

Posudek vedoucího diplomové práce

Michal KASKA

Systém pro sledování a analýzu změn stránek vybraných webových serverů

Cílem práce bylo navrhnout a implementovat systém pro ukládání, sledování a analýzu změn stránek vybraných webových serverů v určité doméně. Předpokládanými uživateli byli především organizace, monitorující obsahy webů z hlediska zákonnosti jejich obsahu (policie), dodržování autorských práv (vlastníci těchto práv či jejich zástupci) a podobně. Dalšími předpokládanými uživateli jsou běžní uživatelé vytvářející historické kopie svých stránek a archivaci jejich změn.

Na rozdíl od konkurenčních systémů, určených primárně ke stažení obsahu, ukládá data do databáze, což sebou přináší řadu výhod z hlediska zabezpečení dat a možností jejich dalšího paralelního zpracování. V tomto směru systém bohužel omezuje fakt, že autor do databáze ukládá právě jen dvě kopie jedné úlohy (aktuální a předešlou) a nelze proto ukládat a sledovat změny stránek v dlouhodobém časovém horizontu. Domnívám se, že povolení neomezeného počtu verzí by nebylo nijak výrazně pracnější a předpokládaným uživatelům by přineslo mnohem vyšší užitek.

Přes ukládání dat v databázi místo přímo do souborového systému program díky promyšlené paralelizaci úloh získává data dostatečnou rychlostí. Duplicitní soubory se ukládají jen jednou, což snižuje požadovaný prostor na disku. Další úspory místa by šlo zřejmě (ve vyšších verzích serveru) dosáhnout zapnutím komprese dat.

Pro porovnávání souborů bylo navrženo a implementováno rozšíření Diffutils, které dovoluje efektivně porovnávat texty, obsahující HTML značky a zobrazovat změny včetně formátování. Z hlediska návrhu aplikace je porovnávání verzí řešeno pluginy, které dovolují přidávat podporu dalších typů souborů v budoucnosti.

Z důvodu obecnosti jsou všechna stažená data ukládána do BLOBů (v MS SQL databázový typ Image). V tomto směru by bylo žádoucí soubory, které pluginy identifikují jako textové, ukládat do CLOBů (v MS SQL databázový typ text). Jednoduše by se tak zpřístupnilo hledání v HTML souborech pomocí vestavěného fulltextového vyhledávání. Opět by měl uživatel tímto zásahem k dispozici aplikaci s větší užitnou hodnotou.

Instalace programu spočívá v přepokopování adresáře z příloženého CD na pevný disk počítače. S konfigurací MS SQL serveru a klientské aplikace podle dokumentace nebyl problém.

Dokumentace návrhu je celkem srozumitelně popsána v páté kapitole. Další podrobnosti je možné najít v generované dokumentaci díky komentářům, ve zdrojových kódech obsažených.

Kvalitu práce tedy sráží především fakt, že autor nedokázal implementačně do všech důsledků využít potenciál vlastního návrhu aplikace. Soustředění se především na vnitřní strukturu programu a některé z hlediska návrhu zajímavější aspekty aplikace vedly k zanedbání řady detailů, které by uživateli zpříjemnily a usnadnily práci a přesvědčily jej, aby do budoucna používal právě a pouze tento program. Přesto se domnívám, že autor splnil požadavky kladené na diplomovou práci a doporučuji ji proto k obhajobě.

V Praze dne 20. 5. 2010

