TOWARDS CACHE OPTIMIZATION IN FINITE AUTOMATA IMPLEMENTATIONS

By

ERNEST KETCHA NGASSAM

Submitted in partial fulfilment of the requirements for the degree Philosophiae Doctor (Computer Science) in the Faculty of Engineering, Built Environment and Information Technology University of Pretoria, Pretoria

September 2006

TOWARDS CACHE OPTIMIZATION IN FINITE AUTOMATA IMPLEMENTATIONS

By

ERNEST KETCHA NGASSAM

Supervisor: Bruce W. Watson

Co-Supervisor: Derrick G. Kourie

Department of Computer Science,
University of Pretoria

ABSTRACT

To the best of our knowledge, the only available implementations of FA-based string recognizers are the so-called conventional table-driven algorithm and, of course, its hardcoded counterpart suggested by Thompson, Penello, and DeRemer in 1967, 1986, and 2004 respectively. However, our early experiments have shown that the performance of both implementations is hampered by the random access nature of the automaton's transition table in the case of table-driven, and also the random access nature of the directly executable instructions that make up each hardcoded state. Moreover, the problem of memory load and instruction load are also performance bottlenecks of these algorithms, since, as the automaton size grows, more space in memory is required to hold data/instructions relevant to the states.

This thesis exploits the notion of cache optimization (that requires good data or instructions organization) in investigating various enhancements of both table-driven and hardcoding.

Functions have been used to formally define the denotational semantics of string recognizers. These functions rely on various so-called *strategy variables* that are integrated into the formal definition of each recognizer. By appropriately selecting these variables, the conventional algorithms may be described, without loss of generality. By specializing these strategy variables, the new and enhanced recognizers can be denotationally described, and resulting algorithms can then be implemented.

We first introduce the so-called *Dynamic State Allocation* (DSA) strategy regarded as a sort of Just-In-time (JIT) implementation of FA-based string recognizers whereby a predefined portion of the memory is reserved for acceptance testing. Then follows the *State pre-Ordering* (SpO) strategy that assumes some prior knowledge on the order in which states would be visited. In this case, acceptance testing takes place once each state have been allocated to its new position in memory. The last strategy referred to as the *Allocated Virtual Caching* (AVC) strategy is based on the premise that a portion of the memory originally occupied by the automaton's states is virtually used as a sort of cache memory in which acceptance testing takes place, enabling therefore, the exploitation of the various performance enhancement notions on which hardware cache memory relies.

It is shown that the algorithms can be classified in a taxonomy tree which is further mapped into a class-diagram that represents the design of a toolkit for FA-based string recognition. Also given in the thesis are empirical results that indicate that the algorithms suggested can, in general, outperform their conventional counterparts when recognizing large and appropriately chosen input strings.

Keywords: Finite automata, algorithmic, taxonomy, toolkit, software construction, cache optimization, locality of reference, implementation strategies, performance, string recognizer.

To Maurice and Madeleine Ngassam;
To Orline Tchamen Ngassam;
To late Pierre Tatchou Tchoukwam;
To late Pauline Tchetgnia;
To late Jean Petji;
To my Ancestors.

ACKNOWLEDGEMENTS

I would like to thank Derrick G. Kourie and Bruce W. Watson, my supervisors, for their many suggestions and constant support during this research. Their support and guidance from the early years of chaos and confusion until the very end of this research will never be forgotten. They have been impressively valuable in enhancing my research capabilities, my thinking abilities, and of course my commitment to hardwork.

During this whole journey, I had the opportunity to interact with various Universities. Such interactions have indeed been fruitful along the path of achieving this goal. To all colleagues and friends of the **FASTAR** (**F**inite **A**utomata **S**ystems, **T**heoretical and **A**pplied **R**esearch) Research group of the University of Pretoria (Department of Computer Science) and the Technical University of Eindhoven (Department of Software Construction), to those at the Czech Technical University in Prague (Department of Computer Science), and to those at the University of South Africa (School of Computing, my employer), I would like to take this opportunity to express my gratitude for their constant support. A special thank to Professor Paula Kotzé of the School of Computing at the University of South Africa for her constant support and encouragement throughout my journey of becoming a *researcher*.

I should also mention that my studies were supported by the University of Pretoria through its postgraduate bursary scheme (Postgraduate Student Award) as well as the National Research Foundation (NRF) through the Postgraduate Supervision Bursary Nomination Scheme graciously motivated by my supervisors Professors Derrick G. Kourie and Bruce W. Watson.

Of course, I am grateful to my parents Maurice and Madeleine, for their patience and *love*. Without them this work would never have come into existence (literally).

My thanks also go to my immediate family: my wife Marie Louise Liliane Yemdam-Ketcha, my children Orline Sorel Ketcha, Karl Ryan Ketcha, Ashlyne Shakirah Ketcha, and my niece Carine Ngami Tchouatchoua. Without them, the motivation, the energy and the passion behind this work would have not been materialized.

Finally I wish to thank the following for their *unconditional* support and love: Lysette Tchatat Ngassam, Guy Tchoukwam Ngassam, Laurent Wandja Ngassam, Mirabelle Soumbé Ngassam, Orline Tchamen Ngassam (*my angel*), and Aimé Floriant Noungo Ngassam. To them I wish to say, *the Ngassam's dream has finally come true, and the real battle has just begun.*

Pretoria, South Africa — Ernest Ketcha Ngassam
Semptember 7, 2006

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**