

Ing. Vlastimil Čavala
VS Ingstav Brno

TYPIZACE ALGORITMŮ V OBLASTI ZPRACOVÁNÍ DAT

1. Úvod

Obsahem příspěvku je sdělení o několikaletých pokusech o zavedení systematického popisu algoritmů z oblasti zpracování hromadných dat. Tato činnost je vedena snahou o jistou úroveň typizace a to jak přímo ve zdrojovém textu programu v jazyce COBOL, tak i v širším slova smyslu, tj. při organizaci jednotlivých programů, datových souborů a tvorbě celého analytického řešení vůbec. V závěru jsou pak příklady možností klasifikace algoritmů a definice některých konkrétních užitečných pojmů.

Celá práce vychází ze zkušeností praxe na magnetopáskově orientovaném počítači TESLA 200, takže je zaměřena na sekvencně uspořádané datové soubory při dávkovém zpracování dat, decentralizovaně pořizovaných z černé pásky.

Hlavní myšlenky, které charakterizují popisované snažení jsou asi tyto:

- Nejen při návrhu programu, ale ani při stanovení koncepce celého automatizovaného řešení nelze obecně spoléhat na intuici, ale je nutno postupovat přísně systematicky.
- I v tak složité oblasti, jakou tato činnost nezporně je, musí existovat nějaké možnosti systematického popisu, definice určitých charakteristik každého řešení a tím i možnosti zevšeobecnění. Při hledání těchto možností je třeba postupovat frekvenčně, tj. pokusit se popsat především nejčastěji se vyskytující způsoby zpracování.

- Součástí pokusů o typizaci musí být definování a zavedení některých jednoznačně chápaných pojmů, které pak nejen umožňují snadnější komunikaci, ale především srozumitelně charakterizují řešení a ukazují jeho náročnost. To ve svých důsledcích zajišťuje podstatně účinnější plánování, řízení a kontrolu celého procesu vývoje automatizovaného řešení.

Vše, co je v tomto materiálu uvedeno si samozřejmě nečiní nárok na úplnost či reprezentativnost, ani na optimální postup - spíše jde o ukázkou jednoho z možných přístupů.

2. Systematický popis řešení problému

Dále jsou popsány typové algoritmy řešení vstupního modulu, jehož funkci vymezíme na "pořizení dat z čírné pásky do základního souboru na magnetické pásce a průběžné započítávání jeho potřebné aktualizace". Algoritmy jsou zpracovávány ve 3 úrovních, počínaje celkovou koncepcí, přes organizaci souborů a programů, až k podmínkám vybraného úseku včetně programového řešení.

Jako metoda vyjádření je použit programový popis, tj. postup po jednotlivých krocích s případnou volbou alternativ. Všechny kroky jsou postupně označovány abecedně, alternativy číselně s možnostmi 1 až 9. Číslo 0 v určitém kroku pak znamená, že jeho náplň platí pro všechny alternativy následujícího kroku shodně. Tato forma hierarchického stromu pak umožňuje stručnou číselnou definici zvoleného konkrétního popisu - bližší by mělo být z uvedených příkladů.

2.1 Typový popis koncepce pořizení dat

/vstupní modul - úroveň 1/

0. Standardní způsob pořizení dat

- čírné pásky /DP/ jsou děrovány úrovněvě, t.j. opakovaně údaje /záhlaví dokladu, částka apod./ jsou děrovány

pouze jedenkrát, a přetřehovány při standardní vstupní konverzi do jednotlivých vět pevné délky

- kontrola, zda byly všechny do VS dodané pásky nahrány do počítače je řešena standardně v rámci provozu VS

A1. Jednorázové pořízení bez možnosti oprav

- provedení není reálné, protože chyby v identifikačních a třídících údajích se vždy vyskytnou a získaný soubor pak není prakticky použitelný
- únosně lze řešit u souboru, který se udržuje na DŠ, a vždy znovu pořizuje, změna položky na určitém DŠ je sice snadná, ale zůstávají problémy s archivací, zadáváním v provozu a s možností ztráty či přeházení DŠ, poněvadž provoz není na zpracování běžně zařízen
- opravy a opětovné nahrávání děrných pásek pracné, nároky na nestandardní individuální postup v provozu VS, mohou vznikat neidentifikovatelné rozdíly při opakovaném nahrávání vinou rozdílného výsledku snímání DP

A2. Postupné pořizování s možností oprav a aktualizace základního souboru /ZS/

- chyby v křivce očí při děrování z DP se zásadně ignorují a chybné položky se do žádné sestavy netisknou, poněvadž při úrovnovém děrování může dojít k libovůli, a není možná jednoznačná identifikace /která by mj. vyžadovala i programátorskou kvalifikaci/
- tištěný protokol o zpracování dávky musí vždy umožňovat kontrolu úplnosti, tj. které datové položky byly a které nebyly pořízeny s tím, že otázka, zda to bude ručním odškrtnutím, nebo na základě nějaké počítačové kontroly, je řešena na dalších úrovních

- není vhodné tisknout ani statistiku ignorovaných položek, poněvadž může vést k nesprávným závěrům /položek může chybět víc, než je signalizováno - a to ať již z důvodu nesprávné křivky očí, nebo i proto, že bylo opomenuto jejich vyděrování/

B21. Pracoviště, které vystavilo prvotní doklad se účastní kontroly a oprav /za takové pracoviště může být považováno i centrální řídicí pracoviště odborných útvarů, které má v náplni příslušnou věcnou problematiku/

C211. Kontrola správnosti vystavení dokladů i pořizení dat se provádí současně /zodpovídá za ni jeden útvar/

- nelze provádět v děrovně, poněvadž ta nemůže zodpovídat za správné zachycení skutečnosti na dokladě
- může provádět pouze odborný útvar, který v tom případě však bere na sebe i opravy případných chyb z pořizení dat, tj. z děrování a snímání /děrování může být pouze přizvána ke spolupráci/

C212. Kontrola správnosti vystavení dokladů i pořizení dat se provádí samostatně /zodpovídají za ni dva útvary, každý za svoji část/

- nutné přesné rozdělení zodpovědnosti mezi děrovnou a odborným útvarem, včetně určení termínů dílčích předávání
- v případě dobré spolupráce možno provádět společně, ale výše uvedenou dílčí zodpovědnost nutno stanovit v každém případě

B22. Pracoviště, které vystavilo doklad se neúčastní kontroly a oprav /ani žádné jiné zodpovědné pracoviště odborných útvarů/

- je možno opravit pouze chyby v pořizení dat, tj. v přenosu informací mezi dokladem a počítačem, tedy chybné nebo neúplné děrování a případné chyby snímání

- chyby ve vystavení dokladu lze spolehlivě opravit jen výjimečně, prakticky pouze případy chybného kontrolního součtu na dokladě, tj. když je jednoznačně určena forma opravy
- případy chyb ve vystavení dokladu, kdy kontrolní dispečerské lístky /průvodky/ udávají jiná čísla nebo počty dokladů, než je jejich fyzický stav již nejsou zcela jednoznačné, poněvadž může případně jít o ztrátu, nebo neoprávněný přebytek dokladů
- zásadně nelze opravovat žádné datové údaje, u kterých je signalizována logická chyba a data jsou z dokladu pořízena správně, poněvadž tím by mohlo dojít k neoprávněné změně jejich zpracování /sem patří i případ chybné kontrolní číslice, poněvadž může jít i o chybný základ - např. přepsání v čísle skladu/
- kontrola a opravy chyb se provádějí prakticky pouze v provozu VS
- realizovat počítačové kontroly na správnost vyatavení dokladů či úplnost ^{veš} dávky z odborného útvaru, které by nebyly využitelné pro kontrolu pořízení dat, ale měly by za cíl odhalovat další logické chyby v tomto případě nemá smysl, poněvadž získané informace o chybách nemohou stejně vyvolat žádnou činnost k nápravě
- odborný útvar musí mít možnost zjistit chybu dodatečně a to zpravidla ručním odkontrolováním, nebo z dalších dopadů, a musí existovat nějaká možnost dodatečné nápravy, je-li to potřeba /požití zvláštního protokolu by vzhledem ke ztrátě aktuálnosti a v souvislosti s ekonomickými hledisky mělo být spíš výjimečné/

Komentář:

Takto zpracovaný popis by měl být ještě doplněn požadavky na definici dalších potřebných charakteristik pořizovaných dat - např. objemových, časových místních apod.

Jak je z popisu zřejmé, jednotlivé formulace jsou stavěny tak, aby uživatel typového řešení kromě daných výchozích podmínek viděl při volbě alternativy na první pohled též důsledky svého rozhodnutí, které nelze pomínout.

Konkrétní popis zvoleného řešení může být proveden buď označením alternativ do předtištěného textu typového popisu nebo výhodněji tzv. číselnou definicí, která využívá stromové struktury označení.

Např. číselná definice O212 znamená kroky:

- O Standardní způsob pořizování dat
- A2 Postupné pořizování s možností oprav a aktualizace ZS
- B21 Pracoviště, které vystavilo doklad se účastní kontroly a oprav
- C212 Kontrola správnosti vystavení dokladů i pořizování dat se provádí samostatně

2.2 Typový popis podmínek organizace datových souborů a programů /vstupní modul - úroveň 2/

0. Základní standardní sestavení

- vstupní konverze do aktualizací souboru /AS/
- třídění AS na základní třídění základního souboru /ZS/
- vlastní aktualizace ZS s vytvořením protokolu o změnovém řízení a 2 shodných kopií nového výstupního základního souboru

A1. Při aktualizaci není povolena změna v popisovači /tj. v pásmu, které je součástí klíče základního třídění/

A2. Povolena změna v popisovači aktualizace

- nutné třídění a zatřídění změněných položek, s aktualizací v další konverzi /do které může být zatřídění případně přímo začleněno/

- B1. Protokol je v základním třídění
- B2. Protokol není v základním třídění
- nutné třídění souboru protokolu a samostatná konverze jeho výstupu
- C1. Soubor protokolu není třeba v počítači archivovat
- C2. Soubor protokolu je třeba archivovat
- nutný archivovaný soubor, ke kterému se postupně přidávají údaje s informacemi o provádění aktualizací
- D1. Při aktualizaci není nutná vazba na další základní kartotéku s jiným tříděním
- D2. Nutná vazba na další jinak tříděnou kartotéku
- nutné třídění na novou základní kartotéku, slučování a zpětné setřídění na základní soubor
- E1. Není nutné řízení většího položkového rozpisu dle výsledku kontroly
- E2. Je nutné řídit položkový rozpis většího počtu údajů s vazbou na výsledek celkové kontroly /nap. úplnosti/
- nutné přetřídění souboru s pracovním výsledkem kontroly, nebo udržování samostatného souboru s kontrolními údaji
- F1. Protokolem je samostatný výpis chyb nebo statistiky
- nutný výstupní soubor na tiskárně nebo mgp
- F2. Není samostatný protokol o chybách
- G1. Protokolem je samostatný opis správných pořizovaných položek
- nutný výstupní soubor na tiskárně nebo mgp
- G2. Není samostatný protokol o pořizovaných položkách

H1. Protokolem je společný výpis správných položek i chyb
/vstupní protokol o aktualizaci/

- nutný výstupní soubor na tiskárně nebo mzp

H2. Není společný protokol o aktualizaci

I1. Další podmínky pro využití ZS nejsou stanoveny

I2. V základním souboru ZS jsou po dobu provádění aktualizace trvale uchovávány chybné nebo kontrolní věty a pro další použití má soubor ^{obsahovat} pouze správné věty

- nutná samostatná konverze s příp. tiskem protokolu

Komentář:

Jednotlivé alternativy v žádném kroku nemají hlubší členění na dalších úrovních, takže bylo možno všude použít pouze jednoduchého označení úrovně a varianty.

Zvláště zde je nutno zdůraznit, že formulace ~~typového~~ popisu musí vycházet z důkladné analýzy všech možností reálného řešení, aby byly vybrány skutečně jednoznačně podložené varianty.

Pro příklad dále uveďme jako číselnou definici konkrétního popisu hodnotu 0111 112 211, která určuje jednoduché změnové řízení, bez možností změn v popisovači, s přímo vystupujícím kompletním protokolem v základním třídění, bez dalších kontrol a vazeb.

2.3 Typový popis podmínek programového řešení vlastní aktualizace /vstupní modul - úroveň 3/

Vzhledem k omezenému rozsahu se zde omezíme pouze na neúplný výčet základních charakteristik.

a/ aktualizace:

základní funkce = vložení a změna s tím, že za zvláštní změnu se považuje též storno

položková změna = jedním nebo více impulzy z AS
se mění pouze 1 věta v ZS

skupinová změna = jedním nebo více impulzy z AS
se mění skupina vět v ZS /určení skupiny
pro skupinovou změnu může být buď řádem, tj.
délkou údaje, nebo od-do/

změna na úrovni věty = způsob zpracování umožňu-
je měnit pouze celé věty nebo skupiny vět

změna na úrovni slova = způsob zpracování umožňu-
je měnit jednotlivá slova v rámci jedné
věty nebo skupiny vět

b/ kontroly:

- duplicita popisovačů v AS při změnách na úrovni věty
- návaznost dávek nebo platnosti
- kontrola jednotlivých údajů nebo jejich vazeb v rámci jedné věty
- kontrolní součtování na počet položek nebo na součet určité hodnoty
- kontrola úplnosti

c/ vystup:

- prostá nebo stupňová konverze
- zpracování po položkách nebo po skupinách, příp. uložených ve vyrovnávací paměti

3. Možnosti klasifikace algoritmů v programech

Během několikaletého využívání algoritmů uspořádaného programování /1/, /2/, /4/ ve VS Ingstav i v řadě dalších výpočetních středisek, kam byly tyto podklady rozšířeny - jako např. /3/ se ukázalo, že takové typové podklady jsou skutečně široce prakticky použitelné. Proto jsou ve VS Ingstav systematicky dále rozpracovávány tak, aby postupně

pokryly realizace všech nejčastěji vyžadovaných funkcí.

V této souvislosti je v dalších odstavcích předkládán pokus o klasifikaci určité třídy úloh z oblasti zpracování hromadných dat a příslušnými definicemi jejich charakteristik. Pro identifikaci jednotlivých vět je zaveden pojem "popisovač" /deskriptor, klíč/, který reprezentuje větu při vyhodnocování logických vztahů, určujících způsob zpracování.

prostá konverze /program nebo podprogram/

- nejjednodušší případ sekvenčního zpracování vstupního souboru, při kterém nezáleží na jeho setřídění
- každá čtená věta se zpracuje stejně, příp. může být proveden i výběr s dalším větvením nebo závěrečná rekapitulace

použití: úprava souboru, nesoučtované výpisy obsahu mgp, výběry a rekapitulace apod.

stupňová konverze /program nebo podprogram/

- zpracování vstupního souboru po skupinách vět se shodným popisovačem na jednotlivých stupních vstup je setříděn dle popisovačů pro vícestupňové zpracování /sestupný sled stupňů/
- při změně v popisovači určitého stupně je prováděno "zakončení" a "zahájení" zpracování příslušné skupiny

použití: obecný algoritmus součtovaného výpisu, stupňovitě členěné výpisy obsahu mgp apod.

řetězená konverze /program nebo podprogram/

- řetěz několika za sebou jdoucích stupňových konverzí z jednoho vstupního souboru
- data každé sestavy jsou označena indikačním klíčem, a zpracování je prováděno nezávisle samostatným algoritmem řízení výstupu

použití: plynulý tisk řetězu několika zcela různých sestav z příslušně uspořádaného souboru

slučování /program nebo podprogram/

- společné zpracování dvou nebo více sekvenčně shodně seříděných souborů
- na základě srovnání jejich popisovačů je pak zpracování větveno do všech vzájemných kombinací přítomnosti a nepřítomnosti odpovídajících vět

použití: obecný algoritmus jakéhokoliv změnového řízení nebo předávání informací mezi soubory

elementární aktualizace /program/

- nejjednodušší případ aktualizace sekvenčně uspořádaného souboru vkládáním, záměnou a stornem na úrovni věty
- v podstatě jde o slučování vstupního aktualizacího a základního souboru v kombinaci s výstupní stupňovou konverzí

použití: obecný algoritmus, zajišťující minimum funkcí, potřebných pro nejjednodušší vstupní modul

4. Literatura

- /1/ Sborníky přednášek ze seminářů DT ČVTS Ostrava, "Metody programování počítačů", Havlíkov 1976 a 1978
- /2/ V. Čevela, Algoritmy UP, VS Ingstav Brno 1975 až 1979
- /3/ V. Škarka, Strukturované programování - metodika, ÚVT Ostrava, 1977
- /4/ V. Čevela, Elementární aktualizace sekvenčního datového souboru, Sborník ze sympózia SVTS Bratislava, Algoritmy '79