

Posudek na diplomovou práci

Paralelní XML parsing

Tomáš Stráník

V současné době jsou již běžně k dispozici počítače disponující schopnostmi pro paralelní vykonávání instrukcí na více jádrech procesoru. Běžné XML parsery přitom těchto možností nevyužívají a parsují XML pouze s využitím jednoho vlákna. Tato práce si kládla za cíl prozkoumat možnosti paralelního parsování XML, vytvořit pilotní implementaci XML parseru využívajícího více výpočetních jader a ověřit škálovatelnost navrženého řešení.

Zatímco kapitola dvě stručně shrnuje známé fakty o XML a programování v paralelním prostředí, kapitola tři již tvoří důležitou část navrženého řešení. Jsou v ní popsány možnosti paralelního zpracování podúloh v závislosti na gramatice jazyka XML.

Kapitola čtyři spadá již do kategorie překladačů, když je v ní podrobněji rozebrána lexikální analýza XML parseru.

Syntaktická analýza s ohledem na paralelní běh je představena v kapitole pět. Tato kapitola je jádrem diplomové práce, neboť v ní je hlavní přínos této diplomové práce – popis paralelního zpracování podúloh XML parseru a jejich následného spojování v jeden naparsovaný dokument.

Kapitola šest pak podrobněji pojednává o pilotní implementaci formou knihovny (což je běžná forma distribuce pro XML parsery) s názvem XPP. Řeší se zde některé technické detaily včetně jednoduché demonstrace použití knihovny.

Zdaleka nejzajímavější je pak kapitola sedm, kde autor měří zrychlení svého řešení v závislosti na počtu využitých procesorů. Jsou zde popsány a vysvětleny zajímavé jevy, které se objevily při měření (např. efekt Amdahlova zákona na malých dokumentech nebo chování na dokumentech s komentáři). Druhou částí měření bylo porovnání s existujícími velmi známými XML parsery. Měření prokázalo, že ačkoliv je knihovna XPP pouze pilotní implementací (srovnání časů XML parserů a sériového běhu knihovny XPP), tak paralelní zpracování knihovnou XPP přináší značné zrychlení při XML parsování.

Jako zajímavost bych uvedl, že v průběhu psaní diplomové práce se objevil článek s podobným řešením uvedený v literatuře jako [7]. Nicméně autorova práce se odlišuje již od počátečního návrhu v několika bodech od navrhovaného řešení, navíc autorem naměřená zrychlení výrazně překonávají hodnoty v článku, kde také mimo jiné měřili zrychlení.

Asi jedinou vadou na kráse je fakt, že výsledná implementace je pouze pilotní, takže není implementován plný XML parser pracující se všemi vlastnostmi XML, jinak by se jednalo o dílo přímo využitelné i v praxi. Nicméně cílem diplomové práce nebylo zhotovit produkt v komerční kvalitě, takže se jedná spíše o povzdechnutí než o nesplnění cíle diplomové práce.

Diplomová práce je velmi kvalitní, navržené řešení předčí všechny zatím známé možnosti paralelního zpracování XML. Určitě je to námět na vědecký článek na konferenci. Doporučuji tuto diplomovou práci k obhajobě.

25.1.2010

RNDf. Jakub Yaghib, Ph.D.