

Ing. Václav Rákosník, VÚMS Praha.

Příspěvek podává stručnou definici nového programovacího prostředí vhodného pro hromadné zpracování dat. GEPAS je vyvinut ve VÚMS Praha a má být jednou z komponent operačního systému DOS 4.

1. Stručná charakteristika hromadného zpracování dat

Hromadné zpracování dat se vyznačuje rozsáhlými vstupními a výstupními informacemi s rozmanitým uspořádáním. Tok informací směřuje od vstupních dat přes datovou základnu k výstupům. Na vstupu bývají data kontrolována z hlediska formální a logické správnosti, chyby jsou protokolovány a na základě těchto protokolů jsou prováděny opravy chybných dat. Ze vstupních dat vzniká nebo je aktualizována datová základna. Nosným prostředkem pro tuto základnu bývá nejčastěji disk a data, která se někdy označují jako stavová, bývají ještě archivována na magnetických páskách. Ze charakteristické operace na datové základně lze označit střádání dat, spojování dat a jejich třídění. Pro výstupní data je většinou používána řádková tiskárna. Data zde vystupují v maticovém uspořádání s mezisoučty za skupiny klíčů a se součty konečnými.

Dále je pro hromadné zpracování dat charakteristické, že probíhá v různých časových periodicitách /např. denních, měsíčních apod./, které mohou a nebo nemusí být časově souměřitelné. Pro každou periodicitu se volí zvláštní zpracovatelský běh. Existují ovšem i takové běhy zpracování, které mohou být zařazeny kdykoliv a tyto běhy můžeme označit za aperiodické.

Data lze z hlediska významu dělit na data běžná a data stavová. Běžná data mají účinnost pouze v rámci jedné periodicity a aktualizují data stavová. Pokud jsou běžná data archivována, je to spojeno s možností pozdějšího opakování dané periodicity zpracování. Data stavová mají v podstatě dvojitý charakter, a to číselníků nebo vnějších střádačů. Liší se jak použitím, tak aktualizací.

Číselníky zpravidla k určitému klíči obsahují podrobnější popis. Jedná-li se např. o osoby, pak k určitému osobnímu číslu jsou uvedeny jméno osoby, adresa bydliště, rodinný stav, hod. plat ap.

Tyto údaje se aktualizují pomocí běžných dat tvrdou záměnou.

Vnější stráně, jsou někdy označovány za data historická/nadále je používána zkratka HD data/, obsahují k danému klíči hodnoty vztažené k určité časové periodě. Tyto hodnoty se aktualizují kumulací dat běžných. Čas od času se data k určité periodicitě nulují a kumulují znova.

Data v datové základně jsou uspořádána ve větách, ty jsou součástí souborů v nečastějších organizacích sekvenčních /SAM/ nebo indexsekvenčních /ISAM/.

Programy, které operují na datové základně lze v podstatě dělit na tři základní typy - kategorie.

Programy I.kategorie zabezpečují vstup dat do datové základny. Data se při konverzi /z pravidla z děrných štítků na disk/ kontrolují a to nejčastěji na numeričnost, správnost klíčů a pomocí kontrolních součtů na správnost dat a úplnost vstupního souboru nebo jeho skupin. Programy zabezpečují možnost protokolování chyb a umožňují opravy chybných dat.

Programy II.kategorie mají reorganizační charakter datové základny. Do této kategorie se nechají zahrnout programy třídící a konverzní. Konverzní existují v několika variantách, ať už jde o konverze ve spojení s kombinacemi organizací souborů /SAM, ISAM/ nebo o konverze s kombinacemi médií /páska, disk/. Pro tyto programy je vstupem i výstupem datová základna.

Programy III.kategorie zabezpečují výstupní informace a algoritmickou část v systému hromadného zpracování dat. " ohledem na tu skutečnost, že právě požadavek produkování výstupních informací je důvodem, proč vůbec dochází ke zpracování dat, jsou tyto programy v celém systému nejdůležitější. Pro tyto programy je vstupem i výstupem datová základna, ale nejčastějším výstupem je řádková tiskárna. V těchto programech jsou zahrnuty všechny výpočty, které vyjadřují funkční vztahy mezi vstupními a výstupními informacemi a dá se říci, že vyjadřují samotnou podstatu řešeného systému zpracování dat. Programy I. a II.kategorie existují vlastně pro to, aby programy III.kateg. mohly účinně splnit svoji funkci. Spojujícím prvkem jednotlivých programů je datová základna a tím je určena i vzájemná součinnost programů.

Činnosti programů jednotlivých kategorií jsou typické a nechají se vyjádřit jakýmsi základními programovými cykly.

2. Princip systému GEPAS

GEPAS je programovací systém, ve kterém se speciálním jazykem popisuje pro celou úlohu komplexně:

- charakteristika vstupních dat
- charakteristika výstupních dat
- modifikace základního programového cyklu programů I. a III.kat.
- součinnost programů
- doplňkové výpočetní procedury k programům III. kateg.

Zdrojový programový text je po prověření na syntaxi podroben generátorem analytickému vyhodnocení. Ke základě tohoto vyhodnocení určí generátor optimální datovou základnu, katalog souborů a jejich vět, katalog výpočetních procedur a nutnost zařazení programů II.kategorie. Pro datové soubory určí prostory na discích a magnetických páskách a vygeneruje:

- všechny programy pro řešenou úlohu /agendu/ v jazyce PL/1, které jsou rozříděny na programy I., II. a III. kategorie
- všechny řídicí štítky monitoru /včetně pokynů určených obsluze počítače/ pod kterými se bude agenda provozovat.

Uživatel nechá kompilátorem PL/1 vygenerované programy přeložit, výsledky uloží do fázové knihovny a může zahájit zkušební provoz na simulovaných datech. V tomto směru je již zmíněný generátor v našem ústavu provozně zkoušen. Později bude generátor doplněn i o generování řídicích dat na kterých může být simulován zkušební provoz agendy.

V datové základně se rozeznávají dva typy souborů a to interní a externí. Externí soubory jsou zde chápány jako soubory, jejichž data mohou do programů vstupovat i z nich vystupovat, ale přísluší cizí datové základně jiné agendy, což umožňuje propojení se cizí data. Interní soubory jsou naopak soubory o které "pečuje" GEPAS od jejich pořízení z děrných štítků až po jejich archivaci na pásky a disky, pro interní soubory navíc děrné štítky a co se týče organizací, jsou zde známá sekvenční a indexsekvenční.

V současné době pracuje generátor pod operačním systémem DOS VS, je převáděn do systému OS VS s cílovými systémy LOS 3 a LOS 4. Podmínkou je aby zmíněné systémy obsahovaly kompilátor jazyka PL/1 optimalizované verze.

3. Stručný popis jazyka GEPAS

Příkazy jazyka jsou obsaženy ve čtyřech základních oddílech. Jedná se o oddíl: - SYSTEM

- FORM
- PRINT
- PROC.

SYSTEM popisuje především druhy a počty periferních zařízení, které mají být pro provoz agendy rezervovány.

FORM popisuje struktury vět souborů /kromě tiskových/ a standardní operace, které mají být provedeny prostřednictvím programů I. a II. kategorie.

PRINT popisuje přiřazení vstupních a výstupních souborů, určuje aktivitu systémových procedur v rámci programového cyklu a popisuje formát výstupních tiskových dat. Tento oddíl je určen především pro popis funkcí programu III. kategorie.

PROC je oddílem výpočetních procedur, pomocí nichž doplňuje uživatel, pokud je třeba, standardní výpočetní operace, které jsou součástí programů III. kategorie. Procedury se popisují v syntaxi jazyka PL/1 a jsou dvojího druhu.

- a/ Procedury v takovém významu, v jakém existují v jazyce PL/1 /klasickém/ a uživatel je pak volá jako funkce v aritmetických nebo logických výrazech nebo příkazem CALL.
- b/ Procedury, které ve vygenerovaném programu III. kategorie samostatně neexistují, ale jsou "vsunuty" do systémových procedur a obklopeny systémovými příkazy. Tyto procedury pak uživatel přímo nevolá, protože jejich aktivita je zabezpečena v rámci systémových procedur a ty jsou volány automaticky ze základního programového cyklu. V těchto procedurách mohou být pak volány procedury klasické /až a//.

Syntaxe a semantika jazyka GEPAS

Vstupní zdrojový text se děruje od 1. do 72. sloupce. Od sloupce č. 73 je text ignorován. Některá klíčová slova musejí začínat na 1. sloupci a tato povinnost je vyznačena v popisu podtržení prvního písmene klíčového slova. Žádné uživatelské jméno nesmí začínat předponou "GEP". Uživatelská jména,

která popisují určitý objekt musí být ve všech oddílech se stejným pojmenováním /tím je zabezpečen např. automatický přenos dat ze vstupů na výstup/. Uživatelská jména se popisují pomocí identifikátorů a délka identifikátoru je /pokud není uvedeno jinak/ omezena na 31 alfanumerických znaků. Klíčové slova mají stanovené pořadí a nesmějí být přehazována. Pro oddíly je také stanoveno pořadí zápisu a to tak, že napřed se zapisuje oddíl SYSTEM, pak oddíly FORM, potom oddíly PRINT a nakonec PROC, které jsou nepovinné. V definici znamenají hranaté závorky nepovinnost, svorky alternativnost a podtržení fakultativnost.

a/ Oddíl SYSTEM má tyto příkazy:

SYSTEM

NAME name - 1;

DEVICEC integ - 1 (number - 1) [, integ - 1 (number - 1)] ;

ZEROHDea name - 2 [, name - 2...] ;

END SYSTEM ;

name - 1 - identifikátor (max. 5 znaků) udávající jméno úlohy a je generován jako předpona jmen jednotlivých rutin.

integ - 1 - číslo diskového nebo páskového zařízení, které bude pro provoz rutiny rezervováno /např. pro pásku může být určeno 2400, pro 100 MB disk 3330 atd./

number - 1 - počet zařízení, který má být rezervován.

aa - číslo v intervalu 01 až 99.

name - 2 - identifikátor určený pro jméno periodicity /např. DEN, MESIC, VPATEK apod./

Příkaz ZEROHD určuje v jaké kombinaci periodicit si uživatel přeje, aby HD data byla nulována současně. Pro každý jednotlivý příkaz ZEROHDea se vygeneruje speciální rutina, která v sobě zahrnuje programy II. kategorie, které provedou vynulování střadačových dat, příslušející k periodicitě name - 2. Zde je nutno poznamenat, že tato data mají zvláštní konstrukci pojmenování a to takovou, že se vlastně skládají ze dvou jmen spojených znakem " _ ". Předpona před tímto znakem je jménem periodicity, přípona za

tímto znakem je jménem zdrojové proměnné
/např. ROK # MZDA, JES # MZDA atd./

b/ Oddíl FORM má příkazy:

FORM name - 3, $\left. \begin{array}{l} \text{EXT} \\ \text{INT} \end{array} \right\}$;

$\left. \begin{array}{l} \text{XTNFD integ-2,name-4,integ-3,integ-4,name-5,integ-5,integ-6;} \\ \text{XTNFI integ-2,name-4,integ-3,integ-4,name-5;} \\ \text{XTNTM integ-2,name-4,integ-3,integ-4,name-5;} \\ \text{XTNPO integ-2,name-4,integ-3,integ-4,name-5;} \\ \text{XTNTP integ-2,name-4,integ-7,integ-8,name-6,integ-5,integ-9;} \end{array} \right\}$
 PER name-7;
 NUMC integ-10;
 [CONNECT name-8;]
 [KEY1 name-9 [,name-9...] ;
 [KEYb name-9 [,name-9...] (name-10);]
 REC

$\left. \begin{array}{l} \text{name-11 } \left\{ \begin{array}{l} \text{FIXED (p [,9])} \\ \text{CHAR (p)} \\ \text{PIC 'FORMAT1'} \end{array} \right\} \text{ [NOPUNCH (name-12)]} \\ \text{GEFILLER CHAR (p) ;} \end{array} \right\}$;

[DEL relace ;]
 [GFORMK name-13 [,name-13...] ;]
 CREC
 $\left. \begin{array}{l} \text{name-14 PIC 'FORMAT1' ;} \\ \text{GEFILLER CHAR (p) ;} \end{array} \right\}$
 END [name-3] ;

name-3 - identifikátor /max. 5 znaků/ označující jméno souboru, pro HD data je předepsaná konstrukce jména G0Hnn, kde nn je pořadové číslo souboru HD a tento soubor pak musí být INT /interní/.

- XTNFD a $\dot{\text{z}}$ XTTFP je určeno výhradně pro externí soubory.
- XTNTP je určeno pro pásku, zbytek pro disky a v seznamech jsou zahrnuty údaje o etiketách a prostorech na kterých jsou soubory uloženy.
- name-7 -- jméno periodicity ve které vstupuje soubor do zpracování.
- integ-10 - uživatelem odhadnutý maximální počet vět, který bude obsažen v souboru.
- name-8 - identifikátor /max. 5 znaků/ označuje jméno souboru na jehož větu bude věta souboru name-3 připojena, toto spojení provede program I. kategorie.
- KEY1 a KEYb /b je číslo od 2 do 9/ jsou vyhrazeny pro soubory jejichž věty jsou svými klíči /name-9/ určeny v souboru jednoznačně. Jestliže je použito KEYb, je ze souboru name-3 odvozen soubor name-10 se stejnou strukturou a obsahem vět, ale je seřazen podle klíčů name-9, které jsou uvedeny u KEYb, což umožňuje vícenásobný přístup k ISAM souborům podle různých klíčů. Vytvoření těchto odvozených souborů provádí program I. kategorie. Soubory HD dat mají KEY1 povinný a KEYb zakázaný.
- REC - uvádí popis struktury věty.
- name-11 - označuje jméno datového prvku věty, pokud se stejné jméno vyskytne ještě v popisu věty jiného souboru, je pokládáno za klíč.
- p - celkový počet míst.
- q - počet míst za desetinou tečkou.
- FORMAT1 - posloupnost znaků S,9,V a opakovacích faktorů podobně jako v jazyce PL/1.
- GEFILLER - na sníženém černém štítku bude při konverzi do datové základny p sloupců ignorováno.
- NOFUNCH - datový prvek na vstupním štítku nebude vůbec děrován a jeho hodnota bude do věty připojena programem II. kategorie ze souboru name-12 /číslníku/, name-11 u kterého je NOFUNCH uveden musí být klíčem a musí být obsažen ve větě souboru name-12.
- relace - posloupnost relačních operátorů a operandů, operandy musí být obsaženy ve struktuře věty /REC/. DEL je po-

voleno pro ISAM soubory. Jestliže data na vstupním štítku vyhoví pravdivostní hodnotě relace, bude věta se stejným klíčem v updatovaném souboru vyloučena a tuto akci provádí program I. kategorie.

- name-13 - identifikátor /max. 5 znaků/ označující jméno ISAM souboru proti kterému bude v programu I. kategorie provedena kontrola při snímání vstupních štítků na povolenou existenci klíčů.
- CREC - uvádí popis kontrolního součtového štítku, který je uživatelem vhodně zařazen v souboru vstupních štítků běžných dat.
- name-14 - uvádí jméno datového prvku ve vstupním štítku, který bude v programu I. kategorie střádan a při sejití kontrolního štítku se provede kontrola na rozdíl.

c/ Příkazy oddílu PRINT:

```

PRINT name-15;
PFA name-16;
CONTROL [FINAL,] name-17, [A/D] [,name-17, [A/D] ...];
MATCH name-18 [,name-18...];
CHAIN name-19 [,name-19...];
TABLE name-20 [,name-20...];
HDI name-21 [,name-21...];
HDO name-22 [,name-22...];
INPUT name-23 [,name-23...];
OUTPUT name-24 [,name-24...];
UPDATE name-25 [,name-25...];
PAGE integ-11,integ-12 [NODIVIDE CH] [NODIVIDE CF] ;

```

```

{1} name-26 TYPE {
  N
  RH
  PH
  CH {FINAL name-27} [NODIVIDE]
  DE [NODIVIDE]
  CF {FINAL name-28} [NODIVIDE]
  FF
  RF
} ;

```



```

{ 2 } name-29 { CHAR (p) [INIT ('alfabetik')] } [integ-13]
{ 02 } { PIC 'FORMAT2' [INIT (integer)] }
        { FIXED (p [q]) [INIT (integer)] }
[REPT PAGE] [LINE { BEFORE 1
                  [BEFORE integ-14]
                  [AFTER integ-15]
                }] [NOPRINT name-30] ;

```

[COPY FROM name-31 TO name-32 IN name-33]

END [name-15] ;

- name-15 - identifikátor /max. 6 znaků/ označuje jméno programu III. kategorie.
- name-16 - jméno periodicity ve které bude program exekuván.
- name-17 - jméno klíče podle kterého budou vstupy a výstupy tříděny přičemž pořadí uvádí hierarchii třídění a A značí vzestupná, D sestupná.
- name-18 - jméno souboru, který bude v programu matchován, alespoň jeden je povinný a pořadí souborů je hierarchické ve smyslu primární, sekundární, terciální atd.
- name-19 - jméno souboru ISAM organizace, který má být řetězen.
- name-20 - jméno souboru organizace STREAM, který zavádí vnější konstanty pro program.
- name-21 - jméno souboru vstupních HD dat.
- name-22 - jméno souboru výstupních HD dat, je podmíněno stejným jménem souboru na vstupu /HDI/.
- name-23 - jméno souboru o jehož okamžiku čtení chce rozhodnout uživatel proti standardu sám /vynucené čtení/.
- name-24 - jméno souboru o jehož okamžiku zápisu rozhoduje uživatel sám /vynucený zápis/.
- name-25 - jméno souboru jehož věty uživatel vynuceně mění /čte a přepisuje/.
- integ-11 - číslo řádku na stránce kde má začít tisk /implicitně je 3/.
- integ-12 - číslo řádku na stránce kde má být tisk ukončen /implicitně je 65/.

- NODIVIDE CH - určuje, že úvodní tiskové hlavičky nesmí být "trženy" perforací papíru na tiskárně.
- NODIVIDE CF - určuje, že tiskové věty uzavírající skupinu nesmí být "trženy" perforací papíru na tiskárně.
- name-26 - jméno věty. V typu M deklaruje uživatel proměnné, které jsou globální pro celý program III. kategorie.
- Ostatní typy RH až RF jsou určeny pro výstupní tiskové věty se současným určením doby výstupu a následujícím významem:
- RH - na začátku programu
- PH - na začátku stránky
- CH - se změnou klíče name-27 /nebo hierarchicky vyššího/ z MATCH souborů
- DE - o přečtení věty z MATCH souborů /je povinný/
- CF - se změnou klíče name-28 /nebo hierarchicky vyššího/
- PF - na konci stránky
- RF - na konci programu
- Pro tyto jednotlivé typy /kromě K/ jsou generovány systémové procedury, které zajistí potřebné přesuny dat k tisku.
- name-29 - jméno prvku věty. Všechny stejného jména např. ve větě typu DE a CF je chápán tak, že na CF je totál a na DE zdroj střádání. Střádání se neprovádí vždy, (např. pro klíče), přitom je zde možné popsat také automatické střádání HD dat přesné pravidlo přesahuje rámec tohoto článku.
- FORMAT2 - posloupnost edičních znaků +, -, S, 9, V spod.
- integ-13 - číslo tiskové pozice na řádku, jestliže je vynecháno, je potlačen tisk prvku.
- name-30 - proměnná, jejíž obsah je testován a je-li rozdílná od nuly, je tisk celého řádku potlačen.
- COPY - popis struktury věty může být kopírován z předchozích vět /name-33/ od jména prvku name-31 do prvku name-33.
- integ-14 - počet volných řádků před tiskem.
- integ-15 - počet volných řádků po tisku.