

Posudek vedoucího na diplomovou práci

Filtrování informací v XML dokumentech

Autorka práce: Michaela Pilátová

Cílem práce bylo použít architekturu filtrování informací v prostředí XML dat. Autorka se zaměřila na jednu z možných variant tématu – tj. systémy vydavatel-předplatitel (publish-subscribe).

Po krátkém úvodu podává autorka v kap. 2 stručný úvod do problematiky a v kap. 3 krátké seznámení s jazykem XML. V kap. 4 nazvané „Práce s XML“ podává (na poněkud nevhodném místě) výklad o XML schématech. Následují informace o parserech a detailnější informace o jazyku XPath, který je použit dále (kap. 5) jako jazyk pro formulaci profilů v navrženém a implementovaném systému. Autorka v kap. 5 diskutuje několik existujících filtrovacích mechanismů a jejich začlenění do architektury systému vydavatel-předplatitel. Uvažovány jsou systémy v síťovém prostředí založené na brokerech. Zde se již dostáváme k jádru práce – strategiím přidělování profilů na hostitelské brokery a přenosům XML dokumentů. Autorka vylepšila jednu z existujících strategií a použila toto vylepšení ve vlastním implementovaném prototypu systému XmlPart. Jeho popisu je věnována kap. 6. Zde je uveden i příslušný algoritmus včetně důkazu správnosti. Kap. 6 také obsahuje návod k práci se systémem (uživatelskou příručku) a některé technické detaily. Poslední sekce kap. 6 jsou věnovány experimentům, jejich výsledky jsou podány v názorných grafech, což umožňuje snadné porovnání. Autorka použila reálná data z databáze INEX, kterou máme k dispozici. Do experimentů vstoupila i komprese. Diskuse porovnání metod ukázala dobrou kvalitu navrženého přístupu.

Práce je napsána na přijatelné úrovni výkladu, i když, zvláště v úvodních kapitolách, poněkud stručně. Také styl resp. čeština se trochu vymyká běžným zvyklostem používaným v odborném textu (např. „Celou práci poté zakončí závěrečná kapitola 7“, „dostane opravdu právě ty zprávy“, „doted“)

Některé připomínky, poznámky a otázky k práci:

8₅ – jaký je rozdíl mezi výrazovou silou a vyjadřovací silou např. na str. 10 nebo dokonce expresivitou na str. 22?

9₄ – jde o dvojici jméno atributu-hodnota a nikoliv atribut-hodnota.

18₄ – ID není hodnota a ani atribut, ale možný typ atributu. Ani výraz „ID atributu“ na str. 19 nahoře nedává smysl.

23 – hierarchický strom – jde o termín s redundancí.

24 – Výsledek dotazu v XPath je (mimo jiné) množina uzlů. To není totéž co elementy či atributy.

40 – Plojem broker by měl být vyložen.

41 – Architektura na obr. 8 by měla korespondovat s obr. 1.

44 – než „exkluzivita“, tak raději rovnou výlučnost.

Závěr: Diplomantka se velmi samostatně chopila tématu, dobře se zorientovala v problematice a tvůrčím způsobem přistoupila k řešení. Práce plně splňuje nároky na diplomovou práci a tvoří dobrý základ pro další vývoj. Doporučuji ji k obhajobě.

V Praze dne 27.1. 2006

Pro
: