

Główne kierunki badawcze w dziedzinie oprogramowania

/Referat Sekcji Oprogramowania przygotowany na posiedzenie Komitetu Informatyki PAN, na podstawie materiałów opracowanych przez J.Borowca, W.Mardala, J.Olszewskiego, W.M.Turskiego i J.Wierzbowskiego/.

Oprogramowanie maszyn cyfrowych stanowi obecnie autonomiczną dziedzinę badań naukowych "o pełnym cyklu rozwojowym": od dociekań podstawowych, ściśle związanych z naukami matematycznymi /lingwistyka formalna, algebra uniwersalna, kombinatoryka analityczna, teoria obliczalności/, poprzez prace konstrukcyjne, związane z naukami systemowymi /teoria systemów, optymalizacja, badania operacyjne/ i prace o charakterze technologicznym /automatyzacja wytwarzania oprogramowania i jej instrumentarium, prakseologia działalności zespołowej/, aż do metodologii wdrożeń i badań jakości /z badaniem niezawodności na czele/.

Szerokość frontu badań w dziedzinie oprogramowania, znaczny stopień powiązań wewnętrznych pomiędzy poszczególnymi kierunkami badawczymi, wyrażający się m.in. tym, że te same problemy stanowią niezależny przedmiot studiów i, jednocześnie, pojawiają się jako zagadnienia wtórne przy innych badaniach, brak ukształtowanych, a tym bardziej dostatecznie uniwersalnych teorii, niedostateczna systematyka przedmiotów badawczych oraz całkowity rozgardiasz dydaktyczny, uniemożliwiający zadowalające kształcenie kadry naukowej dla tej dziedziny, połączone z wyraźnie wzrastającym zapotrzebowaniem społecznym na konkretne rezultaty o praktycznej użyteczności czynią badania w dziedzinie oprogramowania wysoce atrakcyjnym i niezmiernie ważnym działem informatyki, wymagającym szczególnego poparcia ze strony dysponentów środków stymulujących rozwój badań i szczególnej troski ze strony

środowiska naukowego.

W niniejszym referacie Sekcja Oprogramowania przedstawia Komitetowi Informatyki wykaz najważniejszych kierunków badawczych, wybranych ze względu na dwa jednakowo ważne kryteria - znaczenie praktyczne spodziewanych wyników badań, - względne wyodrębnienie się i ugruntowanie w całokształcie problematyki badawczej.

Kryterium pierwsze nie wymaga komentarza, kryterium drugie wybrane zostało jako rękojmia /stosunkowo/ małego ryzyka decyzji koncentracji prac badawczych wokół danej tematyki.

Sekcja Oprogramowania wnioskuje, by Komitet Informatyki PAN zechciał postulować preferowanie tych prac naukowych w dziedzinie oprogramowania, które spełniają powyższe dwa kryteria.

W ujęciu przedmiotowym kierunki badawcze grupują się w czterech tematach:

1. Systemy operacyjne
 2. Banki danych
 3. Translatory
 4. Komunikacja człowiek - maszyna
- Omówmy pokrótce powyższe zagadnienia.

1. SYSTEMY OPERACYJNE

Dziedzina systemów operacyjnych jest najbardziej rozległą w oprogramowaniu maszyn, a w Polsce chyba najmniej uprawianą dyscypliną badawczą. Obejmuje ona oprogramowanie, które w stosunku do programów tzw. użytkowych, a także translatorów /kompilatorów, interpretatorów i generatorów/ języków programowania pełni rolę zespołu programów usługowych, z drugiej zaś strony umożliwia sterowanie procesami liczenia, transmitowania danych, komunikację człowiek - maszyna i prowadzenie wszelkich potrzeb-

nych księgowości i statystyk. Zwróćmy uwagę, że dla niektórych zastosowań maszyn całość oprogramowania może nosić miano systemu operacyjnego. Takim zastosowaniem jest np. system rezerwacji miejsc lotniczych czy kolejowych. W zasadzie nie ma tam mowy o programach użytkowych czy translatorach języków programowania. Jest za to wiele programów /procedur, podprogramów/ transmisji i konwersji danych przekazywanych pomiędzy urządzeniami końcowymi a pamięcią systemu, programów służących do organizacji, tworzenia, powiększania i aktualizacji zbiorów danych, programów redagowania wszelkich potrzebnych zestawień, sprawozdań, kosztów itp., oraz programów z góry przewidzianych na przypadek zmian w konfiguracji samego systemu /więcej i, być może, innych urządzeń końcowych, większa pamięć, większa liczba kanałów transmisji itp./.

Powyższe stwierdzenia należy traktować jako bardzo pobieżną i wstępną ilustrację następującej próby określenia głównych kierunków badawczych w dziedzinie systemów operacyjnych.

1.1. Język operowania systemem liczącym :

Pod tym hasłem należy rozumieć prace w kierunku sformułowania możliwie ogólnego języka operowania systemem liczącym, tzn. języka, w którym będzie możliwe i niezbyt trudne wyrażenie tego, co, kiedy i jak ma być wykonywane. Obecny stan tego kierunku najlepiej chyba ilustruje język rozkazowy /Command Language/ systemów operacyjnych GEORGE dla maszyn firmy ICL, który z jednej strony uważany jest za najlepiej do tej pory sformułowany język operowania systemem liczącym, a z drugiej strony ma się tak do zamierzonego języka operowania systemem liczącym, jak kod wewnętrzny jakiejś maszyny do ALGOLu.

Z tego punktu widzenia system operacyjny powinien być traktowany jako interpretator czy nawet kompilator języka operowania systemem liczącym, co stanowi niejako odwrócenie dotychczasowej praktyki, w której język taki wynika z konstrukcji systemu operacyjnego.

1.2. Metody implementacji systemu operacyjnego

Problem ten obejmuje cały zespół zagadnień badawczych, jak:

- metody projektowania,
- metody programowania /w tym próby w kierunku odpowiedniego języka programowania/,
- metody wiązania z istniejącym już oprogramowaniem /w tym również z działającymi już częściami systemu operacyjnego/,
- metody dokumentacji,
- metody weryfikacji,
- metody wprowadzania zmian wynikających z rozbudowy i zmian sprzętu oraz ze zmian potrzeb użytkowników.

Wiele, aczkolwiek nie wszystkie, z wymienionych problemów znajduje obecnie wyraz w dążeniach do sformułowania języka implementacji systemów operacyjnych w oparciu o język PASCAL, który okazał się już językiem odpowiednim dla dokumentacji niektórych części systemu operacyjnego oraz dla przekazywania wiedzy o ich budowie. W praktyce, udane próby w kierunku języka implementacji systemów operacyjnych przeprowadzono w oparciu o język ALGOL w firmie Burroughs, gdzie język ten odpowiednio rozszerzono, a ponadto sam sprzęt skonstruowano tak, by możliwie najpełniej odwzorować struktury danych tego języka i jego translatorów w technikę adresowania maszyny.

1.3. Metody optymalizacji działania systemu liczącego

Kierunek ten z jednej strony doczekał się już wielu opracowań zarówno praktycznych jak i teoretycznych, obejmujących problemy i techniki podziału czasu procesorów /jednostek arytmetycznych i kanałów transmisji/ i podziału pamięci. Z tego właśnie kierunku wywodzą się zaawansowane prace nad teorią i implementacją procesów współbieżnych, nad metodami ich synchronizacji oraz nad pamięciami wirtualnymi. Z drugiej strony właściwie nie umiemy jeszcze dać rzetelnej odpowiedzi na pytanie dotyczące jakości systemu operacyjnego, czy też szerzej, systemu liczącego. Jest tak dlatego, że nie potrafimy, jak dotychczas, traktować łącznie problemów optymalizacji działania systemu liczącego, widzianych z różnych punktów widzenia, tak, by z tego bezpośrednio wynikały wnioski dotyczące implementacji systemu operacyjnego.

W praktyce, dla badań jakości i porównywania systemów liczących, a tym bardziej projektów ich systemów operacyjnych, wykorzystuje się metody symulacji jako jedyne narzędzie pozwalające uzyskać spojrzenie całościowe; ewentualne korekty polegają zaś na wprowadzaniu zmian do systemu operacyjnego, czy też jego projektu, drogą prób i błędów.

Oczywiście wymienione tutaj trzy główne problemy w dziedzinie systemów operacyjnych nie powinny być traktowane oddzielnie. Wszystkie one są ze sobą ściśle powiązane tak, że praktyczne rozwiązanie jednego z nich będzie mieć wpływ, a nawet może oznaczać rozwiązanie innych problemów w pozostałych kierunkach.

2. BANKI DANYCH

Ten kierunek badań nabiera coraz większego znaczenia, bowiem pojęcie banku danych, niegdyś wiązane przede wszystkim z systemami wyszukiwania informacji, dziś pojawia się niemal we wszyst-

kich typach systemów przetwarzania np. w systemach obliczeń inżynierskich, systemach zarządzania, planowania itd.

Zakres problematyki badawczej związanej z pojęciem banku danych jest bardzo rozległy. Badania w tym zakresie prowadzone w wielu ośrodkach na świecie można podzielić na następujące grupy zagadnień:

- a/ konstrukcja baz danych,
- b/ zagadnienia formalizacji opisu struktur danych i języki opisu danych,
- c/ metody i środki programowe dla definiowania zbioru operacji na bazie danych,
- d/ zagadnienie komunikacji użytkownik - bank danych.

Rozważmy pokrótce powyższe zagadnienia.

2.1. Konstrukcja baz danych

Szybki rozwój informatyki stale zwiększa wymagania w zakresie możliwości przechowywania i przetwarzania coraz większych zbiorów informacji. Podstawowym zagadnieniem jest tu więc konstrukcja baz danych, tj. zbiorów informacji spełniających określone cechy. Do cech tych należy przede wszystkim zaliczyć:

- niepowtarzalność danych, czyli eliminowanie redundancji, dzięki czemu nie tylko zmniejsza się wymagania na pojemność nośników informacji, ale ułatwia się aktualizację zbiorów i unika się sprzeczności wynikających z niejednoczesnej aktualizacji wielokrotnie zapisanych informacji,
- zapewnienie ochrony danych, umożliwiające dostęp do danych tylko upoważnionym użytkownikom.

W chwili obecnej istnieje wiele kierunków badań związanych z konstrukcją baz danych. Z jednej strony prowadzi się próby inter-
pretacji

istniejących sformalizowanych teorii matematycznych dla potrzeb konstrukcji baz danych, z drugiej zaś strony próbuje się tworzyć nowe modele, jak np. modele relacyjne baz danych. Ponadto opracowano w tym zakresie propozycję standardu /prace prowadzone przez grupę CODASYL/, jednakże z teoretycznego punktu widzenia propozycja ta nie jest dostatecznie uzasadniona.

Mówiąc o konstrukcji baz danych należy szczególną uwagę zwrócić na tzw. bazy rozproszone, tj. takie, które są podzielone na części przechowywane w odległych od siebie miejscach. Dostęp do wszystkich części każdej bazy jest zapewniony przez sieć transmisji danych oraz wysoce rozwinięte oprogramowanie specjalne. Użytkownik takiej bazy nie orientuje się nawet z której części bazy korzysta.

2.2. Opis struktur danych

Niezmiernie ważnym zagadnieniem związanym z konstrukcją baz danych jest jednolita klasyfikacja struktur danych oraz język do opisu tych struktur. Istnieje pilna potrzeba opracowania takiego języka, jako narzędzia użytkownika umożliwiającego mu definiowanie dowolnych struktur. Obecnie możliwości te są w istotny sposób ograniczone bądź przez istniejące języki programowania, bądź też przez pracochłonność opisu tych struktur.

2.3. Operacje na zbiorach i metody dostępu

Jest rzeczą oczywistą, że przechowywanie informacji jest tylko środkiem niezbędnym do osiągnięcia przez użytkownika określonego celu. Celem bezpośrednim, przynajmniej w sensie przetwarzania informacji jest wykonywanie określonych operacji na bazie danych. Jak się okazuje intuicyjnie można wyróżnić wiele operacji, które często występują. Konieczne jest jednak przeprowadzenie szczegółowych badań prowadzących do jednolitej klasyfikacji zbioru wszystkich operacji, a następnie do wybrania takiej klasy

operacji, na podstawie której można byłoby generować dowolne inne operacje.

Wykonywanie operacji jest ściśle związane z metodami dostępu do zbiorów. Potrzebne są więc badania metod dostępu i ich wpływu na sposób wykonywania poszczególnych operacji, przy czym pod uwagę należy brać efektywność metod dostępu w zależności od organizacji bazy danych.

2.4. Komunikacja użytkownik - baza danych

Obecnie wraz z rozwojem zastosowań oraz wzrostem stopnia ich skomplikowania coraz większą wagę przywiązuje się do sposobu formułowania żądań przez użytkownika. Stosowane dotychczas formy - pisanie programów nie mogą być w dalszym ciągu akceptowane. Pierwsze prace w tym kierunku objęły systemy wyszukiwania informacji. W wyniku tych prac użytkownik może formułować swoje wymagania, tj. kryteria wyszukiwania informacji w sposób zbliżony do tradycyjnego toku myślenia. Badania powinny wykazać w jaki sposób oraz w jakim zakresie operacji winny być w pierwszym rzędzie udoskonalone formy komunikacji człowiek - maszyna.

3. TRANSLATORY

Prace w tej dziedzinie prowadzone są w Polsce najdłużej, największe jest też zgromadzone doświadczenie. Uzasadnienie celowości tych prac jest doskonale znane, dlatego nie przytaczamy go tutaj.

Aktualnie najważniejsze prace w zakresie konstrukcji translatorów są prowadzone w dwóch kierunkach.

3.1. Tworzenie języków maszynowo zorientowanych /np. BLISS/, dostatecznie wysokiego rzędu, umożliwiających opis procesu translacji, na ogół sterowanego składnią.

Języki te mają szereg zalet:

- a/ umożliwiają produkcję optymalnych translatorów,
- b/ umożliwiają strukturalne konstruowanie translatorów,
- c/ dają precyzyjną dokumentację prowadzonych prac.

Wadą tego rozwiązania jest to, że konstruktor musi nie tylko znać język opisujący, lecz również metody translacji.

3.2. Tworzenie metod formalnego opisu języków, ich składni i semantyki oraz definiowanie systemów, które na podstawie opisu języka automatycznie konstruują jego translator.

Pośród prac prowadzonych w ramach drugiego kierunku można wydzielić następujące grupy tematyczne.

3.2a. Prace związane z warstwą syntaktyczną języków.

Do grupy tej zaliczamy prace nad metodami opisu składni oraz metodami analizy składni; do tych ostatnich zaliczamy również prace nad analizą leksykalną. Jak

Jak dotąd istotną rolę w omawianej grupie zagadnień odgrywały badania nad językami bezkontekstowymi. Najbardziej popularną metodą opisu składni języka jest tzw. notacja BNF oraz jej modyfikacje; inne metody /jak np. tzw. język produkcji Floyda/ nie znalazły szerszego zastosowania. Jeżeli chodzi o metody analizy składni, to opracowano szereg takich metod, jak np. metody z operatorowym, prostym, słabym i uogólnionym pierwszeństwem, metody z ograniczonym i mieszanym pierwszeństwem, metoda LR(k) i jej modyfikacje, metoda LL(k), itd. Obecnie, przede wszystkim dzięki wprowadzeniu gramatyk dwupoziomowych /gramatyki van Wijngaardena, gramatyki afiksowe/ główny akcent przenosi się na badania w zakresie języków kontekstowych. Problemy, które należy tutaj rozwiązać, to opracowanie efektywnych metod analizy składni tych języków, jak również zastosowania w/w

gramatyk do opisu tzw. semantyki statycznej języków.

3.2b. Prace w zakresie semantyki. Są one prowadzone w dwu kierunkach: teoretycznym i realizacyjnym. "Teoretycy" formalizują pojęcie programu w terminach algebry relacji lub funkcji bądź definiują język za pomocą zbioru aksjomatów i dołączonych do nich reguł wywołu, bądź wreszcie budują model języka w oparciu o aparat obiektów abstrakcyjnych. "Praktycy" z kolei proponują różne języki dla definiowania akcji semantycznych sterowanych składnią, bądź definiują atrybuty stałych i zmiennych gramatyki opisywanego języka. Jak się wydaje wyniki teoretyków /z wyjątkiem, być może, prac prowadzonych przez Scott'a i Strachey'a/ nie osiągnęły jeszcze takiego stadium aby można je było stosować w praktyce, zaś propozycje praktyków są za bardzo pragmatyczne, zbyt obciążone zarówno konkretnymi metodami translacji, jak i architekturą maszyn, na których te propozycje są realizowane. Dotyczy to szczególnie semantyki dynamicznej.

Jak dotąd wyniki prac w tym zakresie nie są zadowalające.

Znane dotychczas metody opisu języków nie gwarantują otrzymania optymalnych translatorów tych języków; istniejące metajęzyki są skomplikowane zwłaszcza w części dotyczącej semantyki i pragmatyki.

Wydaje się, że pierwsze z tych rozwiązań jest bardziej dogodne dla konstruowania translatorów dużych, uniwersalnych języków programowania. Natomiast drugie podejście może być dogodne przy konstruowaniu translatorów dla specjalistycznych języków zastosowaniowych.

Warto zwrócić uwagę, że tematyka związana z translatorami obejmuje także zagadnienia mikroprogramowania maszyn cyfrowych.

4. KOMUNIKACJA CZŁOWIEK - MASZYNA

Ten kierunek badań dotyczy nowych technicznych możliwości komunikacji człowiek - maszyna, a szczególnie tych które stwarzają urządzenia komunikacji graficznej. Możliwości te są niezwykle istotne dla wielu dziedzin zastosowań informatyki. Wprowadzają one nowy język wymiany informacji między człowiekiem i maszyną operujący obiektami wielowymiarowymi. Stąd wynika szereg problemów badawczych, których rozwiązanie winno przynieść w szczególności metody i środki programowe dla identyfikacji i generacji obrazów oraz oporowania nimi w maszynie. Tematyka ta niewątpliwie wiąże się z badaniami z zakresu teorii struktur danych.

Obok wymienionych czterech kierunków przedmiotowych, do głównych kierunków badawczych zaliczyć należy również niżej wymienione zagadnienia uniwersalne, dotyczące oprogramowania jako całości, mimo, że konkretne wyniki odniesione do poszczególnych tematów przedmiotowych będą zapewne różniły się między sobą.

5. TECHNOLOGIA WYTWARZANIA OPROGRAMOWANIA

Wśród zagadnień szczegółowych tej dziedziny, na pierwszy plan wysuwają się:

5.1. Stworzenie praktycznie stosowalnej metodologii projektowania oprogramowania. Obiecujące w tym zakresie wydają się zalecenia inżynierii oprogramowania, w sensie zdefiniowanym przez konferencje w Garmisch-Partenkirchen i Rzymie, a w szczególności projektowanie zstępujące /analityczne/, modułarna morfologia, pierwszeństwo poprawności i niezawodności względem optymalności.

5.2. Stworzenie zespołu środków wspomagających projektowanie

i wytwarzanie oprogramowania przez rozwinięcie zasad i praktyczne wdrożenie języków funkcjonalnego opisu oprogramowania, języków częściowej specyfikacji oprogramowania oraz kompilatorów krzyżowych.

6. METODYKA WERYFIKACJI OPROGRAMOWANIA

W dziedzinie tej wyodrębnia się obecnie trzy główne kierunki:

6.1. Metodyka dowodzenia poprawności programów, opierająca się na i ściśle związana z formalną definicją semantyki języków programowania.

6.2. Metodyka testowania i atestowania oprogramowania wiążąca się z implementacją oprogramowania i architekturą systemów liczących. Testowanie dotyczy badania poprawności implementacji poprawnego oprogramowania w ogóle, zaś atestowanie - badania przydatności poprawnie implementowanego poprawnego oprogramowania.

6.3. Metodyka określania stabilności oprogramowania, dotycząca oceny zachowania się oprogramowania przy małych odchyleniach od normy

a/ danych

b/ działania lub konfiguracji sprzętu.

W wykazie najważniejszych kierunków badawczych pominęliśmy kilka tematów tradycyjnie wymienianych w poprzednich materiałach podobnego charakteru, jak również należące do problematyki sprzętowej zagadnienia automatyzacji projektowania maszyn cyfrowych, zawierające szereg tematów badawczych z dziedziny oprogramowania. W szczególności pominęliśmy: konstrukcję nowych języków programowania, nowych pakietów programów, oprogramowanie nowych typów

urządzeń cyfrowych, budowę typowych systemów przetwarzania informacji. Uznając znaczną wagę praktyczną takich prac, a nawet - jak w przypadku typowych systemów przetwarzania informacji do celów zarządzania - ich kluczową rolę w uzyskaniu tak pożądaných sukcesów informatyki, uważamy, że w tych kierunkach osiągnięto już wystarczający poziom wiedzy, pozwalający na prowadzenie prac realizacyjnych bez intensywnych badań naukowych /w każdym razie w zakresie oprogramowania/.