METODIKA KONVERZÍ VSTUPNÍCH DAT

Ing. Jiří Turjanica

V posledních letech došlo v Podniku výpočetní techniky Praha v oblasti vstupních dat ke dvěma významným kvalitativním změnám:

a) podstatně vzrostl počet typů strojů pro pořizování vstupních dat používaných pro pořizování dat projektů rutinně zpracovávaných ve výpočetních střediscích PVT Praha a spolu s tím i počet druhů médií, na které se data pořizují;

b) pořizování vstupních dat se přeneslo z pořizoven výpočetních středisek PVT Praha k zákazníkům.

Tyto změny měly za následek, že u většiny projektů rutinně zpracovávaných v PVT Praha jsou data jednotlivých vstupních souborů pořizována u různých uživatelů na různých pořizovacích strojích, tj. různým způsobem a na různá média. V současnosti je pořizování většiny souborů všech celopodnikových projektů zabezpečeno na SPÚ 800, Videoplexu 3, Laser One/Two a na minimálně jednom druhu pořizovače děrné pásky nebo magnetické kazety. Běžně dochází k situacím, že při zavedení projektu u dalšího uživatele je nutno vyprojektovat a naprogramovat pořizování vstupních souborů na dalším typu pořizovače. Velmi rychle proto přestal být efektivní dosavadní způsob řešení, kdy kontrolní chod přímo čte médium, na které byla vstupní data pořízena. Proto byla v roce 1981 metodickým pokynem č. 25/81 zavedena jednotná vstupní magnetická páska souboru (JMP) jako jediný vstup do kontrolního chodu. Jednotná vstupní magnetická páska souboru obsahuje jeden soubor, který má individuální strukturu. Pro každý vstupní soubor je vytvářena jedna vstupní magnetická páska souboru.

Tímto zásadním řešením se jednoznačně oddělilo pořízení vstupních dat od dalšího zpracování.

Pořízení vstupních dat souboru zahrnuje jejich pořízení na média a konverze dat z médií, na kterých byla pořízena, do souboru na jednotné vstupní magnetické pásce souboru.

Zpracování vstupních dat začíná kontrolním chodem, do kterého vstupuje soubor na jednotné vstupní magnetické pásce souboru.

Rozdělení technologického procesu rutinního zpracování projektu na pořízení vstupních dat a na další zpracování umožnilo komplexní metodické řešení problematiky projekce a programování pořízení vstupních dat jako jedné z nejnáročnějších technologických etap rutinního zpracování projektů. Součástí tohoto řešení bylo ustavení celopodni-

kové skupiny specializované na projekci a programování pořizování vstupních dat, jejíž činnost je specifikována v metodickém pokynu č. 25/81. Dále byly zpracovány metodické materiály, řešící problematiku vstupních dat:

- metodika konverzí vstupních dat včetně příslušného programového vybavení,
- část B - pořizování vstupních dat - metodiky zpracování provozní dokumentace,
- metodika zpracování programové dokumentace pořizovacích strojů,
- knihovní řád knihovny vstupů.

Mezi těmito metodickými materiály patří k nejvýznamějším metodika konverzí vstupních dat. Řeší problematiku konverze dat souboru pořízených na vstupním médiu na jednotnou vstupní magnetickou pásku tohoto souboru. Řešení zavedení metodikou připouští tyto možnosti:

a) jeden projekt se ve výpočetním středisku zpracovává pro několik uživatelů, z nichž každý má svoje vlastní data;

b) zpracování projektu probíhá s daty za určité období najednou, ale vstupní data jsou pořizována a nahrávána v několika dávkách;

c) v rámci jednoho projektu jsou pořizována a nahrávána data několika různých souborů;

d) data každého souboru mohou být pořizována na několika různých typech pořizovačů na stejné nebo různé druhy médií;

e) data na jednom druhu média mohou být pořízena na libovolném počtu médií.

Aby bylo možno problematiku konverzí vstupních dat řešit, bylo třeba zavést určité pojmy, které umožňují popis problematiky a popis způsobu řešení. Byly proto zavedeny pojmy uživatel, soubor období, dávka, skupina dat a médium, které lze charakterizovat takto:

uživatel: jeden projekt může být v rámci jednoho výpočetního střediska rutinně zpracováván pro několik uživatelů, z nichž každý má svoje vlastní data,

soubor: v rámci jednoho projektu existuje jeden nebo několik vstupních souborů. Pro každý z nich se vytváří jiná jednotná vstupní magnetická páska souboru (jiná = s jinou strukturou souboru),

období: rutinní zpracování projektu probíhá s daty za určité období (dekádu, měsíc a pod.),

dávka: data za určité období mohou být nahrávána v jedné nebo několika dávkách. Nahráváním dat jedné dávky jednoho souboru vznikne jedna jednotná vstupní magnetická páska tohoto souboru,

skupina dat: data jednoho souboru pořízená na určitém typu pořizovače. Data jedné dávky jsou tvořena daty tolika skupin dat, na kolika různých typech pořizovačů se soubor pořizuje,

médium: jeden fyzický celistvý nosič dat.

Metodika konverzí vstupních dat rozlišuje z hlediska celkové logické struktury záznamu uloženého na média libovolného druhu tyto části:
- záhlaví média,
- datová část,
- ukončení média.

Datová část média obsahuje věty souboru vstupních dat v hierarchii a logické struktuře předepsané projektem. Ukončení média je zajištěno, buď standardně, tj. pomocí EOF - bloku, nebo je ošetřeno ukončovací řídící větou. Je specifické pro každý druh média.

Pro kontrolu a řízení konverze má největší význam záhlaví média. Identifikuje médium a je tvořeno těmito částmi:
- heslo,
- druh dat,
- pořadové číslo,
- doplňková identifikace.

Pro každý druh média, případně pro typ pořizovače, metodika předepisuje způsob zápisu částí, tvořících záhlaví média.

Části heslo, druh dat a pořadové číslo média jsou povinné části záhlaví a mají předepsáno vyplnění pouze numerickými hodnotami, protože je musí být možno pořídit i na těch pořizovacích strojích, které nemají abecední klávesnici.

Jednotlivé části záhlaví mají následující význam a strukturu:

Heslo je tvořeno deseti znaky a skládá se z položek úloha, soubor, uživatel a období. Všechny tyto položky jsou konstantní pro data jedné zakázky, jednoho souboru a uživatele za určité období. Heslo umožňuje kontrolovat, zda data na médiu přísluší k datům, která operátor nahrává a pro něž hodnotu hesla zadal při zahájení konverze.

Druh dat je tvořen dvěmi znaky a spolu s jednoznakovým označením druhu média (to však není v záhlaví uvedeno) vytváří hodnotu "skupina

dat". Druh dat je tvořen položkami: struktura dat a specifikace dat. Položka struktura dat rozlišuje různé struktury dat na vstupním médiu v rámci téhož souboru na jednotné vstupní magnetické pásce. Na př. data téhož souboru pořízená na VPX 3 a SPÚ 800 budou mít různou hodnotu položky struktura dat, protože struktura jejich zápisu na magnetické pásce a na magnetické kazetě bude různá. Pokud by byla stejná, mohla by položka struktura dat v záhlaví obou médií být stejná. Přidáním znaků označujícím druh média by však i tak vznikly dvě různé hodnoty údaje skupina dat. To je důležité, protože pro každou hodnotu prvních dvou znaků údaje skupina dat je vytvářena jedna výkonná fáze, která umí číst tento druh média s touto strukturou zápisu dat. Z uvedeného vyplývá, že pokud by data určitého souboru byla pořizována sice na různých pořizovačích, ale na stejný druh média a se stejnou strukturou, mohla by mít stejnou hodnotu skupiny dat a být čtena stejnou výkonnou fází. To však neplatí, protože by to znemožnilo kontrolu opakovaného nahrávání téhož média, která se provádí v rámci skupiny dat při nahrávání dávky. Proto musí být v tomto případě zadána taková hodnota položky struktury dat, aby vznikly dvě různé hodnoty skupiny dat. Položka specifikace dat je vyhodnocována výkonnou fází a umožňuje zvolit pracovní režim výkonné fáze, která čte médium této skupiny dat.

Pořadové číslo média je tvořeno pěti znaky a skládá se z položek číslo pořizovny, číslo stroje a číslo média. Tyto položky zabezpečují, že hodnota pořadového čísla média bude navzájem různá na všech médiích téhož uživatele, souboru, skupiny dat a dávky. To umožňuje pomocí pořadového čísla média kontrolovat duplicitu nahrání v rámci médií jedné skupiny dat.

Doplňková identifikace je nepovinná část záhlaví média o rozsahu max. 15 znaků. Je-li použita, potom se její obsah při konverzi nekontroluje, je pouze zabezpečen její opis v protokolu o snímání dávky.

Metodika konverzí vstupních dat člení zpracování z hlediska práce se vstupními daty na tyto úrovně:

- skupina dat,
- dávka,
- kontrolní chod,
- zpracování podsystému

V rámci jednotlivých úrovní jsou prováděny činnosti pro tuto úroveň předepsané a zabezpečené konverzemi stanovenými metodikou konverzí vstupních dat.

Skupina dat: údaj skupina dat slučuje média zpracovatelné jednou větví výkonné fáze konverzního programu.
V rámci skupiny dat kontroluje konverzní program duplicitu média. Duplicita je kontrolována v rámci jednoho vyvolání konverzního programu, tj. i při opakovaném vyvolání téže výkonné fáze. Prakticky to znamená, že jednotlivá média určité skupiny dat nemusí být čtena bezprostředně po sobě. Údaj skupina dat zadává konverznímu programu operátor v souladu s předávacím protokolem snímání dávky.

Dávka: je tvořena médii jedné nebo několika skupin dat, která jsou konvertována při jednom vyvolání konverzního programu. Všechna data dávky mají stejnou hodnotu údaje "heslo" v záhlaví jednotlivých médií. Výsledkem zpracování dávky je vytvoření jedné vstupní magnetické pásky souboru. Tento vytvořený soubor má zapsán v počátečním uživatelském návěští hodnotu údaje "heslo" a číslo dávky.

Kontrolní chod: může být do zpracování začleněn zásadně dvěmi způsoby:

- kontrolní chod navazuje bezprostředně na zpracování každé dávky konverzním programem. Do kontrolního chodu tak vstupuje vždy jediná jednotná vstupní magnetická páska souboru,
- kontrolní chod zpracovává celou množinu jednotlivých vstupních magnetických pásek souboru ze všech nahraných dávek se stejným heslem.

Začlenění kontrolního chodu do systému zpracování je nezávislé na provádění konverzí. Obecně platí, že do kontrolního chodu vstupuje jednotná vstupní magnetická páska souboru (JMP) vzniklá konverzním programem, z kontrolního chodu vystupuje "MP-soubor" pro další zpracování podsystému.

Předávání vstupních dat podsystému do kontrolního chodu na JMP bez ohledu na to, na jakém druhu vstupního média byla pořízena má tyto výhody:

- JMP je dělícím bodem, na který jsou z jedné strany napojeny konverzní programy transformující vstupní data z různých druhů médií na strukturu JMP a z druhé strany programy logických kontrol zabezpečující jednotným způsobem další práci se souborem dat,
- použití principu JMP v podsystému poskytuje možnost nezávislého programování vstupních a konverzních programů a programování dalších programů podsystému,
- u malých počítačů odpadají při oddělení konverze od logických kontrol problémy s kapacitou vnitřní paměti,

- JMP umožňuje kdykoliv zařadit do podsystému novou skupinu dat,
- při změně způsobu pořizování dat se změní pouze konverzní program. Program logických kontrol zastává beze změny, nevyžaduje-li tuto změnu sám projekt.

Logické věty na JMP mají doporučenou pevnou délku. Pro strukturu logické věty nejsou předepsána žádná omezující pravidla.

Zpracování podsystému: do zpracování podsystému vstupuje "MP-soubor", který vznikl zpracováním všech JMP z dávek se shodným obsahem údaje heslo kontrolním chodem.

Metodika konverzí vstupních dat jednoznačně předepisuje struktura a funkce konverzního programu, který musí být strukturován tak, že existuje řídicí program a řada výkonných fází.

Pro nejfrekventovanější operační systémy používané v PVT Praha (MOS, DOS, DOS 3) byl vytvořen standardní řídicí program CONVER jako součást standardního programového vybavení (SPV). Tento program vytváří soubor na JMP s větami formátu FIXBLK resp. FIXUNB s parametricky volitelnou délkou bloku (max. 1800 bytů).
Řídicí program konverze zajišťuje:

a) Řízení celé konverze, tj.:
   - zadání hesla a jeho formální kontrolu,
   - zadání čísla dávky,
   - pokyn operátorovi o zadání skupiny dat,
   - kontrola přípustnosti zadané skupiny dat,
   - zápis bodu opakování před snímáním každého vstupního média.

b) Kontrolu snímaných médií, tj.:
   - kontrola hesla v záhlaví média,
   - při nesouhlasu hesla v záhlaví 1. média dávky možnost opravy zadání hesla,
   - kontrola druhu dat v záhlaví média pro danou výkonnou fázi,
   - kontrola duplicity médií (podle pořadového čísla média v rámci skupiny dat),
   - vyhodnocení počtu chybných vět za médium, v případě překročení povoleného procenta možnost opakovaného snímání nebo vyloučení média, resp. pokračování ve snímání.

c) Výstup dat na JMP, tj.:
   - případná modifikace tabulky DTF výstupního souboru na JMP,
   - otevření výstupního souboru na JMP,
   - generování a zápis věty UHL a UTL,

- zápis vět výstupního souboru,
- uzavření výstupního souboru na JMP.

d) Protokolování konverze, tj.:

- tisk průběžného a závěrečného protokolu konverze.

Program CONVER umožňuje ošetření atypických případů vstupních médií - tj.:
- výkonné fáze mohou zpracovávat média bez záhlaví,
- výkonné fáze zpracovávají média se záhlavím, ale předávají řídícímu programu požadavek, aby nebylo kontrolováno.

V těchto případech ovšem řídící program neprovádí kontrolu duplicit.

Výkonné fáze se programují individuálně s použitím prvků SPV (READVOL1, CHYBBYTE, ICTP8IN, GETREC, soustava DP ap.).

Volbu výkonné fáze a její případné vnitřní členění určuje jednoznačně skupina dat zadaná operátorem v řídícím programu.
Výkonná fáze zajišťuje:

- naplnění údajů potřebných pro řídící program (název snímané skupiny dat pro protokol, maximální povolené procento chybných vět za médium),
- výpis standardní zprávy na OPS (příprava snímání médií příslušného druhu),
- volbu operátora o pokračování (případně ukončení) snímání médií dané skupiny dat,
- otevření vstupního souboru dat (podle druhu média),
- přečtení záhlaví média a předání řídícímu programu, případně indikaci atypických případů (médium bez záhlaví nepo požadavek, aby záhlaví nebylo kontrolováno),
- ošetření stavu, kdy je řídícím programem indikována chyba záhlaví (nesouhlasí heslo nebo nesouhlasí druh dat nebo duplicita),
- čtení vstupních dat a jejich transformaci do tvaru výstupní věty, tj. naplnění výstupní věty, kontrola formální správnosti přečtených údajů, indikace chyb,
- podání informace řídícímu programu, zda je předávána správná či chybná věta,
- uzavření vstupního souboru dat (podle druhu média).

Pro programování výkonné fáze lze s výhodou použít prvek SPV - GETREC, který zabezpečí zpracování vstupních dat parametricky zadaným způsobem.

Konverzní program provádí potřebné formální kontroly údajů a informace o zjištěných chybách ve větě ukládá do jednotlivých bytů chybové byte.

Naplnění a dekódování chybového byte je zabezpečeno příslušným prvkem SPV - CHYBBYTE.

Pro chybový byt byla stanovena závazná struktura, která jednotlivým bitům chybového bytu přiřazuje význam indikace určité chyby.

Kontrolu úplnosti konverze dávky provádějí provozní pracovníci PVT. Kontrola úplnosti spočívá ve dvou základních krocích:

- Kontrola nahrání všech předaných médií prováděná operátorem počítače během zpracování konverze. Operátor v průběhu zpracování poznamenává na předávací protokol, která média zpracoval. Konverzi operátor ukončí po zpracování všech předaných médií.
- Kontrola úplnosti dávky, kterou provádí pracovník provozní kontroly, respektive provozní programátor odpovědný za zakázku. Podklady pro kontrolu úplnosti mu poskytuje konverzní program formou "Protokolu o konverzi vstupních dat". Vyhodnocením zjištěných výsledků jsou získány informace o úplnosti sejmuté dávky.

Konkrétně je postup při provádění kontroly úplnosti vstupního souboru stanoven provozní dokumentací daného projektu. Nedílnou součástí postupu musí být způsob řešení situací, kdy veškerá média nemohla být konverzí zpracována.

Protokol o konverzi vstupních dat tvoří sestava, která se skládá ze dvou částí - průběhu zpracování a rekapitulace výsledku zpracování.

V hlavičce každé stránky protokolu je název podsystému, číslo dávky, datum, čas a číslo stránky.

Průběh zpracování - je protokolováno každé snímané médium.

Rozlišení nahraných a vyloučených médií je kromě textu provedeno opticky (nahraná média se protokolují v levé polovině sestavy, vyloučená v pravé).

Za jednotlivá média se protokoluje:
- název (dodaný výkonnou fází),
- skupina dat (kódově),
- záhlaví média (tj. heslo, druh dat, pořadové číslo média, doplňková identifikace),
- zda bylo médium nahráno, v tom případě počet výstupních vět a z toho chybných,

- zda bylo médium vyřazeno a z jakého důvodu (chybné záhlaví, duplicita, překročen limit počtu chyb).

Rekapitulace výsledku zpracování obsahuje počet sejmutých médií, počet zapsaných vět, z toho chybných vět, procento chybných vět, a to sumárně za:

- jednotlivé skupiny dat,
- dávku.

Metodika konverzních programů byla vydána s platností pro všechny úlohy řešené v PVT Praha a úlohy AIS NV řešené v PVT Bratislava. Uvedené metodické zásady jsou platné pro všechny v PVT provozované operační systémy.

Literatura:

/1/ Metodika konverzí vstupních dat, PVT Praha 1984

/2/ Metodický pokyn č. 25/81 - Zabezpečení jednotného systému programové přípravy pořizování vstupních dat, PVT Praha 1981

/3/ Metodika zpracování programové dokumentace pořizovacích strojů, PVT Praha 1982

/4/ Metodika zpracování provozní dokumentace, část B - pokyny pro pořízení vstupních dat, PVT Praha 1984

/5/ Knihovní řád knihovny vstupů, PVT Praha 1983.