesis

Title: "Indexing Arbitrary Similarity Models"

Candidate: RnDr. Tomas Bartos

1) Contained results

The thesis makes several novel contributions to the similarity search research domain. Specifically, this thesis proposes the SIMDEX framework, which provides a new and systematic way for testing mathematical axioms that may be used for indexing purposes. The proposed methodology is data-driven. This line of research addresses an open problem on similarity search on complex data: finding new axioms, apart from the well-known triangle and Ptolemaic inequalities, for lower-bounding distance functions. The thesis proposes several ways for implementing the SIMDEX framework (an iterative version, a version based on genetic algorithms, and a parallel version), discussing their advantages and disadvantages. Additionally, it shows how the framework can be used in practice for accelerating indexing techniques from the state of the art, by the so-called Smart Pivot Tables. Overall, the presented framework and indexing techniques can be referred as new scientific results. This has been already validated with several publications in renowned international conferences and journals.

2) Importance for the area

The importance of this research for the similarity search domain is two-fold. Firstly, it addresses the efficiency problem, which is critical for the scalability of the indexing techniques and therefore crucial for practical applications in real-world scenarios. Secondly, it also addresses the related effectiveness problem, discussing from a practical point of view the pros and cons of the proposed framework, in terms of the quality of the answer provided by similarity search systems that take advantage of SIMDEX.

3) Possible applications to other areas

A similarity search system is a basic tool that has applications in several areas. The SIMDEX framework could be used to improve, for example, algorithms for data mining based on top-k rankings, thus basically any application (not only from Computer Science but also from other disciplines) that needs to use such algorithms may directly benefit from SIMDEX.

4) Form of the submitted thesis

The thesis is very well written and of good quality. I would recommend revising the future work section on the last chapter (Conclusions), which could have presented a more in-depth discussion of potential research trends.

5) Authors ability for creative scientific work

This thesis proves the author's ability for doing creative scientific work.

6) Final recommendation

I recommend accepting this doctoral thesis, and it is ready to be defended by the candidate for obtaining the Ph.D. degree.