

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

Autor práce: Jiří Setnička
Název práce: Zálohovací systém
Rok odevzdání: 2015
Studijní program a obor: Informatika, obecná informatika
Autor posudku: Mgr. Martin Mareš, Ph.D., vedoucí
Pracoviště: Katedra aplikované matematiky

K celé práci	lepší	OK	horší	nevyh.
Obtížnost zadání	X			
Splnění zadání		X	X	
Rozsah práce		X		

Předmětem práce je automatické zálohování dat klientských počítačů na společný zálohovací server. Navzdory tomu, že tento problém tvoří nedílnou součást správy většiny počítačových sítí, stávající volně šiřitelné zálohovací programy jsou značně nedokonalé.

Tato práce navrhuje nový zálohovací systém, který se snaží daleko lépe hospodařit s prostorem. Především pro reprezentaci historie souborů používá velmi kompaktní formát, který umí obsah souboru odvozovat od jiných verzí téhož souboru, nebo i souborů jiných. Také se vyrovnává s tím, že má k dispozici pouze omezené množství místa, a v závislosti na tom upravuje hustotu historických verzí souborů.

Velké pozornosti se též těší konfigurovatelnost systému: parametry zálohování je možné nastavovat jak vzhledem k hierarchii adresářů, tak pomocí obecnějších podmínek.

Systém je celkově navržen dobře, až na pár míst, kterým bylo věnováno méně pozornosti, než by si zasloužily – viz níže.

Textová část práce	lepší	OK	horší	nevyh.
Formální úprava		X		
Struktura textu		X	X	
Analýza		X	X	
Vývojová dokumentace		X		
Uživatelská dokumentace		X		

V úvodu práce se seznámíme s problémem zálohování dat a s různými situacemi, kdy dochází k obnově dat ze záloh. Jednotlivé situace přitom kladou na zálohování různé požadavky.

První kapitola představuje základní přístupy k zálohování od prostého kopírování souborů až po různé způsoby detekce, přenosu a kompaktní reprezentace změn. Tato část hraje také roli analýzy problému: jednotlivé přístupy hodnotí a vybírá vhodné vlastnosti

pro navrhovaný systém. Všechny důležité věci jsou zde řečeny, jen postrádám rešerši stávajících zálohovacích programů.

Poté následuje návrh nového systému. Problém je zde rozdělen na několik nezávislých částí: konfigurační pravidla, model ukládání dat a adaptéry pro různé způsoby přístupu ke zdrojovým datům. Chybí zde rozbor strategie pro udržování historie; ta je popsána později v implementační části, ale bez zdůvodnění. Také by bylo užitečné zmínit, co všechno můžeme ve struktuře UNIXového souborového systému potkat: symbolické linky jsou zmíněny pouze okrajově, o hard-lincích, socketech, rourách a speciálních souborech se mlčí. Též se považuje za samozřejmost, že adresářový strom se vejde do operační paměti, což nejspíš platí, ale zasloužilo by si to alespoň ověření hrubým výpočtem.

Další kapitoly přináší uživatelskou a programátorskou dokumentaci. Obě jsou napsané stručně, ale zcela dostatečně.

Závěr dílo hodnotí a uvádí mnoho námětů na budoucí vylepšení.

Celkově nemám proti textové části práce zásadní výtky. Jen bych uvítal, kdyby vlastnosti systému byly ověřeny alespoň jednoduchým experimentem, který by ukázal, jak je navržená reprezentace dat efektivní a jak dobře funguje navržený algoritmus pro „ředění“ historie na skutečných datech. Také by bylo zajímavé vědět, zda je zálohování bržděno spíš diskem, nebo rychlostí procesoru při výpočtu hešovacích funkcí.

Práce je psaná příjemnou, čtivou češtinou s minimem jazykových a typografických chyb. Použité zdroje jsou korektně citovány.

Implementační část práce	lepší	OK	horší	nevyh.
Kvalita návrhu		X		
Kvalita zpracování		X	X	
Stabilita implementace		X		

Implementace je kvalitní. Praktickou použitelnost omezuje to, že kvůli nedostatku času nebyl implementován žádný adaptér pro vzdálený přístup, takže lze zálohovat pouze lokální soubory, případně soubory dostupné po některém ze síťových filesystemů (NFS, SSHFS, apod.).

Také nepovažuji za vhodnou volbu formátu uložení meta-dat. Zde je použita obecná knihovna pro serializaci objektů, takže formát je příliš spjatý s konkrétní implementací. Zejména u zálohovacího systému mi přijde důležité, aby mohlo existovat více nezávislých implementací používajících stejný formát dat, což se tímto rozhodnutím dosti komplikuje.

Použití knihovny `boost` pro platformně nezávislý přístup k souborovému systému se pravděpodobně časem ukáže být příliš těžkopádné, a přesto nedostatečně obecné.

Celkově oceňuji, že se autor zhostil nesnadného úkolu navrhnout univerzální zálohovací systém a vytvořil program, který je ve většině ohledů lepší než ty stávající.

Celkové hodnocení: výborně

Práci navrhuji na zvláštní ocenění: ne

V Praze dne 28. srpna 2015

Martin Mareš