

Abstract

On the Size of Deterministic Finite Automata Obtained from XPath Expression

AKIMASA MORIHATA^{1,a)}

Presented: November 5, 2015

XPath is fundamental to tree processing. It is well known that a core part of XPath, navigational top-down XPath, can be implemented by a top-down finite-state transition. However, to the author's knowledge, it is not known how large the minimum deterministic finite automaton corresponding to an XPath expression can be. In this presentation, we show the following results. Given an XPath expression of size n , one can obtain a deterministic finite automaton of $O(n)$ states if it does not contain predicates or descendant axes, whereas the number of states can be doubly-exponential to n otherwise. These results are not only theoretically interesting but also practically useful, especially for distributed XPath processing. They enable us to estimate sizes of messages exchanged between computation nodes. The results show that in the worst case, the message size should be exponential to the size of XPath expression, which indicates an inherent difficulty of distributed XPath processing.

¹ Graduate School of Arts and Sciences, The University of Tokyo, Meguro, Tokyo 153-8902, Japan

^{a)} morihata@graco.c.u-tokyo.ac.jp