

Context-Free-Grammar based Token Tagger in Reconfigurable Devices

Young H. Cho, James Moscola, and John W. Lockwood

Washington University in St. Louis
Department of Computer Science and Engineering
St. Louis, Missouri 63130
{young, jmm5, lockwood}@arl.wustl.edu

We present a high performance reconfigurable hardware architecture for detecting patterns as well as their contextual meaning. By analyzing for both semantic content and structure, the accuracy of content-level processing systems can be improved. Our system is built using semantics defined by context-free-grammar (CFG) to tag the streaming data. Unlike the traditional table look up and stack based engines used in CFG parsers, we explore a new method that maps the grammar structure on to the Field Programmable Gate Arrays (FPGA) hardware. The structure is a direct translation of the grammar which enables the meaning of the patterns to be determined based on the location of its detection. The parallel pattern detection engines are instantiated using FPGA resources. Our implementation scans for the regular expression patterns and determines their semantics as defined by a grammar. This highly parallel and fine grained pipelined engine with 8 bit input bus can operate in bandwidth above 2 Gbps. For a simple XML grammar example, our engine can detect and tag the patterns at 1.57 Gbps on Xilinx Virtex E FPGA and 4.26 Gbps on the new Virtex 4 devices.