Querying XML: benchmarks and recursion
Afanasiev, L.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
In this thesis, we pursue two main research themes: performance evaluation and optimization of XML query processing. Our focus is on the XQuery query language. These themes are tightly connected, since performance evaluation is aimed at measuring the success of optimization techniques. More specifically, we pursue *benchmarking as a performance evaluation technique* on the one hand, and *optimization techniques for recursion* in XQuery, on the other hand.

In the first part of the thesis, we develop benchmarking methodology and tools for XQuery. We start by analyzing XQuery benchmarks published by 2006, namely XMach-1, XMark, X007, MBench, and XBench, and survey their usage. We analyze and compare their workloads (i.e., the data and query sets) and measures. We also execute the benchmarks on four XQuery engines and analyze the results. Next, we discuss how to achieve the repeatability of experimental evaluations of computer systems in the database domain. As part of setting up a methodology for repeatability, we perform a review of articles submitted to the research conference SIGMOD 2008 and measure the repeatability of the presented experimental evaluations. Further, we address the problems and challenges of automating the execution of performance evaluation benchmarks on many XML query engines and comparison of their performance. We present a software tool, XCheck, as a solution to these problems. As a result of our analysis of XQuery benchmarks, we identify a lack of suitable tools for precise and comprehensive performance evaluations. We address this problem by developing a methodology for micro-benchmarking XML query engines, which we refer to as MemBeR. MemBeR also comprises a framework for collecting and storing micro-benchmarks. Finally, we present a MemBeR-style micro-benchmark for testing performance of value-based joins expressed in XQuery. We evaluate the micro-benchmark by analyzing the performance of four XQuery engines.

In the second part of the thesis, we investigate declarative means of obtaining recursion in XQuery. Namely, we add an inflationary fixed point operator to XQuery. We propose an optimization technique for processing this opera-
This optimization relies on a distributivity property of XQuery expressions. Further, we implement this technique on top of the XML database system, MonetDB/XQuery, and evaluate its performance using the tools developed in the first part of the thesis. Finally, we investigate the theoretical aspects of this inflationary fixed point operator in the context of Core XPath, the XML tree navigational core of XPath and XQuery. We prove that the satisfiability problem of Core XPath extended with the inflationary fixed point operator is undecidable.