Podstawy informatyki

dr inż. Sławomir Koczubiej

Politechnika Świętokrzyska
Wydział Zarządzania i Modelowania Komputerowego
Katedra Technologii Informatycznych

(13 października 2024)

Kontakt

Budynek C, pokój 3.26
sk@tu.kielce.pl

Materiały do pobrania, aktualności, terminy zaliczeń

http://staff.tu.kielce.pl/sk

Organizacja wykładów

- Wykłady są nieobowiązkowe (zgodnie z postanowieniami regulaminu), ale...
- Na wykłady czasem warto zajrzeć.

Warunki zaliczenia wykładu

Test zaliczeniowy po zakończeniu wykładów.

Organizacja laboratoriów

- Zajęcia laboratoryjne są obowiązkowe.
- Dopuszcza się jedną nieobecność.
- Większa liczba nieobecności powoduje zmniejszenie oceny do niedostatecznej włącznie (3 lub więcej nieobecności).
- W przypadku usprawiedliwionej nieobecności zajęcia można odrobić z inną grupą (jeśli istnieje taka możliwość).

Warunki zaliczenia laboratoriów

Wykonanie ćwiczeń i zaliczenie sprawdzianów kontrolnych.

Treść wykładów

- Wprowadzenie i istota informatyki. Maszyny liczące, elementy kodowania informacji, reprezentacja informacji w komputerze. Kompresja i szyfrowanie.
- Struktura sprzętowa komputerów. Zasoby komputera. Współczesne architektury komputera. Ograniczenia numeryczne prowadzenia obliczeń. Ergonomia pracy przy komputerze.
- System operacyjny, zadania i klasyfikacja systemu operacyjnego. Budowa systemu operacyjnego. Systemy i typy plików. Operacje na plikach. Cechy wybranych współczesnych systemów operacyjnych. Darmowe oprogramowanie, przykłady.
- Internet, historia, zagrożenia, zasoby, narzędzia do przeszukiwania (przeglądarki i wyszukiwarki). Komunikacja internetowa. Usługi internetowe.
- Podstawy obliczeń z wykorzystaniem komputera. Arkusz kalkulacyjny, systemy algebry komputerowej, oprogramowanie statystyczne.

Treść laboratoriów

- Wykorzystanie arkusza kalkulacyjnego do pracy z danymi - sortowanie danych, formatowanie danych, formatowanie warunkowe, import i eksport danych.
- Obliczenia i przetwarzanie danych, formuły, funkcje wbudowane: matematyczne, statystyczne i finansowe.
- Prezentacja danych i wyników obliczeń – tabele, formatowanie tabel, zróżnicowane typy wykresów, formatowanie wykresów.
- Oprogramowanie użytkowe typu CAS – środowisko, zapis wyrażeń arytmetycznych i podstawowych funkcji. Generowanie wykresów funkcji.
- Oprogramowanie CAS. Operacje na wektorach i macierzach. Rozwiązywanie równań, układów równań, nierówności. Analiza statystyczna.

Literatura

- Karpisz D., Wojnar L. *Podstawy informatyki.* Wydawnictwo Politechniki Krakowskiej, Kraków 2005.
- Gonet M. *Zrozumieć Excela.* Helion, Gliwice 2019.
- Kopertowska M. *Arkusze kalkulacyjne.* PWN, Warszawa 2006.
- Komorowski T., Borawska A., Cypryjański J. *Excel dla menedżera - Casebook.* PWN, Warszawa 2016.
- Gonet M. *Excel w obliczeniach naukowych i inżynierskich.* Helion, Gliwice 2011.
- Lembas J., Kawa R. *Wstęp do informatyki.* PWN, Warszawa 2017.
- Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. *Wprowadzenie do algorytmów.* WNT, Warszawa 2004.

Informacja i środki jej przetwarzania

Informacja

Informacja (łac. *informatio*) – przedstawienie, wizerunek; termin interdyscyplinarny, definiowany różnie w różnych dziedzinach nauki. Najogólniej — właściwość pewnych obiektów, relacja między elementami zbiorów pewnych obiektów.

Cechy informacji:

- wielkość abstrakcyjna,
- może być przechowywana,
- może być przesyłana między obiektami,
- może być przetwarzana przez pewne obiekty.

Informacje mogą stanowić różnorodne dane, takie jak liczby, tekst, obiekty multimedialne (grafika, dźwięki).

Informatyka

Dyscyplina nauki zaliczana do nauk ścisłych oraz techniki zajmująca się przetwarzaniem informacji, w tym również technologiami przetwarzania informacji oraz technologiami wytwarzania systemów przetwarzających informację.

Początkowo stanowiła część matematyki, później rozwinęła się do odrębnej dyscypliny – pozostaje jednak nadal w ścisłej relacji z matematyką, która dostarcza informatyce podstaw teoretycznych.

W języku polskim termin **informatyka** zaproponował w październiku 1968 Romuald Marczyński w Zakopanem na ogólnopolskiej konferencji poświęconej *maszynom matematycznym* na wzór (fr.) *informatique* i (niem.) *informatik*.

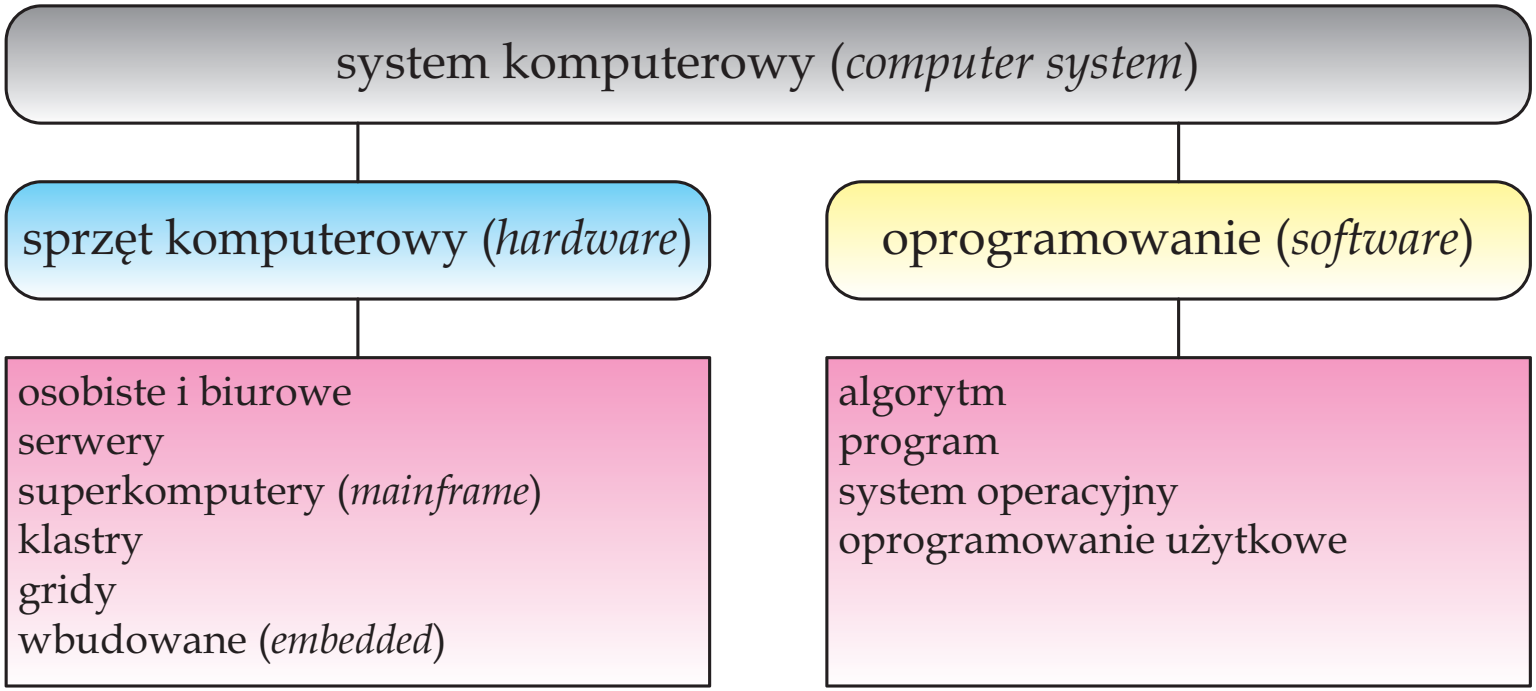
Bardziej znane działy informatyki to przede wszystkim:

- **administracja sieciowa** – zarządzanie siecią komputerową,
- **administracja systemem** – zarządzanie systemem informatycznym,
- **algorytmika** – tworzenie i analizowanie algorytmów,
- **architektura procesorów** – projektowanie procesorów,
- **grafika komputerowa** – wykorzystuje technikę komputerową w celu wizualizacji rzeczywistości,
- **inżynieria oprogramowania** – produkcja oprogramowania,
- **języki programowania** – tworzenie języków programowania,
- **programowanie komputerów** – czyli tworzenie kodu źródłowego programów komputerowych,
- **sprzęt komputerowy** – komputery i ich urządzenia peryferyjne,
- **symulacja komputerowa** – komputerowa symulacja z wykorzystaniem modelowania matematycznego,
- **bazy danych** – tworzenie, zarządzanie systemami baz danych,
- **sztuczna inteligencja** – komputerowe symulowanie inteligencji,
- **teoria informacji** – dyscyplina zajmująca się problematyką przetwarzania i przesyłania informacji,
- **webmastering** – projektowanie, programowanie serwisów internetowych.

System komputerowy

System komputerowy

Układ dwóch składowych: **sprzętu komputerowego** oraz **oprogramowania**. System komputerowy coraz częściej działa w sieci komputerowej. Można mówić o różnych poziomach takiego systemu: sprzęt komputerowy, system operacyjny (oprogramowanie systemowe), oprogramowanie użytkowe (aplikacje).



Sprzęt komputerowy – materialna część komputera. Ogólnie hardware'em nazywa się sprzęt komputerowy jako taki i odróżnia się go od software'u – czyli oprogramowania.

Podział ten jest nieostry, gdyż współcześnie wiele elementów sprzętu komputerowego posiada *wszyte* weń na stałe oprogramowanie, stanowiące jego integralną część, bez którego elementy te nie mogłyby funkcjonować. Np. większość drukarek komputerowych posiada w swojej pamięci zestaw komend, przy pomocy których realizuje proces drukowania i których odpowiednik znajduje się w pamięci komputera stanowiąc programowy sterownik tego urządzenia. Wiele urządzeń – typu karty graficzne, płyty główne posiada własne oprogramowanie nazywane **BIOS**-em. W stosunku do oprogramowania niektórych urządzeń używa się słowa **firmware**.

Oprogramowanie (*software*) – całość informacji w postaci zestawu instrukcji, zaimplementowanych interfejsów i zintegrowanych danych przeznaczonych dla komputera do realizacji wyznaczonych celów. Celem oprogramowania jest przetwarzanie danych w określonym przez twórcę zakresie.

Oprogramowanie tworzą programiści w procesie programowania. Oprogramowanie jako przejaw twórczości jest chronione prawem autorskim.

Oprogramowanie pisane jest zazwyczaj przy użyciu różnych języków programowania z wykorzystaniem algorytmów.

Często składowe systemu komputerowego (sprzęt i oprogramowanie) przedstawia się w postaci kilku warstw:

- komputer,
- oprogramowanie systemowe,
- oprogramowanie narzędziowe,
- oprogramowanie użytkowe,
- użytkownicy.

Komputer – zapewnia podstawowe możliwości obliczeniowe (procesor, pamięć, urządzenia wejścia/wyjścia), czyli są to podstawowe zasoby systemu komputerowego.

Oprogramowanie systemowe – kontroluje i koordynuje działanie zasobów sprzętowych przez zastosowanie różnych programów użytkowych dla różnych użytkowników.

Oprogramowanie narzędziowe – dogodne interfejsy użytkowe wspomagające zarządzanie zasobami sprzętowymi oraz usprawniające, modyfikujące oprogramowanie systemowe.

Oprogramowanie użytkowe – określają sposoby użycia zasobów systemowych do rozwiązywania problemów obliczeniowych zadanych przez użytkownika (kompilatory, systemy baz danych, gry, oprogramowanie biurowe), tworzone przez programistów.

Użytkownicy – ludzie, maszyny, inne komputery, mający bezpośredni kontakt z oprogramowaniem użytkowym.

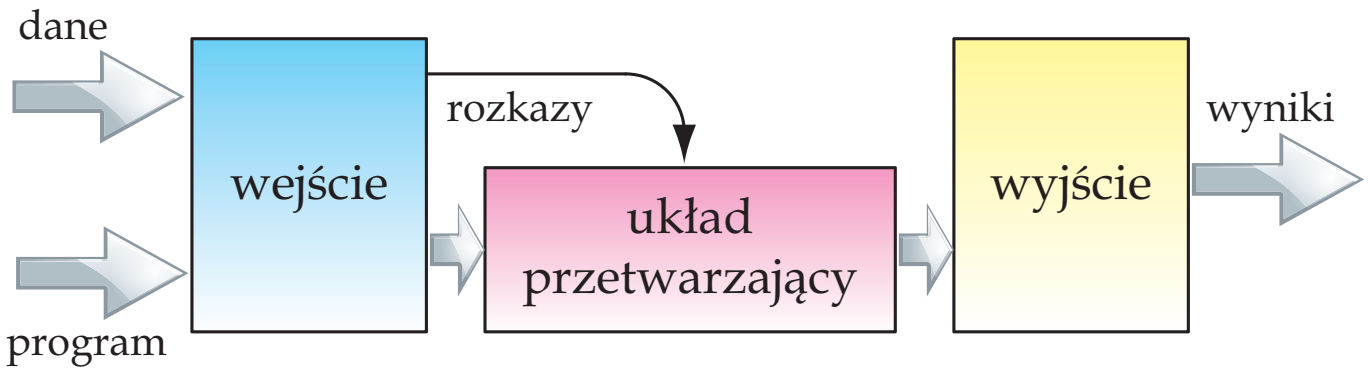
Komputer

Komputer

Urządzenie do przetwarzanie danych, umożliwiające wprowadzanie, przechowywanie i wyprowadzanie danych. Charakteryzuje je zdolność wykonywania wielokrotnie, automatycznie powtarzanych obliczeń, według algorytmicznego wzorca zwanego programem.

Granica jest tu umowna, ponieważ taką definicję komputera spełniają też kalkulatory programowalne (naukowe, inżynierskie), jednak kalkulatory służą tylko do obliczeń numerycznych, podczas gdy nazwa komputer najczęściej dotyczy **urządzeń wielofunkcyjnych**.

Jakkolwiek istnieją mechaniczne urządzenia liczące, które potrafią realizować całkiem złożone programy, zazwyczaj nie zalicza się ich do komputerów. Warto jednak pamiętać, że prawzorem komputera jest abstrakcyjny model zwany **maszyną Turinga**, a pierwsze urządzenia ułatwiające obliczenia były znane w starożytności, np. **abakus** z 440 p.n.e.



Architektura komputera

Sposób organizacji elementów tworzących komputer. Pojęcie to używane jest dosyć luźno. Może ono dzielić systemy komputerowe ze względu na wiele czynników, zazwyczaj jednak pod pojęciem architektury komputera rozumie się organizację połączeń pomiędzy pamięcią, procesorem i urządzeniami wejścia-wyjścia.

Innym, stosowanym potocznie znaczeniem terminu *architektura komputera* jest typ procesora wraz z zestawem jego instrukcji. Właściwszym określeniem w tym przypadku jest **model programowy procesora** (*ISA*, *Instruction Set Architecture*).

Model programowy procesora to określenie dotyczące organizacji, funkcjonalności i zasad działania procesora, widoczne z punktu widzenia programisty jako dostępne mechanizmy programowania.

Na model programowy procesora składają się m.in.:

- lista rozkazów procesora,
- obsługiwane typy danych,
- dostępne tryby adresowania,
- zestaw rejestrów dostępnych dla programisty,
- zasady obsługi wyjątków i przerwań.

Procesory posiadające ten sam model programowy są ze sobą **kompatybilne**, co oznacza, że mogą wykonywać te same programy i generować te same rezultaty. Przykładami modeli programowych:

- **IA-32**: i386, Pentium, K6, Athlon (RISC),
- **SPARC**: UltraSPARC, SPARC64,
- **AMD64**: Athlon 64 i wzwyż, Pentium 4 Prescott i wzwyż.

Taksonomie architektur komputerowych

Jednym z rodzajów klasyfikacji jest **Taksonomia Skillicorna**, zaproponowana ok. 1988 roku. Zakłada ona, że każda architektura stanowi połączenie pewnej liczby abstrakcyjnych składników. W efekcie, taksonomia syntetyzuje architekturę, zamiast ją klasyfikować.

Klasyfikacja Skillicorna posługuje się elementami architektury:

- procesory instrukcji (sterujące) – **IP** (*Instruction Processor*),
- procesory danych – **DP** (*Data Processor*),
- hierarchia pamięci instrukcji – **IM** (*Instruction Memory Hierarchy*),
- hierarchia pamięci danych – **DM** (*Data Memory Hierarchy*),
- układy łączące powyższe elementy.

Zadania **procesora instrukcji**:

- wyznaczenie adresu następnej instrukcji do wykonania,
- sterowanie adresacją pamięci instrukcji,
- odczyt i dekodowanie instrukcji,
- sterowanie procesorem danych – wysłanie mu instrukcji do wykonania,
- wyznaczanie adresów operandów,
- odbieranie informacji stanu od procesorów danych.

Zadania **procesora danych**:

- odbiór od układu sterującego instrukcji i adresów operandów,
- odczyt operandów z pamięci danych,
- wykonanie instrukcji operandach,
- wyznaczenie informacji stanu dla układu sterującego,
- zapamietanie wyników obliczeń w pamięci danych.

Hierarchiczna pamięć instrukcji jest zbudowana podobnie jak **hierarchiczna** pamięć danych.

Hierarchiczna pamięć instrukcji lub danych zawiera:

- pamięć podręczną instrukcji/danych,
- pamięć operacyjną,
- pamięć dodatkową (np. dyskową).

Pamięć operacyjna oraz pamięć dodatkowa może być wspólna lub oddzielna dla danych i instrukcji. Tego problemu klasyfikacja Skillicorna nie rozstrzyga.

Układy łączące obejmują następujące typy:

- połączenie 1 do 1 – pojedyncze połączenie, (1 - 1),
- połączenie n do n – *n* połączeń pojedynczych równoległych (n - n),
- połączenie 1 do n – rozesłanie informacji do *n* odbiorców (1 - n),
- połączenie n na n – przełącznik krzyżowy łączący *n* wejść z *n* wyjściami (n$\times$n).

Same procesory danych nie zawierają żadnych elementów pamiętających.

W modelach architektur przyjmuje się, że liczba hierarchii pamięci jest równa liczbie procesorów danego typu. Oznacza to, że model architektury ze wspólną hierarchią pamięci dla kilku procesorów jest przedstawiany jako model z kilkoma hierarchiami pamięci i możliwością dostępu każdego procesora do każdej hierarchii pamięci.

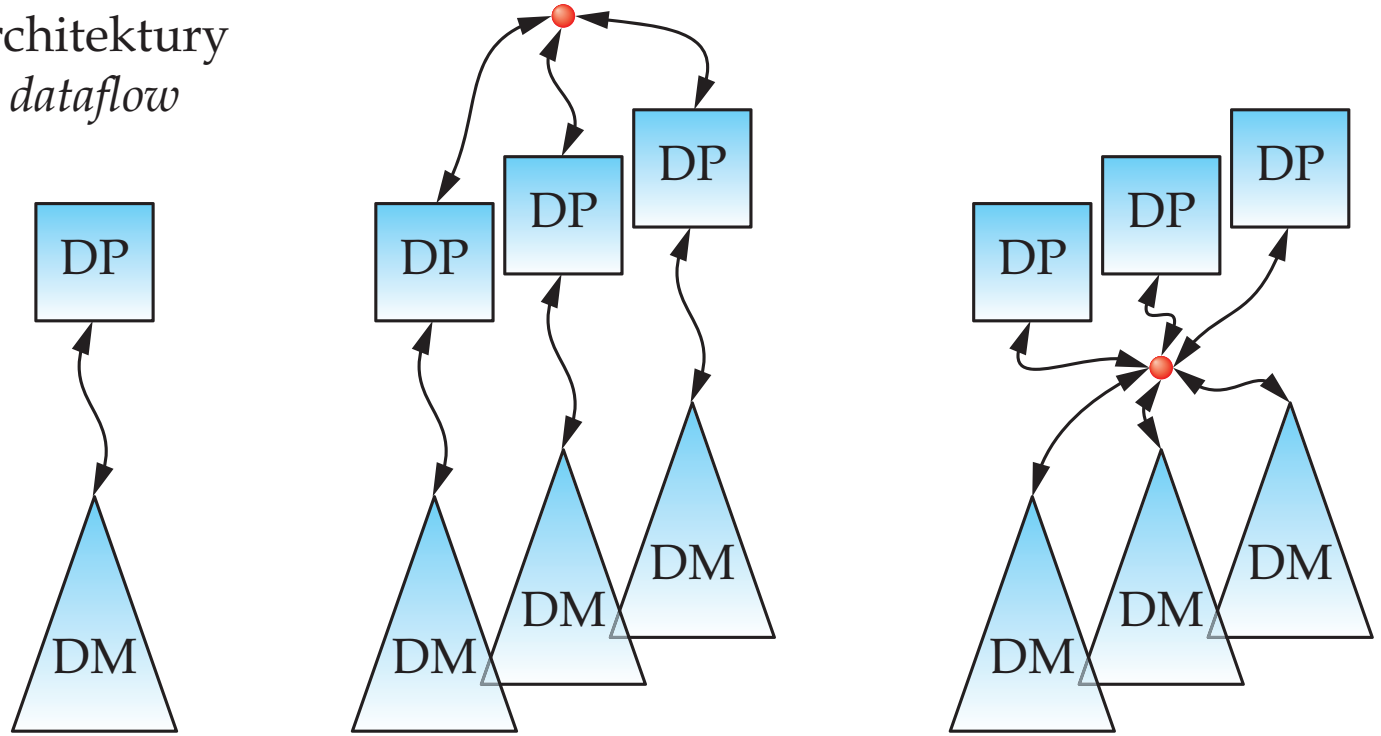
Komputer musi zawierać przynajmniej jeden procesor danych. Dozwolone są połączenia pomiędzy procesorami i hierarchiami pamięci tego samego rodzaju (kodu albo danych) oraz połączenia pomiędzy procesorami (tego samego lub różnych typów).

Używając taksonomii Skillcorna można zbudować około 30 różnych modeli architektur. Sześć z tych modeli ma sensowne znaczenie:

- uniprocesor dataflow (w taksonomii Flynna będzie to model bez strumienia instrukcji i jednym strumieniem danych, NISD),
- wieloprocesor dataflow (w taksonomii Flynna będzie to model bez strumienia instrukcji i N strumieniami danych, NIMD),
- uniprocesor von Neumanna (SISD),
- procesor wektorowy (SIMD),
- wieloprocesor słabo sprzężony (MIMD),
- wieloprocrsor silnie sprzężony (MIMD).

Sprzężenie pomiędzy procesorami tego samego typu ma sens tylko dla procesorów danych. Sprzężenie słabe będzie realizowane przez kanał komunikacyjny a silne przez wzajemny dostęp do hierarchii pamięci (praktycznie wspólna hierarchia pamięci).

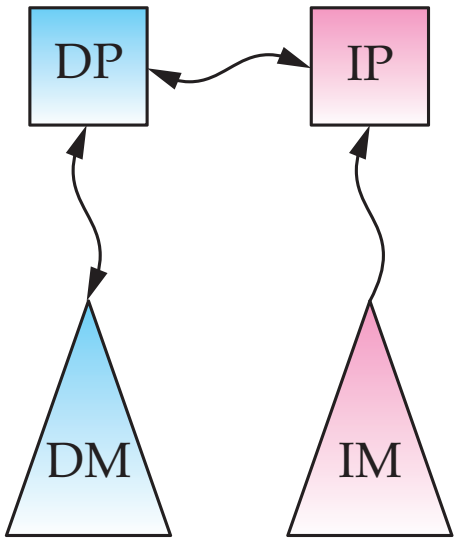
architektury
dataflow



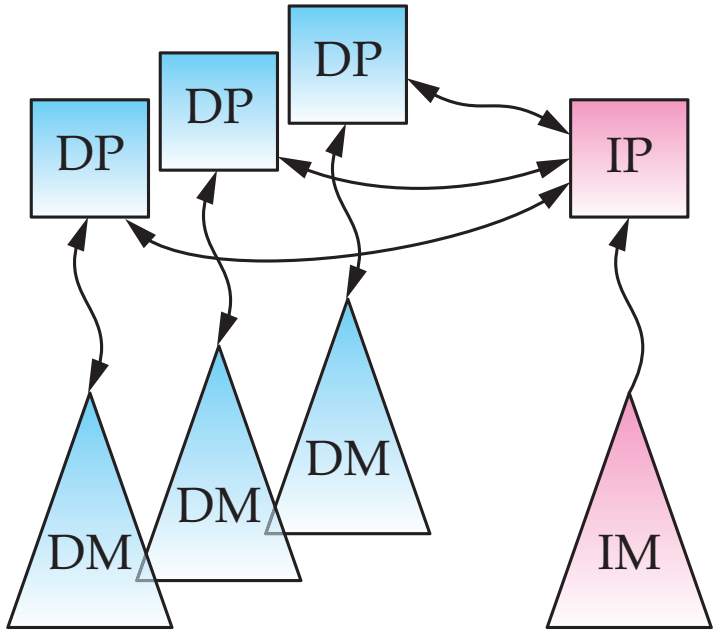
uniprocesor

wieloprocesor
słabo sprzężony

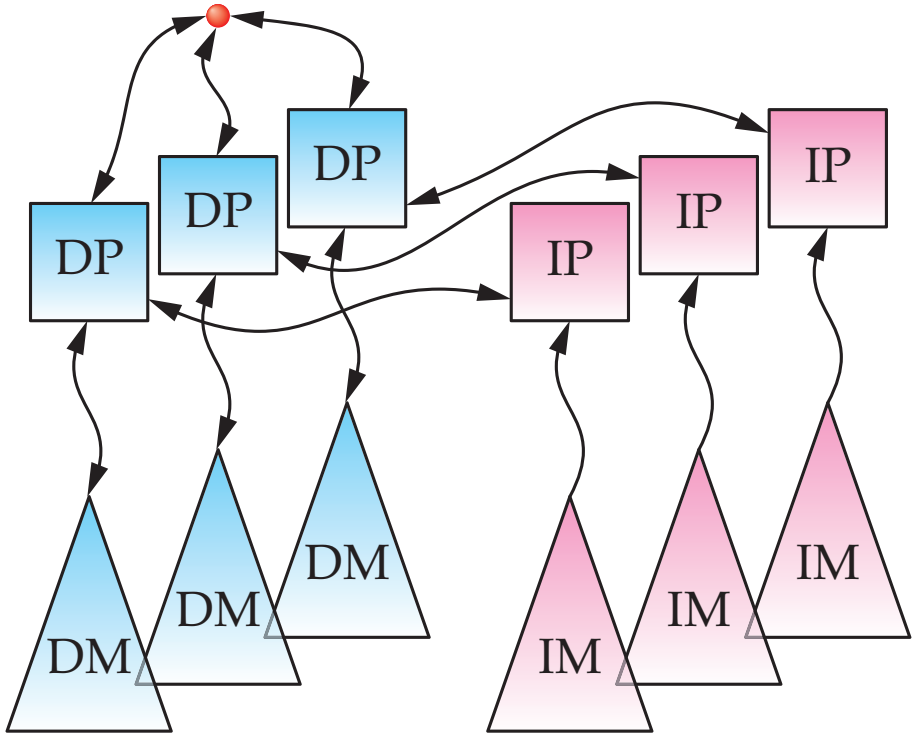
wieloprocesor
silnie sprzężony



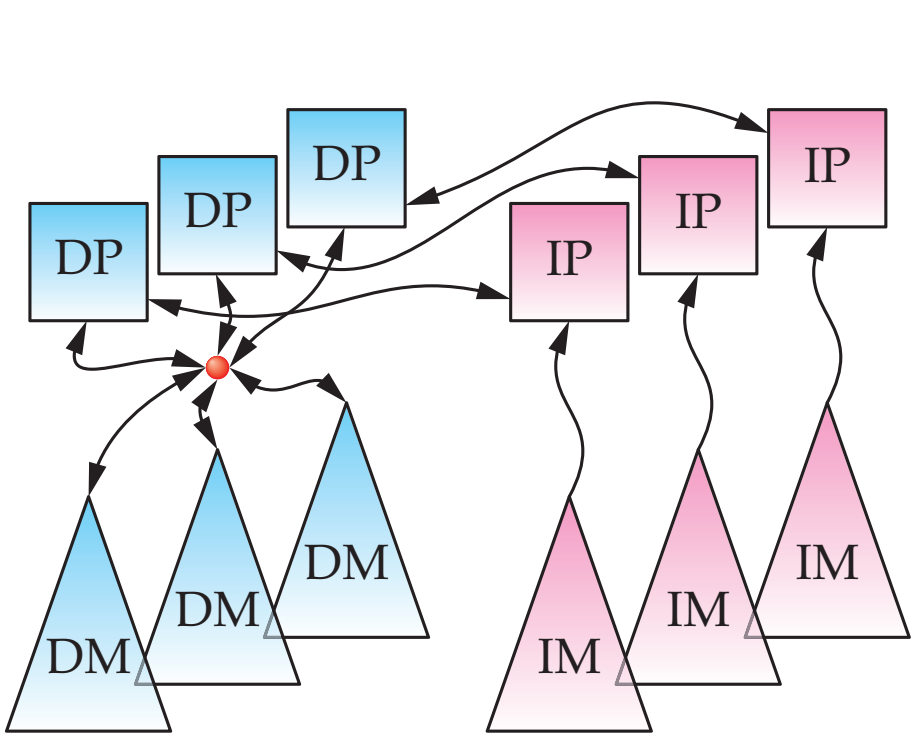
uniprocesor
von Neumanna



procesor
wektorowy



wieloprocesor
słabo sprzężony



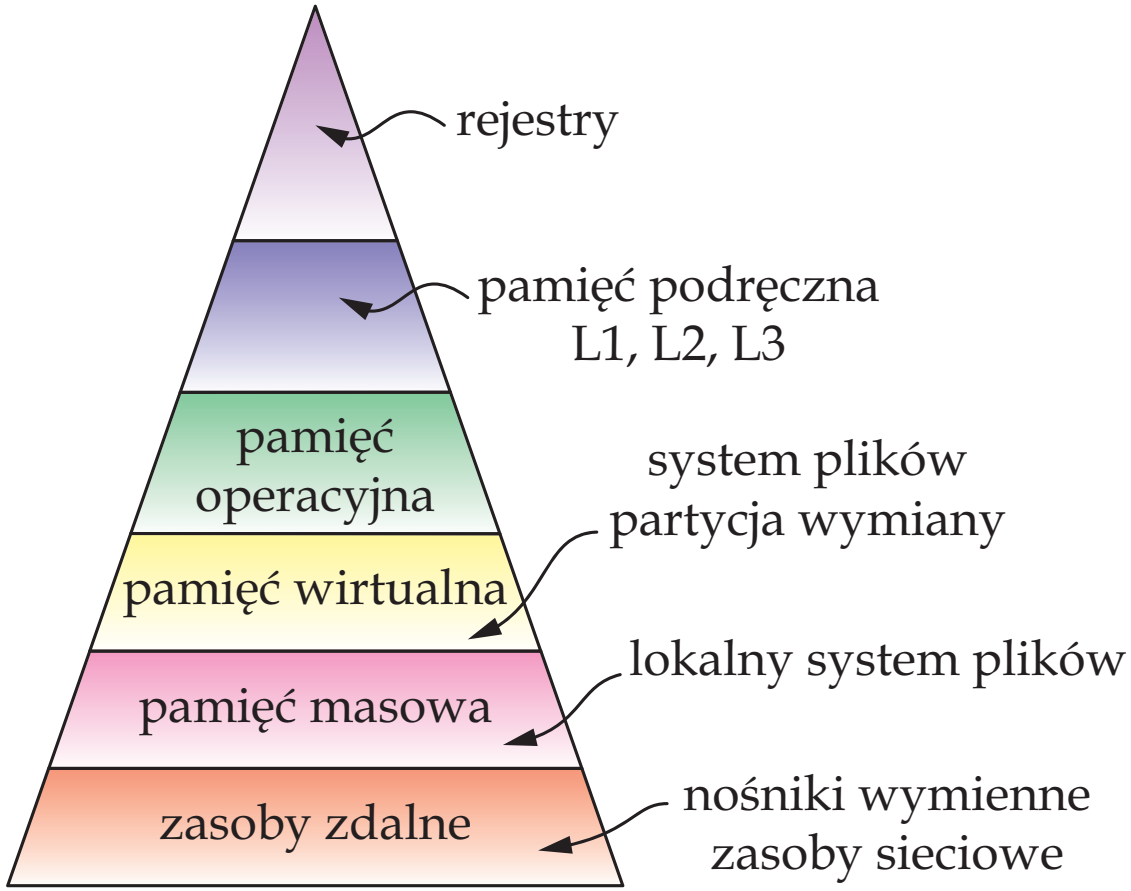
wieloprocesor
silnie sprzężony

Hierarchia pamięci

W taksonomia Skillicorna występuje pojęcie hierarchii pamięci (a nie pamięć). Słowo *hierarchia* dobrze oddaje budowę pamięci współczesnego komputera, w którym znajduje się kilka bloków funkcjonalnych służących do przechowywania programów i danych.

Idealny komputer powinien mieć jak największą i jak najszybszą pamięć. Pojemność pamięci wpływa na jej fizyczne rozmiary, a te – na czas dostępu. Nie można więc zbudować dowolnie dużej i jednocześnie szybkiej pamięci.

Problem ten rozwiązuje się przez wyodrębnienie wielu warstw o zróżnicowanej pojemności i szybkości, tworzących razem hierarchię pamięci. Kolejne warstwy w miarę oddalania się od procesora mają coraz większe pojemności i coraz dłuższe czasy dostępu.


rejestry
pamięć podręczna
L1, L2, L3
pamięć
operacyjna
system plików
partycja wymiany
pamięć wirtualna
lokalny system plików
pamięć masowa
nośniki wymienne
zasoby sieciowe
zasoby zdalne

Hierarchia pamięci współczesnego komputera, z punktu widzenia konstrukcji komputera, składa się z czterech warstw.

Rejestry fizycznie znajdują się wewnątrz procesora, dzięki czemu dostęp do nich jest bardzo szybki.

Pamięć podręczna (kieszeń, *cache*), wprowadzone po raz pierwszy około 1968 roku, zapewniają buforowanie danych pomiędzy procesorem i pamięcią operacyjną w celu przyspieszenie dostępu do pamięci.

Warstwa pamięci **wirtualnej**, powstała również około 1968 roku, zapewnia rozszerzenie pamięci operacyjnej.

Z punktu widzenia użytkownika do hierarchii pamięci należy zaliczyć wszelkie zasoby służące przechowywaniu danych. Logiczne staje się więc uzupełnienie rysunku o **lokalny system plików** komputera oraz o **zasoby zdalne**, w postaci nośników wymiennych i serwerów sieciowych.

Mechanizmy sterujące przemieszczaniem danych pomiędzy poszczególnymi warstwami są różne.

O umieszczeniu danych w rejestrach decyduje **programista** piszący program w asemblerze lub **kompilator** języka wysokiego poziomu.

Styk warstwy pamięci podręcznej i pamięci operacyjnej jest sterowany na **poziomie sprzętu**.

Stykiem pamięci operacyjnej i wirtualnej steruje **system operacyjny** przy użyciu jednostki zarządzania pamięcią.

O umieszczeniu danych w pamięci wirtualnej decyduje **użytkownik** – otwierając plik danych lub uruchamiając program.

Przemieszczaniem danych pomiędzy lokalnym systemem plików i nośnikami wymiennymi lub zasobami sieciowymi steruje **użytkownik**.

Tabela przedstawia orientacyjne parametry poszczególnych warstw hierarchii pamięci. Należy zwrócić uwagę na dużą różnicę czasów dostępu pamięci podręcznej i pamięci operacyjnej lub wirtualnej – czas podany dla pamięci dotyczy pojedynczego, losowego dostępu do pamięci dynamicznej typu DDR.

warstwa, pojemność, czas dostępu		
rejestry	< 1 kB	< 1 ns
pamięć podręczna L1	$\leqslant 128$ kB	≈ 1 ns
pamięć podręczna L2	$\leqslant 12$ MB	$1 \ldots 2$ ns
pamięć podręczna L3	$\leqslant 256$ MB	$2 \ldots 5$ ns
pamięć operacyjna	$\leqslant 64$ GB	$10 \ldots 50$ ns
pamięć wirtualna, system plików	$\geqslant 128$ GB	$\leqslant 10$ ms
nośniki wymienne, sieć komputerowa	∞	s, min.

Maszyna von Neumanna

Maszyna von Neumanna – model architektury komputera, opracowany przez Johna von Neumanna, Johna W. Mauchly'ego oraz Johna P. Eckerta w 1945 roku. Jej cechą charakterystyczną jest taki sam sposób przechowywania instrukcji i danych w hierarchii pamięci. Model maszyny von Neumanna wprowadza specyficzny mechanizm dostępu do pamięci – poprzez adres.

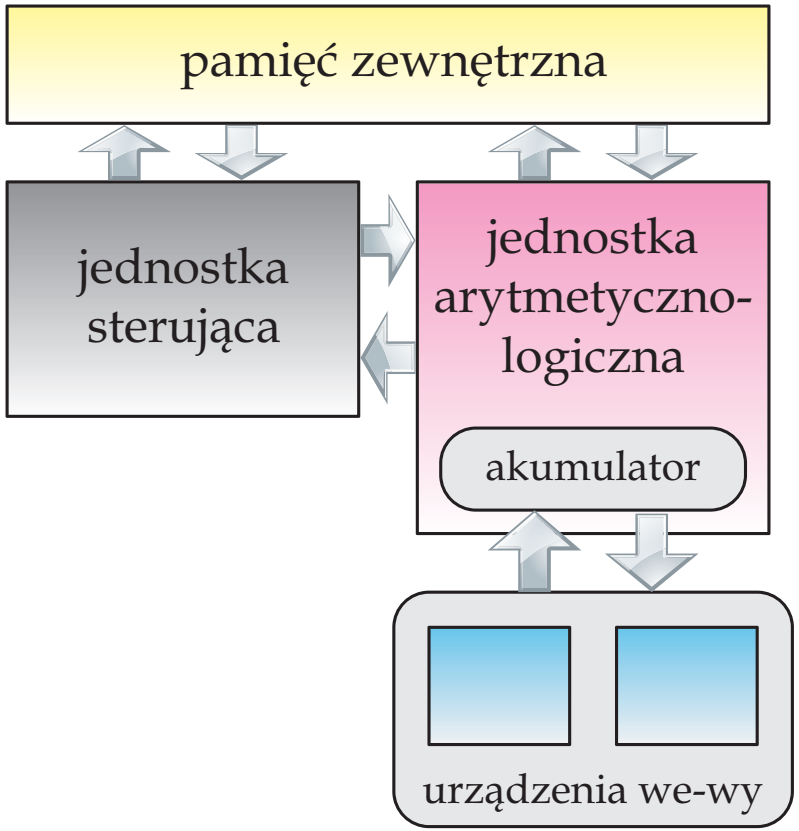
Maszyna von Neumanna następujące cechy:

- wykorzystuje model obliczeń zaproponowany przez Turinga, wykonuje obliczenia zgodnie z programem,
- program jest przechowywany w pamięci razem z danymi,
- pamięć składa się z pewnej liczby ponumerowanych komórek,
- dostęp do pamięci następuje poprzez podanie numeru komórki, czyli **adresu**,
- adres jest przechowywany i inkrementowany w specjalnym rejestrze procesora, zwanym **licznikiem instrukcji** (PC, *Program Counter*).

W architekturze tej komputer składa się z czterech głównych komponentów:

- pamięci komputerowej przechowującej dane programu oraz instrukcje programu, każda komórka pamięci ma unikatowy adres,
- jednostki sterującej odpowiedzialnej za pobieranie danych i instrukcji, pamięci oraz ich sekwencyjne przetwarzanie,
- jednostki arytmetyczno-logicznej odpowiedzialnej za wykonywanie podstawowych operacji arytmetycznych,
- rządzeń wejścia-wyjścia służących do interakcji z operatorem.

Jednostka sterująca wraz z jednostką arytmetyczno-logiczną tworzą procesor.


pamięć zewnętrzna
jednostka sterująca
jednostka arytmetyczno-logiczna
akumulator
urządzenia we-wy

System komputerowy zbudowany w oparciu o architekturę von Neumanna powinien:

- mieć skończoną i funkcjonalnie pełną listę rozkazów,
- mieć możliwość wprowadzenia programu do systemu komputerowego poprzez urządzenia zewnętrzne i jego przechowywanie w pamięci w sposób identyczny jak danych,
- dane i instrukcje w takim systemie powinny być jednakowo dostępne dla procesora,
- informacja jest tam przetwarzana dzięki sekwencyjnemu odczytywaniu instrukcji z pamięci komputera i wykonywaniu tych instrukcji w procesorze.

Podane warunki pozwalają przełączać system komputerowy z wykonania jednego zadania (programu) na inne bez fizycznej ingerencji w strukturę systemu, a tym samym gwarantują jego uniwersalność.

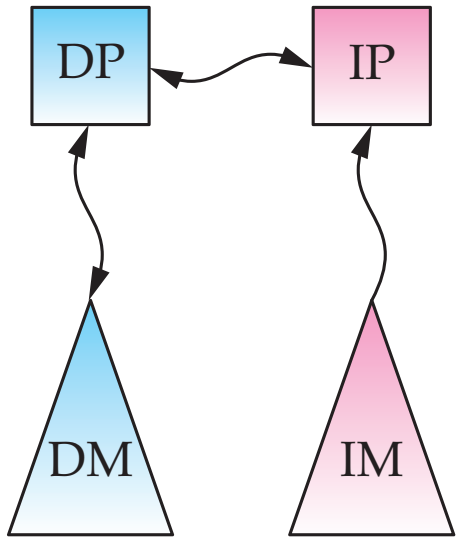
Dwa warianty architektury von Neumanna różnią, się sposobem przechowywania instrukcji i danych:

- architektura **Harvard** – oddzielne hierarchie pamięci danych i rozkazów,
- architektura **Princeton** – wspólna hierarchia pamięci danych i rozkazów.

Architektura Harvard jest niekiedy uważana za architekturę nie spełniającą postulatów von Neumanna wobec faktu oddzielnego przechowywania instrukcji i danych.

Cechy charakterystyczne architektury Harvard:

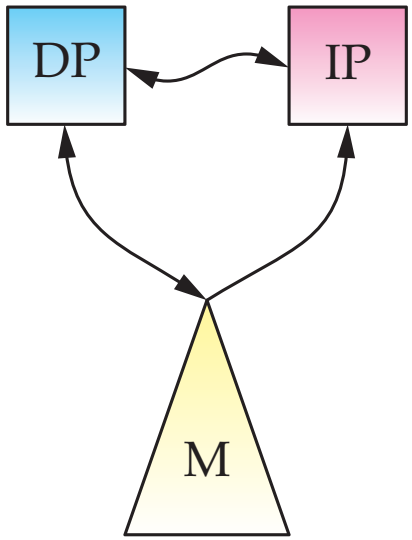
- oddzielne hierarchie pamięci danych i instrukcji,
- wysoka wydajność dzięki możliwości równoczesnego pobierania instrukcji i wykonywania operacji na hierarchii pamięci,
- brak możliwości programowania – nie ma możliwości zapisu instrukcji do pamięci instrukcji,
- komputer jest dostarczany ze stałym programem,
- jest wykorzystywana w procesorach sygnałowych oraz mikrokomputerach jednoukładowych (zastosowania wbudowane).



Uwaga! Rysunek posługuje się symbolami zapożyczonymi z taksonomii Skillicorna w sposób sprzeczny z zasadami budowy modeli wprowadzonymi przez tę taksonomię.

Cechy charakterystyczne architektury Princeton:

- *wzorcowa* realizacja maszyny von Neumanna ze wspólną hierarchią pamięci instrukcji i danych,
- nie można równocześnie pobierać danych i rozkazów (*von Neumann bottleneck*),
- nieograniczone możliwości modyfikacji programu,
- obiekt zapisany jako dana może być pobrany jako instrukcja,
- wykorzystywana w komputerach uniwersalnych.

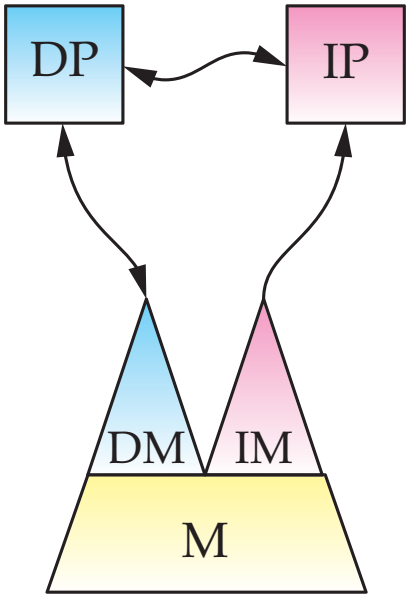


Uwaga! Rysunek posługuje się symbolami zapożyczonymi z taksonomii Skillicorna w sposób sprzeczny z zasadami budowy modeli wprowadzonymi przez tę taksonomię.

Zmodyfikowana architektura harwardzka – architektura mieszana, łączy w sobie cechy architektury harwardzkiej i architektury von Neumanna. Oddzielone zostały częściowo hierarchie pamięci na dane i instrukcje.

Cechy charakterystyczne architektury Harvard-Princeton:

- realizacja maszyny von Neumanna z oddzielonymi *górnymi* warstwami hierarchii pamięci i wspólnymi *dolnymi*,
- przynajmniej jeden poziom pamięci podręcznej jest oddzielny dla procesorów instrukcji i danych,
- szybkie działanie dzięki równoległości dostępów jak w architekturze Harvard,
- możliwość programowania niezbędna w komputerach uniwersalnych, jak w architekturze Princeton,
- program użytkowy nie ma pełnej kontroli nad położeniem obiektów w hierarchii pamięci, brak możliwości modyfikacji,
- w/w kontrolę może mieć wyróżniony program (proces) – system operacyjny, większość współczesnych komputerów uniwersalnych, w tym komputery PC ma architekturę Harvard-Princeton.

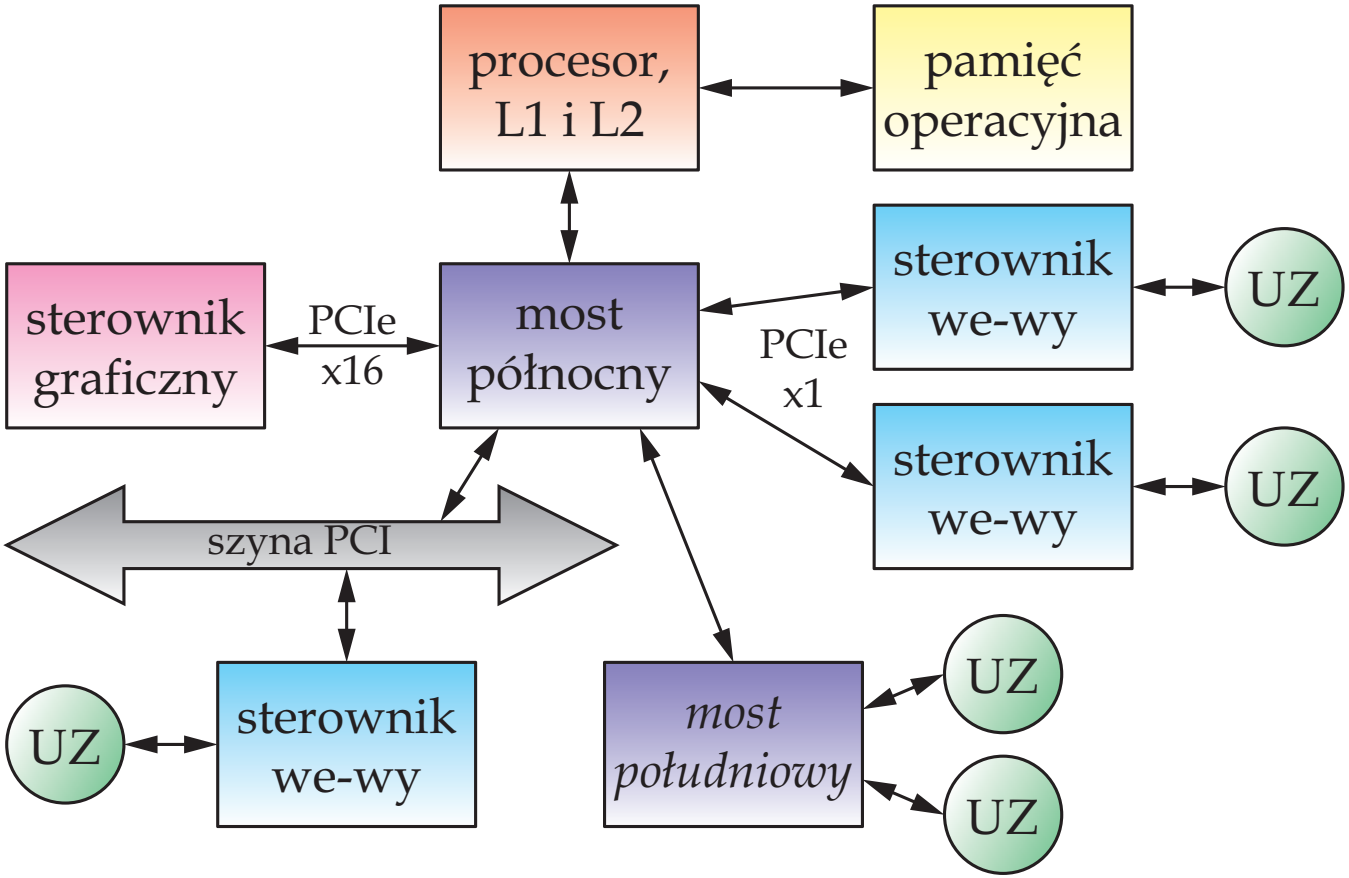


Uwaga! Rysunek posługuje się symbolami zapożyczonymi z taksonomii Skillicorna w sposób sprzeczny z zasadami budowy modeli wprowadzonymi przez tę taksonomię.

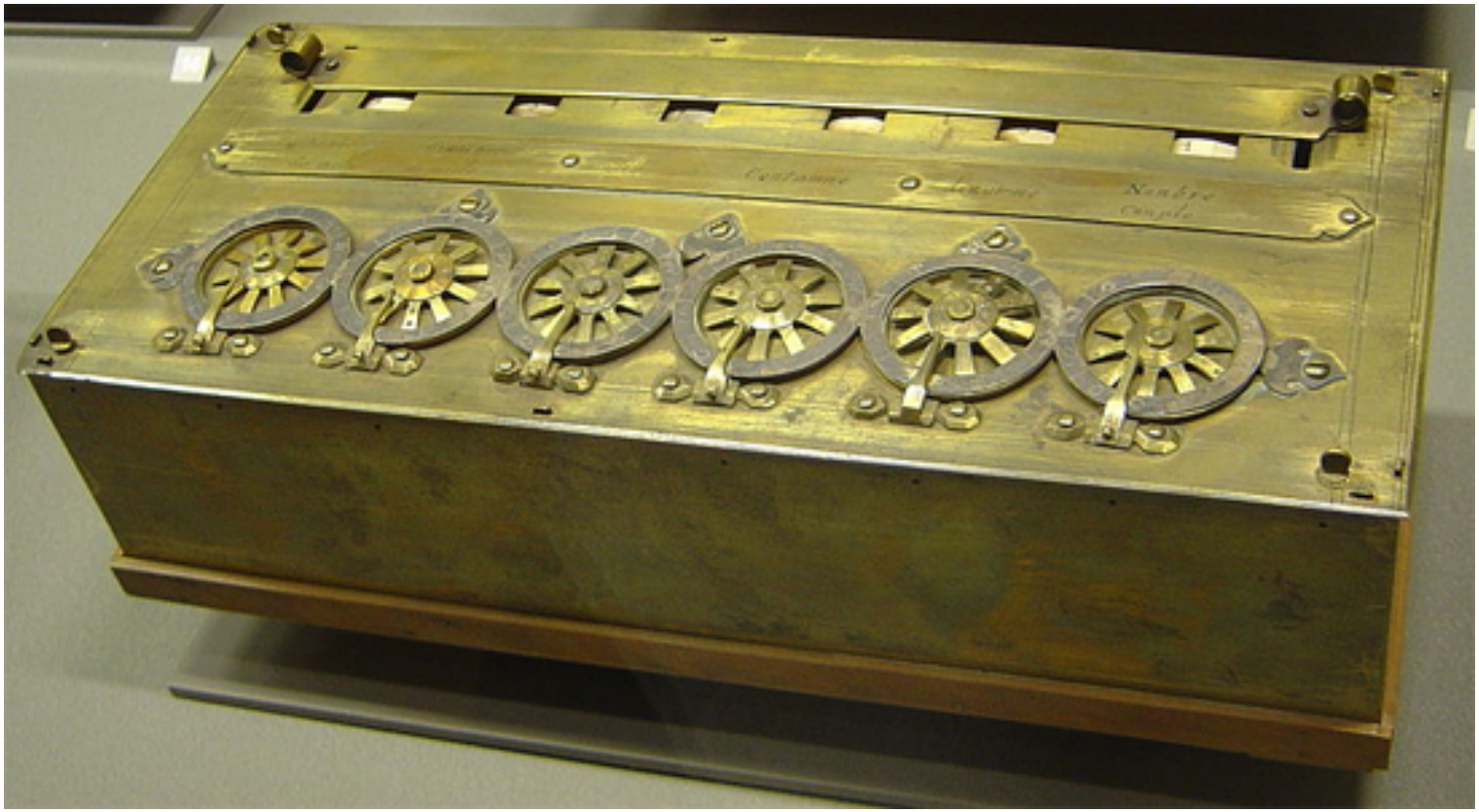
Architektura z połączeniami punkt-punkt

Architektura obecnie konstruowanych komputerów (klasy PC) jest bardziej złożona. We współczesnych komputerach **sterownik pamięci** umieszczony jest w **procesorze**. **Most północny** jest wyposażony w indywidualne łącza dla sterowników urządzeń zewnętrznych, zrealizowane w standardzie **PCI express**.

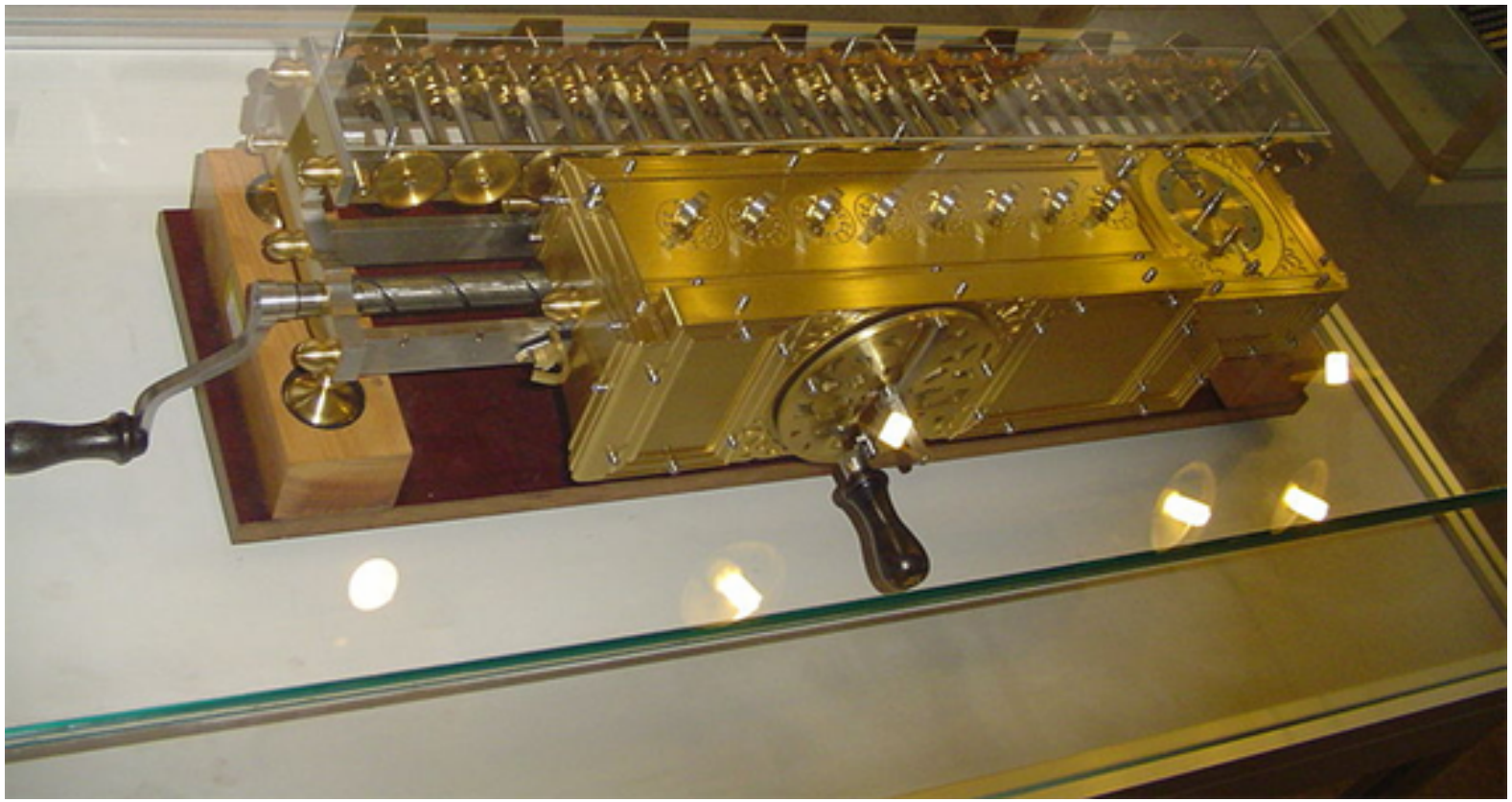
Most południowy jest zintegrowanym **sterownikiem urządzeń zewnętrznych**. Szyna **PCI** została zachowana w celu umożliwienia podłączenia starszych sterowników urządzeń. Jest ona przeznaczona do usunięcia.


procesor,
L1 i L2
pamięć
operacyjna
sterownik
graficzny
PCIe
x16
most
północny
PCIe
x1
sterownik
we-wy
UZ
sterownik
we-wy
UZ
szyna PCI
UZ
sterownik
we-wy
most
południowy
UZ
UZ

Pascalina – maszyna licząca zaprojektowana przez Blaise Pascala około 1645 roku. Pascalina umożliwiała jedynie dodawanie i odejmowanie liczb – była więc mechanicznym sumatorem.

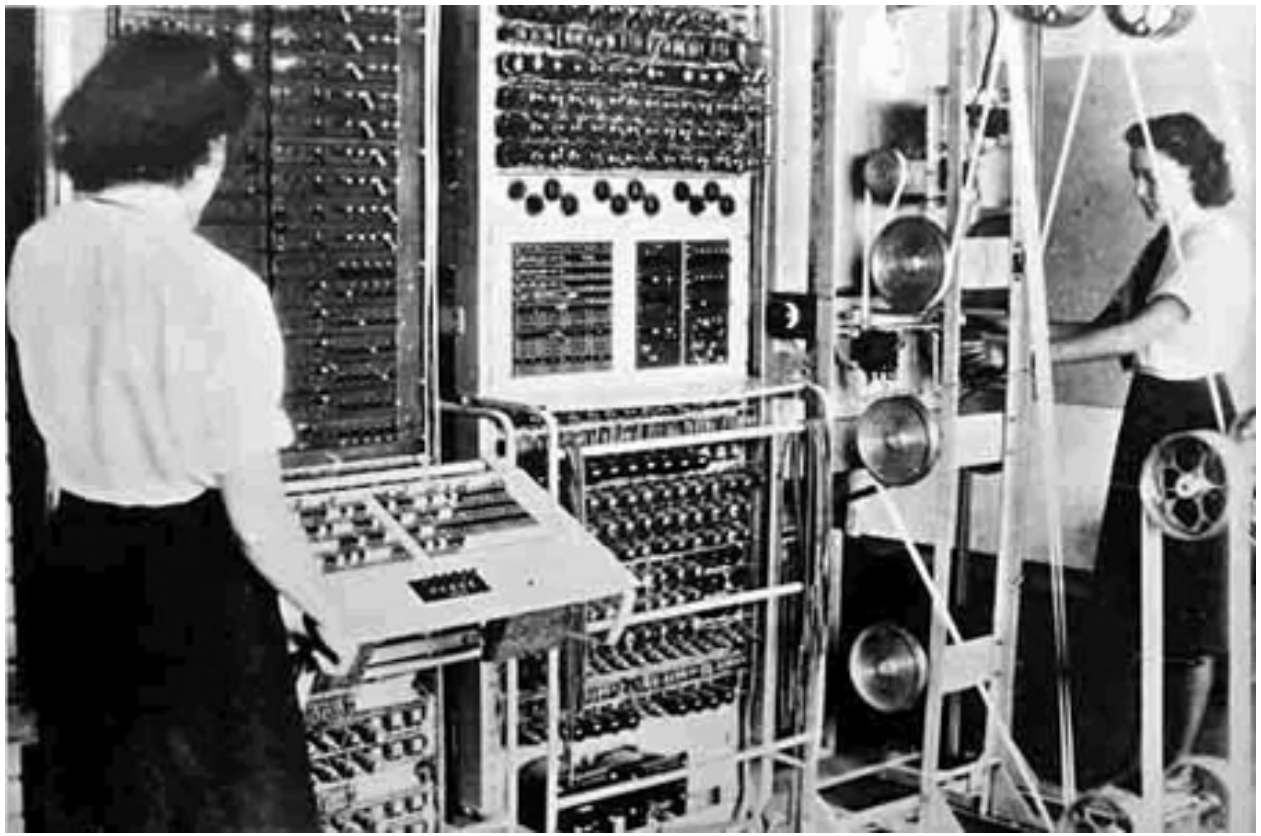


Lebendige Rechenbank, dosł. *żywa ława do obliczeń* – maszyna licząca zaprojektowana i wykonana przez Gottfrieda Wilhelma Leibniza w 1671. Maszyna ta była formą kalkulatora mechanicznego. Była zdolna do odejmowania, mnożenia, dzielenia i wyprowadzania pierwiastków kwadratowych.

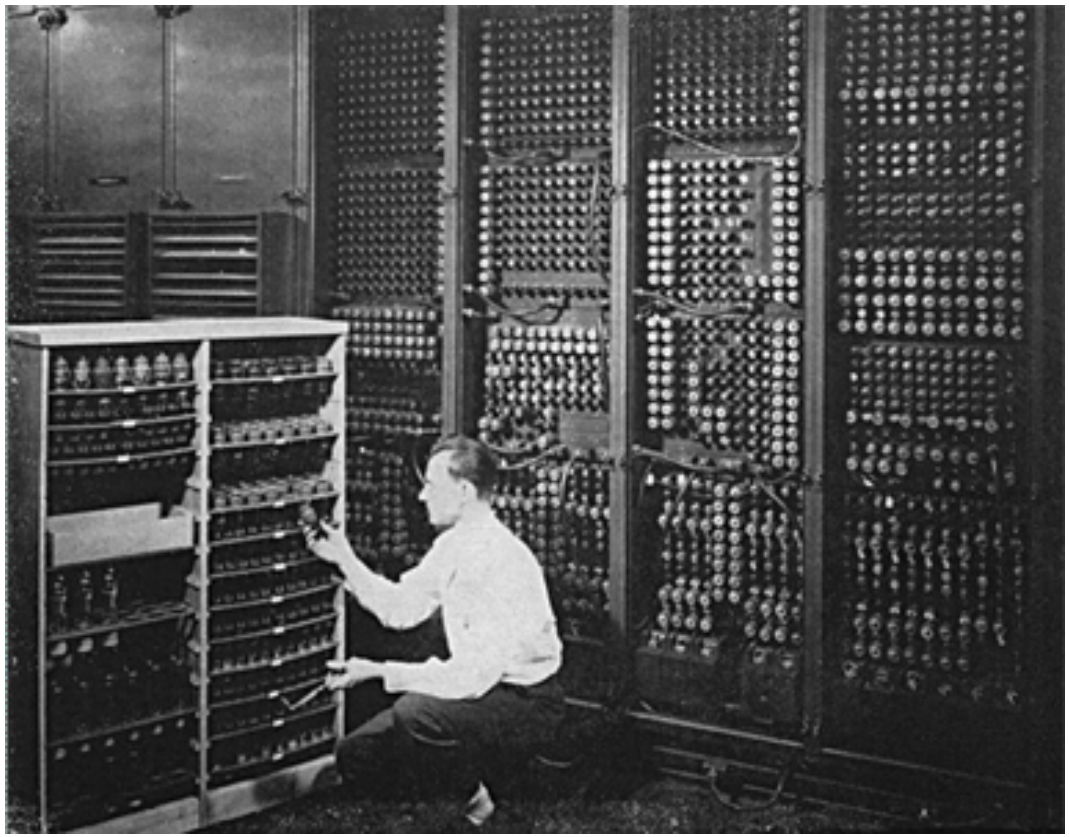


Z3 – to pierwszy działający, w pełni automatyczny komputer o zmiennym programie zbudowany przez niemieckiego inżyniera Konrada Zuse w 1941 roku na bazie jego wcześniejszej, mechanicznej konstrukcji, Z1. Maszyna była wykorzystywana w czasie wojny do obliczeń niezbędnych przy projektowaniu skrzydeł.

Colossus – seria programowalnych maszyn cyfrowych oparta na teoretycznych podstawach prac Alana Turinga. Projektem Colossus kierowali Max Newman i Tommy Flowers, uczestniczył w nim również Alan Turing. Colossus został zbudowany 14 kwietnia w 1943 roku w brytyjskim ośrodku kryptograficznym Bletchley Park i przeznaczony był do zastosowań wojskowych. Służył do rozpracowania sposobu działania niemieckiej Maszyny Lorenza i łamania jej szyfrów.



ENIAC, ang. *Electronic Numerical Integrator And Computer* – komputer skonstruowany w latach 1943-1945 przez J. P. Eckerta i J. W. Mauchly'ego na Uniwersytecie Pensylwanii w USA. Zaprzestano jego używania w 1955. Do roku 1975 powszechnie uważany był za pierwszy komputer na świecie, jednak teraz o miano to ubiegają się również – po odtajnieniu danych brytyjskich maszyny Colossus oraz niemieckie Konrada Zuse.



Informacja a komputer

Wewnętrzna budowa komputera wymusza używanie ściśle określonych sposobów kodowania (systemów liczbowych) i przetwarzania informacji.

Komputer ENIAC pracował w **systemie dziesiętnym**! Przełączniki elektryczne musiały rozpoznawać 10 stanów napięcia – jedna z przyczyn wolnej pracy i małej odporności na zniekształcenia elektryczne.

System dziesiętny nie jest jedynym sposobem przedstawiania liczb.

System binarny (dwójkowy) – urządzenia korzystające z niego muszą rozpoznawać tylko dwa stany, czyli **1** (jest sygnał) lub **0** (nie ma sygnału) a nie dziesięć stanów jak to ma miejsce w systemie dziesiętnym.

Zaletami systemu binarnego, w kontekście budowy komputera sa:

- znacznie prostszy układ elektroniczny,
- poziom sygnałów elektrycznych, rozpatrywanych jako odrębne stany mogą należeć do szerszego przedziału napięć,
- większa odporność na zniekształcenia.

Dla dowolnego systemu liczenia istnieje zbiór cyfr, z których tworzone są liczby.

Systemy liczbowe dzieli się na:

- pozycyjne,
- niepozycyjne.

Systemy niepozycyjne

W systemach **niepozycyjnych** poszczególne cyfry zachowują swą wartość liczbową bez względu na miejsce jakie zajmują w liczbie (np. system rzymski – siedem znaków: I, V, X, L, C, D, M).

XII = 10 (X) + 1 (I) + 1 (I) = 12
IX = 10 (X) - 1 (I) = 9
MCDX = 1000 (M) + 500 (D) - 100 (C) + 10 (X) = 1410

Systemy pozycyjne

W systemach **pozycyjnych** wartość liczbowa cyfry zależy od umiejscowienia (pozycji) w liczbie (np. system dziesiętny, dwójkowy).

Liczba różnych cyfr systemu nazywa się jego **podstawą** P.

Wartość liczbowa cyfry określona jest przez **wagę**. Waga na pozycji n równa jest podstawie P podniesionej do potęgi n.

Sposób zapisu liczby L w dowolnym systemie pozycyjnym jest bardzo prosty:

$$L = a_{n-1} \cdot P^{n-1} + a_{n-2} \cdot P^{n-2} + \ldots + a_0 \cdot P^0 = \sum_{i=0}^{n-1} a_i \cdot P^i.$$

Dla systemu **dziesiętnego** (decymalnego):

$$P = 10, \quad a_n = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\},$$

$$142 = 1 \cdot 10^2 + 4 \cdot 10^1 + 2 \cdot 10^0 = 1 \cdot 100 + 4 \cdot 10 + 2 \cdot 1.$$

Dla systemu **dwójkowego** (binarnego):

$$P = 2, \quad a_n = \{0, 1\},$$

$$10010_2 = 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 1 \cdot 16 + 0 \cdot 8 + 0 \cdot 4 + 1 \cdot 2 + 0 \cdot 1 = 18.$$

Dla systemu **szesnastkowego** (heksadecymalnego):

$$P = 16, \quad a_n = \{0, 1, 3, 4, 5, 6, 7, 8, 9, \mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}\},$$

$$3e8_{16} = 3 \cdot 16^2 + 14 \cdot 16^1 + 8 \cdot 16^0 = 3 \cdot 256 + 14 \cdot 16 + 8 \cdot 1 = 1000.$$

Konwersja liczb naturalnych

$$190 = 1 \cdot 10^2 + 9 \cdot 10^1 + 0 \cdot 10^0,$$

$$190 = ?_2$$

dzielenie, wynik		
$190 / 2$	$= 95$	reszta $= 0$
$95 / 2$	$= 47$	reszta $= 1$
$47 / 2$	$= 23$	reszta $= 1$
$23 / 2$	$= 11$	reszta $= 1$
$11 / 2$	$= 5$	reszta $= 1$
$5 / 2$	$= 2$	reszta $= 1$
$2 / 2$	$= 1$	reszta $= 0$
$1 / 2$	$= 0$	reszta $= 1$

$$190 = 10111110_2.$$

$$10111110_2 = 1 \cdot 2^7 + 0 \cdot 2^6 + 1 \cdot 2^5 + 1 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 =$$

$$= 1 \cdot 128 + 0 \cdot 64 + 1 \cdot 32 + 1 \cdot 16 + 1 \cdot 8 + 1 \cdot 4 + 1 \cdot 2 + 0 \cdot 1.$$

Używając jednego bajtu, można zapisać liczę z zakresu $0 \ldots 255$, czyli $00000000_2 \ldots 11111111_2$ i $00_{16} \ldots ff_{16}$.

Działania arytmetyczne na liczbach w systemie dwójkowym i szesnastkowym są odpowiednikiem działań w systemie dziesiętnym i opierają się na elementarnych działaniach, np.:

- $1_2 + 0_2 = 1_2$,
- $1_2 + 1_2 = 10_2$,
- $1_2 \cdot 0_2 = 0_2$,
- $1_2 \cdot 1_2 = 1_2$,
- $10_2 - 1_2 = 1_2$.

Przykłady:

$$\begin{array}{r} \ 1\ 1\ 1\ 1\ 1\ 1 \\ \ 1\ 1\ 1\ 1\ 1\ 1\ 1 \\ +\quad\ \ 1\ 0\ 0\ 1\ 1 \\ \hline 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0 \end{array}$$

$$\begin{array}{r} \ 1\ 1\ 1\ 1\ 1\ 1\ 1 \\ -\quad\ \ 1\ 0\ 0\ 1\ 1 \\ \hline 1\ 1\ 0\ 1\ 1\ 0\ 0 \end{array}$$

$$\begin{array}{r} \ \ 1\ 1\ 0\ 1 \\ \times\quad\ \ 1\ 0\ 1\ 1 \\ \hline \ \ 1\ 1\ 0\ 1 \\ \ 1\ 1\ 0\ 1 \\ \ 0\ 0\ 0\ 0 \\ +\ 1\ 1\ 0\ 1 \\ \hline 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1 \end{array}$$

$$\begin{array}{r|l} \ 1\ 1\ 0 & \\ \hline \ 1\ 1\ 0\ 1 & / \ 1\ 0 \\ -\ 1\ 0 & \\ \hline \ 0\ 1\ 0\ 1 & \\ -\quad 1\ 0 & \\ \hline \ 0\ 0\ 0\ 1 & \\ -\quad\ \ 1\ 0 & \\ \hline \ 0\ 0\ 0\ 1 & \end{array}$$

Jednostki informacji

Najmniejszą jednostką informacji potrzebna do określenia, który z dwóch równie prawdopodobnych stanów przyjął układ jest **bit**, **b** (*BInary digiT*). Bit odpowiada informacji Tak – Nie, Prawda – Fałsz, 1 – 0.

Jest to również najmniejsza jednostka informacji używana w odniesieniu do sprzętu komputerowego. Bit przyjmuje wartości wtedy **0** lub **1**.

Wyższe jednostki to informacji:

- półbajt (*nybble*) – 4 bity,
- **oktet** (*octet*) – 8 bitów,
- **bajt** (*byte*) **B** – pierwotnie liczba bitów przetwarzana jednocześnie przez komputer lub adresowana przez procesor, obecnie używany wyłącznie do oznaczania 8 bitów (czyli oktetu).

Często używa się też:

- **słowo** (*word*) to zwykle 16 bitów (2 bajty) – jednostka informacji, na której operuje komputer, mogą zdarzać się słowa o innej długości, np.: 4, 8, 16 bajtów,
- **słowo procesora** – jednostka informacji o długości naturalnej dla procesora,
- **słowo pamięci** – jednostka informacji możliwa do przesłania w jednym cyklu transmisji do/z pamięci, zwykle 64 b, czasem 128 b.

Do określenia większej liczby bajtów stosuje się często (**niepoprawnie**) przedrostki **kilo**, **mega**, **giga** itd. Są to przedrostki dziesiętne układu SI, będące wielokrotnościami liczby 10 ($10^{3 \cdot n}$) a nie liczby 2:

- **kilobajt** (*kilobyte*), **kB** – $10^3 = 1000$ bajtów,
- **megabajt** (*megabyte*), **MB** – $10^6 = 1000^2 = 1$ milion bajtów,
- **gigabajt** (*gigabyte*), **GB** – $10^9 = 1000^3 = 1$ miliard bajtów,
- **terabajt** (*terabyte*), **TB** – $10^{12} = 1000^4 = 1$ bilion bajtów.

W celu odróżnienia przedrostków o mnożniku 1000 od przedrostków o mnożniku 1024, pojawiła się propozycja ujednoznacznienia, opracowana przez IEC, polegająca na dodawaniu litery **i** po symbolu przedrostka dwójkowego oraz **bi** po jego nazwie.

Nowe przedrostki nazywane zostały przedrostkami **dwójkowymi** (binarnymi). Jednak ta propozycja rozwiązania problemu niejednoznaczności przedrostków nie została przyjęta przez wszystkie środowiska. Przedrostki dwójkowe są wielokrotnościami liczby 2 ($2^{10 \cdot n}$):

- **kibibajt** (*kibibyte*), **KiB** – $2^{10} = 1024$ bajty,
- **mebibajt** (*mebibyte*), **MiB** – $2^{20} = 1024^2 = 1$ milion 48 tysięcy 576 bajtów,
- **gibibajt** (*gibibyte*), **GiB** – $2^{30} = 1024^3 = 1$ miliard 73 miliony 741 tysięcy 824 bajtów,
- **tebibajt** (*tebibyte*), **TiB** – $2^{40} = 1024^4 = 1$ bilion 99 miliardów 511 milionów 627 tysięcy 776 bajtów.

Współczesne komputery są używane do przetwarzania danych różnych typów – liczbowych, logicznych, tekstowych, a także obrazów i dźwięków. Działają w oparciu o system binarny.

- Wartości logiczne (prawda/fałsz).
- Znaki pisarskie.
- Liczby
  - całkowite
    - nieujemne
    - ze znakiem
  - niecałkowite
    - stałopozycyjne
    - zmiennopozycyjne
- Dźwięki, sygnały jednowymiarowe.
- Obrazy rastrowe.

Wszystkie dane, na których operuje komputer, są zapisane w postaci tbciągów cyfr binarnych – bitów, interpretowanych najczęściej jako liczby binarne.

Wszelkie dane o charakterze innym niż liczby, muszą być zapisane (zakodowane) w postaci liczb lub grup liczb.

Komputer przetwarza wyłącznie grupy bitów, tworząc liczby binarne, których długość wynosi $8 \cdot 2^n$ bitów (np. 8, 16, 32, 64).

Dane alfanumeryczne

Dane alfanumeryczne, tekstowe – mają postać znaków pisarskich – liter, cyfr, znaków przestankowych i innych symboli. W komputerze są one reprezentowane przez liczby, określające pozycję danego symbolu w tablicy kodowej.

We współczesnych komputerach używa się kilku standardów kodowania znaków pisarskich, najpopularniejsze to:

- ASCII,
- UNICODE.

ASCII

ASCII – (*American Standard Code for Information Interchange*), 7-bitowy kod przyporządkowujący liczby z zakresu 0 . . . 127 literom (alfabetu angielskiego), cyfrom, znakom przestankowym i innym symbolom oraz poleceniom sterującym. Został opracowany dla urządzeń dalekopisowych.

Na przykład litera **a** jest kodowana liczbą 97, a znak spacji jest kodowany liczbą 32.

Litery, cyfry oraz inne znaki drukowane tworzą zbiór znaków ASCII. Jest to 95 znaków o kodach 32 . . . 126. Pozostałe 33 kody (0 . . . 31 i 127) to tzw. kody sterujące służące do sterowania urządzeniem odbierającym komunikat, np. drukarką lub terminalem.

- Kody sterujące zajmują pozycje: $0 \ldots 31$, w tym:
  - **CR** - powrót na początek wiersza: kod 13,
  - **LF** – przejście do następnego wiersza: kod 10,
  - inne ważne: **HT** (tabulacja pozioma), **FF** (rozpoczęcie nowej strony), **BEL** (sygnał dźwiękowy), **VT** (tabulacja pionowa), **BSP** (cofnięcie o jeden znak).
- **Spacja**: kod 32 ($0x20$).
- **Cyfry** $0 \ldots 9$: kody $48 \ldots 57$ ($0x30 \ldots 0x39$).
- **Litery** w kolejności alfabetycznej:
  - wielkie: kody $65 \ldots 90$ ($0x41 \ldots 0x5a$),
  - małe: kody $97 \ldots 122$ ($0x61 \ldots 0x7a$).
- **Kasowanie znaku**: kod 127 ($0x7f$).

Ponieważ kod ASCII jest 7-bitowy, a większość komputerów operuje na 8-bitowych bajtach, dodatkowy bit został wykorzystany na powiększenie zbioru kodowanych znaków do 256 symboli.

Powstało wiele różnych rozszerzeń ASCII wykorzystujących ósmy bit. W kodach tych pierwsze 128 pozycji jest identyczne, jak w kodzie ASCII, a następne 128 pozycji zawiera znaki dodatkowe, np. litery akcentowane, rozszerzony zestaw symboli matematycznych, litery alfabetów narodowych (słowiańskie, skandynawskie, cyrylica, etc.).

Istnieje wiele rozszerzeń tej rodziny, używanych w różnych częściach świata. W Polsce najpowszechniej używa się kodów **ISO8859-2** oraz Microsoft **CP1250**, nazywanych stronami kodowymi.

BIN	DEC	HEX	znak	BIN	DEC	HEX	znak
00000000	0	00	NUL (*Null*)	01000001	65	41	A
00000001	1	01	SOH (*Start Of Heading*)	01000010	66	42	B
00000010	2	02	STX (*Start of Text*)	01000011	67	43	C
00000011	3	03	ETX (*End of Text*)	01000100	68	44	D
...				...			
00110000	48	30	0	01110111	119	77	w
00110001	49	31	1	01111000	120	78	x
00110010	50	32	2	01111001	121	79	y
00110011	51	33	3	011 1010	122	7A	z
...				...			
00111101	61	3D	=	01111100	124	7C	|
00111110	62	3E	>	01111101	125	7D	}
00111111	63	3F	?	01111110	126	7E	~
01000000	64	40	@	01111111	127	7F	DEL (*Delete*)
...				...			

Unicode

Unicode – uniwersalny komputerowy zestaw znaków mający w zamierzeniu obejmować wszystkie znakówi pisarskie zapisu fonetycznego (głoskowego) używanych na całym świecie. Liczba pozycji kodowych jest praktycznie nieograniczona, obecnie jest zdefiniowanych kilkadziesiąt tysięcy znaków.

Definiują go dwa standardy – Unicode oraz **ISO10646**. Znaki obu standardów są identyczne. Standardy te różnią się w drobnych kwestiach, m.in. Unicode określa sposób składu.

Unicode jest rozwijany jest przez konsorcjum, w którego skład wchodzą firmy komputerowe, producenci oprogramowania, instytuty naukowe, agencje międzynarodowe oraz grupy zainteresowanych użytkowników. Konsorcjum współpracuje z organizacją ISO.

Standard Unicode obejmuje przydział przestrzeni numeracyjnej poszczególnym grupom znaków, nie obejmuje zaś sposobów bajtowego kodowania znaków.

Jest kilka metod kodowania, oznaczanych skrótowcami **UCS** (*Universal Character Set*) i **UTF** (*Unicode Transformation Format*). Do najważniejszych należą:

- UTF-32/UCS-4,
- UTF-16,
- UTF-8.

Kody pierwszych 256 znaków Unicode pokrywają się z kodami ISO Latin 1 (czyli ISO8859-1), przez co kody pierwszych 128 znaków pokrywają się z kodami ASCII.

Należy jednak pamiętać, że jest to zbieżność wyłącznie numerów przyporządkowanych konkretnym znakom, wartości bajtów użytych do ich zapisania mogą się różnić od tych, które uzyska się stosując Latin 1 lub ASCII.

UTF-8 – zmiennej długości system kodowania znaków dla Unicode. Może reprezentować każdy znak w systemie Unicode. Kodowanie zostało zaprojektowane dla zapewnienia zgodności z ASCII.

Zalety kodowania UTF-8:

- każdy tekst w ASCII jest tekstem w UTF-8,
- żaden znak spoza ASCII nie zawiera bajtu z ASCII,
- typowy tekst ISO Latin-x rozrasta się w bardzo niewielkim stopniu po przekonwertowaniu do UTF-8,
- znaki o kodzie różnym od 0 nie zawierają bajtu 0, co pozwala stosować UTF-8 w ciągach zakończonych zerem,
- o każdym bajcie wiadomo, czy jest początkiem znaku, czy też leży w jego środku,
- nie zawiera bajtów 0$x$ff i 0$x$fe, więc łatwo można go odróżnić od tekstu UTF-16.
- brak problemów z uporządkowaniem bajtów (*endianness*),
- jest domyślnym kodowaniem w ML (również w jego aplikacjach: HTML, SVG, XSL, CML, MathML).

Dźwięk i obraz

Dźwięk jest zapamiętywany w postaci wartości napięcia reprezentującego chwilowe ciśnienie akustyczne. Wartość ta jest kwantowana z użyciem rozpiętości zwykle od 8 bitów do 32 bitów z częstotliwością zależną od potrzeb, zwykle od 8 kHz do 48 kHz (próbkowanie).

Obraz rastrowy jest zapisany w postaci prostokątnej macierzy punktów (**pikseli**). Każdemu pikselowi odpowiada jeden kolor, zapisany w postaci składowych – jasności świateł (barw) podstawowych. Wartości jasności zapisane są w postaci liczb.

Dane logiczne

Najprostszy typ danych stanowią dane **logiczne**. Mogą one przyjmować dwie wartości. Bajtowe adresowanie danych używane w komputerach oraz fakt, że wiele komputerów traktuje jako podstawowy format danych słowo 32-bitowe powodują, że dane logiczne są zwykle zapisywane w postaci bajtów lub słów, pomimo, że do ich zapisu wystarczyłby pojedynczy bit.

Należy zwrócić uwagę na reprezentację wartości *prawda* i *fałsz* w różnych językach programowania. Wartość *fałsz* jest zwykle reprezentowana przez słowo o wszystkich bitach równych 0.

Wartość *prawda* może być reprezentowana przez liczbę całkowitą równą 1, liczbę całkowitą o dowolnej wartości różnej od zera (również ujemną), słowo o wszystkich bitach równych 1.

Liczby całkowite nieujemne

W dalszej będziemy posługiwali się założeniem, że dane reprezentowane są przez słowo komputera, w którym poszczególne bity zostały ponumerowane od prawej do lewej strony.

NBC

Kod **NBC** (*Natural Binary Code*, NKB, naturalny kod binarny) jest w zasadzie tym samym co pozycyjny system dwójkowy. Numer bitu jest równy wykładnikowi jego wagi binarnej.

$$L_{NCB} = \sum_{i=0}^{n-1} b_i \cdot 2^i.$$

Wartości wag w kodzie NBC (dla bajtu)

waga	$2^7 = 128$	$2^6 = 64$	$2^5 = 32$	$2^4 = 16$	$2^3 = 8$	$2^2 = 4$	$2^1 = 2$	$2^0 = 1$
cyfra	b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0

Liczby całkowite ze znakiem

2C

Zapis **2C** (*Two's Complement*, U2, uzupełnień do 2) jest najczęściej stosowanym zapisem liczb całkowitych. Jest on podobny do NKB, z tą różnicą, że najbardziej znaczący bit ma wagę ujemną. Typ `int` jest we współczesnych komputerach implementowany jako zapis 2C. Negacja liczby w tym systemie polega na negacji bitowej i inkrementacji.

$$L_{2C} = -b_{n-1} \cdot 2^{n-1} + \sum_{i=0}^{n-2} b_i \cdot 2^i.$$

Wartości wag w kodzie 2C

waga	$-(2^8) = -128$	$2^6 = 64$	$2^5 = 32$	$2^4 = 16$	$2^3 = 8$	$2^2 = 4$	$2^1 = 2$	$2^0 = 1$
cyfra	b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0

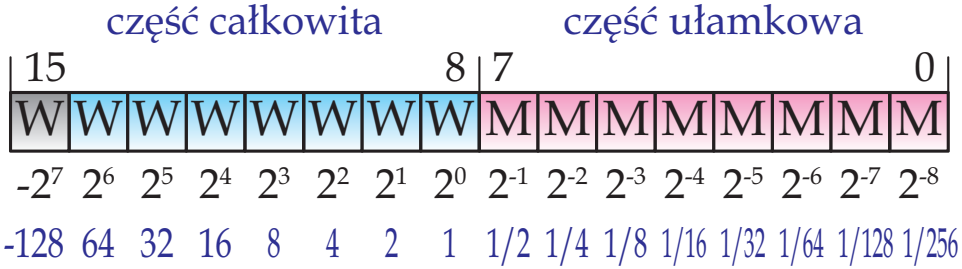
Liczby niecałkowite

System stałopozycyjny

Do zapisywania liczb ułamkowych i mieszanych można użyć zapisu **stałopozycyjnego**. W zapisie tym liczba jest reprezentowana przez słowo binarne, w którym pewne, z góry określone liczby bitów reprezentują część całkowitą i część ułamkową liczby.

Odpowiada to interpretacji zapisu całkowitoliczbowego, pomnożonej przez wartość będącą ujemną potęgą liczby 2.

Do zapisu liczb bez znaku używa się jako bazowej postaci NKB, a do zapisu liczb ze znakiem kodu U2 (zwykle).



Komputery zazwyczaj nie obsługują w szczególny sposób zapisów stałopozycyjnych. Podstawowe operacje są wykonywane tak samo, jak na liczbach całkowitych, odmienna jest jedynie interpretacja zapisu, za którą jest odpowiedzialny wyłącznie programista.

Zapis zmiennopozycyjny dla systemu dziesiętnego

Zapis **zmiennopozycyjny** umożliwia zapisywanie liczb całkowitych i ułamkowych o bardzo dużym zakresie dynamiki wartości bezwzględnych.

Każda liczba może być zapisana na kilka sposobów, różniących się położeniem przecinka oddzielającego część całkowitą od ułamkowej i wartością wykładnika:

$$-1,2345 \cdot 10^5 \qquad -0,12345 \cdot 10^6 \qquad -12,345 \cdot 10^4.$$

Wartość liczby zmiennoprzecinkowej jest obliczana według wzoru:

$$L = Z \cdot M \cdot B^W,$$

gdzie:

- Z (*sign*) – znak liczby, 1 lub -1,
- M (*mantissa*) – mantysa, liczba ułamkowa,
- B (*base*) – podstawa systemu liczbowego,
- W (*exponent*) – wykładnik (cecha), liczba całkowita.

Postacią **znormalizowaną** nazwiemy liczbę, która ma część całkowitą części znaczącej wyrażoną przez pojedynczą cyfrę różną od zera.

Aby zapisać (przechować) liczbę, musimy zapisać jej znak, część znaczącą oraz wykładnik, który jest liczbą całkowitą ze znakiem. Podstawa systemu jest ustalona i jej zapis jest zbędny.

$$-1,2345e5$$

W postaci znormalizowanej nie da się zapisać zera, ponieważ zero nie ma żadnej cyfry znaczącej różnej od 0.

Binarny zapis zmiennopozycyjny

Obecnie niemal wszystkie komputery posługują się binarnym zapisem zmiennopozycyjnym zgodnym ze standardem IEEE754.

Standard zakłada, że, o ile tylko jest to możliwe, liczby zapisuje się w postaci znormalizowanej. **Bazą** systemu jest liczba 2, **wykładnik** określa potęgę liczby 2.

Ponieważ w systemie binarnym jedyną cyfrą różną od zera jest jedynka, każda liczba w postaci znormalizowanej ma część całkowitą równą 1, nie ma więc potrzeby zapisywania jej, zapisuje się tylko część ułamkową.

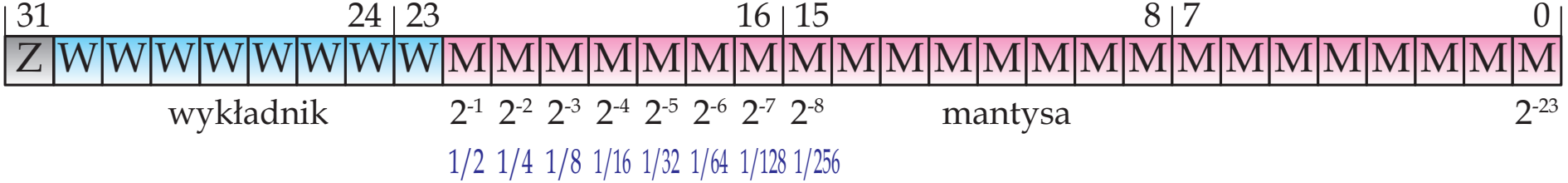
Wykładnik jest liczbą całkowitą ze znakiem. W IEEE754 wykładnik jest zapisywany w kodzie **spolaryzowanym** (biased), w którym wartość podkładu jest określona wzorcem bitowym $01\ldots11$, o liczbie bitów równej szerokości pola bitowego wykładnika. Dwie wartości pola wykładnika są zarezerwowane i oznaczają, że zapis nie reprezentuje postaci znormalizowanej.

Bity wykładnika o wzorcu $00\ldots00$ oznaczają zapis zdenormalizowany. Wartość wykładnika jest w tym przypadku taka sama, jak przy zapisie znormalizowanym z wzorcem wykładnika $00\ldots01$, a część całkowita części znaczącej ma wartość 0 (a nie 1 jak w postaci znormalizowanej).

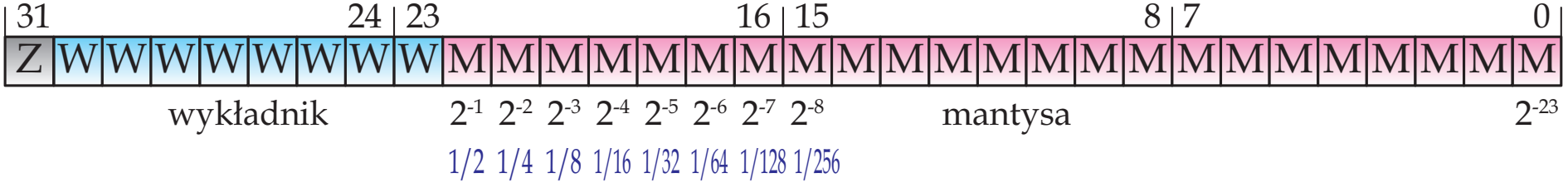
Pole wykładnika o wzorcu $11\ldots11$ oznacza nie-liczby: wartości **nieskończone** i wartości **błędne**.

Liczba pojedynczej precyzji zajmuje 4 bajty (32 bity).

Mantysa ma 23 bity, wykładnik ma 8 bitów, znak 1 bit, a nadmiar wynosi 127.



Mantysa jest **znormalizowana**, tj. należy do przedziału $[1, 2)$ (przedział prawostronnie otwarty). Jeżeli M jest stałe, a W zmienia się, wówczas przesunięciu ulega przecinek – stąd właśnie pochodzi nazwa tej reprezentacji.



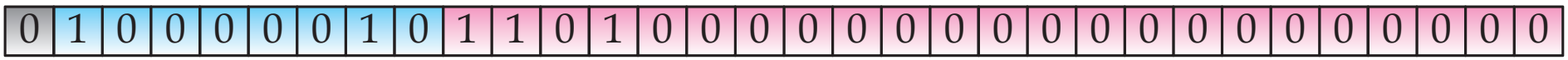
Procedura zapisu liczby L:

- określenie znaku, bit 31 $= 1$ jeżeli L jest ujemna, 0 jeśli dodatnia,
- znalezienie największej liczby postaci 2^W, mniejszej niż L,
- zapisanie wykładnika z nadmiarem $W + 127$,
- dzielenie liczby $L/2^W$ (wynik 1.xxxx),
- odjęcie 1 i szukanie mantysy,
- zaznaczenie bitu jako 1 jeśli po odjęciu $1/2$ (bit $+ 1$) mamy wartość nieujemną, jeśli wartość jest ujemna, bit oznaczany jest jako 0, i ignorowana operacja odejmowania,
- procedura jest powtarzana aż w wyniku odejmowania otrzymamy 0 lub dojdzie do bitu nr 0.

Przykład dla liczby 14,5:

- liczba jest dodatnia $\Rightarrow$ bit 31 $= 0$,
- największa liczba 2^W mniejsza niż 14,5 to $3^3 = 8 \Rightarrow W = 3$,
- zapisanie wykładnika $127 + W = 130 \Rightarrow 10000010$,
- dzielenie $14,5/2^3 = 1,8125$,
- odejmowanie $1,8125 - 1 = 0,8125$,
  - $0,8125 - 1/2 = 0.3125 \Rightarrow$ bit 23 $= 1$,
  - $0,3125 - 1/4 = 0.0625 \Rightarrow$ bit 22 $= 1$,
  - $0,0625 - 1/8 = -0.0625 \Rightarrow$ bit 21 $= 0$ (ignorowanie operacji),
  - $0,0625 - 1/16 = 0 \Rightarrow$ bit 20 $= 1$,
- pozostałe bity mantysy $= 0$.

14,5



Niektóre wartości specjalne

wartość,	wykładnik,	mantysa,	uwagi
NaN	11...11	xx...xx	symbol nie reprezentuje wartości liczbowej, powstaje zazwyczaj w wyniku niedozwolonej operacji (np. pierwiastkowanie liczby ujemnej)
INF	11...11	00...00	bit znaku decyduje o znaku wartości, rozróżnia się $+\infty$ i $-\infty$
0	00...00	00...00	bit znaku decyduje o znaku wartości, jest wynikiem operacji w przypadku wystąpienia nadmiaru (przepełnienia), przy dzieleniu przez 0, itp., może być dodatnia lub ujemna

Standard IEEE754, definiuje dwie klasy liczb:

- pojedynczej precyzji (`single`),
- podwójnej precyzji (`double`).

Podstawowym formatem jest format typu `double` – 64-bitowy (podwójnej precyzji).

Format 32-bitowy jest używany w zastosowaniach, gdzie wymagana precyzja jest niewielka, ma on tylko 23 bity znaczące (typ `single`).

Liczba **pojedynczej** precyzji: mantysa 23 bity, wykładnik 8 bitów, znak 1 bit, nadmiar $2^8 - 1 = 127$.

Liczba **podwójnej** precyzji: mantysa 52 bity, wykładnik 11 bitów, znak 1 bit, nadmiar $2^{11} - 1 = 1023$.

Format 80-bitowy (`extended`) był używany w starszych jednostkach zmiennopozycyjnych procesorów rodziny x86. Obecnie wychodzi z użycia.

Format 128-bitowy jest formatem *przyszłościowym* dla liczb o dużej precyzji.

Precyzja – liczba miejsc po przecinku jest określona przez mantysę.

Pojedyncza precyzja
Mantysa ma 23 bity, $1/2^{23} \approx 1,2 \cdot 10^{-7} \Rightarrow$ 7 cyfr po przecinku.

Podwójna precyzja
Mantysa ma 52 bity, $1/2^{52} \approx 2,2 \cdot 10^{-16} \Rightarrow$ 15 - 16 cyfr po przecinku.

Rozszerzona precyzja
Mantysa ma 64 bity, $1/2^{64} \approx 5,4 \cdot 10^{-20} \Rightarrow$ 19 cyfr po przecinku.

Operacje arytmetyczne

Mamy dwie liczby zmiennoprzecinkowe: $L_1 = M_1 \cdot B^{W_1}$ i $L_2 = M_2 \cdot B^{W_2}$, przy czym $L_1 > L_2$, wówczas:

- dodawanie
$$L_1 + L_2 = \left(M_1 + M_2 \cdot B^{W_2 - W_1}\right) \cdot B^{E_1},$$
- odejmowanie
$$L_1 - L_2 = \left(M_1 - M_2 \cdot B^{W_2 - W_1}\right) \cdot B^{E_1}.$$

Mamy dwie liczby zmiennoprzecinkowe: $L_1 = Z_1 \cdot M_1 \cdot B_1^W$ i $L_2 = Z_2 \cdot M_2 \cdot B_2^W$, wówczas:

- mnożenie
$$L_1 \cdot L_2 = (Z_1 \cdot Z_2) \cdot (M_1 \cdot M_2) \cdot B^{W_1 + W_2},$$
- dzielenie
$$L_1 / L_2 = (Z_1 \cdot Z_2) \cdot (M_1 / M_2) \cdot B^{W_1 - W_2}.$$

Jeśli liczby mają różne wykładniki, to podczas dodawania mantysa liczby o mniejszym wykładniku musi zostać **zdenormalizowana** – we wzorze jest to przemnożenie M_2 przez czynnik $B^{W_2-W_1}$.

W szczególnym przypadku, jeśli $W_1 - W_2$ jest większe niż liczba cyfr mantysy, to po denormalizacji mantysa będzie miała wartość 0, a liczba o mniejszym wykładniku nie wpłynie na wynik dodawania bądź odejmowania.

Liczby o skończonej reprezentacji dziesiętnej mogą mieć nieskończoną reprezentację binarną (np.: 0,1; 0,3).

Reprezentacja zmiennopozycyjna jest reprezentacją przybliżoną, a wyniki operacji są w rzeczywistości przybliżeniami. Oznacza to, że w praktyce nie można stosować relacji równości w odniesieniu do liczb zmiennopozycyjnych ($|a - b| < \epsilon$).

Arytmetyka liczb zmiennopozycyjnych **nie jest łączna**, dla x i y może zachodzić:

$$(x + y) + z \neq x + (y + z),$$
$$(x \cdot y) \cdot z \neq x \cdot (y \cdot z),$$

nie jest też rozdzielna, czyli:

$$x \cdot (y + z) \neq (x \cdot y) + (x \cdot z).$$

Innymi słowy, kolejność wykonywania operacji arytmetycznych ma znaczenie i wpływa na końcowy wynik.

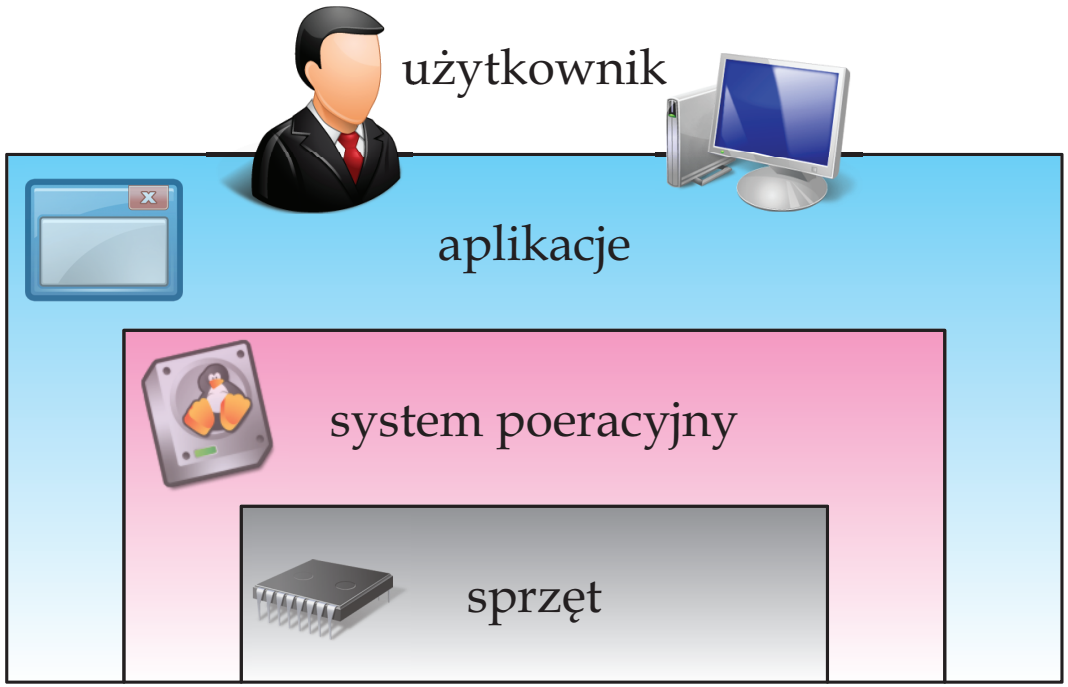
Pojęcie **systemu operacyjnego** jest trudne do zdefiniowania w zwartej formie. Krótki opis jest zbyt ogólny, żeby określić rolę i sposób działania tego specyficznego oprogramowania. Podobnie, trudne jest określenie elementów składowych systemu operacyjnego, czyli jednoznaczne oddzielenie oprogramowania systemowego od aplikacyjnego.

System operacyjny

System operacyjny jest warstwą oprogramowania operującą bezpośrednio na sprzęcie, której celem jest zarządzanie zasobami systemu komputerowego i stworzenie użytkownikowi środowiska łatwiejszego do zrozumienia i wykorzystania.

Należy zwrócić uwagę na dwie zasadnicze kwestie:

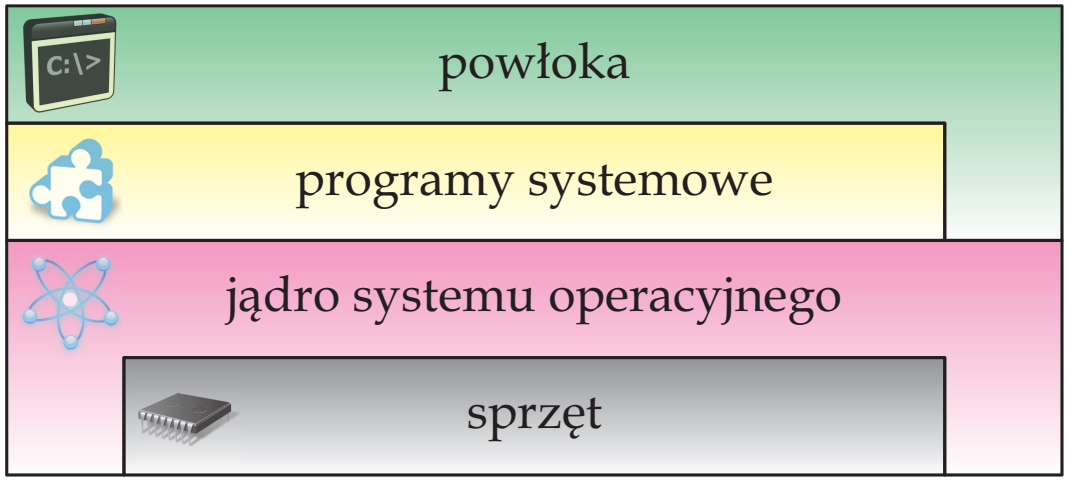
- zarządzanie zasobami systemu komputerowego,
- stworzenie środowiska dla użytkownika.



System operacyjny pośredniczy pomiędzy **użytkownikiem** a **sprzętem**, dostarczając **wygodnego środowiska** do **wykonywania programów**. Użytkownik końcowy korzysta z programów (aplikacji), na potrzeby których przydzielane są zasoby systemu komputerowego.

Przydziałem tym zarządza system operacyjny, dzięki czemu można uzyskać stosunkowo duży stopień niezależności programów od konkretnego sprzętu oraz odpowiedni poziom bezpieczeństwa i sprawności działania.

Nie ma precyzyjnego określenia, które składniki wchodzą w skład systemu operacyjnego jako jego części. Najczęściej akceptuje się definicję *marketingową*, zgodnie z którą to wszystko, co producent udostępnia w ramach zbioru oprogramowania nazywanego systemem operacyjnym, stanowi jego cześć (podejście takie było przyczyną wielu problemów prawnych firmy Microsoft).



W ogólnym przypadku w strukturze systemu operacyjnego wyróżnia się **jądro** oraz **programy systemowe**, które dostarczane są razem z systemem operacyjnym, ale nie stanowią integralnej części jądra. Jądro jest zbiorem **modułów**, które ukrywają szczegóły sprzętowej realizacji systemu komputerowego, udostępniając pewien zestaw **usług**, wykorzystywanych między innymi do implementacji programów systemowych. W dalszej części system operacyjny będzie rozumiany głównie jako jądro, i inne elementy oprogramowania integralnie związane z funkcjonowaniem jądra.

Z punktu widzenia kontaktu z użytkownikiem istotny jest **interpreter poleceń**, który może być częścią jądra lub programem systemowym (np. w systemach UNIX/Linux). Interpreter wykonuje pewne polecenia **wewnętrznie**, tzn. moduł lub program interpretera dostarcza implementacji tych poleceń. Jeśli interpreter nie może wykonać wewnętrznie jakiegoś polecenia, uruchamia odpowiedni program (polecenie **zewnętrzne**), jako odrębny proces.

Zadania systemu operacyjnego

- Definicja interfejsu użytkownika.
- Udostępnianie systemu plików.
- Udostępnianie środowiska do wykonywania programów użytkownika:
  - mechanizm ładowania i uruchamiania programów,
  - mechanizmy synchronizacji i komunikacji procesów.
- Sterowanie urządzeniami wejścia-wyjścia.
- Obsługa podstawowej klasy błędów.

Definiowanie **interfejsu użytkownika** – system operacyjny dostarcza użytkownikom zbiór poleceń lub system okienkowy wraz z odpowiednim menu, który umożliwia interakcję z systemem komputerowym.

Udostępnianie systemu plików – system operacyjny organizuje i ułatwia dostęp do informacji np. w postaci hierarchicznego systemu plików.

Udostępnianie środowisko do wykonywania programów: system operacyjny dostarcza struktur danych do organizacji wykonywania programu oraz zachowywania i odtwarzania stanu przetwarzania (procesy i przełączanie kontekstu). System operacyjny udostępnia też programistom mechanizmy komunikacji pomiędzy procesami (kolejki komunikatów, strumienie, pamięć współdzielona) oraz mechanizmy synchronizacji procesów (semafory, muteksy).

Sterowanie urządzeniami wejścia-wyjścia – odpowiednie moduły sterujące, integrowane z systemem operacyjnym, inicjalizują pracę urządzeń zewnętrznych oraz pośredniczą w efektywnym przekazywaniu danych pomiędzy jednostką centralną a tymi urządzeniami.

Obsługa podstawowej klasy błędów – system operacyjny reaguje na błędy użytkowników (niedostępność zasobów, brak prawa dostępu), programistów (błąd dzielenia przez 0, naruszenie ochrony pamięci, nieprawidłowy kod rozkazu) oraz systemu (błąd braku strony pamięci, błąd magistrali).

Efektywność zarządzania zasobami oraz wygodny interfejs dla użytkownika są dwoma ogólnymi, niezależnymi celami projektowymi systemów operacyjnych. Pierwszy z tych celów był kluczowy w rozwoju rodziny systemów uniksowych. Dopiero w późniejszych etapach ich rozwoju pojawił się intuicyjny okienkowy interfejs użytkownika.

Systemy rodziny MS Windows zorientowane były natomiast przede wszystkim na interfejs użytkownika, na bazie którego w późniejszych etapach rozwoju powstawał *pełnowartościowy* system operacyjny, uwzględniający szerzej rozumiane zarządzanie zasobami.

Zarządzanie zasobami systemu operacyjnego

- Przydział zasobów.
- Planowanie dostępu do zasobów.
- Ochrona i autoryzacja dostępu do zasobów.
- Odzyskiwanie zasobów.
- Rozliczanie – gromadzenie danych o wykorzystaniu zasobów.

Na nasze potrzeby, **zasób** będzie rozumiany intuicyjnie jako **element systemu komputerowego** istotny, czy wręcz kluczowy dla realizacji przetwarzania. Funkcja zarządzania zasobami nie jest bezpośrednio wykorzystywana przez użytkownika (czasami nie jest przez niego w ogóle dostrzegana). Jej celem jest optymalizacja wykorzystania zasobów przez użytkowników.

Przydział zasobów – realizacja żądań dostępu do zasobów w taki sposób, że zasoby używane są zgodnie z intencją użytkowników (np. zagwarantowanie wyłącznego dostępu drukarki).

Planowanie dostępu do zasobów – strategia przydziału zasobów gwarantująca bezpieczeństwo, żywotność, brak zakleszczenia, sprawiedliwość oraz optymalność ich wykorzystania. Warto zwrócić uwagę na odróżnienie planowania od samego przydziału – przydział oznacza powiązanie zasobu z realizowanym zadaniem, podczas gdy planowanie wiąże się z odejmowaniem decyzji odnośnie wyboru zdania, któremu zasób będzie przydzielony.

Ochrona i autoryzacja dostępu do zasobów – dopuszczanie możliwości użytkowania zasobu tylko przez osoby uprawnione i w zakresie przydzielonych im uprawnień.

Odzyskiwanie zasobów – dołączanie zwolnionych zasobów do zbioru zasobów wolnych po zakończeniu ich użytkowania.

Rozliczanie – rejestrowanie i udostępnianie informacji o wykorzystaniu zasobów w celach kontrolnych i rozrachunkowych.

Zasoby zarządzane przez system operacyjny

Typowymi zasobami zarządzanymi przez system operacyjny są: procesor, pamięć, urządzenia wejścia-wyjścia (w tym system plików, stanowiący tzw. wirtualne wejście-wyjście). Zależnie od zasobu zarządzanie charakteryzuje się pewną specyfiką.

- Procesor – przydział czasu procesora.
- Pamięć:
  - alokacja przestrzeni adresowej dla procesów,
  - ochrona i transformacja adresów.
- Urządzenia wejścia-wyjścia:
  - udostępnianie i sterowanie urządzeniami pamięci masowej,
  - alokacja przestrzeni dyskowej,
  - udostępnianie i sterownie drukarkami, skanerami itp.
- Informacja (system plików):
  - organizacja i udostępnianie informacji,
  - ochrona i autoryzacja dostępu do informacji.

Procesor jest zasobem współdzielonym przez wiele procesów, w związku z czym zadaniem systemu operacyjnego jest przydział kwantu czasu procesora i wywłaszczanie zadania, które:

- wykorzystało już swój czas,
- nie może kontynuować przetwarzania ze względu na brak innych zasobów (brak gotowości urządzeń wejścia-wyjścia),
- ma zbyt niski priorytet.

Pamięć jest zasobem, który przydzielany jest na wyłączność danego przetwarzania. Zadaniem systemu jest zatem utrzymywanie informacji o zajętości przestrzeni adresowej, znajdowanie i przydzielanie odpowiednich fragmentów wolnej pamięci na żądanie aplikacji użytkownika lub innych modułów systemu operacyjnego oraz reagowanie na naruszenie ochrony pamięci.

System operacyjny pośredniczy również w transformacji adresów wirtualnych na fizyczne w systemach z segmentacją lub stronicowaniem przez organizację tablicy segmentów lub stron oraz obsługę błędów strony.

Urządzenia zewnętrzne są stosunkowo wolne, w związku z czym, w celu poprawy efektywności, zarządzanie tymi urządzeniami wymaga odpowiedniej organizacji systemu przerwań oraz buforowania danych, ewentualnie spoolingu.

Osobnym zagadnieniem jest zarządzanie urządzeniami pamięci masowej, zwłaszcza odpowiednia organizacja przestrzeni dyskowej oraz optymalizacja ruchu głowic dyskowych.

Informacja jest z punktu widzenia użytkownika najważniejszym zasobem, gdyż jej przetwarzanie jest celem systemu komputerowego. System operacyjny odpowiada za organizację gromadzenia i udostępniania informacji, jej ochronę przed nieuprawnioną ingerencją, spójność w przypadku awarii itp.

Klasyfikacja systemów operacyjnych (sposób przetwarzania)

- Systemy przetwarzania bezpośredniego (*on-line processing systems*) – systemy interakcyjne:
  - występuje bezpośrednia interakcja pomiędzy użytkownikiem a systemem,
  - wykonywanie zadania użytkownika rozpoczyna się zaraz po przedłożeniu.
- Systemy przetwarzania pośredniego (*off-line processing systems*) – systemy wsadowe:
  - występuje znacząca zwłoka czasowa między przedłożeniem a rozpoczęciem wykonywania zadania,
  - niemożliwa jest ingerencja użytkownika w wykonywanie zadania.

W przypadku **systemu przetwarzania bezpośredniego** użytkownik wprowadza zadanie do systemu i oczekuje na wyniki. W trakcie przetwarzania jest zatem możliwa interakcja pomiędzy użytkownikiem a systemem (aplikacją). Użytkownik może być na przykład poproszony o wprowadzenie jakiś danych na terminalu, wybranie czegoś z menu itp.

W przypadku **przetwarzania pośredniego** zadanie jest realizowane w czasie wybranym przez system. Po przedłożeniu zadania ingerencja użytkownika jest niemożliwa. Wszystkie dane muszą być zatem dostępne w momencie przedkładania zadania, a jakikolwiek błąd programowy lub niekompletność danych oznacza konieczność przedłożenia i wykonania zadania ponownie.

Klasyfikacja systemów operacyjnych (liczba wykonywanych programów)

- Systemy jednozadaniowe – niedopuszczalne jest rozpoczęcie wykonywania następnego zadania użytkownika przed zakończeniem poprzedniego.
- Systemy wielozadaniowe – dopuszczalne jest istnienie jednocześnie wielu zadań (procesów), którym zgodnie z pewną strategią przydzielany jest procesor.

Systemy jednoprogramowe, zwane też **jednozadaniowymi**, umożliwiają uruchomienie jednego zadania użytkownika, które ewentualnie może być wykonywane współbieżnie z pewnymi procedurami systemu operacyjnego.

Systemy wieloprogramowe (**wielozadaniowe**) dostarczają mechanizm przełączania kontekstu, umożliwiając w ten sposób zachowanie stanu wykonywania określonego programu (stanu procesu), a następnie odtworzenie stanu wykonywania innego programu (w szczególności innego wykonywania tego samego programu). Przełączenie kontekstu jest skutkiem zwolnienia procesora, które z kolei następuje w wyniku:

- żądania przydziału dodatkowego zasobu,
- zainicjowania operacji wejścia-wyjścia,
- przekroczenia ustalonego limitu czasu (kwantu czasu),
- uzyskania gotowości przez inne zadanie (proces) o wyższym priorytecie.

Klasyfikacja systemów operacyjnych (liczba użytkowników)

- Systemy dla jednego użytkownika – zasoby przeznaczone są dla jednego użytkownika, nie ma mechanizmów autoryzacji, a mechanizmy ochrony informacji są ograniczone.
- Systemy wielodostępne – wielu użytkowników może korzystać z zasobów systemu komputerowego, a system operacyjny gwarantuje ich ochronę przed nieupoważnioną ingerencją.

W systemach dla **jednego użytkownika** nie ma problemu autoryzacji, czyli konieczności identyfikowania zleceniodawcy poszczególnych zadań. Mechanizmy ochrony są ograniczone w tym sensie, że nie ma potrzeby ochrony zasobów jednego użytkownika przed drugim użytkownikiem tego samego systemu operacyjnego, ale w czasie powszechności sieci rozległych istnieje jednak problem ochrony zasobów przed ingerencją z zewnątrz.

System operacyjny w przypadku **wielodostępu** musi zagwarantować, że jeden użytkownik nie jest w stanie zakłócić pracy innych użytkowników. Jest to problem właściwego udostępniania zasobów oraz dostępności mechanizmów ochrony prywatnych zasobów jednego użytkownika przed ingerencją innego.

Inne rodzaje systemów operacyjnych

- Systemy czasu rzeczywistego (*real-time systems*) – zorientowane na przetwarzanie z uwzględnieniem czasu zakończenie zadania, tzw. linii krytycznej (*deadline*).
- Systemy sieciowe i rozproszone (*network and distributed systems*) – umożliwiają zarządzanie zbiorem rozproszonych jednostek przetwarzających, czyli zbiorem jednostek (komputerów), które są zintegrowane siecią komputerową i nie współdzielą fizycznie zasobów.
- Systemy operacyjne komputerów naręcznych – tworzone dla rozwiązań typu PDA, czy telefonów komórkowych, podlegają istotnym ograniczeniom zasobowym.

W **systemach czasu rzeczywistego** priorytetem jest minimalizacja czasu odpowiedzi (reakcji) lub czasu realizacji zadania, gdyż po przekroczeniu pewnego czasu wartość wyników albo jest znacznie mniejsza (przewidywanie kursów akcji na giełdzie) albo są one całkowicie bezużyteczne (prognozowanie pogody).

Szczególnym przypadkiem, gdzie czas jest krytyczny, są wszelkiego rodzaju systemy sterowania w czasie rzeczywistym (w komputerach pokładowych samochodów lub samolotów, urządzeniach medycznych). Systemy operacyjne czasu rzeczywistego są więc budowane pod kątem szybkości reakcji na zdarzenie zewnętrzne. Ich zadaniem jest minimalizować czas oczekiwania na zasoby dla czasowo krytycznych zadań, dlatego unika się w ich przypadku rozwiązań, które zmniejszają przewidywalność tego czasu (np. pamięci wirtualnej).

Rozproszone systemy operacyjne zapewniają, że system komputerowy, złożony z autonomicznych jednostek przetwarzających połączonych siecią komputerową, postrzegany jest jako całość. Zasoby tego systemu udostępniane są w jednolity sposób niezależnie od ich fizycznej lokalizacji – niezależnie od tego, czy są to zasoby lokalne danej jednostki, czy zasoby związane integralnie z jednostką zdalną.

Cecha ta odróżnia systemy rozproszone od systemów sieciowych, które również umożliwiają dostęp do zdalnych zasobów, ale nie ukrywają faktu fizycznego rozproszenia tych zasobów. Inaczej mówiąc, w systemie sieciowym odróżnia się dostęp lokalny i dostęp zdalny do zasobów.

Rozwiązania dla **systemów naręcznych** nie muszą tworzyć środowiska dla zaawansowanego przetwarzania wielozadaniowego, ale ze względu na niewielkie rozmiary urządzeń podlegają dość rygorystycznym ograniczeniom zasobowym.

W przypadku tego typu rozwiązań, jak również rozwiązań dla innych urządzeń mobilnych, dość istotnym zasobem, którym należy odpowiednio zarządzać jest energia.

Budowa i zasada działania systemu operacyjnego

System operacyjny jest **programem**, jednak jego działanie jest dość specyficzne, gdyż musi on **nadzorować** (monitorować) **pracę komputera** nawet wówczas, gdy wykonywany jest jakiś program aplikacyjny.

System operacyjny musi **reagować** na **błędy** w programach aplikacyjnych, **porządkować** system komputerowy po awariach, z kolei błędy w kodzie jądra systemu operacyjnego mogą zdestabilizować funkcjonowanie całego systemu komputerowego.

Działanie współczesnych systemów operacyjnych jest rezultatem ewolucji w architekturze sprzętowo-programowej, w której potrzeby w zakresie implementacji pewnych mechanizmów systemu operacyjnego wymuszały wprowadzanie stosownych rozwiązań na poziomie architektury komputera (procesora, jednostki zarządzania pamięcią, układu bezpośredniego dostępu do pamięci, układów wejścia-wyjścia).

Działania procesora, w zasadzie ogranicza się do **wykonywania rozkazów** programów (systemu operacyjnego też), które powtarzają się cyklicznie. W związku z czym określa się je jako cykle rozkazowe. Realizacja cyklu rozkazowego wymaga na ogół kilku interakcji procesora z pamięcią. Każdą taką interakcję określa się mianem cyklu maszynowego.

W każdym **cyklu rozkazowym** występuje cykl maszynowy pobrania kodu rozkazu. W zależności od trybu dostępności operandów mogą też wystąpić cykle pobrania operandu z pamięci (albo rejestrów wejścia-wyjścia) lub składowania operandu w pamięci (albo rejestrach wejścia-wyjścia). Każdy cykl maszynowy oznacza zatem zapis lub odczyt pamięci, przy czym cykl pobrania kodu rozkazu oznacza zawsze odczyt.

W czasie wykonywania rozkazu mogły nastąpić pewne **zdarzenia zewnętrzne** w stosunku do procesora, nie związane z bieżącym cyklem rozkazowym, ale wymagające od procesora jakiejś reakcji. Konieczność reakcji zgłaszana jest poprzez sygnał na odpowiedniej linii wejściowej procesora. Ostatnia faza cyklu rozkazowego polega zatem na sprawdzeniu, czy wystąpiło takie zgłoszenie.

Jeśli nie było zgłoszenia, rozpoczyna się następny cykl maszynowy. Jeśli jednak było zgłoszenie – nazywane **przerwaniem**, następuje ciąg działań, zmierzających do zidentyfikowania źródła przerwania, a następnie przekazania sterowania do stosownej **procedury obsługi**.

Procedury obsługi przerwań są częścią programu jądra systemu operacyjnego.

Przerwania można podzielić na:

- **zewnętrzne** – pochodzące z układów na zewnątrz procesora, czyli od urządzeń wejścia-wyjścia, czasomierzy, układu bezpośredniego dostępu do pamięci itp.,
- **wewnętrzne** (diagnostyczne) – zgłaszane przez procesor, będące następstwem wykrycia jakiegoś stanu wyjątkowego,
- **programowe** – wynikające z wykonania specjalnej instrukcji procesora, umożliwiające programom użytkownika dostęp do wybranych funkcji jądra systemu operacyjnego.

Przerwania od urządzeń zewnętrznych zgłaszane są po zakończeniu operacji wejścia-wyjścia i przekazywane na specjalne wejście procesora najczęściej przez sterownik przerwań. Tą samą ścieżką zgłaszane są również przerwania od układów ściśle współpracujących z procesorem – czasomierzy, układów bezpośredniego dostępu do pamięci itp.

Są to typowe przerwania, gdyż ich źródło jest poza procesorem i jest od niego niezależne.

W przeciwieństwie do przerwań zewnętrznych, przerwania programowe są wynikiem wykonania specjalnej instrukcji procesora.

Przerwania diagnostyczne są z kolei generowane wewnętrznie przez procesor w sytuacji zajścia określonego stanu. Są zatem pośrednim skutkiem wykonania określonego ciągu rozkazów prowadzących do osiągnięcia tego stanu. Tego typu przerwania można traktować jak pułapki lub wyjątki.

System przerwań umożliwia **niesekwencyjne** (współbieżne) wykonywanie programów. Zmiana sekwencji wykonywania instrukcji polega na tym, że w reakcji na przerwanie następuje **zapamiętanie bieżącego stanu** przetwarzania (najważniejszych rejestrów procesora), **przekazanie sterowania** do ustalonej procedury obsługi i rozpoczęcie wykonywania instrukcji tej procedury.

W szczególności może to prowadzić do **przełączenia kontekstu**, czyli przekazania sterowania po zakończeniu procedury obsługi przerwania do innego przetwarzania, niż to które zostało przerwane.

Sposób **przechowywania informacji** jest specyficzny dla każdego rodzaju pamięci (dysk CD, dysk twardy), dlatego też systemy operacyjne oferują **specjalne struktury** danych nazywane **plikami**.

Pliki pozwalają uzyskać **dostęp** do **informacji** w ten sam sposób, niezależnie do tego gdzie ta informacja jest zapisana. Aby ułatwić zarządzanie plikami systemy operacyjne utrzymują informacje o nich w katalogach, które z kolei są organizowane w struktury katalogów.

Struktury katalogów wraz z plikami tworzą **system plików**. System plików jest najbardziej dostrzegalna dla użytkownika częścią systemu operacyjnego. Często systemy operacyjne obsługują więcej niż jeden system plików.

Definicja pliku, typy plików

Plik jest abstrakcyjnym typem danych dostarczanym przez system operacyjny w celu umożliwienia spójnej metody dostępu do informacji umiejscowionych w pamięciach różnego typu. Pliki posiadają szereg cech, z których najważniejsza, dla użytkownika jest jednoznacznie je identyfikująca **nazwa**.

Format pliku może być **swobodny** lub **ściśle określony**. Do plików o swobodnym formacie należą pliki **tekstowe**, w których dane są przechowywane w postaci ciągów znaków zakończonych znakiem (-kami) końca wiersza. Pliki o określonym formacie zawierają najczęściej **dane numeryczne binarne** lub **rekordy danych**.

Ogólnie, pliki mogą przechowywać programy, zarówno w postaci źródłowej, jak i wynikowej oraz dane. Informacje w plikach umieszczane są **porcjami**, które nazywamy rekordami **logicznymi**. Rozmiar takiego rekordu jest zmienny, w szczególności może on wynosić jeden bajt. Ponieważ rekordy logiczne są rozmieszczane w **blokach alokacji**, a rzadko się zdarza aby rozmiar rekordu logicznego był wielokrotnością rozmiaru bloku alokacji, jest to przyczyna powstawania fragmentacji.

Typ pliku jest cechą pliku, która określa jego **strukturę**. Systemy operacyjne mogą, ale nie muszą interpretować typy plików.

Przykładem systemów dysponujących taką cechą są systemy rodziny Windows, w których użytkownik może skojarzyć określoną akcję z konkretnym typem pliku.

Inne podejście może zakładać, że system operacyjny w ogóle nie będzie interpretował typów plików. Takie podejście zastosowano w systemie UNIX. Zawartość plików w tym systemie jest traktowana jako ciąg bajtów, a ich interpretacja jest pozostawiana aplikacjom użytkownika.

Operacje na plikach

Ponieważ plik jest abstrakcyjną strukturą danych, to system operacyjny musi dostarczyć definicji **operacji**, które mogą być na niej wykonywane. Istnieje pięć takich operacji, które są uznawane za podstawowe: **tworzenie pliku**, **zapis pliku**, **odczyt pliku**, **ustawienie wskaźnika pliku**, **usuniecie pliku**. Wszystkie te operacje wiążą się z przeszukaniem katalogu i znalezieniem pozycji odpowiadającej plikowi, na którym operacja ma zostać przeprowadzona (w przypadku tworzenia pliku musi to być pozycja pusta).

Aby wyeliminować przeszukiwanie katalogu z tych operacji wprowadza się dodatkową operacje nazywaną **otwieraniem pliku**. Jeśli plik jest otwierany, to system operacyjny tworzy w pamięci operacyjnej kopie jego wpisu w katalogu. Operacją uzupełniającą do otwierania jest **zamykanie pliku**, które powoduje zapisanie wpisu z pamięci operacyjnej na dysk, jeśli te wersje się różnią i usunięcie wpisu z pamięci. Systemy wielozadaniowe dostarczają także operacji **blokowania** (*lock*) części lub całości plików, celem ochrony ich przed współbieżną **modyfikacją**.

Semantyka spójności

Semantyka spójności nazywana również semantyką spoistości lub semantyką integralności określa **reguły dostępu** do plików **współdzielonych** w systemach wielozadaniowych, dzięki którym zmiany w takich plikach mogą być dokonywane w sposób bezpieczny.

Semantyka ta określa również zakres widoczności zmian dokonywanych w pliku przez jego pozostałych użytkowników.

W semantyce spójności systemów UNIX/Linux, zasób jakim jest plik ma tylko jeden obraz, widoczny dla każdego z użytkowników. Może to powodować opóźnienia w realizacji operacji wykonywanych na pliku, ze względu na konieczność modyfikacji na zasadach wyłączności.

Semantyka systemu UNIX/Linux definiuje reguły:

- jeśli kilku użytkowników ma otwarty plik współdzielony i jeden z nich dokona modyfikacji jego zawartości, to jej skutki nie są widoczne natychmiast dla pozostałych,
- zmiany w pliku są widoczne tylko dla tych użytkowników, którzy zamknęli plik i ponownie go otworzyli.

Te reguły wymagają, aby każdy z użytkowników pliku otrzymywał jego kopie, unikalny obraz tego pliku dostępny dla każdego z nich osobno.

Katalogi logiczne

W systemach plików są definiowane **katalogi logiczne**. Katalog logiczny jest plikiem, którego zawartość stanowią wykazy informacji o innych plikach, czyli tak zwane metadane plików. Do tych metadanych mogą zaliczać się:

- unikatowa nazwa pliku,
- informacja o typie pliku (np. trzyliterowe rozszerzenie nazwy),
- wskaźnik pliku,
- rozmiar pliku (w bajtach lub jednostkach alokacji),
- lokalizacja pliku, czyli np. numer pierwszego bloku alokacji przedzielonego plikowi,
- informacje o ochronie, czyli o trybie dostępu,
- licznik użycia, określający ilu użytkowników korzysta w danej chwili z pliku,
- identyfikator właściciela lub twórcy pliku,
- czasy utworzenia, ostatniej modyfikacji lub ostatniego dostępu do pliku.

Katalogi tworzą strukturę katalogów. Taka struktura może obejmować nie tylko jedno urządzenie fizyczne, ale również kilka takich urządzeń (np. struktura katalogów systemu UNIX/Linux). Dzięki temu system operacyjny ukrywa przed użytkownikiem faktyczną lokalizację plików.

Kontrola dostępu

Z każdym plikiem lub katalogiem można związać **wykaz dostępów** (*access control list*), czyli opis praw jakie przysługują poszczególnym użytkownikom do tego pliku (czy użytkownik może dokonywać zapisu, odczytu itd) po ich identyfikacji przez system operacyjny.

Jeśli użytkownik próbuje wykonać jakąś operację, to system operacyjny odnajduje jego identyfikator w wykazie dostępów i sprawdza jakie prawa mu przysługują. Jeśli w wykazie znajduje się operacja, którą użytkownik chce wykonać, to jest ona wykonywana, w przeciwnym razie system odmawia dostępu użytkownikowi. Wadą tego rozwiązania może okazać się wielkość takiego wykazu.

W **dostępie grupowym** z każdym plikiem lub katalogiem związanych jest dziewięć bitów. Każda trójka bitów opisuje prawa dostępu do pliku pozwalające na **odczyt**, **zapis** i **wykonanie** (rwx w notacji uniksowej). Również znaczenie każdej z tych trójek jest inne.

Pierwsza opisuje prawa **właściciela pliku** (*user*), druga prawa **grupy** użytkowników do której należy właściciel pliku (*group*), a trzecia określa prawa dostępu dla **pozostałych** użytkowników w systemie (*others*).

Grupy użytkowników są zdefiniowane w specjalnie chronionym pliku systemowym. Dzięki dostępowi grupowemu możliwe jest współdzielenie plików. Również ilość informacji związanych z ochroną plików, które trzeba zapamiętać jest mała w stosunku do wykazu dostępów.

Wolne i Otwarte Oprogramowanie (*Free Libre/Open Source Software, FOSS, FLOSS*) – neutralny skrót pozwalający objąć jednym mianem zarówno **Wolne Oprogramowanie** (*Free Software*) jak i **Otwarte Oprogramowanie** (*Open Source*).

Wolne Oprogramowanie – oprogramowanie, które może być uruchamiane, kopiowane, rozpowszechniane, analizowane oraz zmieniane i poprawiane przez użytkowników.

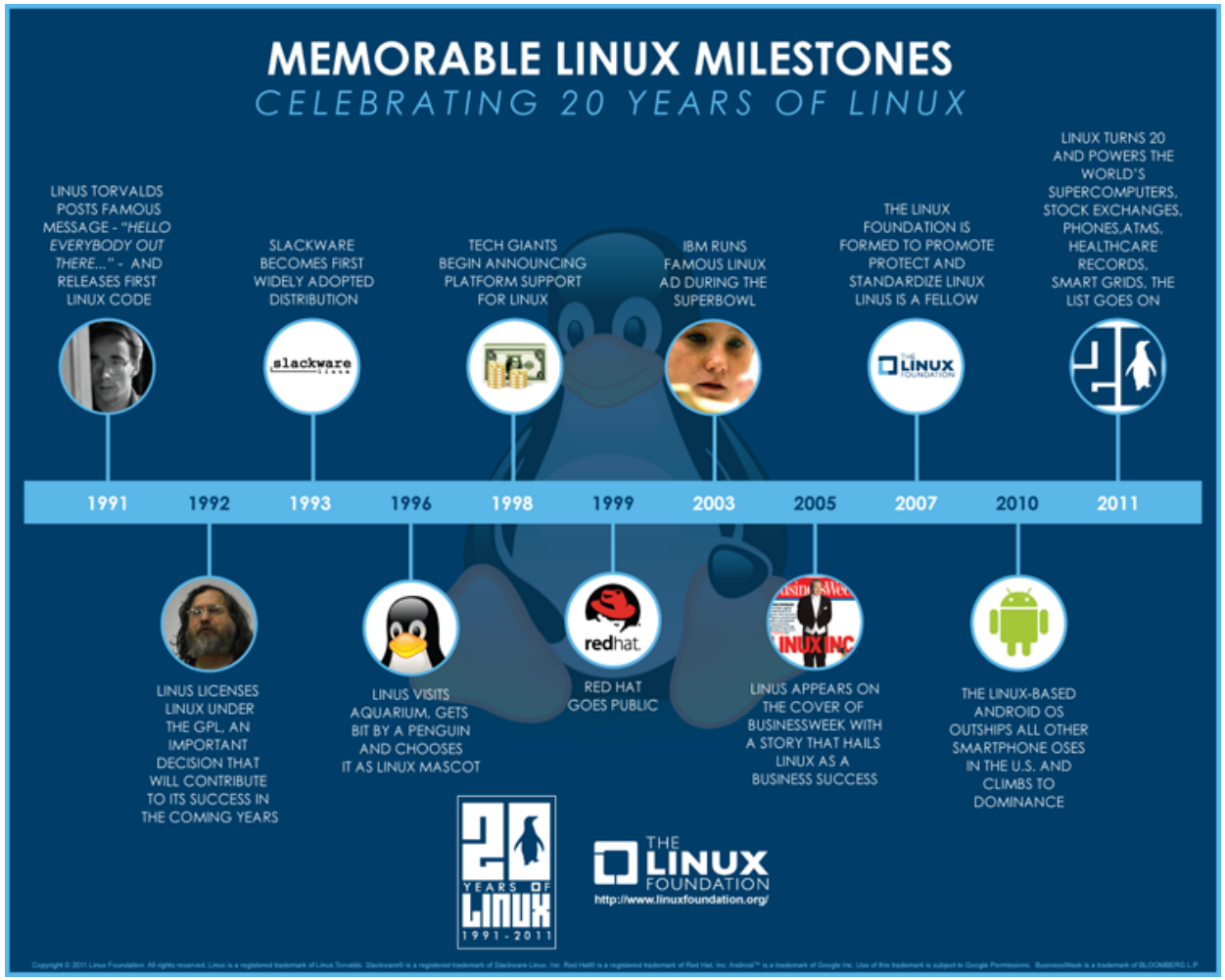
Definicja Wolnego Oprogramowania została opublikowanej przez **Free Software Foundation** i wynikają z niej następujące wolności:

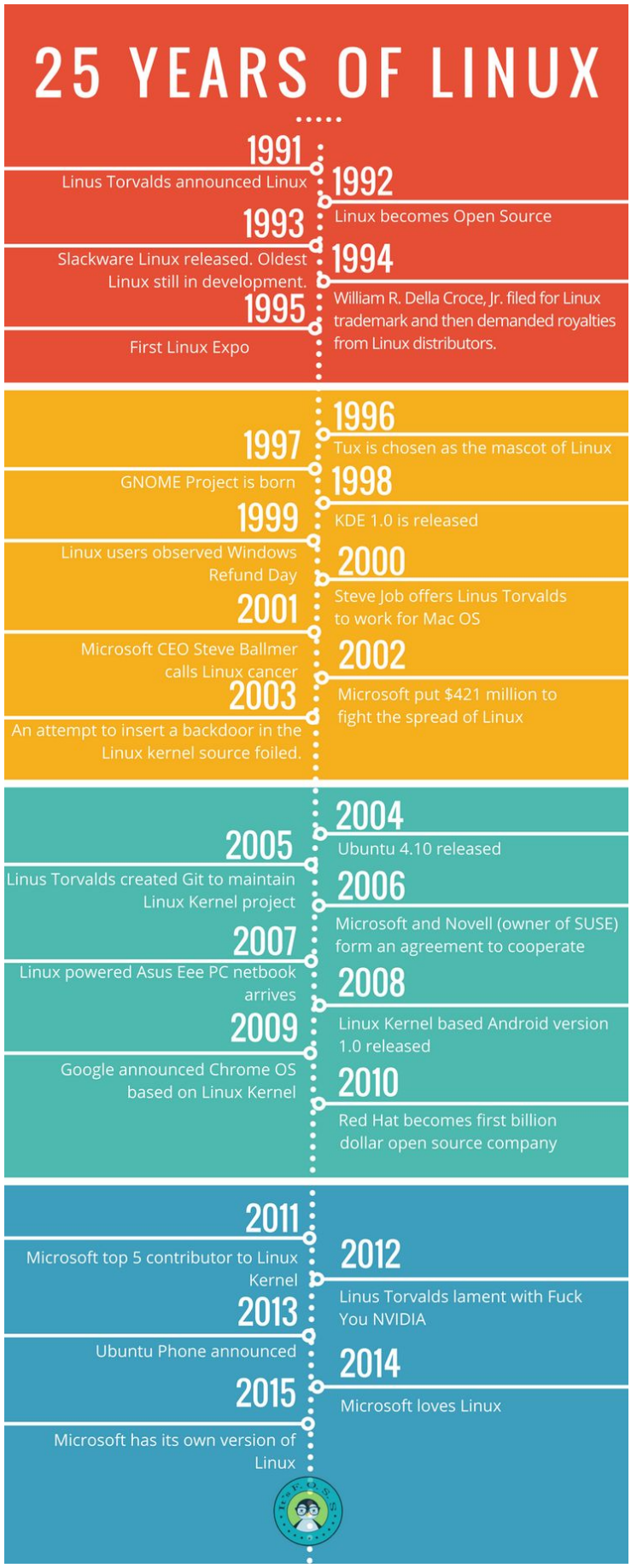
- wolność uruchamiania programu, w dowolnym celu,
- wolność analizowania programu oraz dostosowywania go do swoich potrzeb,
- wolność rozpowszechniania kopii programu,
- wolność udoskonalania programu i publicznego rozpowszechniania własnych ulepszeń.

Otwarte Oprogramowanie (*Open Source*) – odłam ruchu Wolnego Oprogramowania, alternatywny dla Free Software, głównie z przyczyn praktycznych.

W praktyce każdy program na licencji zgodnej z definicją Free Software Foundation jest jednocześnie zgodny z bardziej liberalną definicją **Open Source Movement**.

GNU/Linux – system operacyjny.

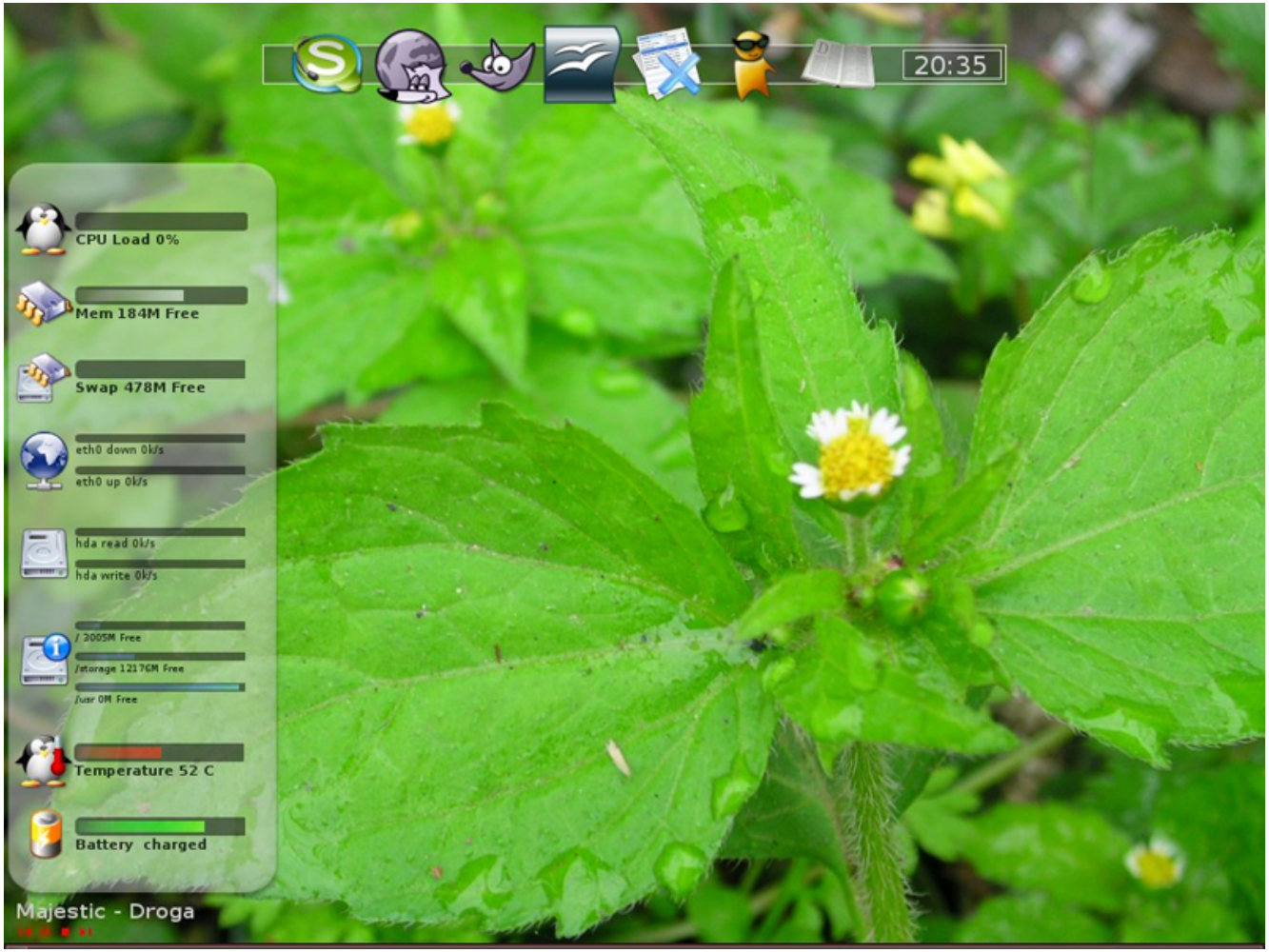



25 YEARS OF LINUX
1991
Linus Torvalds announced Linux
1992
Linux becomes Open Source
1993
Slackware Linux released. Oldest Linux still in development.
1994
William R. Della Croce, Jr. filed for Linux trademark and then demanded royalties from Linux distributors.
1995
First Linux Expo
1996
Tux is chosen as the mascot of Linux
1997
GNOME Project is born
1998
KDE 1.0 is released
1999
Linux users observed Windows Refund Day
2000
Steve Job offers Linus Torvalds to work for Mac OS
2001
Microsoft CEO Steve Ballmer calls Linux cancer
2002
Microsoft put $421 million to fight the spread of Linux
2003
An attempt to insert a backdoor in the Linux kernel source foiled.
2004
Ubuntu 4.10 released
2005
Linus Torvalds created Git to maintain Linux Kernel project
2006
Microsoft and Novell (owner of SUSE) form an agreement to cooperate
2007
Linux powered Asus Eee PC netbook arrives
2008
Linux Kernel based Android version 1.0 released
2009
Google announced Chrome OS based on Linux Kernel
2010
Red Hat becomes first billion dollar open source company
2011
Microsoft top 5 contributor to Linux Kernel
2012
Linus Torvalds lament with Fuck You NVIDIA
2013
Ubuntu Phone announced
2014
Microsoft loves Linux
2015
Microsoft has its own version of Linux

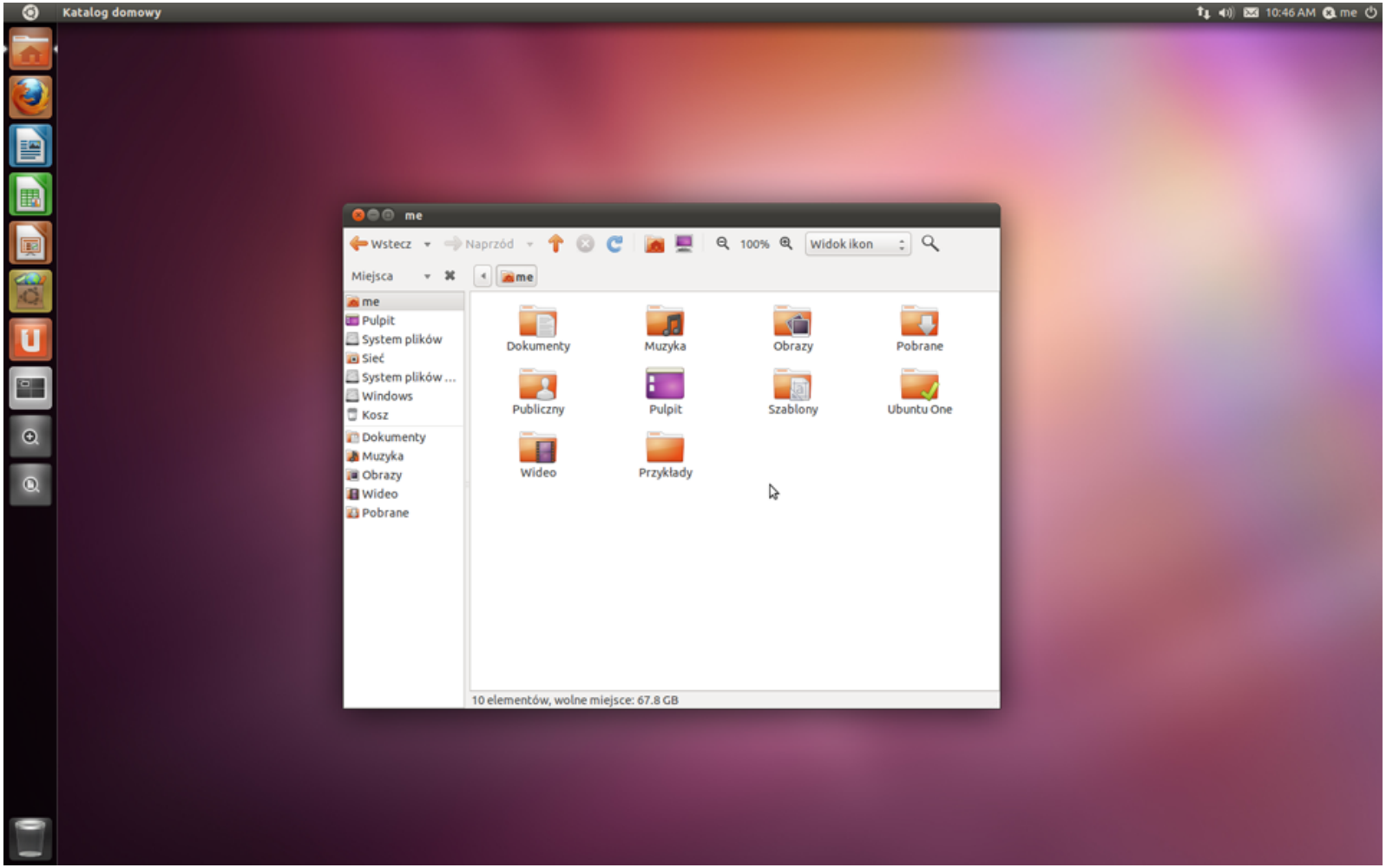
Krótka charakterystyka i ważniejsze cechy systemu GNU/Linux:

- efektywność i stabilność systemu,

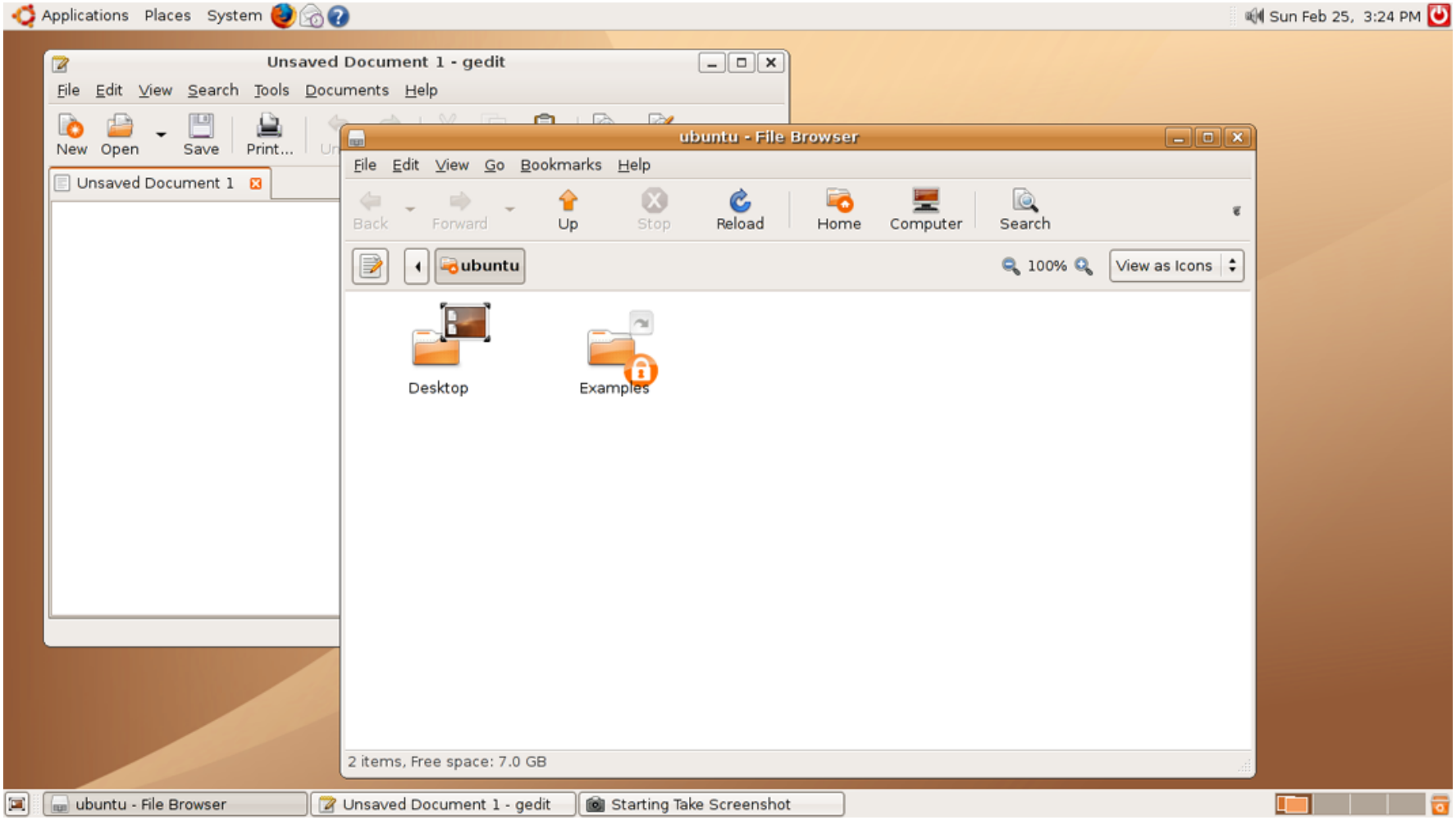
- powszechna dostępność bez jakichkolwiek opłat licencyjnych,
- bogaty zestaw oprogramowania umożliwiający szeroki zakres zastosowań,



- możliwość pracy na wielu platformach sprzętowych przy stosunkowo niewielkich wymaganiach,
- możliwość łatwej współpracy z innymi popularnymi systemami operacyjnymi,



- bogata dokumentacja w wersji elektronicznej,
- dostępność kodu źródłowego.

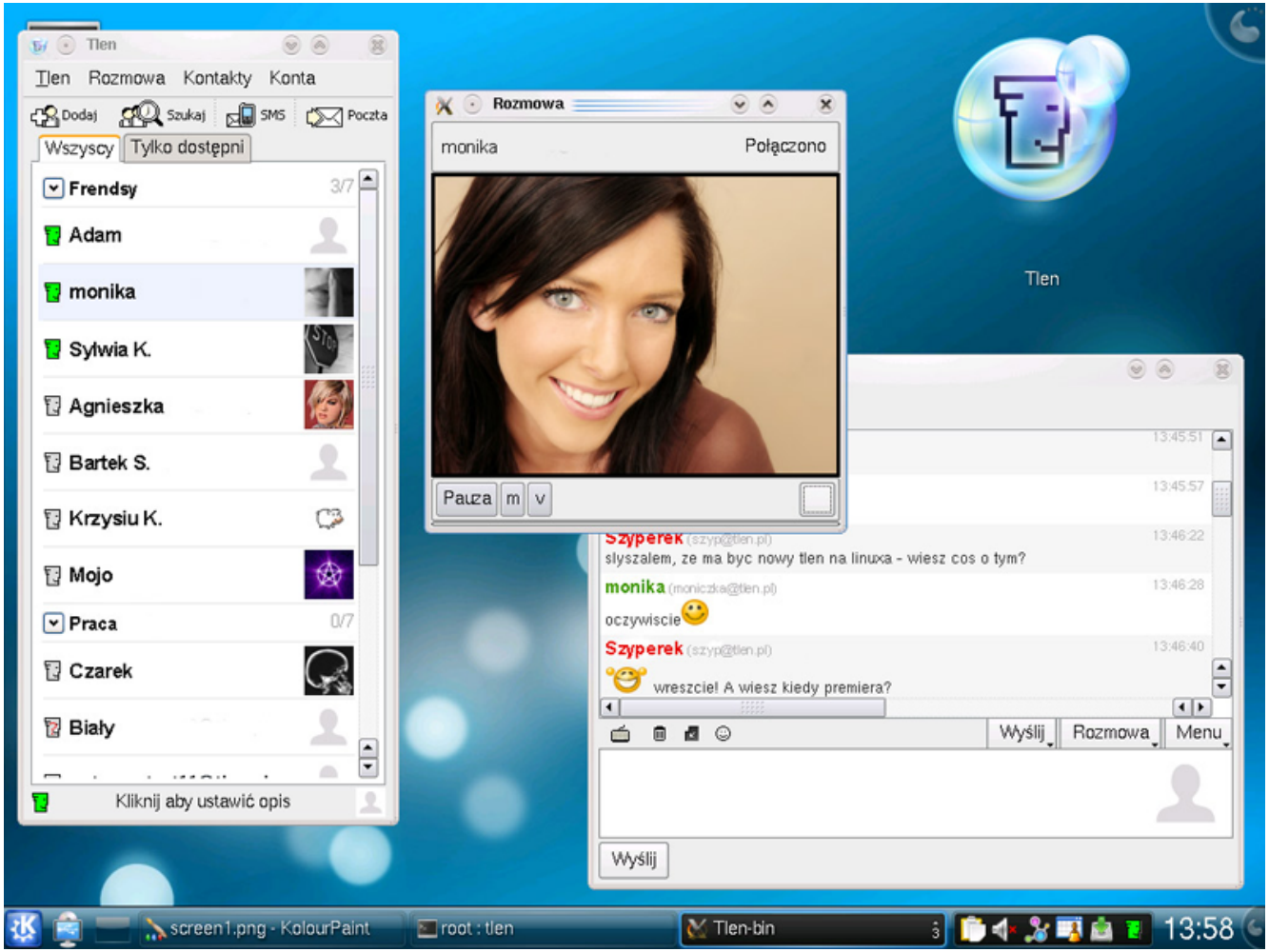


- wielodostęp,
- wielozadaniowość, czyli praca z podziałem czasu procesora pomiędzy wiele zadań,

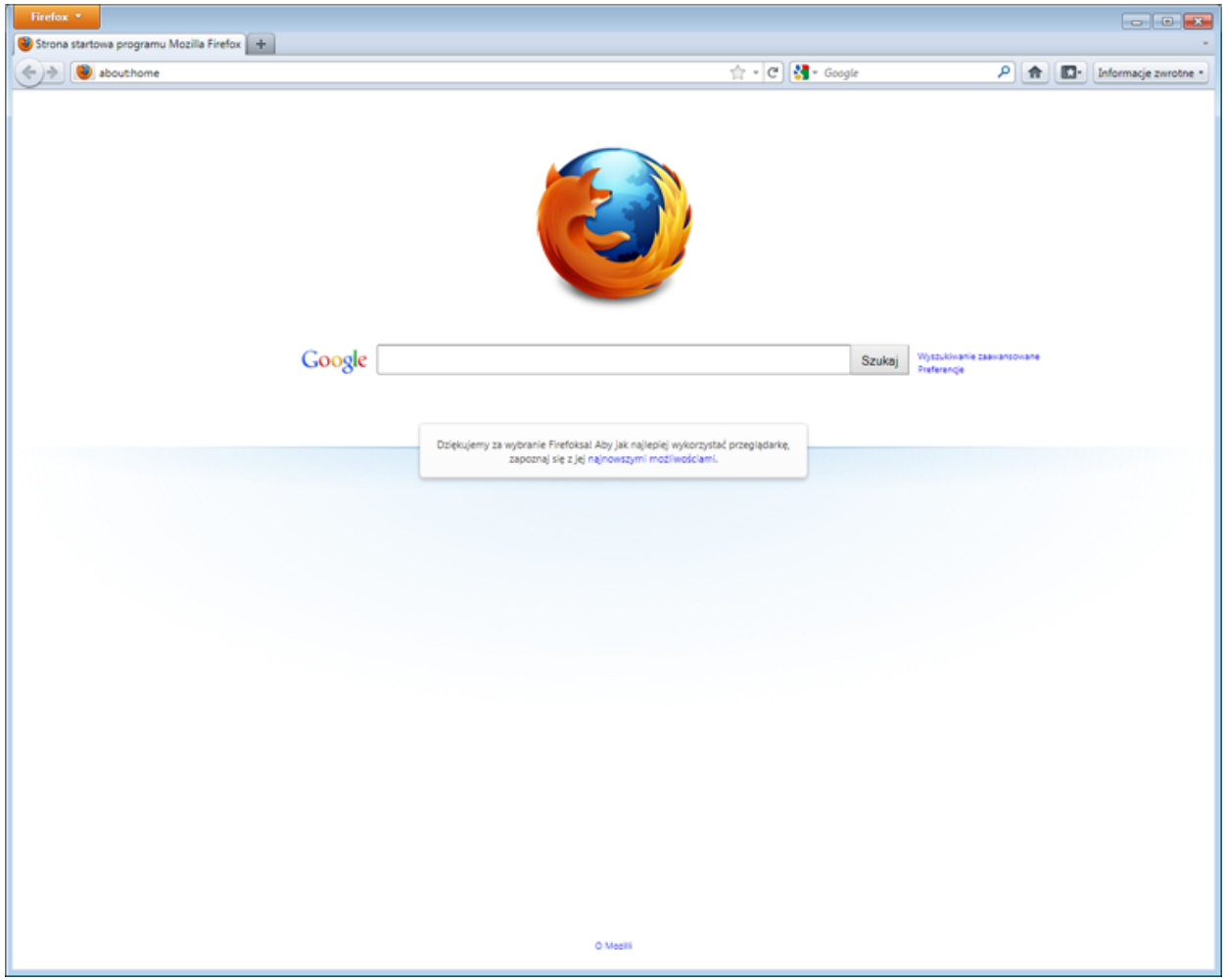
- wieloprzetwarzanie, czyli praca wieloprocesorowa,
- możliwość uruchamiania zadań w łagodnym czasie rzeczywistym,

- obsługa różnych typów systemów plików,
- obsługa różnych protokołów sieciowych,

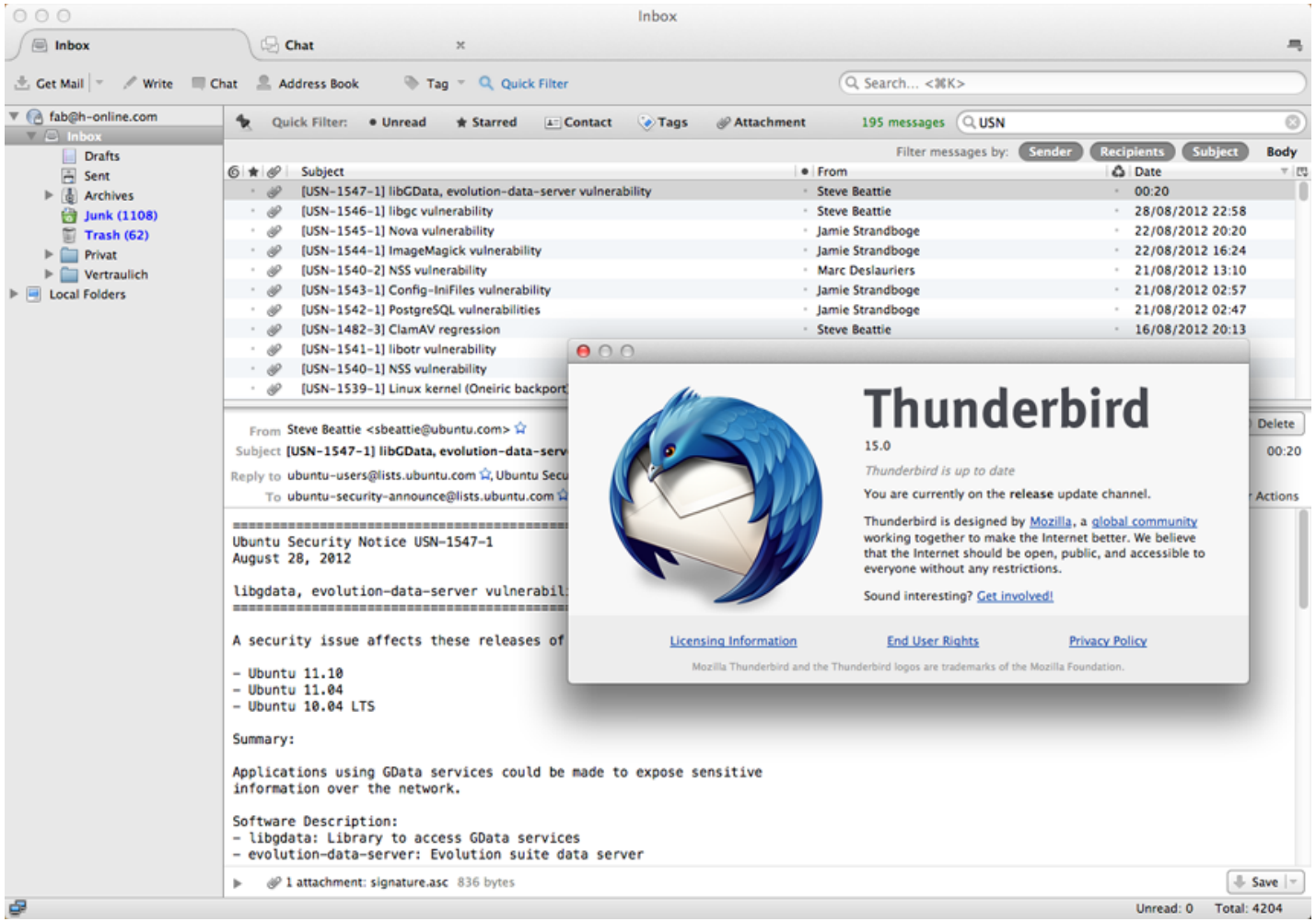
- obsługa różnych formatów plików wykonywalnych,
- wykorzystanie współdzielonych bibliotek.



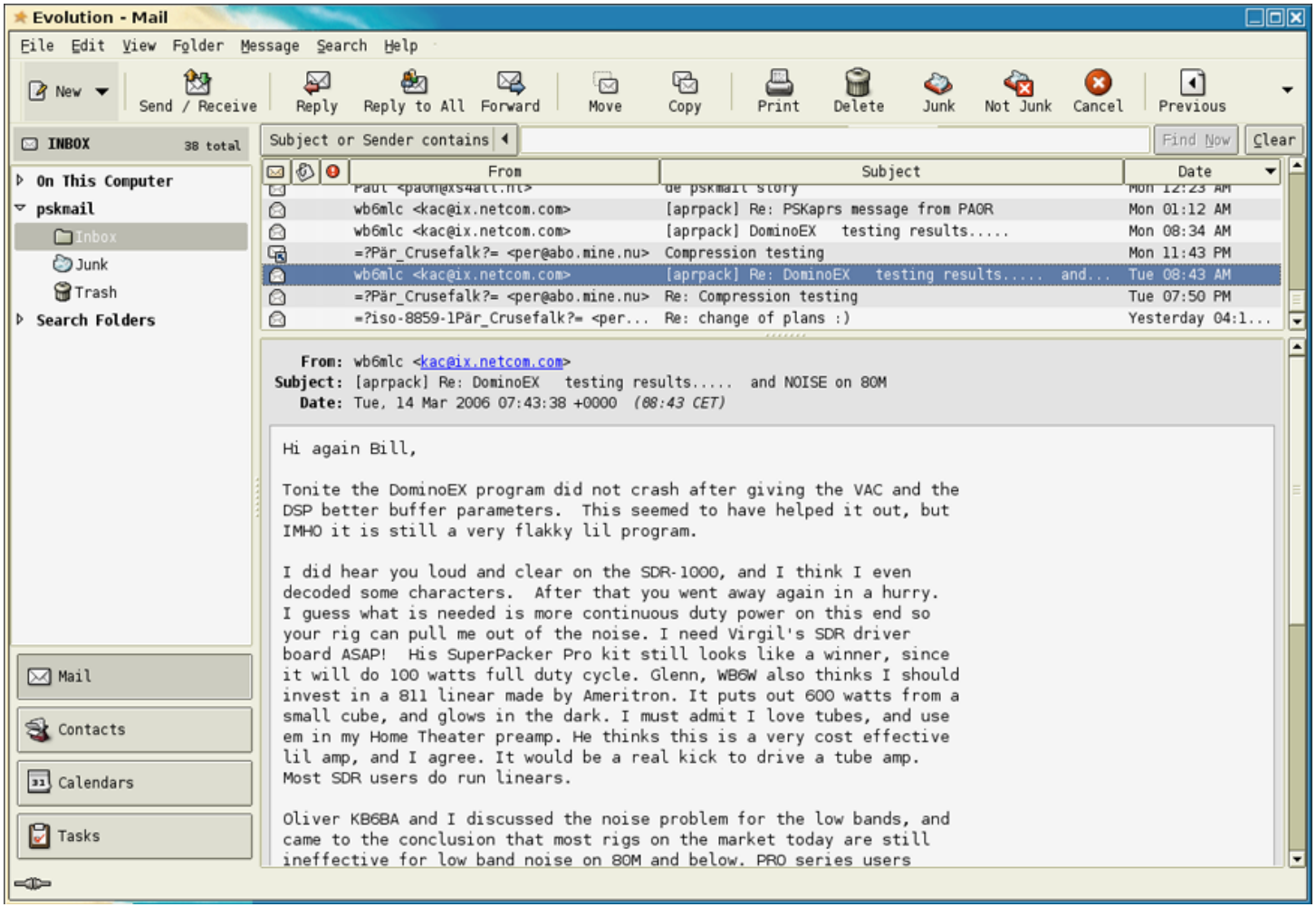
Firefox – przeglądarka internetowa.



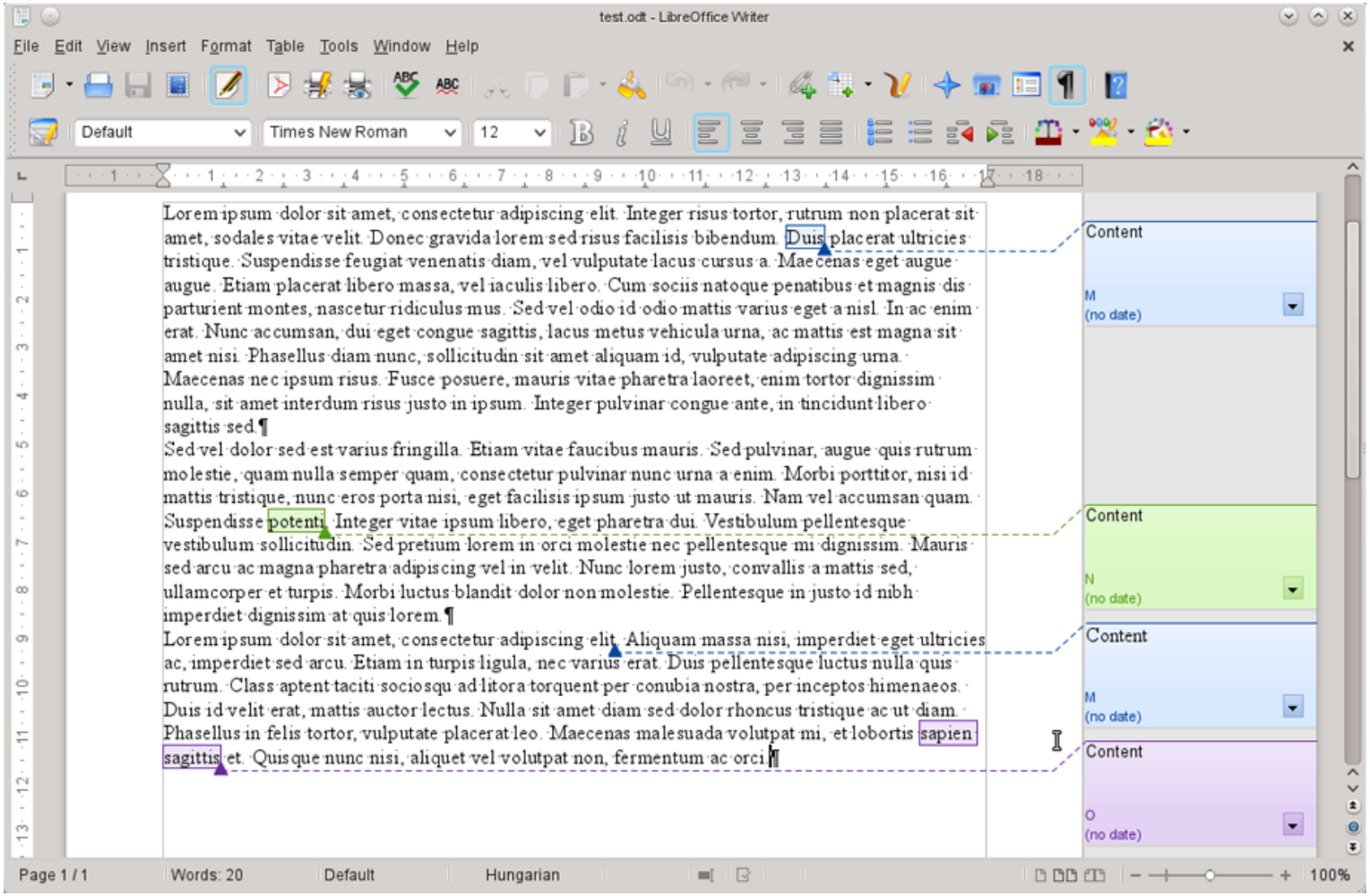
Thunderbird – klient poczty, czytnik grup dyskusyjnych i kanałów informacyjnych.



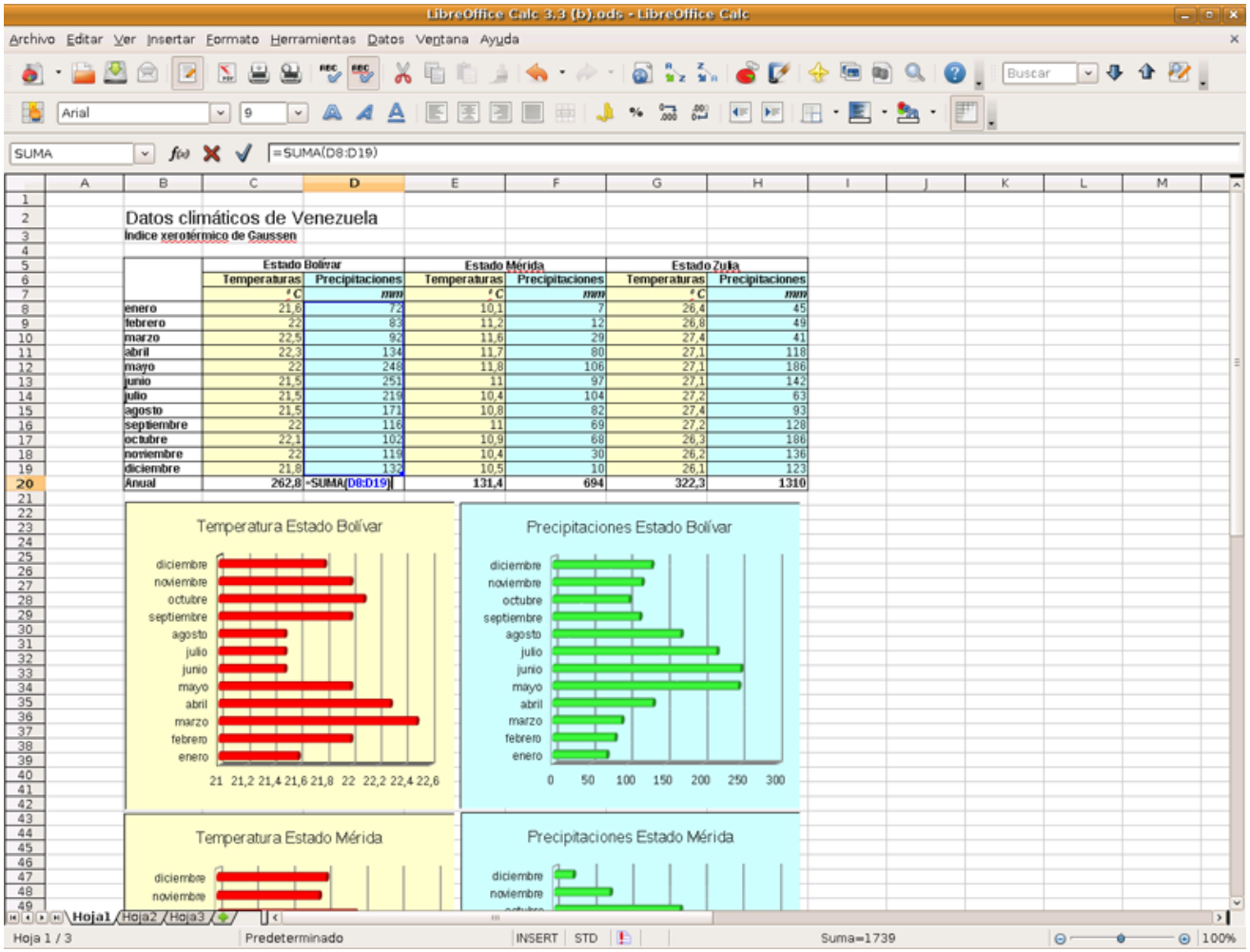
Evoliution – klient poczty, organizer.



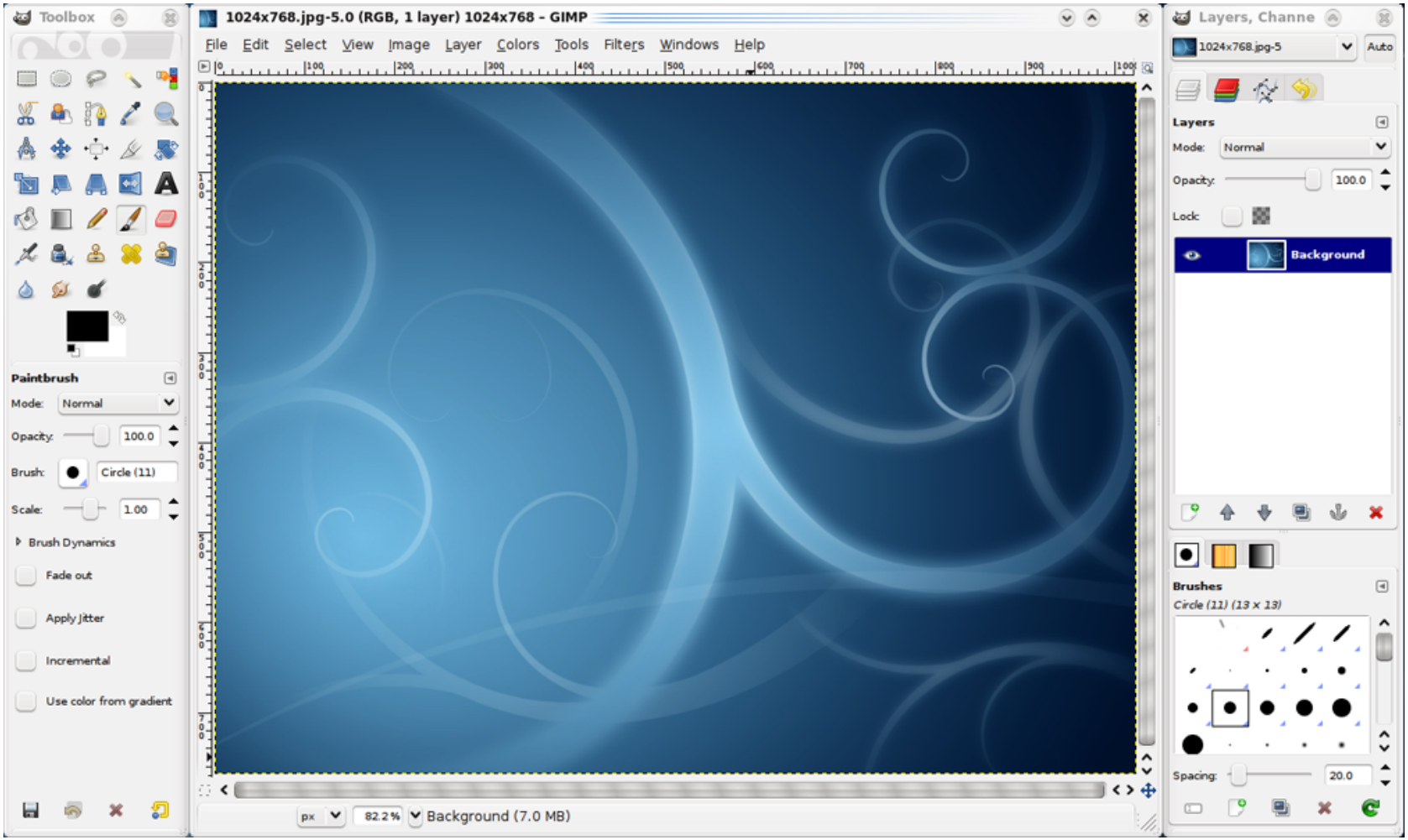
LibreOffice Writer – edytor tekstu.



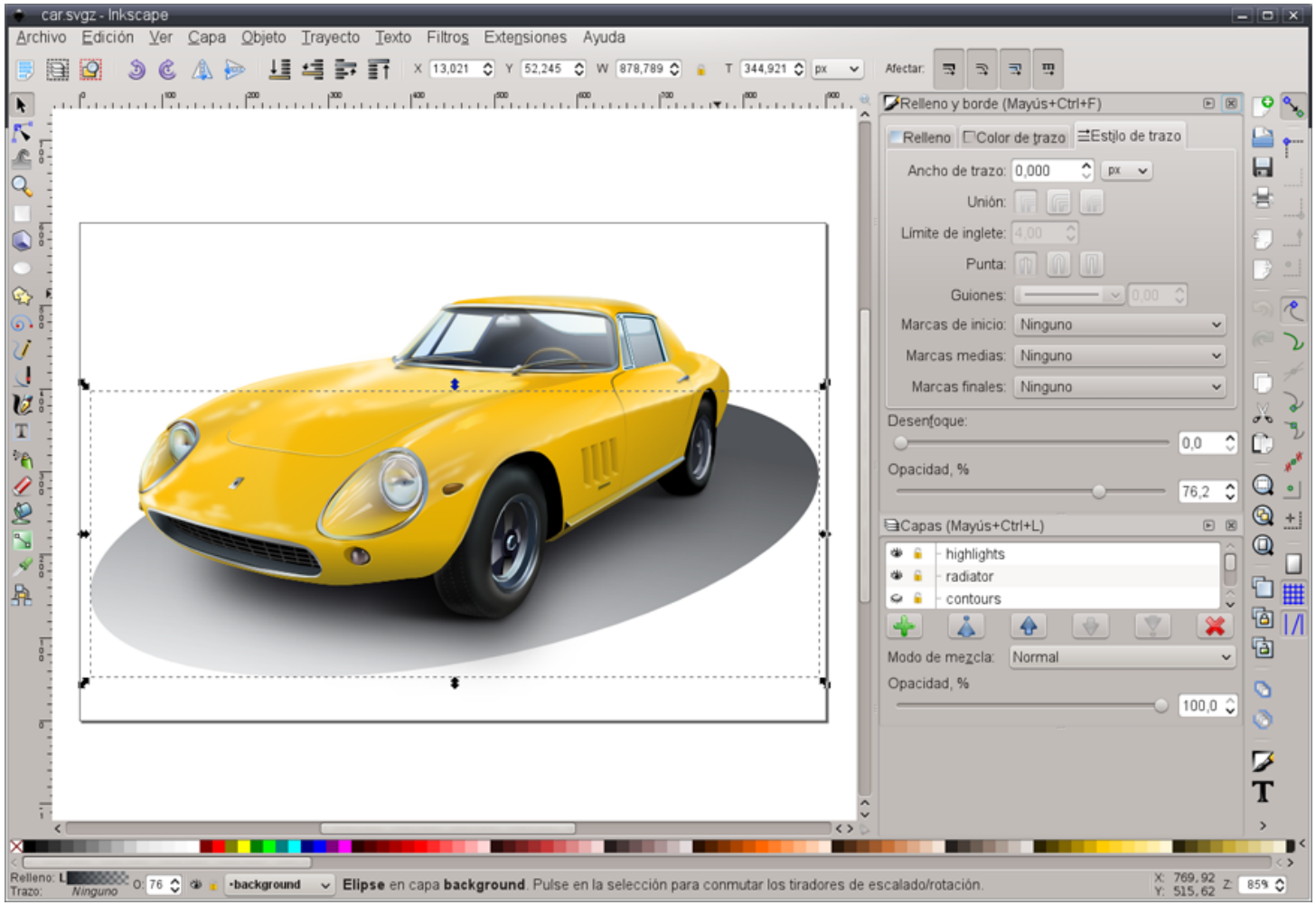
LibreOffice Calc – arkusz kalkulacyjny.



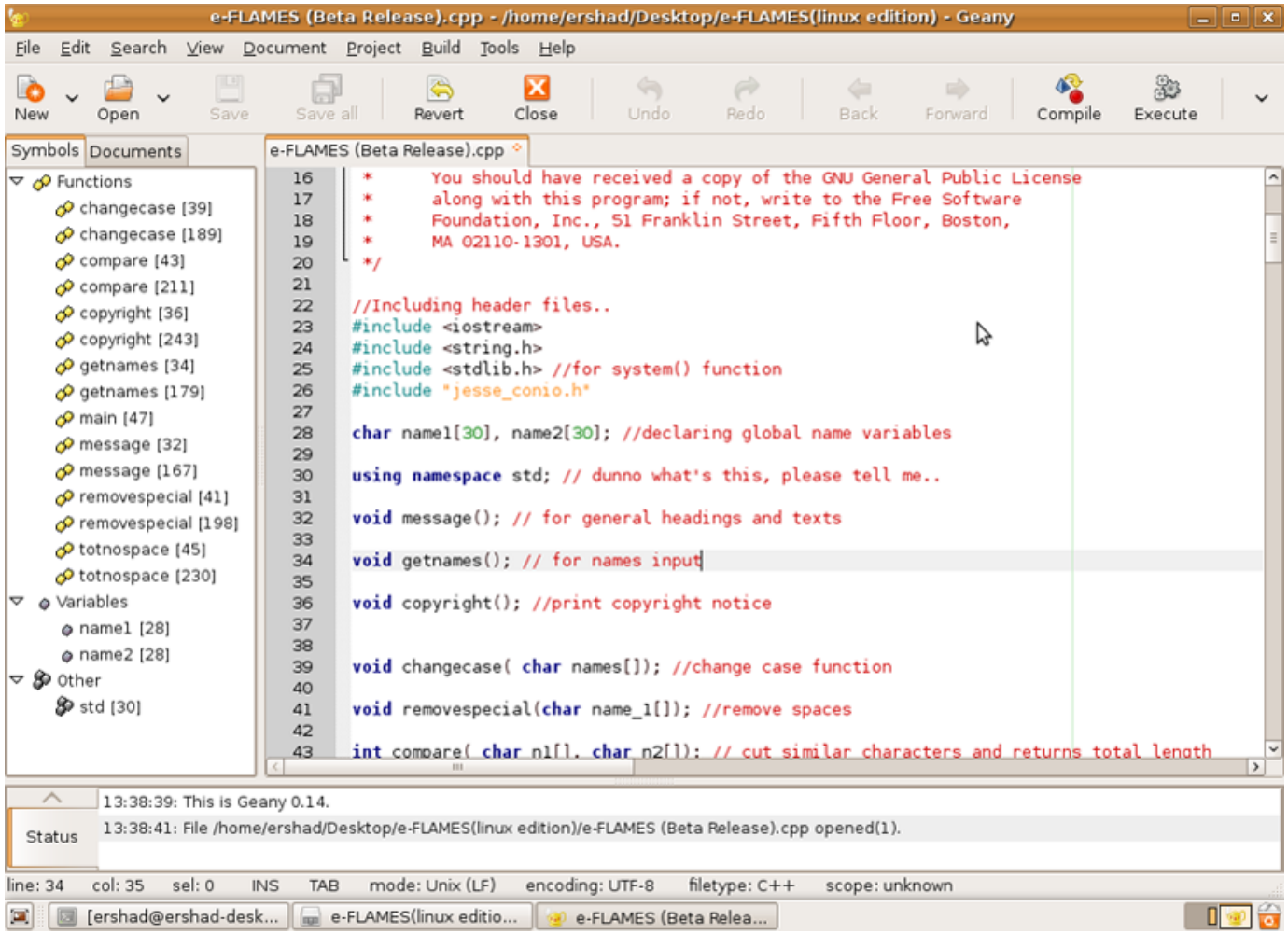
GIMP – edytor grafiki rastrowej.



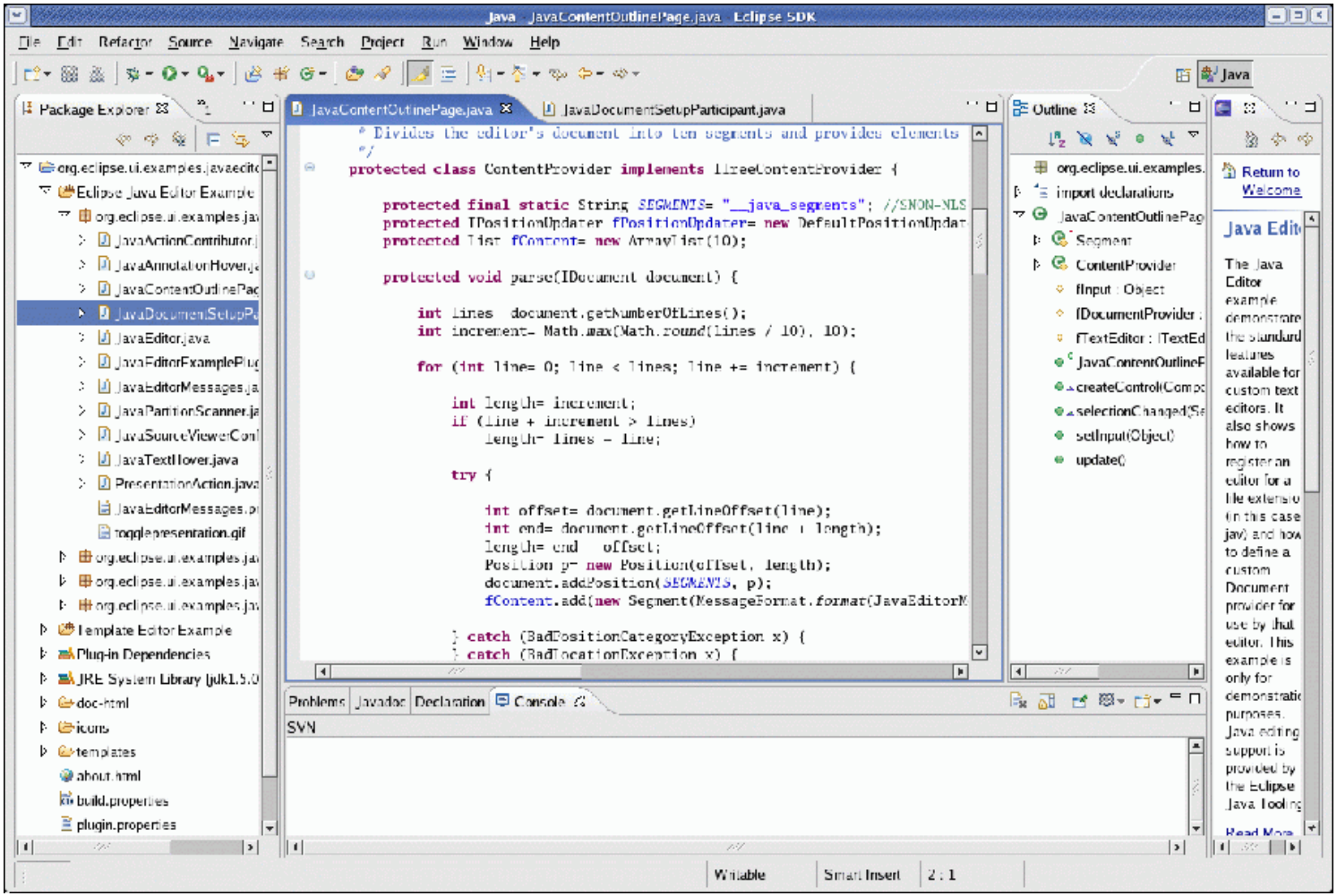
Inkscape – edytor grafiki wektorowej.



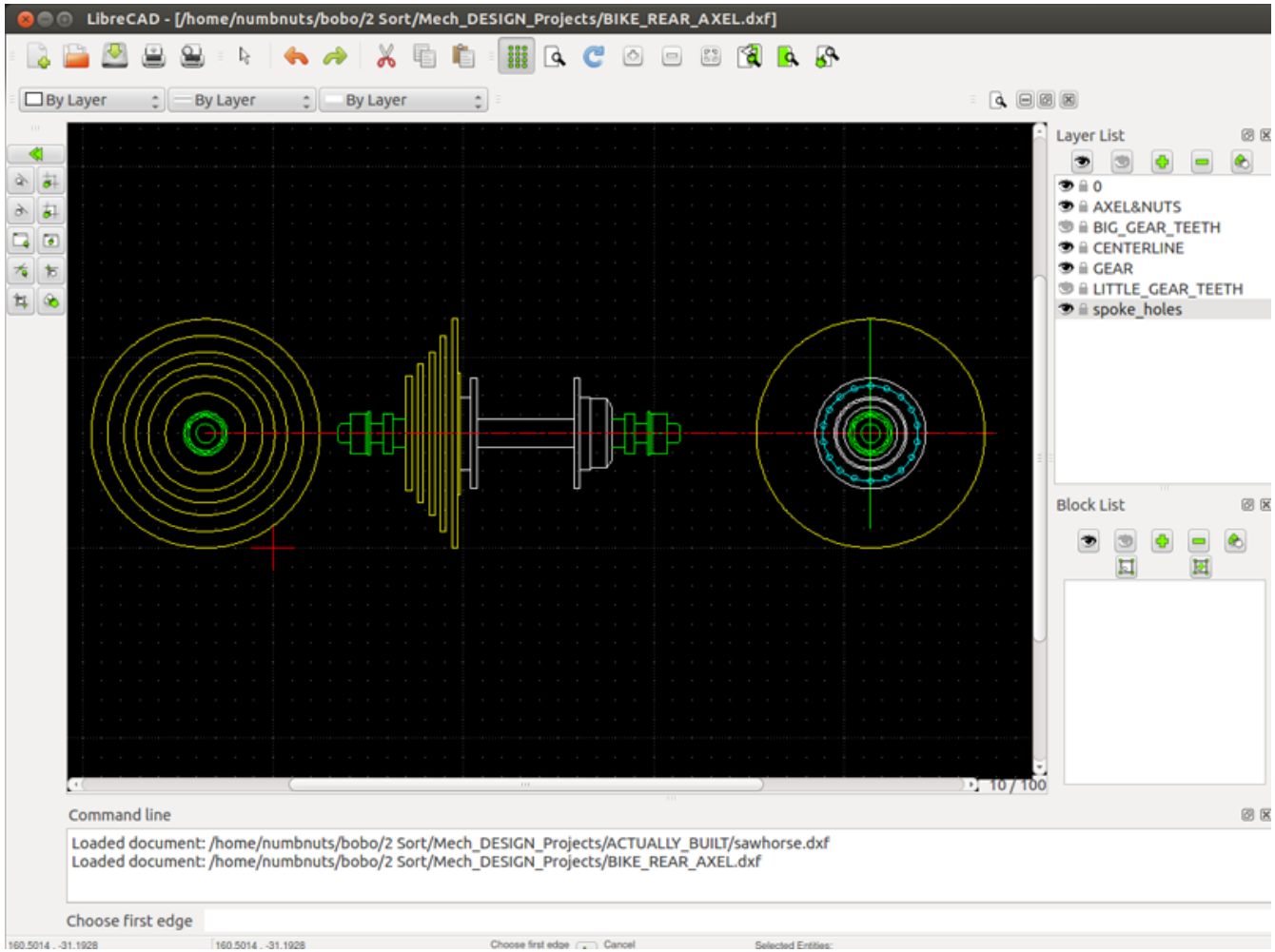
Geany – edytor tekstu programisty.



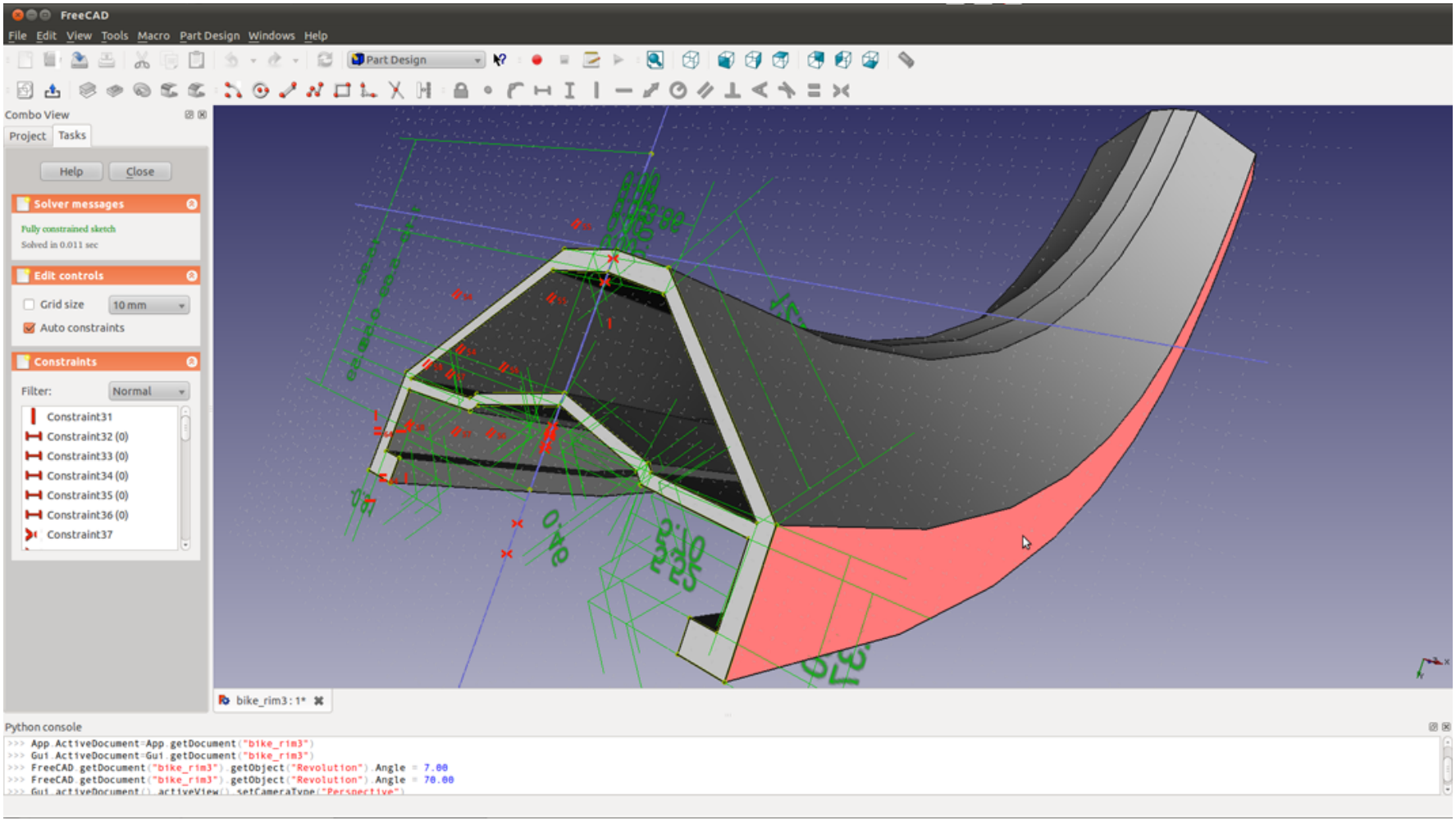
Eclipse – zintegrowane środowisko programistyczne.



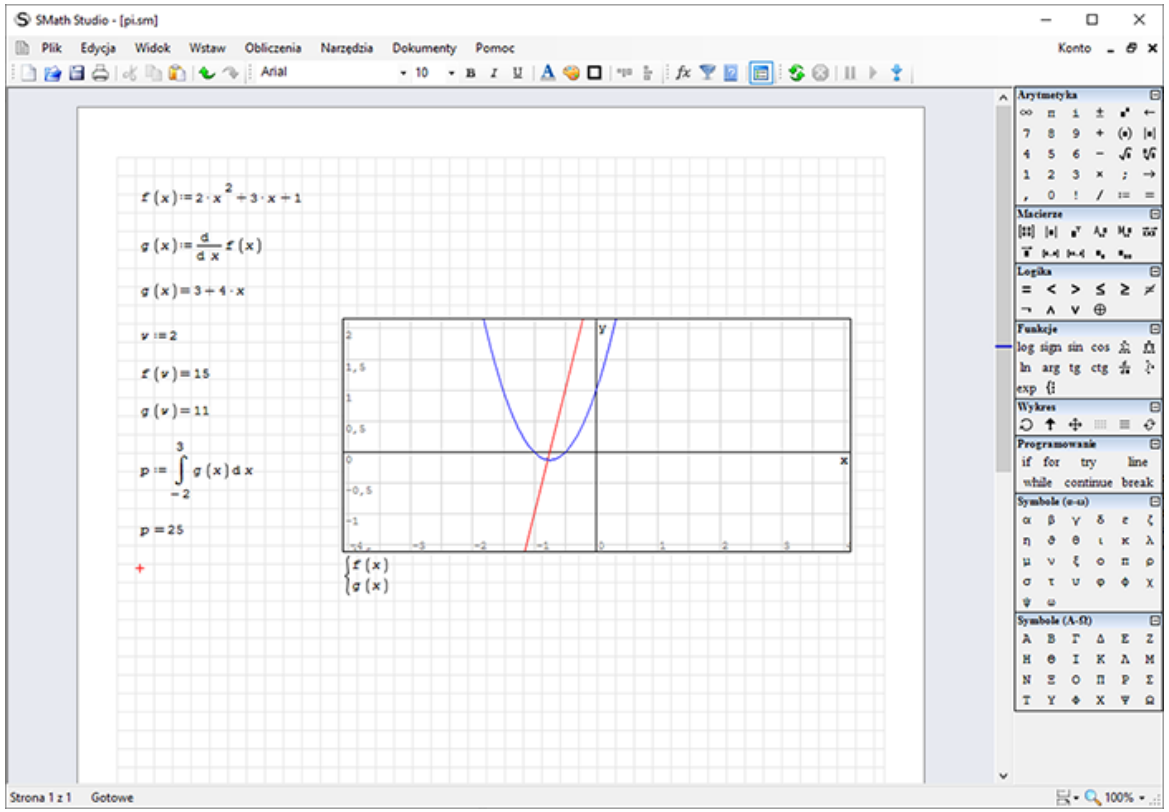
LibreCAD – aplikacja CAD 2D.



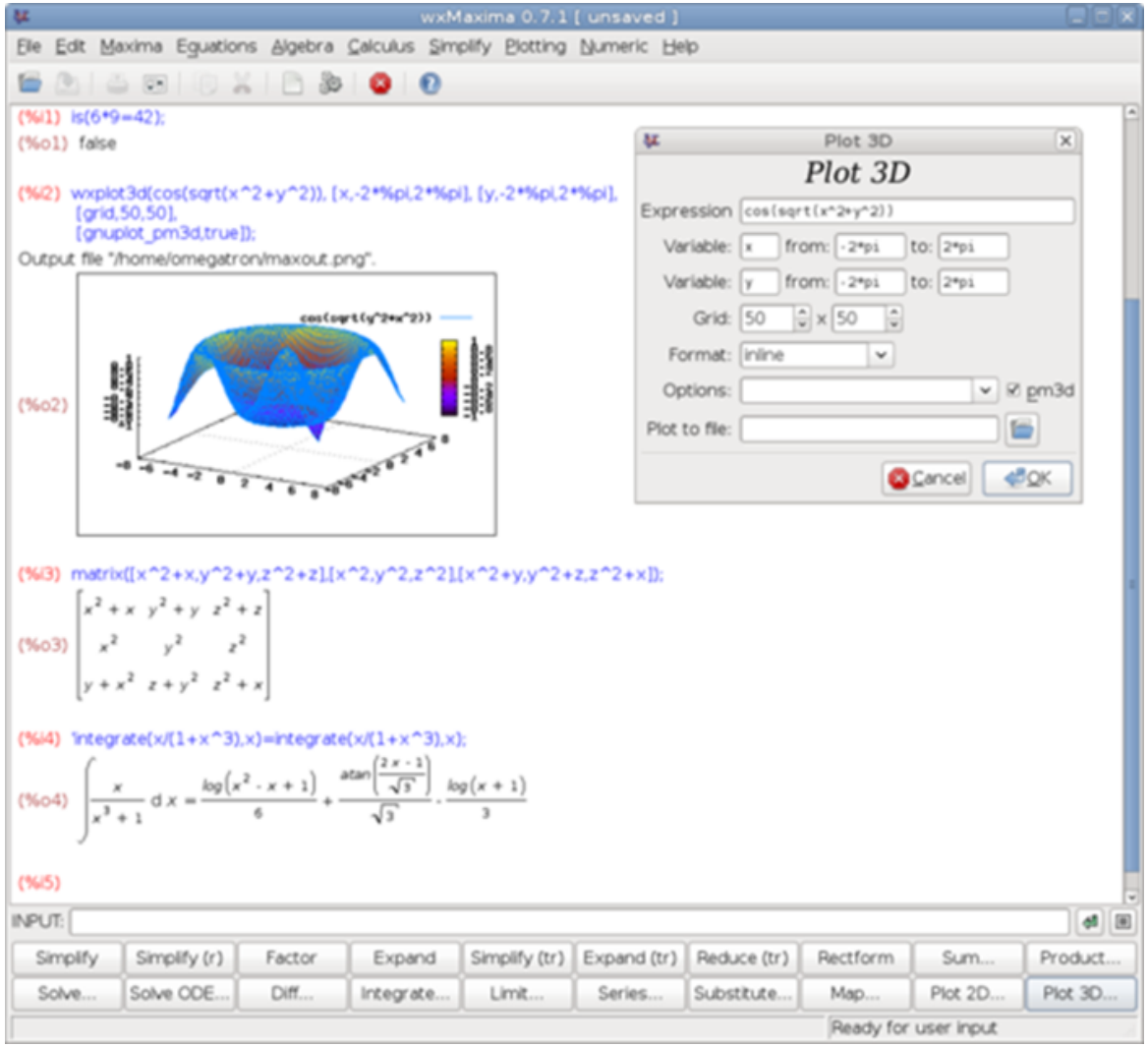
FreeCAD – aplikacja CAD 3D.



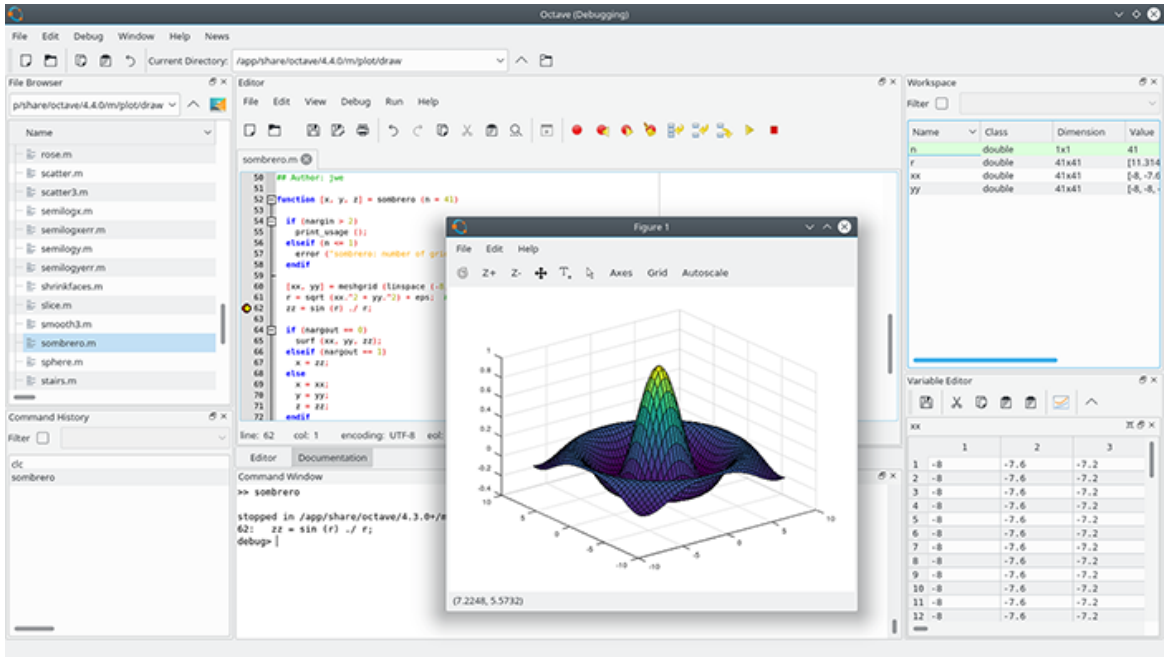
SMath Studio – aplikacja CAS (zamknięte źródło).



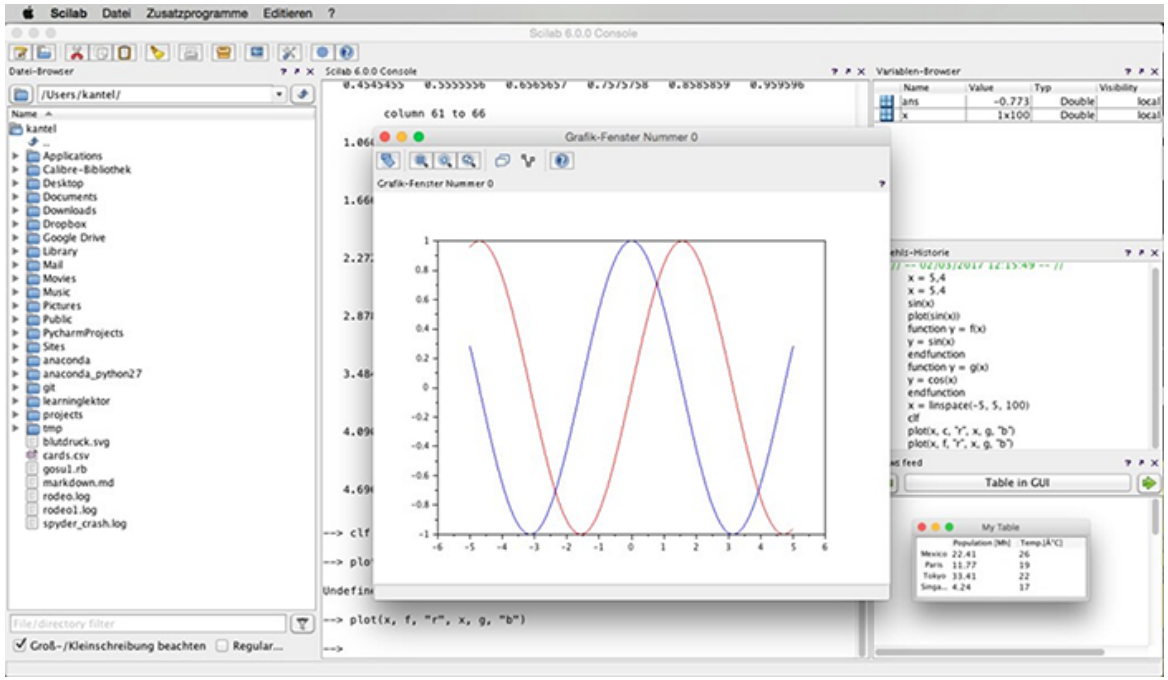
Maxima – aplikacja CAS.



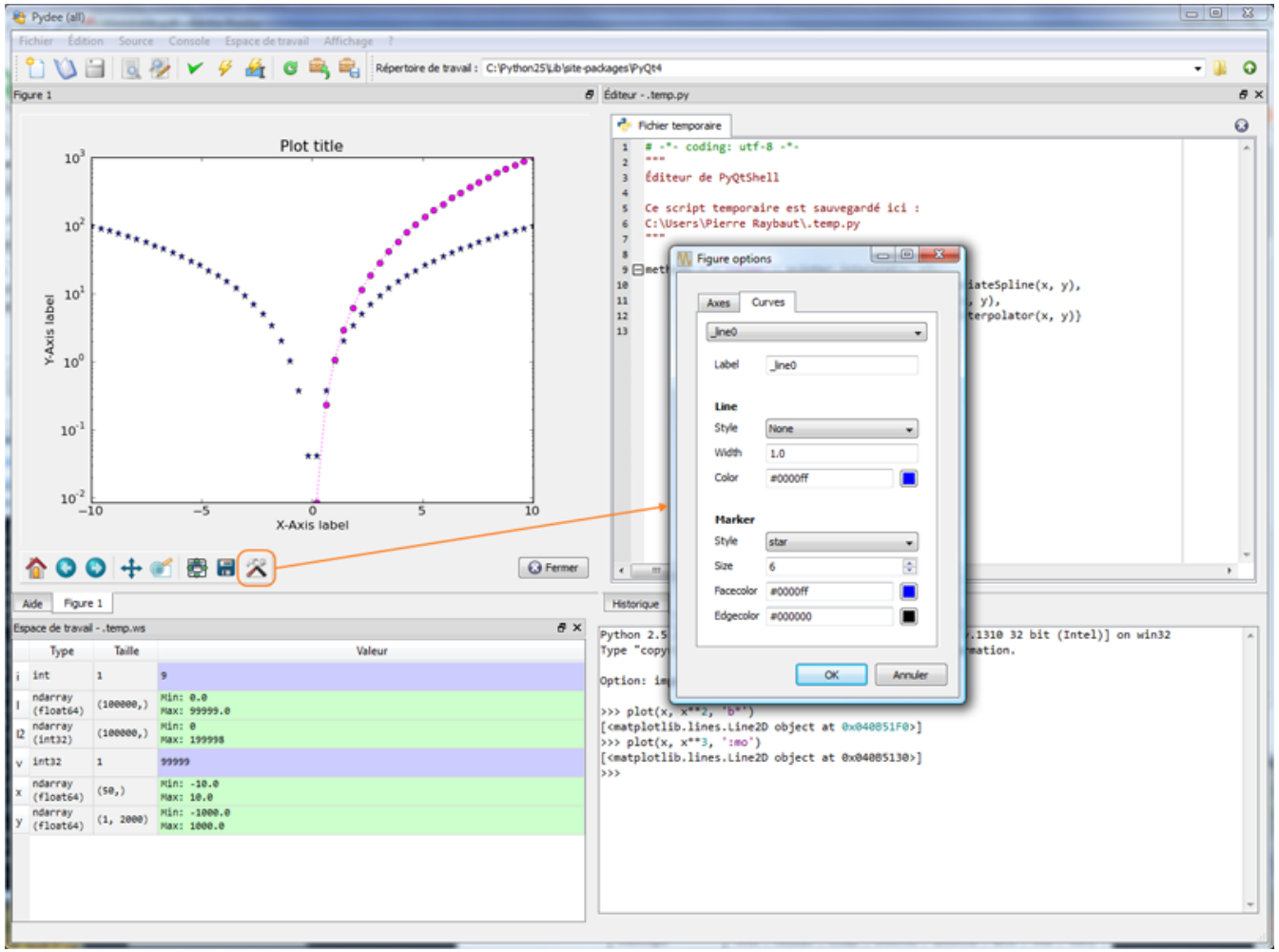
Octave – aplikacja CAS.



SciLab – aplikacja CAS.



Python(x,y) – aplikacja CAS, język programowania.



Potrzeby:

- projektowanie,
- modelowanie,
- symulacja,
- analiza wyników.

Narzędzia:

- obliczenia algebraiczne,
- rozwiązywanie równań i układów równań,
- interpolacja i aproksymacja,
- symulacja,
- prezentacja wyników.

System algebry komputerowej

System algebry komputerowej lub komputerowy system obliczeń symbolicznych (*computer algebra system*, *CAS*) – program komputerowy wspomagający obliczenia symboliczne w matematyce, fizyce i dyscyplinach technicznych.

Typowe wyrażenia z jakimi operują tego rodzaju programy zbudowane są z wielomianów jednej lub wielu zmiennych, funkcji elementarnych, macierzy lub macierzy całek i pochodnych takich wyrażeń.

Typowe operacje wykonywane na wyrażeniach:

- upraszczanie wyrażeń,
- podstawianie wyrażeń symbolicznych za zmienne i redukcja wyrazów podobnych,
- rozwijanie iloczynów,
- rozkład wyrażeń na czynniki,
- różniczkowanie symboliczne,
- całkowanie symboliczne – całki oznaczone i nieoznaczone,
- symboliczne rozwiązywanie niektórych typów równań i ich układów,
- rozwiązywanie równań różniczkowych określonych typów,

- obliczanie granic funkcji i ciągów,
- obliczanie sum szeregów,
- rozwijanie funkcji w szereg,
- operacje na macierzach – mnożenie, odwracanie, obliczanie wyznacznika,
- obliczenia związane z teorią grup,
- obliczenia związane ze statystyką matematyczną,
- operacje na listach i zbiorach elementów,
- eksport wyników obliczeń do formatu TeX-a i EPS.

Większość programów CAS umożliwia rysowanie wykresów funkcji (jednej i dwu zmiennych oraz zmiennej zespolonej) i przeprowadzanie obliczeń z praktycznie dowolną dokładnością.

Wiele z nich ma wbudowane języki programowania, dzięki czemu użytkownik może wykorzystywać do rozwiązywania zadań własne algorytmy i zwiększać w ten sposób funkcjonalność programu.

Systemy algebry komputerowej należy właściwie odróżnić od programów przeznaczonych do obliczeń **numerycznych i inżynierskich** lub języków programowania wyposażonych w dodatkowe biblioteki do obliczeń i symulacji komputerowych.

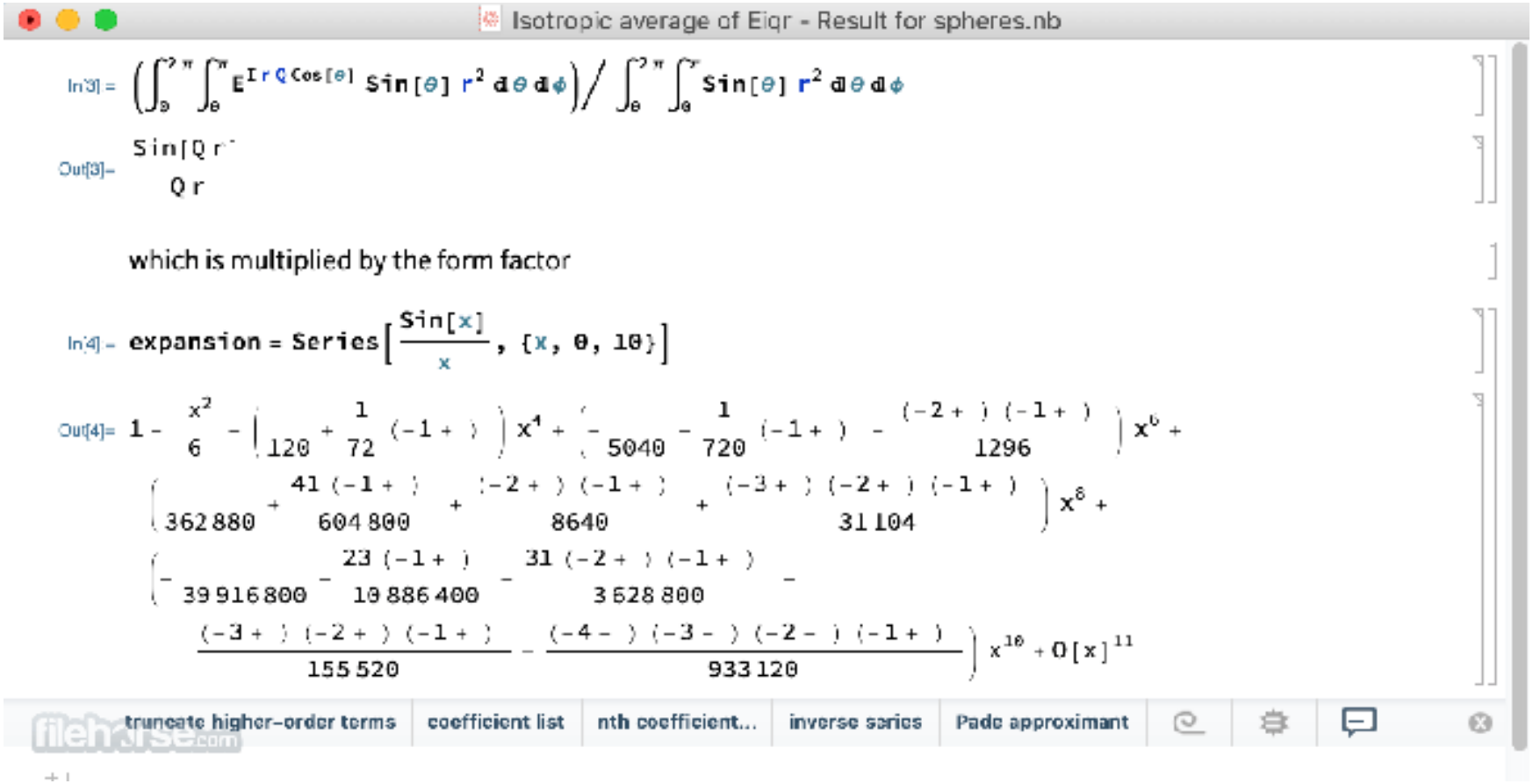
Rozgraniczenie nie jest jednakże jednoznaczne, gdyż niektóre systemy algebraiczne zawierają moduły umożliwiające szybkie obliczenia numeryczne i na odwrót.

Wspomaganie obliczeń matematycznych

rodzaj rachunku	numeryczny	symboliczny
rozwiązywanie trudnych zadań praktycznych	zazwyczaj tak	zazwyczaj nie
dostępność metod o różnej skuteczności	tak	tak
wymaga wiedzy wykraczającej poza rozwiązywane zadanie	najczęściej tak	najczęściej nie
wynik	skończony zestaw liczb lub rysunek	wzór lub informacja o rozwiązaniu

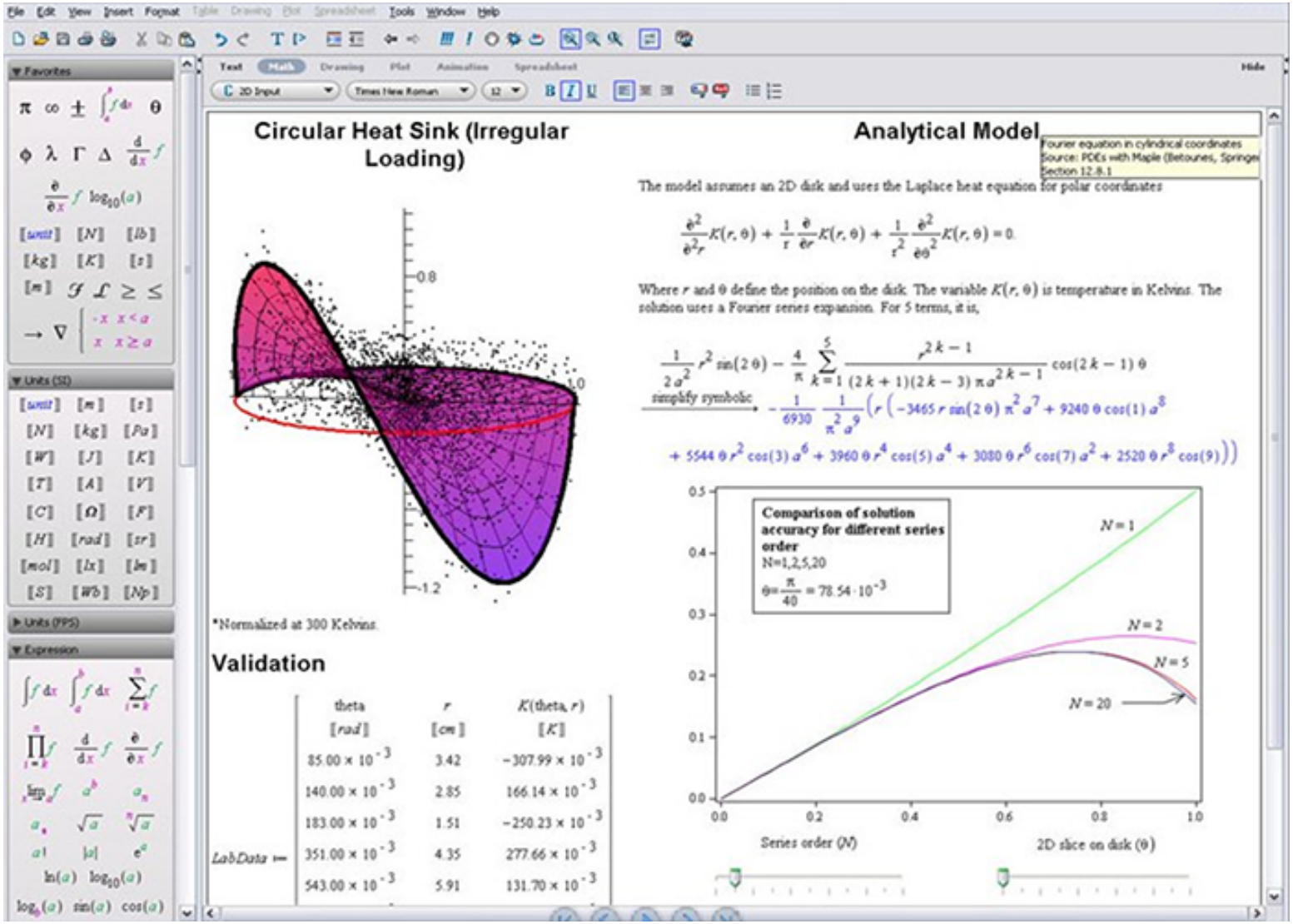
rodzaj rachunku	numeryczny	symboliczny
potrafi działać na abstrakcyjnych obiektach	nie	tak
dobrze radzi sobie z nieskończonościami	zazwyczaj nie	zazwyczaj tak
dobrze radzi sobie z mnogością parametrów	tak	nie
precyzja wyniku	ograniczona	nieskończona
ostateczna jakość wyniku	niepewna	niepewna

Mathematica



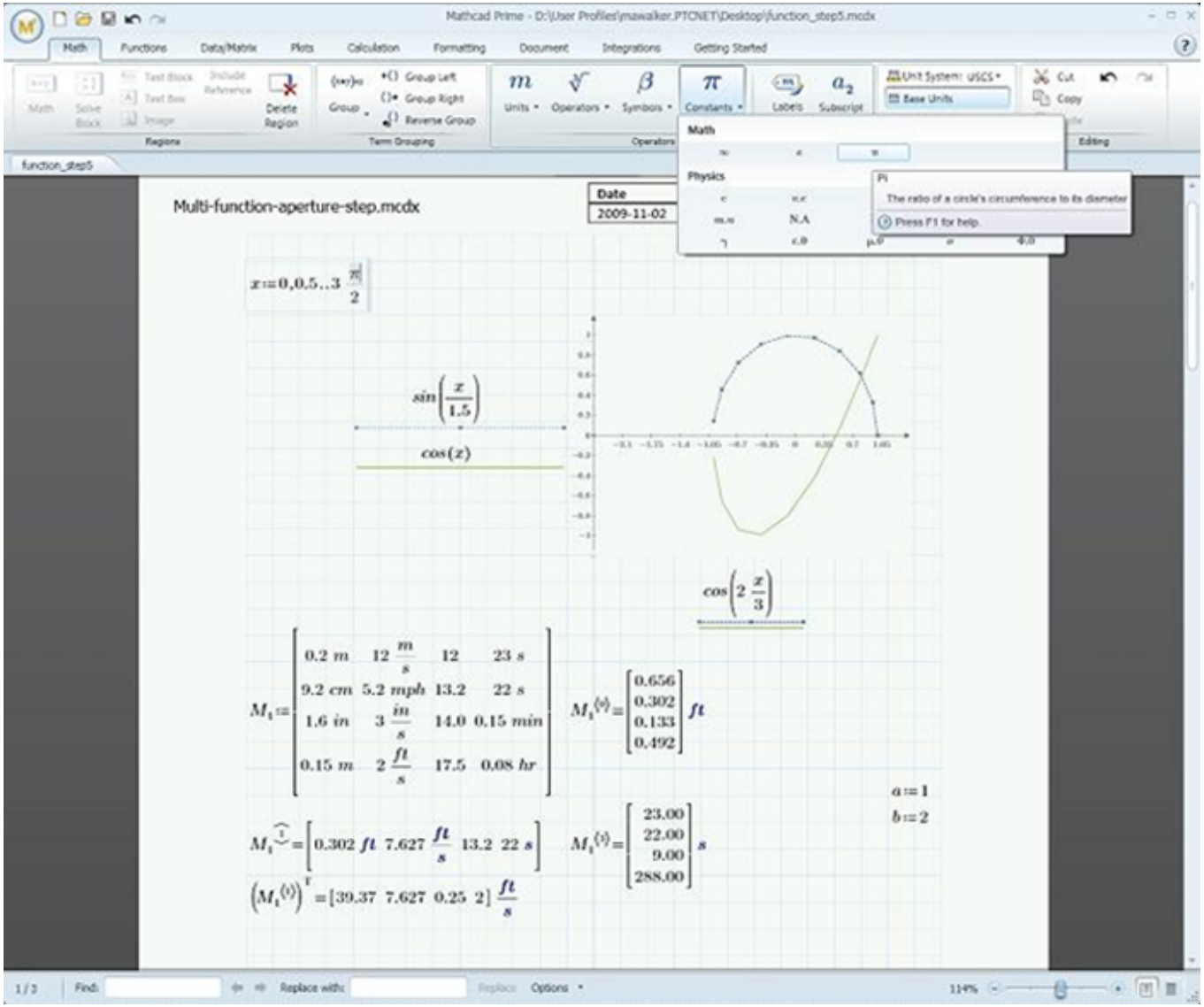
rachunek symboliczny, komercyjny

Maple



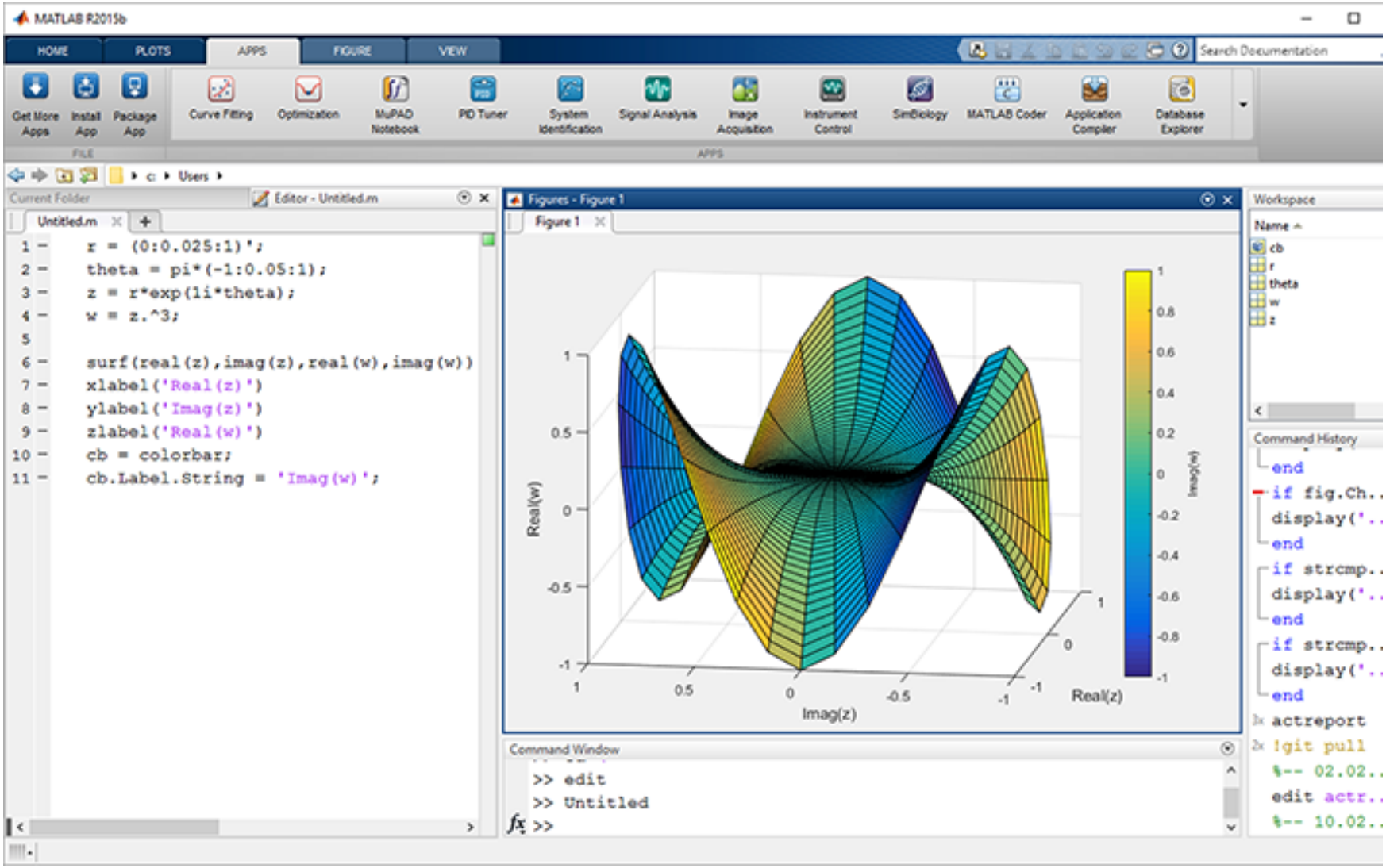
rachunek symboliczny, komercyjny

Mathcad



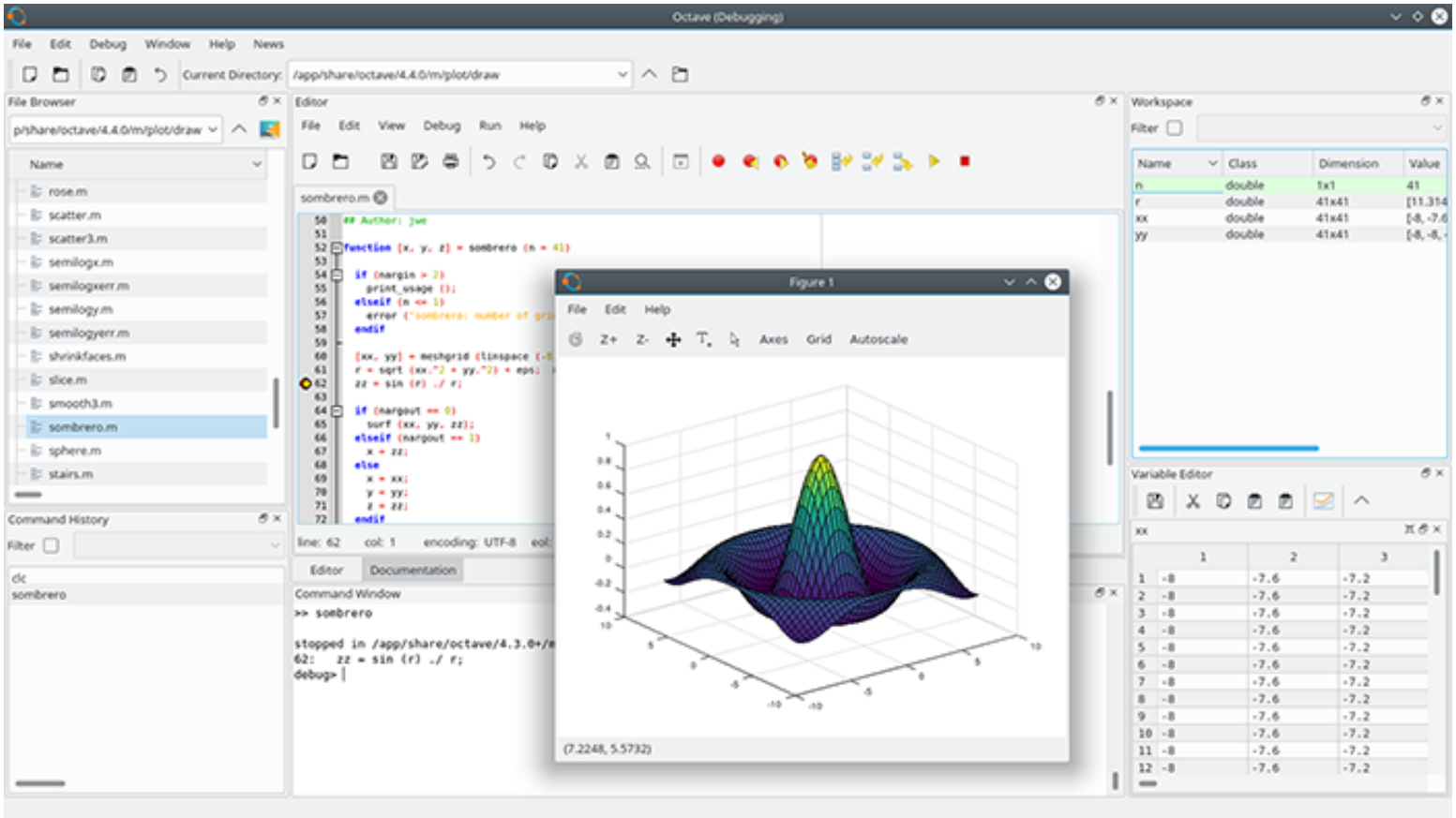
rachunek symboliczny, metody numeryczne, jednostki, komercyjny

MATLAB



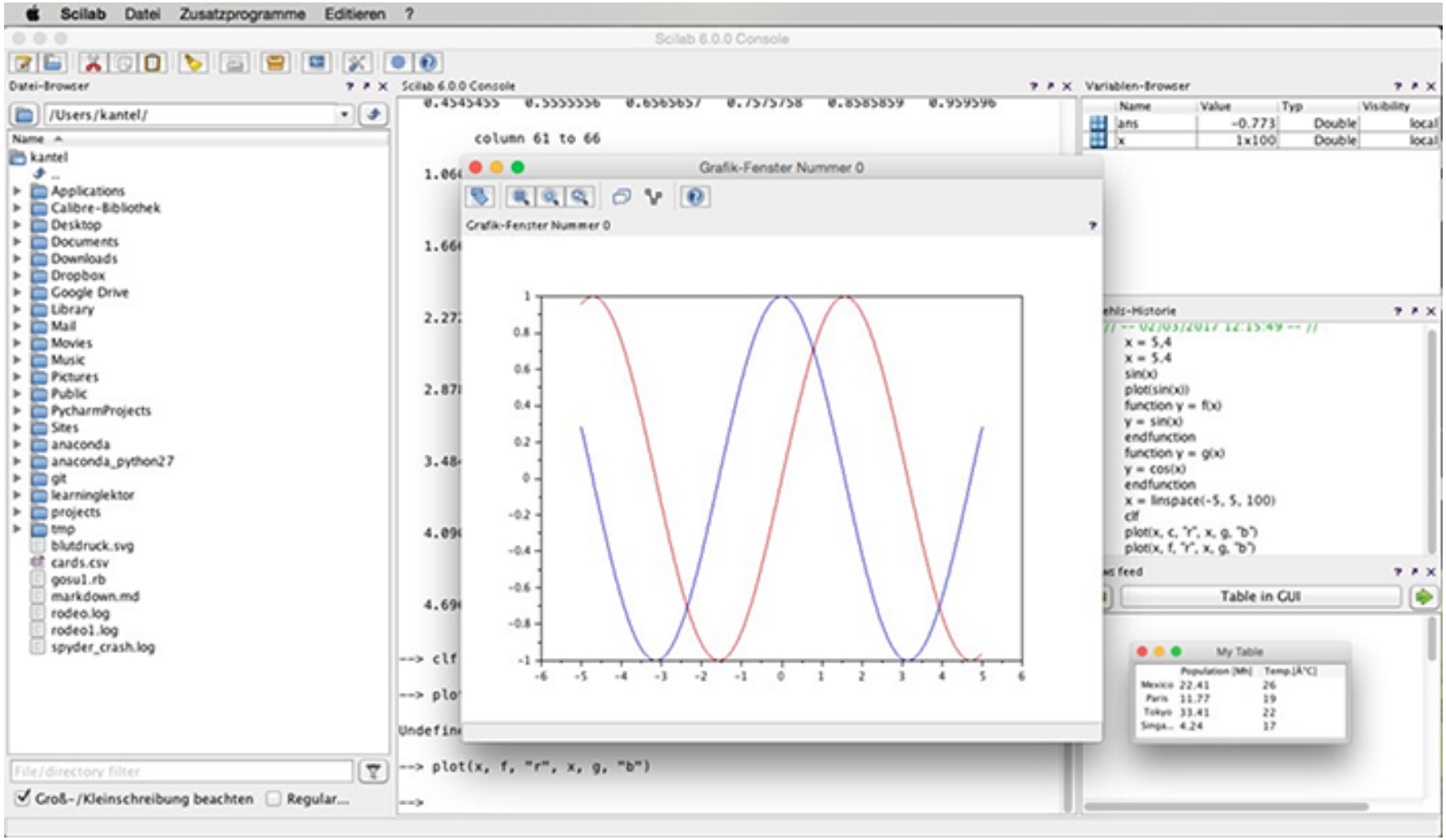
rachunek numeryczny, rachunek symboliczny, modele symulacyjne, własny język programowania, komercyjny

Octave



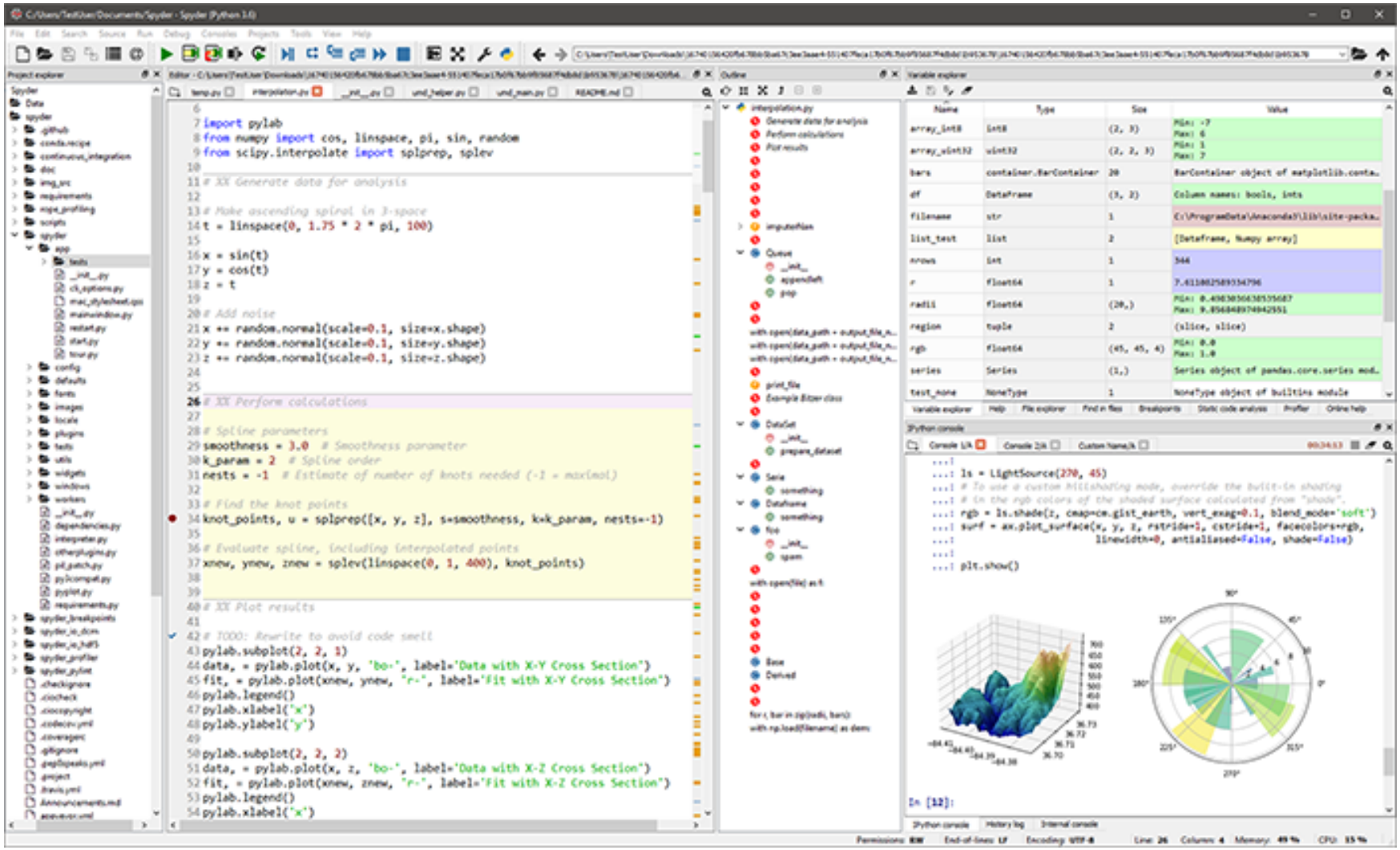
rachunek numeryczny, częściowo zgodny z systemem MATLAB, darmowy

Scilab



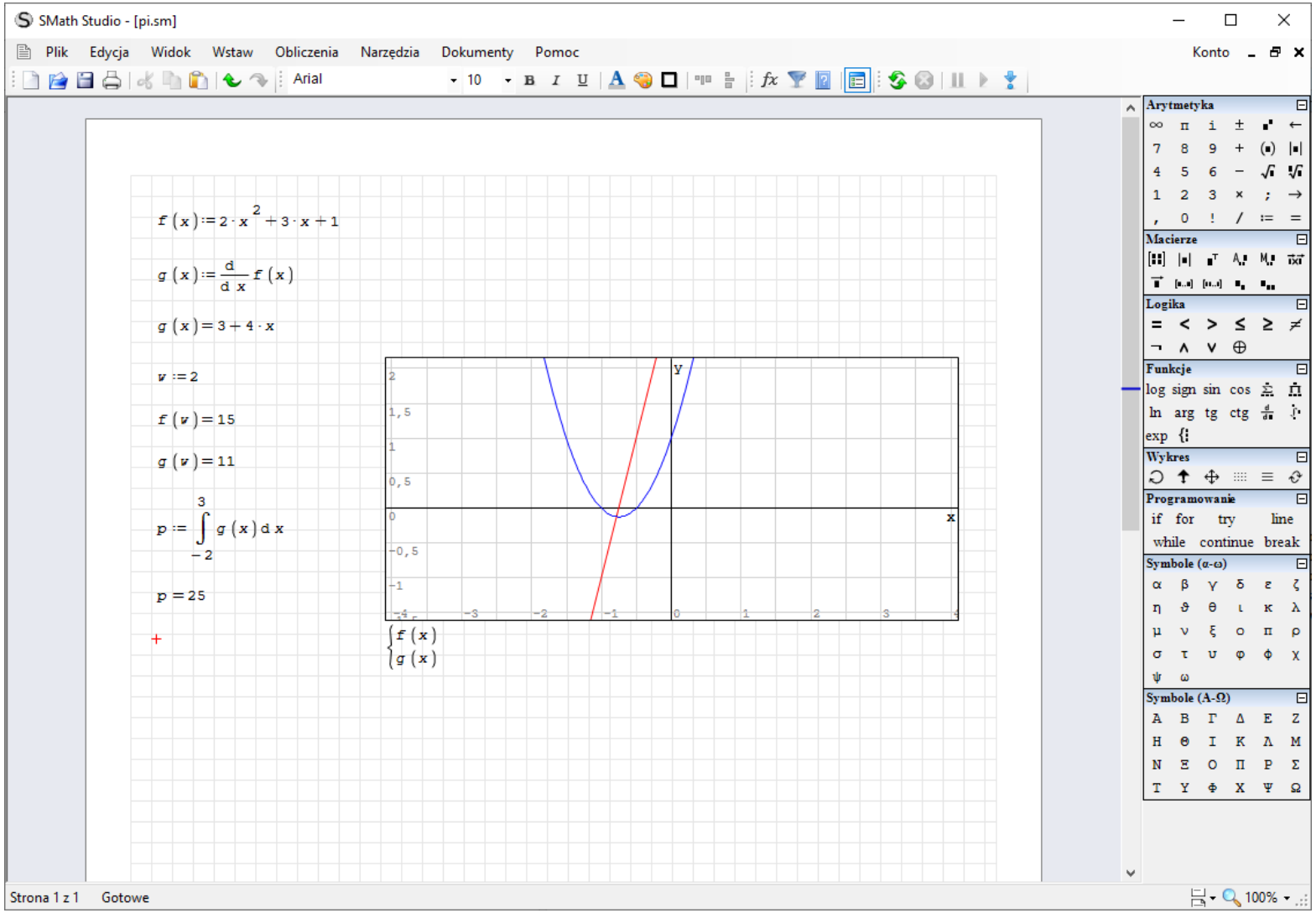
rachunek numeryczny, darmowy

Python (NumPy, SciPy, Matplotlib, Spyder)



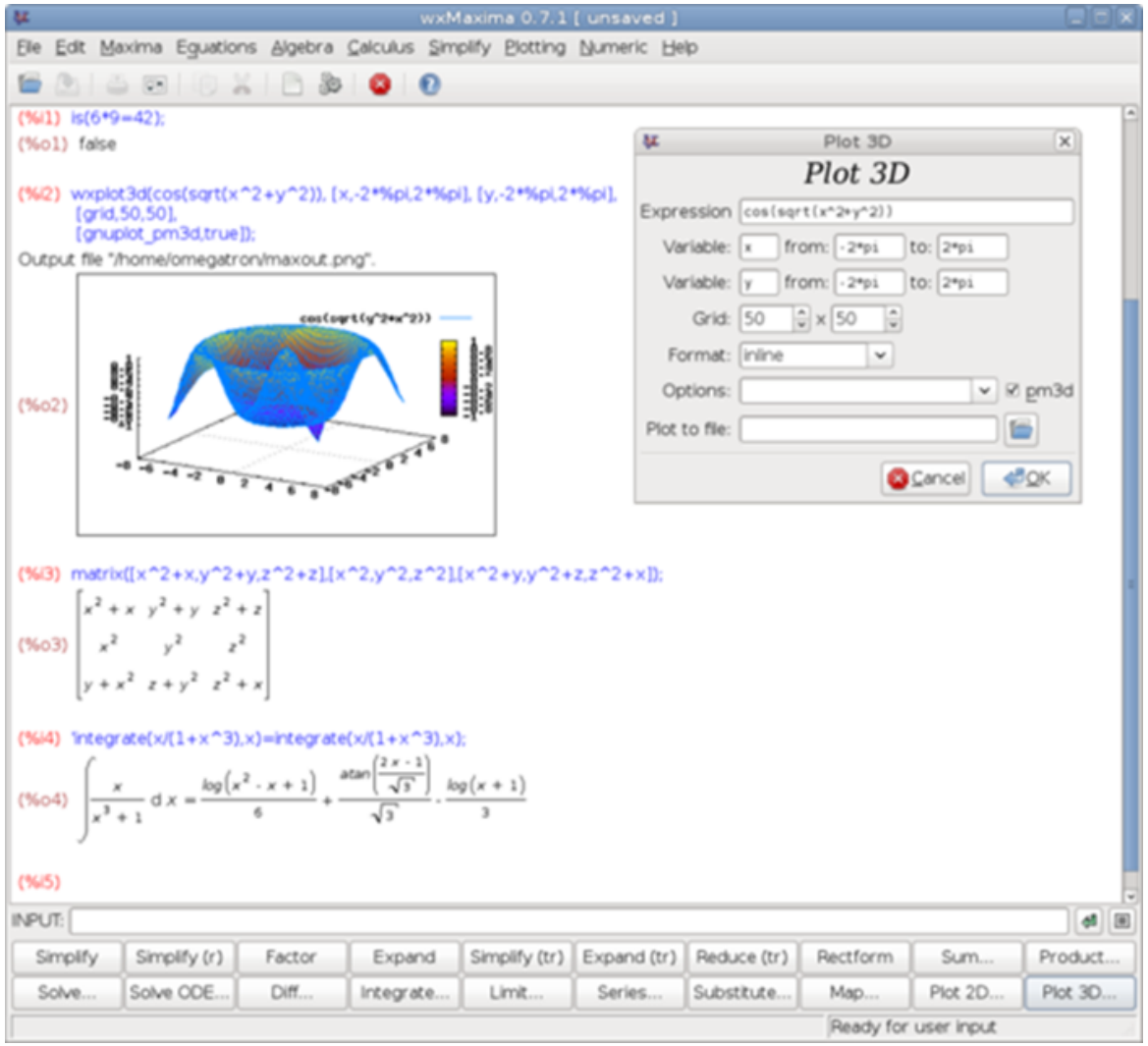
język programowania ogólnego przeznaczenia, biblioteki numeryczne, darmowy

SMath Studio



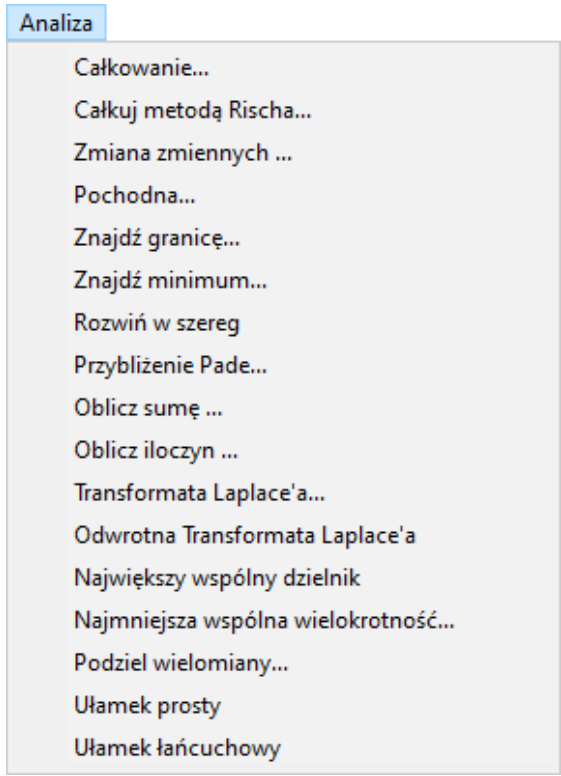
rachunek symboliczny, darmowy

Maxima



rachunek symboliczny, darmowy


Komórka
Wykonaj komórki
Evaluate All Visible Cells Ctrl+R
Evaluate All Cells Ctrl+Shift+R
Evaluate Cells Above Ctrl+Shift+P
Evaluate Cells Below
Usuń wszytskie komórki wyjściowe
Kopiuj ostatnie wejście Ctrl+I
Copy Previous Output Ctrl+U
Complete Word Ctrl+Space
Show Template Ctrl+Shift+Space
Insert Input Cell Ctrl+0
Wstaw komórkę tekstowa Ctrl+1
Wstaw komórkę tytułową Ctrl+2
Wstaw komórkę rozdziału Ctrl+3
Wstaw komórkę podrozdziału Ctrl+4
Insert Subsubsection Cell Ctrl+5
Insert heading5 Cell Ctrl+6
Insert heading6 Cell Ctrl+7
Wstaw znak końca strony
Wstaw obrazek...
Fold All Ctrl+Alt+[
Unfold All Ctrl+Alt+]
Poprzenie polecenie Alt+Up
Następne polecenie Alt+Down
Merge Cells Ctrl+M
Divide Cell Ctrl+D
Automatically answer questions

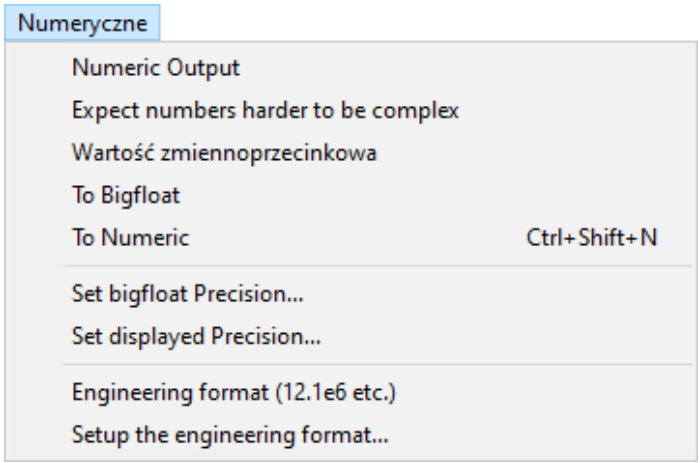

Analiza
Całkowanie...
Całkuj metodą Rischa...
Zmiana zmiennych ...
Pochodna...
Znajdź granicę...
Znajdź minimum...
Rozwiń w szereg
Przybliżenie Pade...
Oblicz sumę ...
Oblicz iloczyn ...
Transformata Laplace'a...
Odwrotna Transformata Laplace'a
Największy wspólny dzielnik
Najmniejsza wspólna wielokrotność...
Podziel wielomiany...
Ułamek prosty
Ułamek łańcuchowy

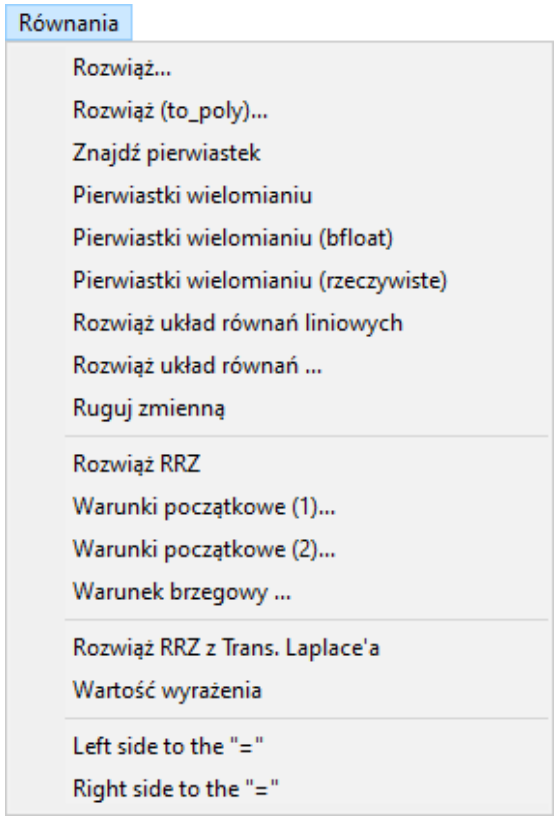
List

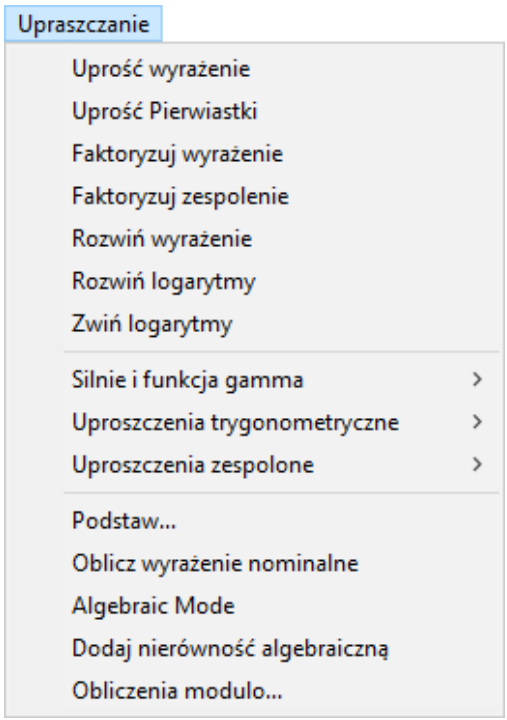
- Create list ›
- Use list ›
- Extract Elements ›
- Append ›
- Length
- Reverse
- Sort
- Remove duplicates
- Push
- Pop
- Nested list to Matrix
- Matrix to nested List

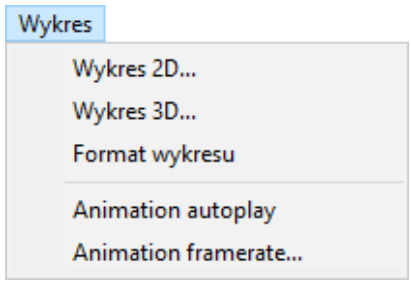
Matrix

- Generuj macierz
- Generuj macierz z wyrazu...
- Wprowadź macierz
- Nested list to Matrix
- Matrix from csv file
- Matrix to csv file
- Macierz odwrotna
- Wielomian charakterystyczny ...
- Wyznacznik
- Wartości własne
- Wektory własne
- Macierz dołączona
- Rank
- Transponuj macierz
- Extract Row
- Extract Column
- Remove Rows or Columns
- Convert Row to list
- Convert Column to list
- Multiply matrices
- Matrix exponent
- Hadamard (element-by-element) product
- Hadamard exponent
- Utwórz listę...
- Zastosuj do listy ...
- Map to List(s)...
- Mapuj macierz...


Numeryczne
Numeric Output
Expect numbers harder to be complex
Wartość zmiennoprzecinkowa
To Bigfloat
To Numeric
Ctrl+Shift+N
Set bigfloat Precision...
Set displayed Precision...
Engineering format (12.1e6 etc.)
Setup the engineering format...


Równania
Rozwiąż...
Rozwiąż (to_poly)...
Znajdź pierwiastek
Pierwiastki wielomianiu
Pierwiastki wielomianiu (bfloat)
Pierwiastki wielomianiu (rzeczywiste)
Rozwiąż układ równań liniowych
Rozwiąż układ równań ...
Ruguj zmienną
Rozwiąż RRZ
Warunki początkowe (1)...
Warunki początkowe (2)...
Warunek brzegowy ...
Rozwiąż RRZ z Trans. Laplace'a
Wartość wyrażenia
Left side to the "="
Right side to the "="


Upraszczanie
Uprość wyrażenie
Uprość Pierwiastki
Faktoryzuj wyrażenie
Faktoryzuj zespolenie
Rozwiń wyrażenie
Rozwiń logarytmy
Zwiń logarytmy
Silnie i funkcja gamma
Uproszczenia trygonometryczne
Uproszczenia zespolone
Podstaw...
Oblicz wyrażenie nominalne
Algebraic Mode
Dodaj nierówność algebraiczną
Obliczenia modulo...


Wykres
Wykres 2D...
Wykres 3D...
Format wykresu
Animation autoplay
Animation framerate...

SMath Studio

To darmowy program algebry komputerowej z zamkniętym kodem.

Cechą charakterystyczną programu **SMath Studio** jest łatwość obsługi. Interfejs programu imituje notatnik i jest intuicyjny w użyciu, a wiele operacji daje się realizować za pomocą myszy. Równania i wyrażenia algebraiczne wyświetlane są w postaci graficznej, a nie tekstowej.

Niektóre możliwości programu:

- obliczenia symboliczne,
- operacje na wektorach i macierzach,
- rozwiązywanie układów równań,
- wykreślanie wykresów funkcji jednej i dwu zmiennych,
- znajdowanie pierwiastków wielomianów i innych funkcji.

SMath Studio - [pi.sm]

Plik Edycja Widok Wstaw Obliczenia Narzędzia Dokumenty Pomoc Konto

Arial 10

$f(x) := 2 \cdot x^2 + 3 \cdot x + 1$

$g(x) := \frac{d}{dx} f(x)$

$g(x) = 3 + 4 \cdot x$

$w := 2$

$f(w) = 15$

$g(w) = 11$

$p := \int_{-2}^{3} g(x)\,dx$

$p = 25$

$\begin{cases} f(x) \\ g(x) \end{cases}$

Arytmetyka
Macierze
Logika
Funkcje
log sign sin cos
ln arg tg ctg
exp
Wykres
Programowanie
if for try line
while continue break
Symbole (α-ω)
Symbole (A-Ω)

Strona 1 z 1 Gotowe 100%

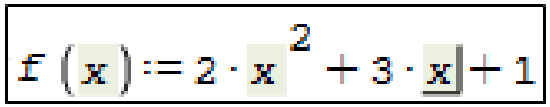
Arkusz SMath Studio

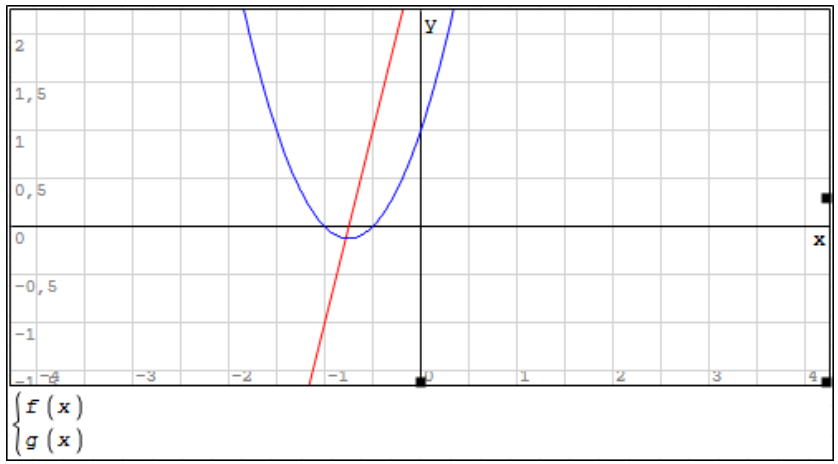
Regiony:

- region wyrażenia matematycznego,
- region tekstu,
- region wykresu.

Kursory:

- czerwony krzyżyk – miejsce wstawienia regionów,
- czerwona linia – kursor edycji tekstu,
- niebieskie linie – kursor edycji wyrażeń matematycznych.



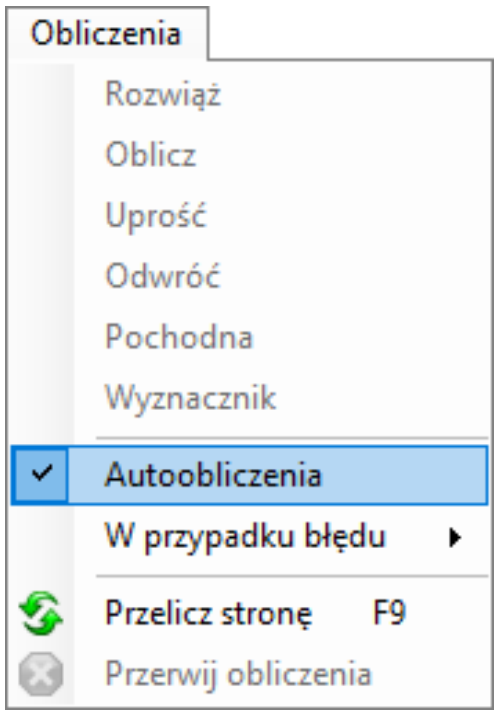


Wyrażenia

$a := 4,321$ $\quad B := \frac{1}{3}$ $\quad f(x) := x^2 - 2 \cdot x + 3$

$\frac{1}{2} + 2,34 \cdot \sqrt[3]{2} = 3,4482$ $\quad f(3) = 6$

- definicja zmiennych i funkcji,
- wyznaczanie wartości wyrażeń,
- edycja wyrażeń,
- automatyczne i ręczne wyznaczanie wartości wyrażeń,



Analiza matematyczna i algebra liniowa

- formatowanie wyników obliczeń,
- rachunek różniczkowy i całkowy,
- definicja zmiennej zakresowej,
- tworzenie wykresów funkcji,

$a := 4,321$ $\quad B := \frac{1}{3}$ $\quad f(x) := x^2 - 2 \cdot x + 3$

$\frac{1}{2} + 2,34 \cdot \sqrt[3]{2} = 3,4482$ $\quad f(3) = 6$

$f(x) := 2 \cdot x^2 + 3 \cdot x + 1$

$g(x) := \frac{d}{dx} f(x)$

$g(x) = 3 + 4 \cdot x$

$w := 2$

$f(w) = 15$

$g(w) = 11$

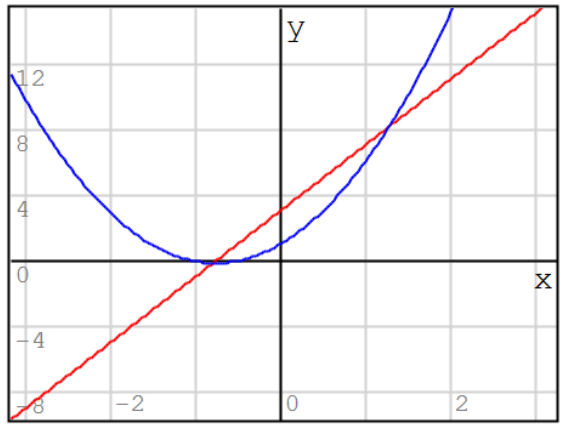
$p := \int_{-2}^{3} g(x)\,dx$

$p = 25$

$10000 \cdot \pi = 31415,9265$

$10000 \cdot \pi = 3,142 \cdot 10^4$

$10000 \cdot \pi = 31415,9265359$



$\begin{cases} f(x) \\ g(x) \end{cases}$

- transpozycja macierzy,
- mnożenie macierzy przez skalar,
- mnożenie macierzy,

$$A := \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$$

$$B := A^{T}$$

$$B = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

$$A + B^{T} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \\ 20 & 22 & 24 \end{bmatrix}$$

$$k := 3$$

$$k \cdot B = \begin{bmatrix} 3 & 12 & 21 & 30 \\ 6 & 15 & 24 & 33 \\ 9 & 18 & 27 & 36 \end{bmatrix}$$

- wyznacznik macierzy,
- macierz odwrotna,
- odwołanie do kolumny i wiersza,
- łączenie macierzy,

$$A \cdot B = \begin{bmatrix} 14 & 32 & 50 & 68 \\ 32 & 77 & 122 & 167 \\ 50 & 122 & 194 & 266 \\ 68 & 167 & 266 & 365 \end{bmatrix} \qquad |A \cdot B| = 0$$

$$F := \begin{bmatrix} 1 & 2 & 3 \\ -2 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad A := \begin{bmatrix} 3 & 5 & 7 \\ 4 & 6 & 8 \\ 5 & 7 & 9 \end{bmatrix}$$

$$|F| = -36 \qquad |A| = 0$$

$$\mathrm{col}(F; 2) = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \qquad \mathrm{row}(A; 3) = \begin{bmatrix} 5 & 7 & 9 \end{bmatrix}$$

$$\mathrm{stack}(A; F) = \begin{bmatrix} 3 & 5 & 7 \\ 4 & 6 & 8 \\ 5 & 7 & 9 \\ 1 & 2 & 3 \\ -2 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\mathrm{augment}(A; F) = \begin{bmatrix} 3 & 5 & 7 & 1 & 2 & 3 \\ 4 & 6 & 8 & -2 & 5 & 6 \\ 5 & 7 & 9 & 7 & 8 & 9 \end{bmatrix}$$

- sortowanie według kolumny,
- sortowanie według wiersza,
- podmacierz,
- element minimalny,
- element maksymalny,
- średnia,

$$AF := \text{augment}(A; F)$$

$$AF = \begin{bmatrix} 3 & 5 & 7 & 1 & 2 & 3 \\ 4 & 6 & 8 & -2 & 5 & 6 \\ 5 & 7 & 9 & 7 & 8 & 9 \end{bmatrix}$$

$$\text{csort}(AF; 4) = \begin{bmatrix} 4 & 6 & 8 & -2 & 5 & 6 \\ 3 & 5 & 7 & 1 & 2 & 3 \\ 5 & 7 & 9 & 7 & 8 & 9 \end{bmatrix}$$

$$\text{rsort}(AF; 2) = \begin{bmatrix} 1 & 3 & 2 & 5 & 3 & 7 \\ -2 & 4 & 5 & 6 & 6 & 8 \\ 7 & 5 & 8 & 7 & 9 & 9 \end{bmatrix}$$

$$\text{submatrix}(AF; 2; 3; 3; 6) = \begin{bmatrix} 8 & -2 & 5 & 6 \\ 9 & 7 & 8 & 9 \end{bmatrix}$$

$$\min(AF) = -2$$

$$\max(AF) = 9$$

- miejsca zerowe funkcji,
- miejsca zerowe wielomianów,

$$f(x) := \mathbf{e}^{x} - x^{3}$$

$$\text{roots}(f(\_x); \_x; 2) = 1,8572$$

$$\text{roots}(f(\_x); \_x; 4) = 4,5364$$

$$w(x) := -3 \cdot x^{3} - 2 \cdot x^{2} + 5 \cdot x + 1$$

$$a := \begin{bmatrix} 1 \\ 5 \\ -2 \\ -3 \end{bmatrix} \qquad \text{polyroots}(a) = \begin{bmatrix} -0,1897 \\ 1,1084 \\ -1,5853 \end{bmatrix}$$

- rozwiązywanie układów równań liniowych,
- rozwiązywanie układów równań nieliniowych,

$$A := \begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 2 \\ 5 & 3 & 1 \end{bmatrix} \qquad B := \begin{bmatrix} 13 \\ -3 \\ 50 \end{bmatrix}$$

$$|A| = 25 \qquad X := A^{-1} \cdot B$$

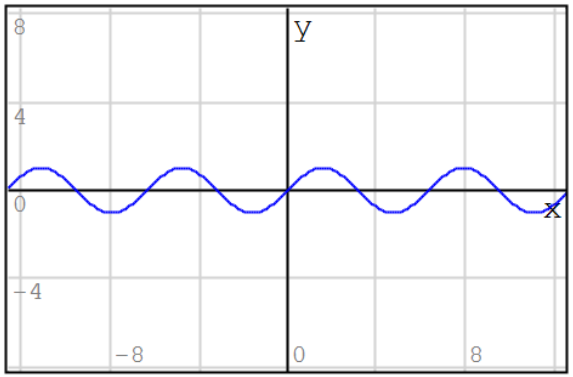
$$X = \begin{bmatrix} 0,28 \\ 16,84 \\ -1,92 \end{bmatrix} \qquad A \cdot X = \begin{bmatrix} 13 \\ -3 \\ 50 \end{bmatrix}$$

$$R(x; y; z) := \begin{bmatrix} x^3 + 3 \cdot x^2 + 9 \cdot x - 2 \cdot y = 4 \cdot z \\ e^x = y + 1 \\ z^2 = 3 - \cos(y) \end{bmatrix}$$

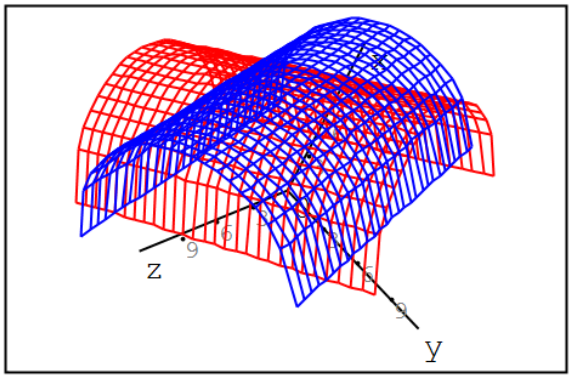
$$R_x := \text{roots}\left(R(x; y; z); \begin{bmatrix} x \\ y \\ z \end{bmatrix}; \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}\right)$$

$$R_x = \begin{bmatrix} 0,7239 \\ 1,0624 \\ 1,5853 \end{bmatrix}$$

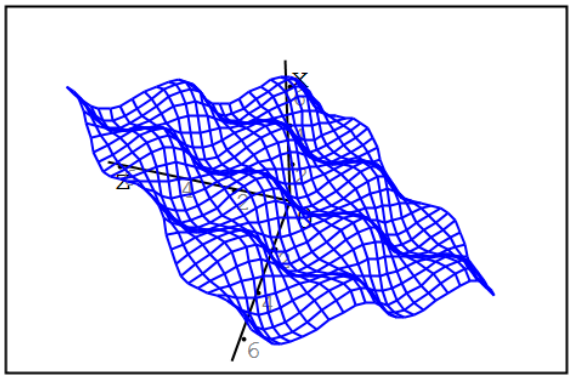
- wykresy 3D.



$\sin(x)$



$$\begin{cases} \sqrt{25 - x^2} \\ \sqrt{25 - y^2} \end{cases}$$

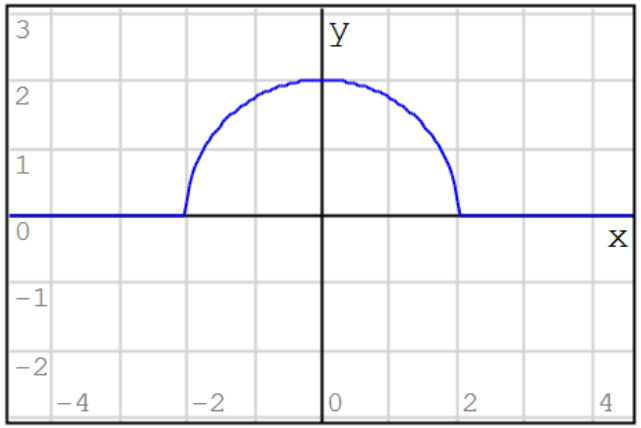


$\sin(x) \cdot \cos(y)$

Programowanie

- warunek **if**,

$$f(x) := \text{if } |x| > 2 \quad 0 \quad \text{else} \quad \sqrt{4 - x^2}$$



$f(x)$

- pętla **for**,

$n := 5 \qquad s := 0$

for $i \in [1..n]$
  $s := s + i$

$s = 15$

$$\sum_{i=1}^{5} i = 15$$

$n := 9 \qquad s := 0$

for $i \in [1..n]$
  $s := s + i$

$s = 45$

$$\sum_{i=1}^{9} i = 45$$

- pętla **while**.

$x := 2 \qquad \varepsilon := 10^{-6}$

$r := \frac{x}{2} \qquad r_n := \frac{r}{2} + \frac{x}{2 \cdot r}$

while $|r_n - r| > \varepsilon$

$\begin{cases} r := r_n \\ r_n := \frac{r}{2} + \frac{x}{2 \cdot r} \end{cases}$

$r_n = 1,4142$

$\sqrt{2} = 1,4142$

$x := 5 \qquad \varepsilon := 10^{-6}$

$r := \frac{x}{2} \qquad r_n := \frac{r}{2} + \frac{x}{2 \cdot r}$

while $|r_n - r| > \varepsilon$

$\begin{cases} r := r_n \\ r_n := \frac{r}{2} + \frac{x}{2 \cdot r} \end{cases}$

$r_n = 2,2361$

$\sqrt{5} = 2,2361$

Po co mi programowanie?

Linus Torvalds

Computer programming is not for everyone. I think it's reasonably specialized, and nobody really expects most people to have to do it.

Steve Jobs

Everybody in this country should learn how to program a computer, because it teaches you how to think.

Celem nauki programowania jest wykształcenie umiejętności **myślenia komputacyjnego** (*computational thinking*), które obejmuje myślenie algorytmiczne w rozwiązywaniu problemów oraz umiejętność programowania rozszerzone na wszystkie obszary działalności ludzkiej. Takie podejście przede wszystkim pozwala zwiększyć efektywność i ułatwić pracę.

Nawet życiowe problemy i decyzje można potraktować jako problem algorytmiczny.

Języki programowania lub **makropolecenia** udostępniają szereg narzędzi pozwalających na przyśpieszenie i zwiększenie efektywności pracy, zrobienie czegoś w łatwiejszy sposób, a nawet utworzenie kompletnie nowych narzędzi.

Rozwój języków programowania znacznie obniżył próg umiejętności jakie należy posiadać, żeby zacząć naukę. Nie trzeba posiadać żadnych informacji na temat budowy komputera, nie trzeba mieć solidnych podstaw matematycznych (choć te są przydatne), nie trzeba mieć superkomputera ani kupować dodatkowego drogiego oprogramowania.

Przypomnienie podstawowych terminów

Program komputerowy (aplikacja) – sekwencja symboli (zrozumiałych dla komputera rozkazów) przeznaczonych do przetworzenia zgodnie z pewnymi regułami, zwanymi **językiem programowania**.

Program w postaci języka *zrozumiałego* dla człowieka nazywany jest **kodem źródłowym**, podczas gdy program wyrażony w postaci zrozumiałej dla maszyny (to jest za pomocą ciągu liczb, a bardziej precyzyjnie zer i jedynek) nazywany jest kodem maszynowym bądź postacią binarną (wykonywalną).

Program jest zazwyczaj wykonywany przez komputer, bezpośrednio – jeśli wyrażony jest w języku zrozumiałym dla danej maszyny lub pośrednio – gdy jest interpretowany przez inny program (interpreter).

Programy komputerowe można zaklasyfikować według ich zastosowań. Wyróżnia się zatem aplikacje użytkowe, systemy operacyjne, gry, kompilatory i inne. Programy wbudowane wewnątrz urządzeń określa się jako **firmware**.

W najprostszym modelu wykonanie programu (zapisanego w postaci zrozumiałej dla maszyny) polega na umieszczeniu go w pamięci operacyjnej komputera i wskazaniu procesorowi adresu pierwszej instrukcji.

Po tych czynnościach procesor będzie wykonywał kolejne instrukcje programu, aż do jego zakończenia.

Program komputerowy będący w trakcie wykonania nazywany jest **procesem** lub **zadaniem**.

Tworzenie programu komputerowego można podzielić na dwa etapy:

- Po zrodzeniu się **pomysłu** powinien powstać **algorytm**. Algorytm wymusza stosowanie podziału programu na funkcje. zmienne, obiekty, na których program będzie operował, jak również wprowadzenie procedur, które opisują wykonywane operacje.
- Algorytm należy zapisać w języku programowania, stosując dostępne struktury danych i funkcje – tworzenie **kodu źródłowego**. W trakcie tworzenia programu kod jest poddawany **debugowaniu** – wyszukiwanie błędów.

Językiem programowania nazywamy zestaw zasad tekstowego lub graficznego opisu algorytmu za pomocą przyjętych elementów języka.

Podobnie jak języki naturalne, język programowania składa się ze zbiorów reguł syntaktycznych oraz semantyki, które opisują, jak należy budować poprawne wyrażenia oraz jak komputer ma je rozumieć

Język programowania pozwala na precyzyjny zapis algorytmów oraz innych zadań, jakie komputer ma wykonać.

Podział języków programowania

Kod maszynowy (język maszynowy) – język programowania, w którym zapis programu wymaga **instrukcji** bezpośrednio jako liczb, które są rozkazami i danymi bezpośrednio pobieranymi przez procesor wykonujący ten program.

Jest dopasowany do konkretnego typu procesora i przeznaczony do bezpośredniego wykonania przez procesor. Analiza kodu maszynowego jest praktycznie niemożliwa przez człowieka.

Języki niskopoziomowe – przedstawiają one instrukcje udostępniane przez system komputerowy w postaci prostych oznaczeń (o ograniczonej liczbie, zakodowane w procesorze). Do języków niskopoziomowych należą **assemblery**.

Języki wysokopoziomowe – składnia i słowa kluczowe mają maksymalnie ułatwić rozumienie kodu programu dla człowieka, tym samym zwiększając poziom abstrakcji i dystansując się od budowy sprzętu komputerowego.

Kod napisany w języku wysokiego poziomu nie jest bezpośrednio *zrozumiały* dla komputera – większość kodu stanowią tak naprawdę normalne słowa (najczęściej w języku angielskim).

Języki wysokopoziomowe dzielimy na dwie grupy:

- interpretowane,
- kompilowane.

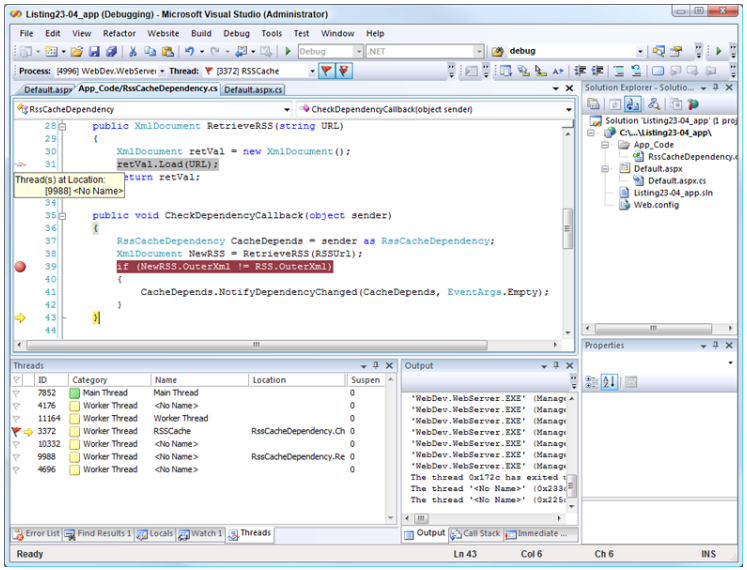
Języki interpretowane nie wymagają kompilacji tylko **interpretera**. Są przechowywane w postaci kodu źródłowego i dopiero podczas uruchomienia wczytywane, analizowane i wykonywane przez interpreter języka – **PHP**, **JavaScript**, **Python**, **PERL**. Programy przeznaczone do interpretacji często nazywane są **skryptami**.

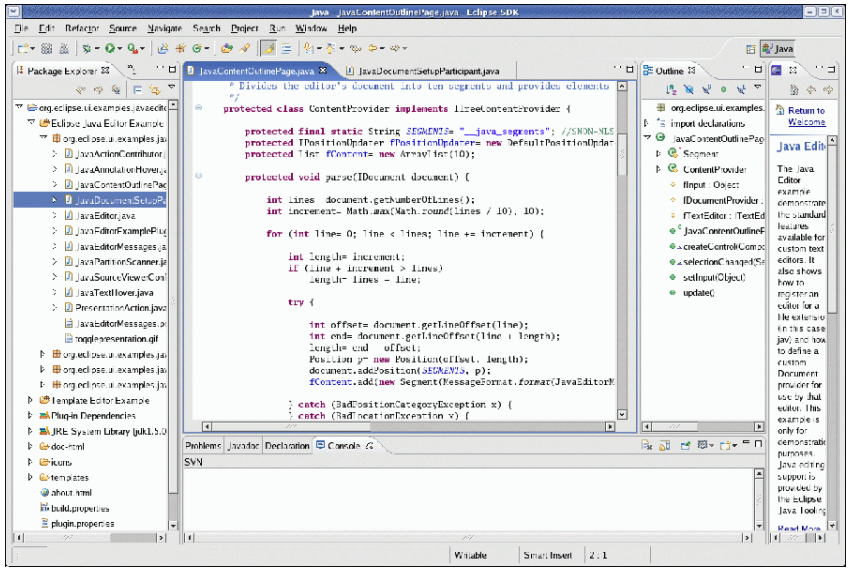
Języki kompilowane wymagają procesu kompilacji kodu źródłowego do postaci kodu maszynowego (postaci binarnej). Robi to specjalny program zwany **kompilatorem**, dzięki czemu możliwe staje się jego późniejsze uruchomienie. Języki kompilowane: **Pascal**, **C**, **C++**, **Fortran**, **Java**.

Do utworzenia programu w danym języku niezbędne są **edytor** tekstu, **debugger** i **kompilator**. Programy te mogą tworzyć integralne środowisko pracy, udostępniające kombinacje tych funkcji – mówimy wówczas o **środowisku programistycznym**.

Aby przyśpieszyć tworzenie aplikacji, szczególnie z interfejsem graficznym, tworzy się narzędzia **RAD** (*Rapid Application Development*), które umożliwiają tworzenie programów z gotowych komponentów (w sensie elementów interfejsu i funkcji z implementacją typowych algorytmów) na przykład:

- Microsoft VisualStudio,
- Eclipse IDE.





Syntaktyka i semantyka

Aby ciąg znaków mógł być rozpoznany jako program napisany w danym języku, musi spełniać reguły **syntaktyki** (składni). Składnia opisuje:

- rodzaje dostępnych symboli,
- zasady, według których symbole mogą być łączone w większe struktury.

Należy zauważyć, że na etapie przetwarzania składni w ogóle nie jest brane pod uwagę znaczenie poszczególnych symboli. W praktyce kod poprawny składniowo nie musi być poprawny semantycznie. Występuje tu analogia do języków naturalnych. Zdanie „Bźdźiągwy się muciosząl" jest poprawne pod względem gramatycznym, lecz nie posiada żadnego znaczenia, ponieważ zostały w nim użyte nieistniejące słowa.

Semantyka języka programowania definiuje precyzyjnie znaczenie poszczególnych symboli oraz ich funkcję w programie. Semantykę najczęściej definiuje się słownie, ponieważ większość z jej zagadnień jest trudna lub wręcz niemożliwa do ujęcia w jakikolwiek formalizm.

Część błędów semantycznych można wychwycić już w momencie wstępnego przetwarzania kodu programu, np. próbę odwołania się do nieistniejącej funkcji, lecz inne mogą ujawnić się dopiero w trakcie wykonywania.

Błędy

W trakcie pisania kodu nie da się uniknąć błędów. Błędy mogą wynikać z niepoprawnego wpisania instrukcji, pominięcia kropek, przecinków, nawiasów, itp. Takie błędy noszą nazwę **syntaktycznych** (składniowych).

Oprócz błędów syntaktycznych, można spotkać jeszcze błędy **semantyczne** (znaczeniowe, wykonania) i **logiczne**.

Błędy wykonania występują w chwili odtwarzania procedur (programu) i mogą wynikać z próby uruchomienia nieistniejącej procedury, otwarcia nieistniejącego pliku lub złego typowania zmiennych.

Błędy logiczne zwykle nie wywołują żadnych komunikatów błędu. Chociaż program jest poprawny pod względem syntaktycznym i semantycznym, nie powoduje żadnych problemów kompilacji i uruchamiania to sam rezultat działania może być błędny.

Błędy logiczne powodują otrzymanie wyników innych niż się spodziewano. Są to zwykle błędy bardzo trudne do odnalezienia.

Języki programowania

Dla początkujących, problemem jest mnogość dostępnych obecnie języków programowania. Najbardziej klasyczne języki programowania to **C** i **C++**, popularne są **Java** i **C#**, modny jest **Python**, **JavaScript**, **TypeScript**, **Go**.

Analitycy często używają jżyków pozwalających na szybkie pisanie aplikacji np. **Visual Basic** i eksploracji baz danych: **SQL**, czy dedykowane do analizy danych **R** lub do obliczeń numerycznych (analizy danych też) **MATLAB**, **Scilab**.

- https://www.tiobe.com/tiobe-index/ – Wskaźnik popularności języków programowania.
- https://insights.stackoverflow.com/survey/ – Ankiety serwisu społecznościowego programistów.

Paradygmat programowania

Paradygmat programowania – wzorzec programowania komputerów, który definiuje sposób patrzenia programisty na przepływ sterowania i wykonywanie programu komputerowego.

Przykładowo, w programowaniu **obiektowym** jest on traktowany jako zbiór współpracujących ze sobą obiektów, podczas gdy w programowaniu **funkcyjnym** definiujemy, co trzeba wykonać, a nie w jaki sposób.

Różne języki programowania mogą wspierać różne paradygmaty programowania. Przykładowo, **Smalltalk** i **Java** są ściśle zaprojektowane dla potrzeb programowania obiektowego, natomiast **Haskell** jest językiem funkcyjnym. Istnieją także języki wspierające kilka paradygmatów, np. **Python**.

Wiele paradygmatów jest dobrze znanych z tego, jakie praktyki są w nich zakazane, a jakie dozwolone.

Na przykład, ścisłe programowanie funkcyjne nie pozwala na tworzenie skutków ubocznych (dowolny efekt wyrażenia, lub wywołania funkcji, który wykracza poza zwrócenie wartości). W programowaniu strukturalnym nie korzysta się z instrukcji skoku.

Zależności między paradygmatami programowania mogą przybierać skomplikowane formy, ponieważ jeden język może wspierać wiele różnych paradygmatów. Na przykład, **C++** posiada elementy programowania proceduralnego, obiektowego oraz uogólnionego, co stanowi o nim, że jest hybrydowym językiem. To projektanci i programiści decydują, jak zbudować z nich w pełni działający program.

Przykłady paradygmatów programowania:

- programowanie **imperatywne**,
- programowanie **deklaratywne**,
- programowanie **proceduralne**,
- programowanie **strukturalne**,
- programowanie funkcyjne,
- programowanie **obiektowe**,
- programowanie uogólnione,
- programowanie **sterowane zdarzeniami**,
- programowanie logiczne,
- programowanie aspektowe,
- programowanie agentowe,
- programowanie modularne.

Programowanie imperatywne – paradygmat programowania, który opisuje proces wykonywania jako sekwencję instrukcji zmieniających stan programu, podobnie jak tryb rozkazujący, wyraża żądania jakichś czynności do wykonania.

Programy imperatywne składają się z ciągu komend do wykonania przez komputer. Rozszerzeniem (w sensie wbudowanych funkcji) i rodzajem (w sensie paradygmatu) programowania imperatywnego jest programowanie proceduralne.

Programowanie deklaratywne – rodzina paradygmatów programowania, które nie są z natury imperatywne. W przeciwieństwie do programów napisanych imperatywnie, programista opisuje warunki, jakie musi spełniać końcowe rozwiązanie (co chcemy osiągnąć), a nie szczegółową sekwencję kroków, które do niego prowadzą (jak to zrobić).

Przykłady języków: **XSLT**, **Prolog**.

Programowanie proceduralne to paradygmat programowania zalecający dzielenie kodu na procedury, czyli fragmenty wykonujące ściśle określone operacje.

Procedury nie powinny korzystać ze zmiennych globalnych (w miarę możliwości), lecz pobierać i przekazywać wszystkie dane (czy też wskaźniki do nich) jako parametry wywołania.

Programowanie strukturalne to paradygmat programowania zalecający hierarchiczne dzielenie kodu na bloki, z jednym punktem wejścia i jednym lub wieloma punktami wyjścia.

Chodzi przede wszystkim o nieużywanie instrukcji skoku. Dobrymi strukturami są np. instrukcje: warunkowe, pętle, wyboru.

Strukturalność zakłócają instrukcje typu: `break`, `continue`, `switch`, które jednak w niektórych przypadkach znacząco podnoszą czytelność kodu.

Praktycznie w każdym języku można programować strukturalnie, jednakże w niektórych jest to styl naturalny (np. **Pascal**).

Programowanie obiektowe (*Object-Oriented Programming*) – paradygmat programowania, w którym programy definiuje się za pomocą obiektów – elementów łączących stan (czyli dane, nazywane polami lub właściwościami) i zachowanie (czyli procedury, tu: metody). Obiektowy program komputerowy wyrażony jest jako zbiór takich obiektów, komunikujących się pomiędzy sobą w celu wykonywania zadań.

Podejście to różni się od tradycyjnego programowania proceduralnego, gdzie dane i procedury nie są ze sobą bezpośrednio związane. Programowanie obiektowe ma ułatwić pisanie, konserwację i wielokrotne użycie programów lub ich fragmentów.

Największym atutem programowania, projektowania oraz analizy obiektowej jest zgodność takiego podejścia z rzeczywistością – mózg ludzki jest w naturalny sposób najlepiej przystosowany do takiego podejścia przy przetwarzaniu informacji. Przykłady języków: **C++**, **JAVA**.

Programowanie sterowane zdarzeniami – metodologia tworzenia programów komputerowych, która określa sposób ich pisania z punktu widzenia procesu przekazywania sterowania między poszczególnymi modułami tej samej aplikacji.

Programowanie sterowane zdarzeniami jest mocno powiązane ze środowiskami wieloprocesowymi, z graficznymi środowiskami systemów operacyjnych oraz z programowaniem obiektowym.

Paradygmat zakłada, że program jest cały czas bombardowany zdarzeniami, na które musi odpowiedzieć, i że przepływ sterowania w programie jest całkowicie niemożliwy do przewidzenia z góry.

Programowanie zdarzeniowe jest dominującym typem programowania związanego z graficznym interfejsem użytkownika (*Graphical User Interface*) – zdarzenia to naciśnięcia myszy, klawiszy, żądania odświeżenia przez system okienkowy, różne zdarzenia sieciowe i inne.

W programowaniu zdarzeniowym ważne jest żeby nie obsługiwać zbyt długo danego zdarzenia, bo blokuje się w ten sposób obsługę innych. W przypadku serwerów obniżyło by to znacznie wydajność, w przypadku GUI program zbyt wolno odpowiadałby na akcje użytkownika.

Można to osiągnąć za pomocą asynchronicznego I/O, wielowątkowości, rozbijania zdarzenia na podzdarzenia i wielu innych mechanizmów.

Program komputerowy działa według określonego algorytmu.

Algorytm – ciąg jasno zdefiniowanych czynności, koniecznych do wykonania pewnego rodzaju zadania.

Zapis algorytmu działania w wybranym języku programowania nazywamy **implementacją algorytmu**.

Jako przykład stosowanego w życiu codziennym algorytmu podaje się często przepis kulinarny. Dla przykładu, aby ugotować bigos należy w określonej kolejności oraz odstępach czasowych (imperatyw czasowy) dodawać właściwe rodzaje kapusty i innych składników.

Przykład ten ma wyłącznie charakter poglądowy, ponieważ język przepisów kulinarnych nie został jasno zdefiniowany. Algorytmy zwykle formułowane są w sposób ścisły w oparciu o język matematyki.

Algorytm prowadzi do rozwiązania zadania w **skończonej liczbie kroków**.

Do danego celu prowadzi zwykle więcej niż jedna droga. Jak więc oceniać alternatywne sposoby rozwiązania problemu?

Podstawowe parametry algorytmu to jego **złożoność czasowa** i **złożoność pamięciowa**. Oprócz tego przy algorytmach działających na liczbach trzeba pamiętać o **stabilności numerycznej**.

Złożoność czasowa mówi, ile kroków obliczeniowych i ile czasu wymaga zakończenie algorytmu dla danej porcji danych.

Złożoność pamięciowa mówi, jaką maksymalnie część danych i wyników pośrednich trzeba w ramach danego algorytmu przechowywać w pamięci operacyjnej.

Stabilność numeryczna określa wrażliwość wyniku końcowego na błędy zaokrągleń w trakcie obliczeń oraz na dokładność danych początkowych.

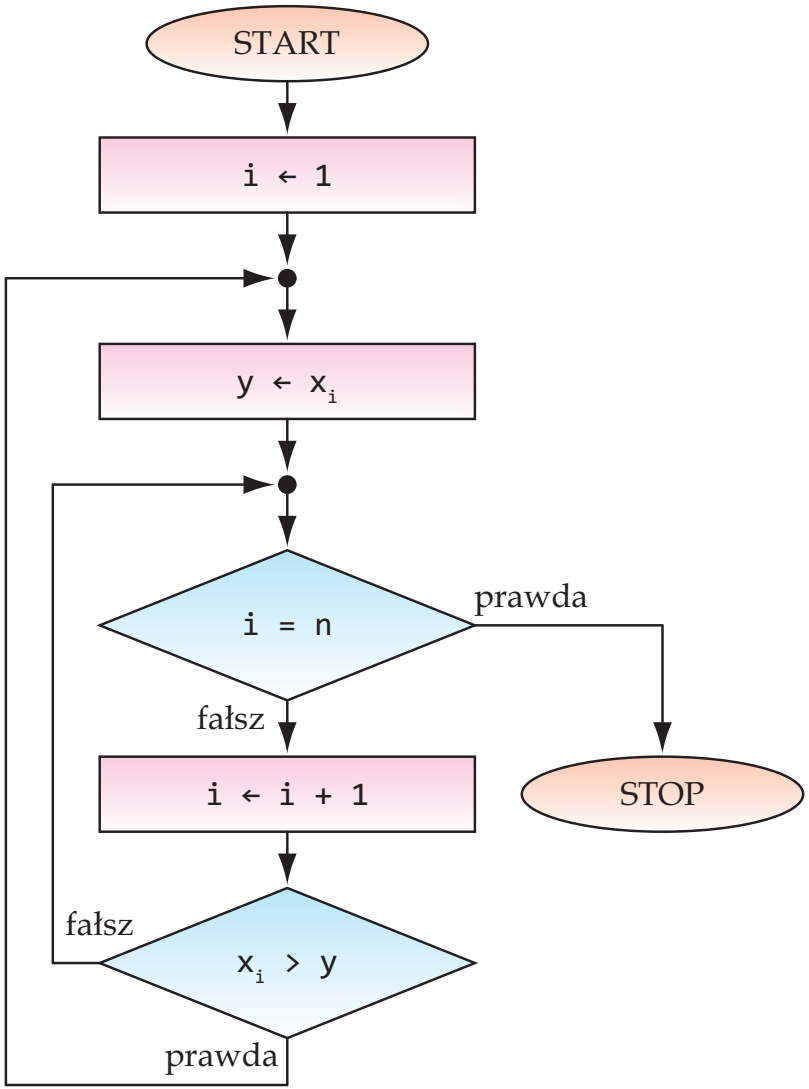
Sposoby zapisu algorytmów

Algorytm znajdowania wartości $y = \max\{x_i\}$, gdzie $1 \leqslant i \leqslant n$.

- **Język naturalny**.
- Schemat blokowy.
- Język formalny.

1. $i \leftarrow 1$, idź do 2,
2. $y \leftarrow x_i$, idź do 3,
3. Czy $i = n$? Jeśli tak – koniec, jeśli nie – idź do 4,
4. $i \leftarrow 1 + 1$, idź do 5,
5. Czy $x_i > y$? Jeśli tak – idź do 2, jeśli nie – idź do 3.

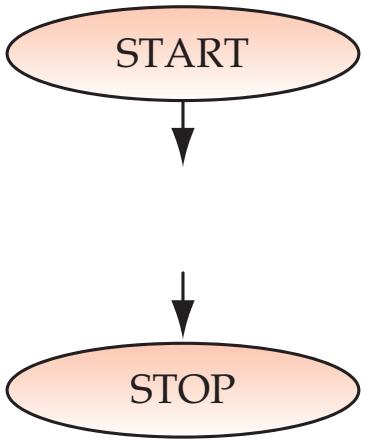
- Język naturalny.
- **Schemat blokowy**.
- Język formalny.



- Język naturalny.
- Schemat blokowy.
- **Język formalny**, C.

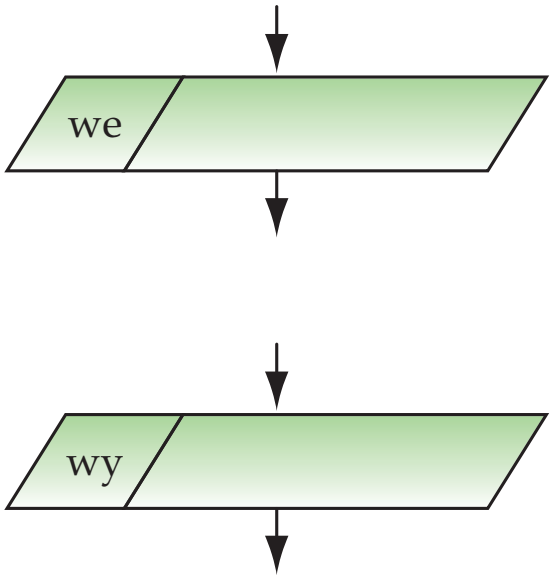
```c
int maxValue(int array[], int size)
{
  int i, max;

  max=array[0];

  for(i=1; i<size; i++)
    {
      if(array[i] > max)
        max = array[i];
    }

   return max;
}
```

Schematy blokowe



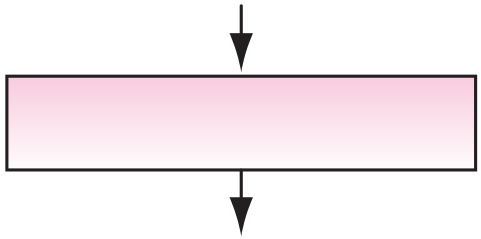
Start – początek programu

Stop – koniec programu

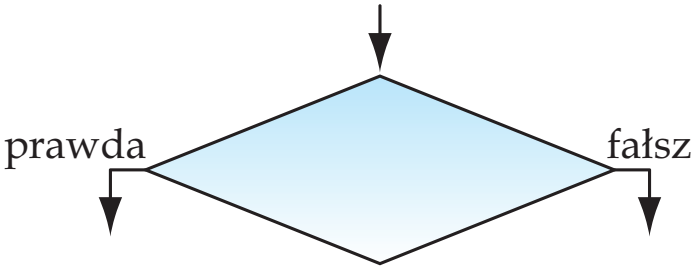


Blok wejścia – wprowadzanie danych

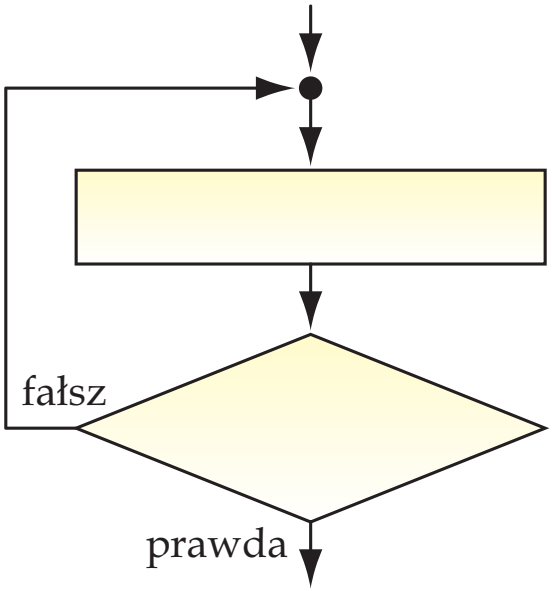
Blok wyjścia – wyprowadzanie wyników



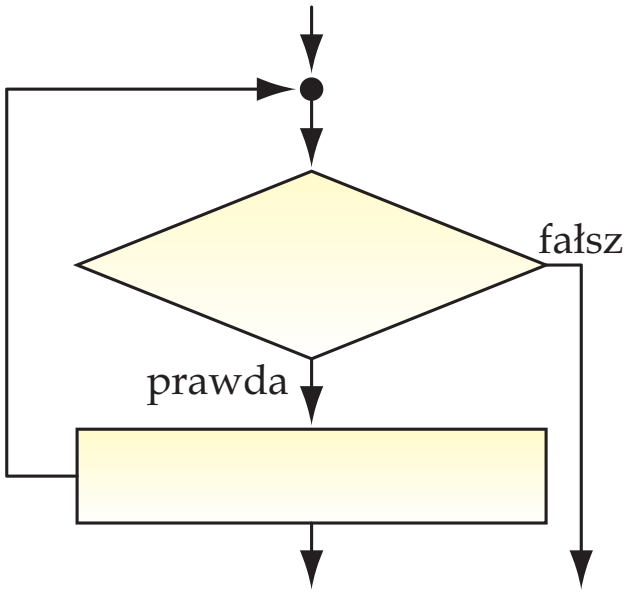
Blok operacyjny – przetwarzanie danych



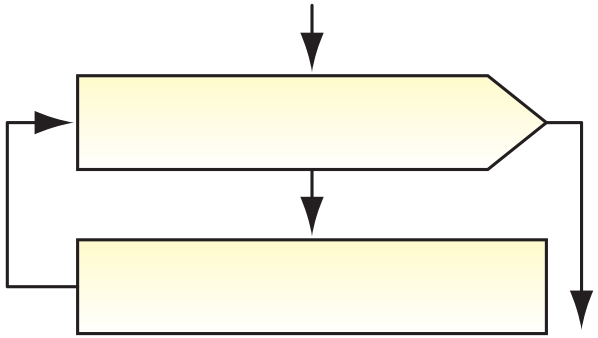
Blok decyzyjny – instrukcja warunkowa



Pętla *powtórz* – z warunkiem po bloku operacyjnym



Pętla *dopóki* – z warunkiem przed blokiem operacyjnym

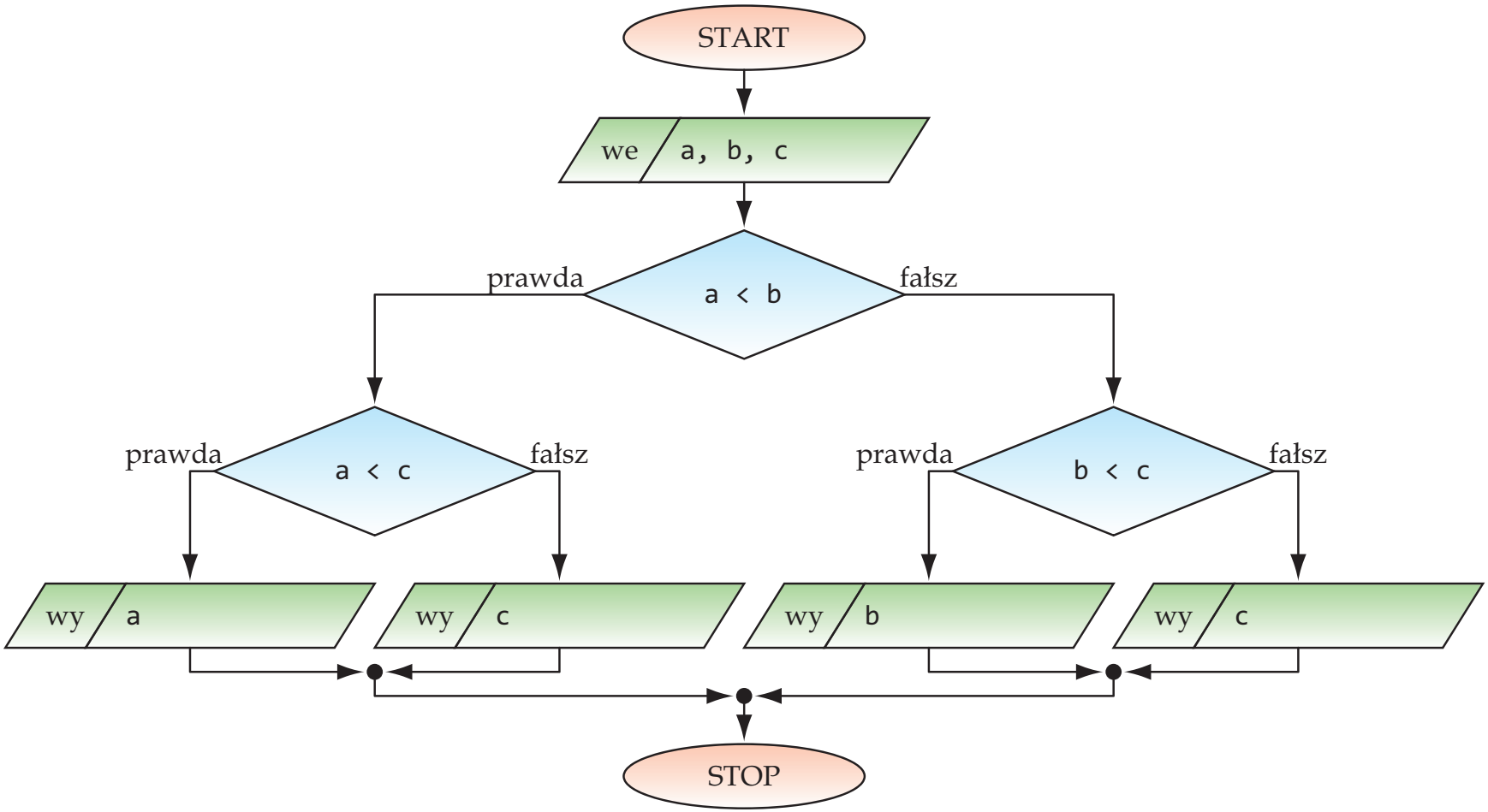


Pętla *dla* – z wiadomą liczbą iteracji

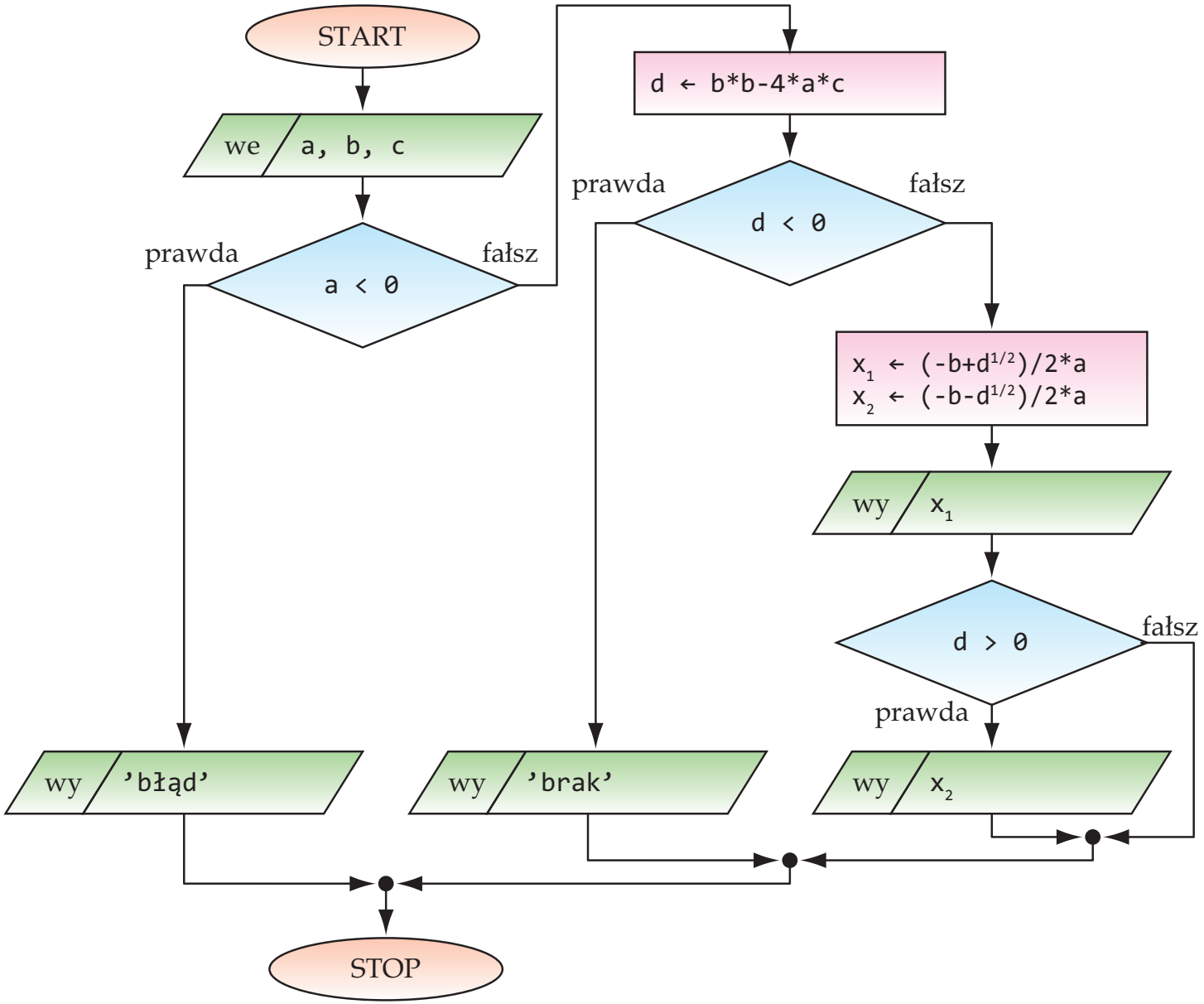
Algorytmy można podzielić na:

- algorytmy liniowe, jeśli wykorzystują operacje bezpośredniego następstwa,
- algorytmy warunkowe, jeśli wykorzystują operacje warunkowe,
- algorytmy iteracyjne, jeśli wykorzystują pętle.

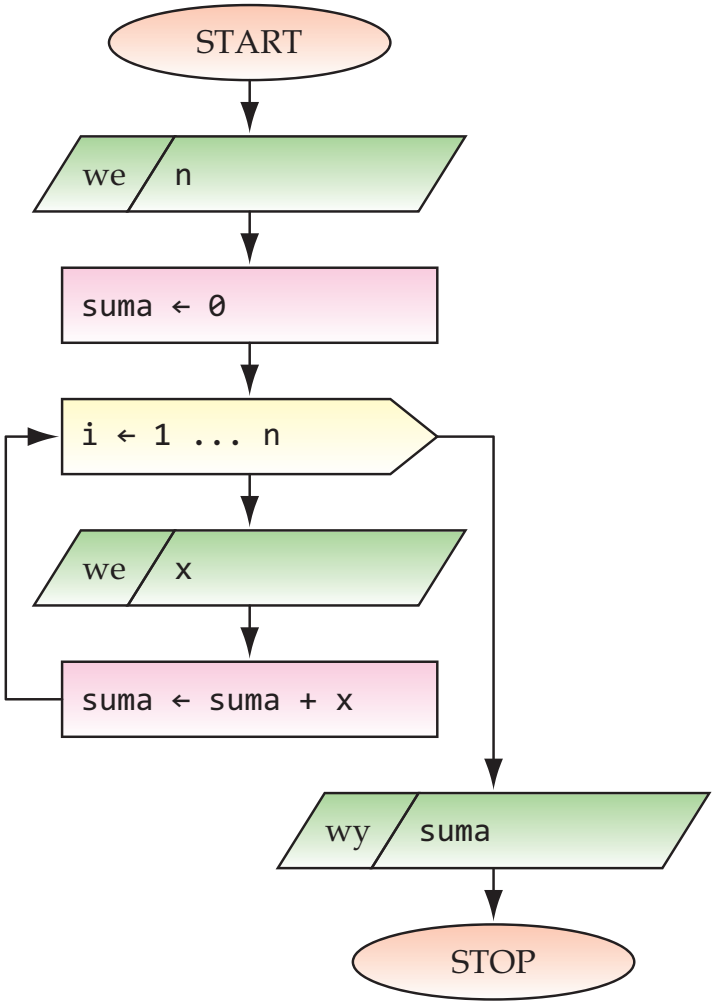
Algorytm szukania wartości minimalnej z trzech



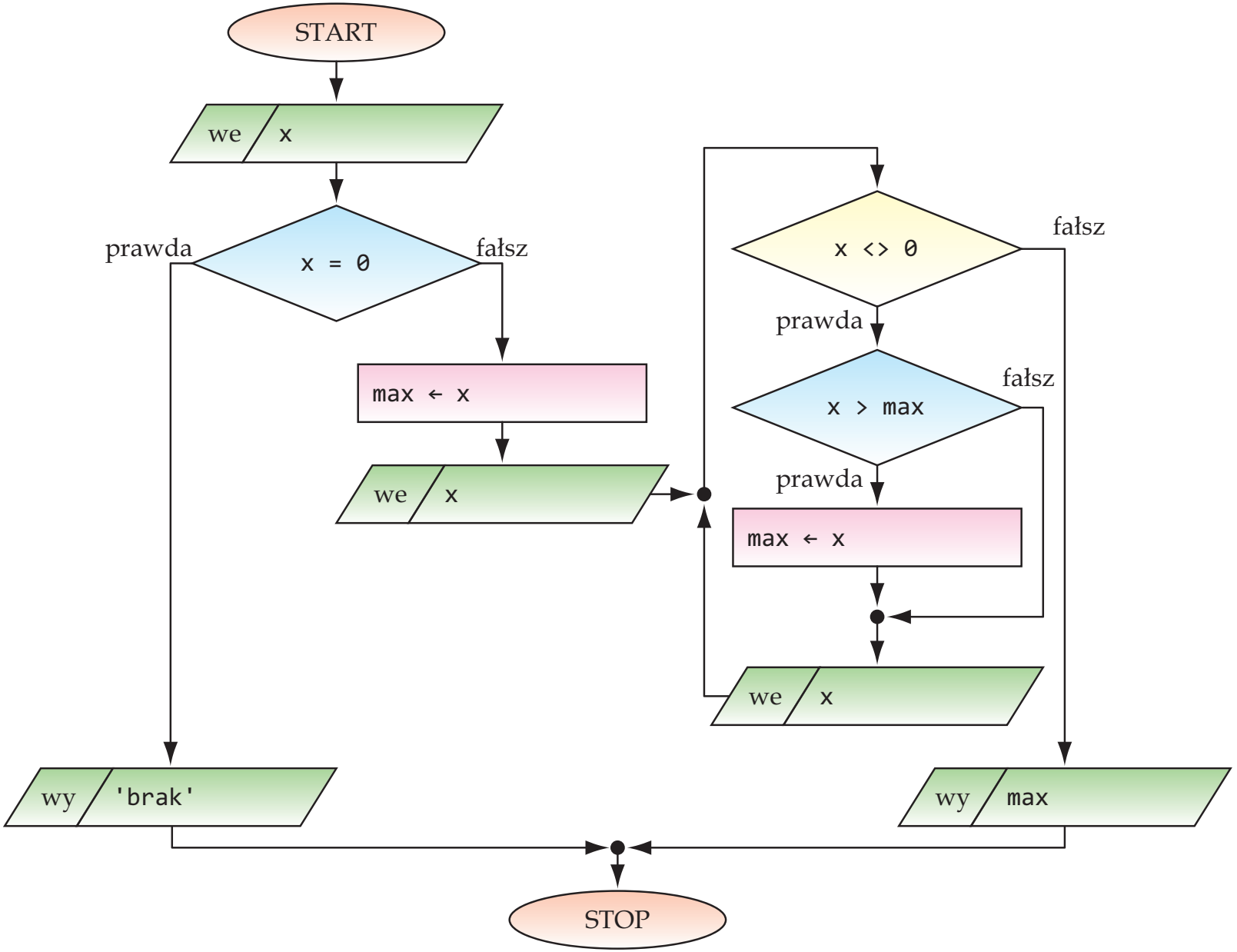
Algorytm obliczania pierwiastków równania kwadratowego



Algorytm sumowania



Algorytm szukania wartości ekstremalnej

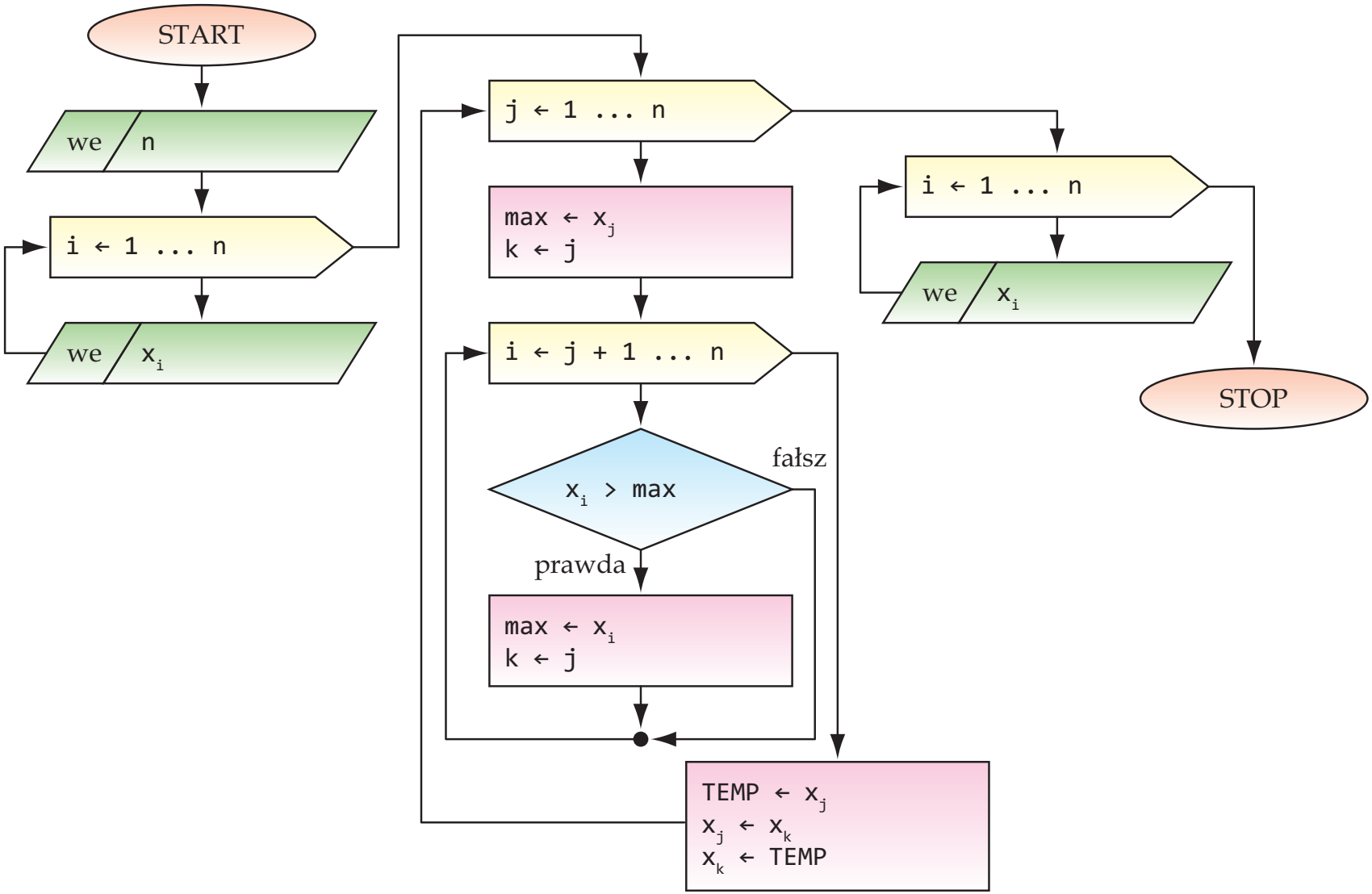


Algorytm sortowania przez wybór

Jest to chyba najbardziej intuicyjny algorytm sortowania. Polega on na wielokrotnym wyborze minimalnego elementu z coraz krótszego podciągu danych.

Przebieg:

- wybieranie minimum z ciągu elementów na pozycjach od 1 do n i zamienianie go z pierwszym elementem,
- wybieranie minimum z ciągu elementów na pozycjach od 2 do n i zamienianie go z drugim elementem (po tym kroku elementy na pozycjach od 1 do 2 są uporządkowane),
- wybieranie minimum z ciągu elementów na pozycjach $n-1$ i n i zamienianie z elementem na pozycji $n-1$ (po tej operacji elementy na pozycjach od 1 do $n-1$ są uporządkowane, a element na pozycji n jest maksymalny, czyli ciąg elementów na pozycjach od 1 do n jest uporządkowany).


START
we n
i ← 1 ... n
we x_i
j ← 1 ... n
max ← x_j
k ← j
i ← j + 1 ... n
x_i > max
fałsz
prawda
max ← x_i
k ← j
TEMP ← x_j
x_j ← x_k
x_k ← TEMP
i ← 1 ... n
we x_i
STOP

Podsumowanie

Budując algorytmy należy przestrzegać następujących zasad:

- algorytmy trzeba rozłożyć na jak najprostrze operacje podstawowe,
- każdy krok algorytmu dokładnie określić i przemyśleć,
- rozważyć wszystkie możliwe przypadki funkcjonowania algorytmu w zależności od danych wejściowych oraz operacji wewnętrznych,
- każdy algorytm powinien prowadzić do rozwiązania generując jedną lub wiele odpowiedzi na zadany problem.

Algorytm zdefiniowany z wykorzystaniem powyższych zasad powinien mieć jasno określony punkt startowy i punkt końcowy.

Metoda numeryczna

Jest to metoda, która umożliwia sprowadzenie najrozmaitszych obliczeń matematycznych, z dowolnej dziedziny ludzkiej działalności, do wykonywania skończonej liczby **najprostszych działań arytmetycznych** (czyli idealne do obliczeń komputerowych).

Skuteczne rozwiązywanie jakiegokolwiek problemu czy zadania, za pomocą odpowiedniej metody numerycznej, jest możliwe tylko wtedy, **gdy a priori wiadomo, że poszukiwane rozwiązanie istnieje i jest jedyne** (jednoznaczne).

Obliczenia numeryczne z reguły prowadzą do **wyników przybliżonych** i do nas należy ocena jakości rozwiązania, według przez nas wybranych kryteriów.

Przyczyny powstawania błędów w obliczeniach numerycznych:

- **błąd reprezentacji** – wynika ze skończonej reprezentacji liczby w komputerze,
- **błąd zaokrąglenia** – powstaje wtedy, gdy zapamiętanie dokładnego wyniku pewnego działania arytmetycznego wymaga więcej miejsca w pamięci niż jest to dla niego przeznaczone,
- **błąd redukcji** – występuje przy dodawaniu liczb różniących się znakami, o dokładnie tych samych cechach i niewiele różniących się mantysach,
- **błędy modelu** – wynikają z przybliżenia rzeczywistości przez pewne modele teoretyczne (numeryczne),
- **błędy obcięcia** – np.: obliczanego szeregu nieskończonego.

Z błędami redukcji wiąże się problem dokładności reprezentacji liczby za pomocą skończonej liczby cyfr.

Będziemy nazywać **liczbą cyfr znaczących** tę liczbę cyfr występujących z zapisie liczby, która pozostaje po odrzuceniu tzw. lewych zer (czyli występujących na lewo od pierwszej cyfry niezerowej w zapisie liczby).

$$0{,}00033333$$

$$0{,}33333$$

$$333{,}33$$

$$33333{,}$$

Każda z powyższych liczb ma liczbę cyfr znaczących równą LCZ $= 5$. LCZ decyduje o **dokładności (względnej) reprezentacji liczb**, a co za tym idzie od dokładności, z jaką wykonywane są operacje arytmetyczne przez konkretny komputer (czyli zależy od wielkości mantysy liczb zapamiętanych w komórkach jego pamięci).

Algorytmy standardowych metod numerycznych

Metody aproksymacji i interpolacji funkcji:

- metoda najmniejszych kwadratów,
- interpolacja Lagrange'a.

Całkowanie numeryczne:

- kwadratury Newtona–Cotesa (trapezów, Simpsona),
- złożone kwadratury całkowania,
- kwadratura Gaussa.

Metody rozwiązywania równań nieliniowych:

- metoda bisekcji (połowienia),
- metoda Newtona (stycznych),
- metoda siecznych.

Metody rozwiązywania układów równań liniowych:

- metody bezpośrednie (metoda eliminacji Gaussa, metoda Choleskiego),
- metody iteracyjne (metoda Jacobiego, metoda Gauss-Seidela).

Metody rozwiązywania problemów początkowych i brzegowych:

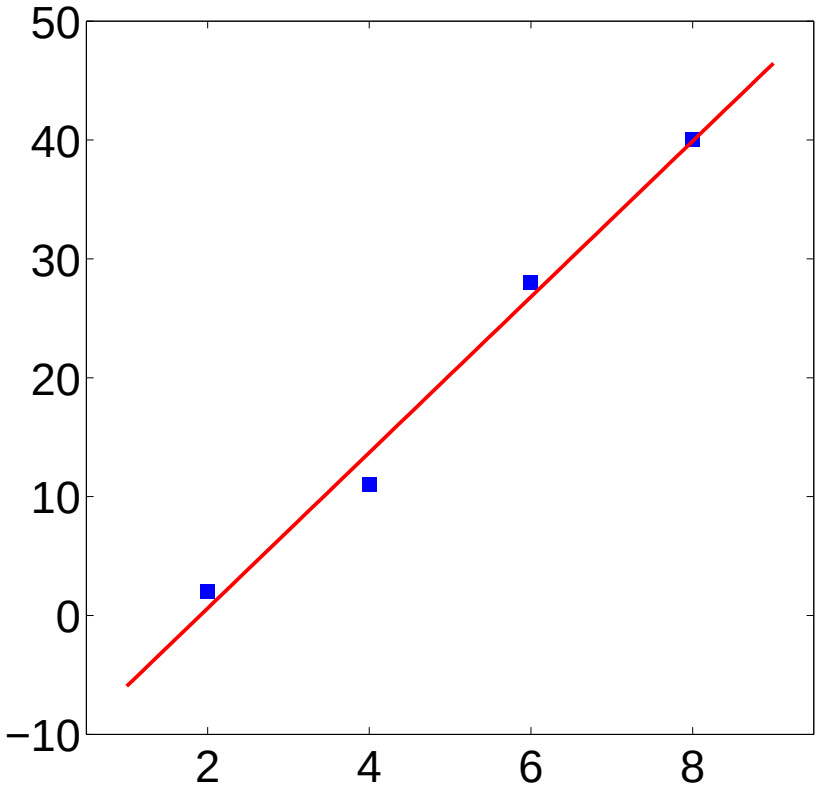
- metoda Eulera,
- metoda Rungego-Kutty,
- metoda różnic skończonych,
- metoda elementów skończonych,
- metoda elementów brzegowych.

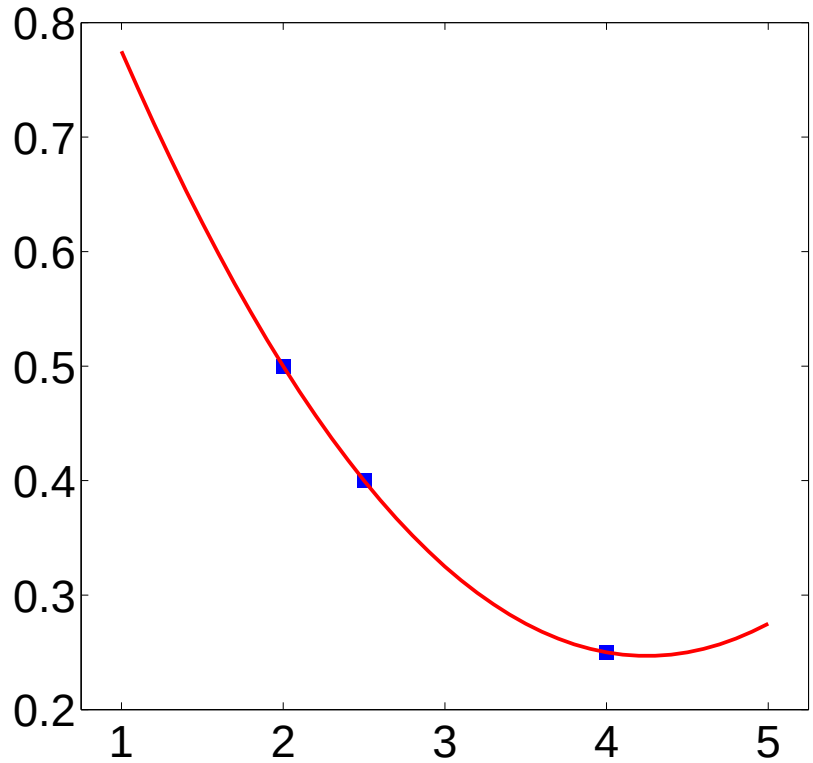
Aproksymacja i interpolacja

Aproksymacją nazywamy procedurę zastępowania jednej funkcji (funkcja aproksymowana) inną funkcją (funkcja aproksymująca) w taki sposób, aby funkcje te niewiele się różniły w sensie określonej normy.

Jeżeli dana funkcja określona jest w całym przedziale, w którym dokonujemy aproksymacji, mamy do czynienia z aproksymacją **funkcji ciągłej**, natomiast, gdy funkcja dana określona jest wyłącznie na skończonym zbiorze punktów zwanych **węzłami**, mówimy o aproksymacji **funkcji dyskretnej**. Jest to częsty przypadek w praktycznych zastosowaniach inżynierskich. Funkcja aproksymowana zwykle reprezentującą dane pomiarowe znane dla n różnych wartości zmiennej niezależnej.

Celem aproksymacji jest więc wyznaczenie funkcji (zwykle ciągłej), która będzie przebiegała w pobliżu danych punktów. W szczególnym przypadku funkcja aproksymująca może przebiegać przez niektóre z tych punktów. Jeżeli natomiast funkcja będzie przebiegać przez wszystkie podane punkty, będziemy mieć do czynienia z **interpolacją**.





Przyczyną stosowania aproksymacji mogą być między innymi:

- chęć zastąpienia funkcji niedogodnej do obliczeń numerycznych inną, dogodniejszą funkcją, która będzie niewiele odbiegać od funkcji wyjściowej (taką funkcją może być odpowiednio dobrany wielomian),
- potrzeba wyznaczenia wartości funkcji danej dyskretnie w innym punkcie obszaru,
- konieczność znalezienia dostatecznie gładkiej funkcji ciągłej przechodzącej w pobliżu zadanych punktów.

Dyskretna aproksymacja w bazie jednomianów

Rozwiązanie problemu aproksymacji danej dyskretnej funkcji $f(x_i)$, gdzie $i = 0, 1, \ldots, n$ wymaga przyjęcia bazy funkcyjnej aproksymacji (tutaj bazy jednomianów) i zapisaniu funkcji aproksymującej ($k = 0, 1, \ldots, m$)

$$P_m(x) = a_0 \cdot x^0 + a_1 \cdot x^1 + a_2 \cdot x^2 + \ldots + a_m \cdot x^m = \sum_{k=0}^{m} a_k \cdot x^k,$$

oraz na ustaleniu kryterium oceny jakości aproksymacji, które posłuży do jednoznacznego określenia wartości a_k, którego najprostsza postać jest następująca (metoda najmniejszych kwadratów):

$$r = \sum_{i=0}^{n} |f(x_i) - P_m(x_i)|^2.$$

Wartość funkcji r jest pewną miarą odchylenia funkcji aproksymującej $P_m(x)$ od aproksymowanej $f(x)$ (błąd aproksymacji). W przypadku aproksymacji $m < n$, wówczas $r > 0$ i osiąga minimum, gdy:

$$\frac{\partial r}{\partial a_k} = 0, \quad k = 0, 1, \ldots, m.$$

Warunków tych jest $m + 1$, a więc tyle ile niewiadomych współczynników wielomianu aproksymującego a_k.

Obliczając pochodne cząstkowe funkcji r otrzymamy

$$\frac{\partial}{\partial a_k} \left(\sum_{i=0}^{n} (f(x_i) - P_m(x_i))^2 \right) = 0,$$

$$2 \sum_{i=0}^{n} (f(x_i) - P_m(x_i)) \frac{\partial}{\partial a_k} P_m(x_i) = 0, \quad k = 0, 1, \ldots, m.$$

I po pewnych przekształceniach otrzymamy układ równań normalnych

$$a_0 \sum_{i=0}^{n} x_i^0 + a_1 \sum_{i=0}^{n} x_i^1 + a_2 \sum_{i=0}^{n} x_i^2 + \ldots + a_m \sum_{i=0}^{n} x_i^m = \sum_{i=0}^{n} f(x_i) x_i^0,$$

$$a_0 \sum_{i=0}^{n} x_i^1 + a_1 \sum_{i=0}^{n} x_i^2 + a_2 \sum_{i=0}^{n} x_i^3 + \ldots + a_m \sum_{i=0}^{n} x_i^{m+1} = \sum_{i=0}^{n} f(x_i) x_i^1,$$

$$\ldots$$

$$a_0 \sum_{i=0}^{n} x_i^m + a_1 \sum_{i=0}^{n} x_i^{m+1} + a_2 \sum_{i=0}^{n} x_i^{m+2} + \ldots + a_m \sum_{i=0}^{n} x_i^{2m} = \sum_{i=0}^{n} f(x_i) x_i^m.$$

Przykład

Obliczmy liniowy wielomian aproksymacyjny $P_1 = a_0 + a_1 \cdot x$ dla danych:

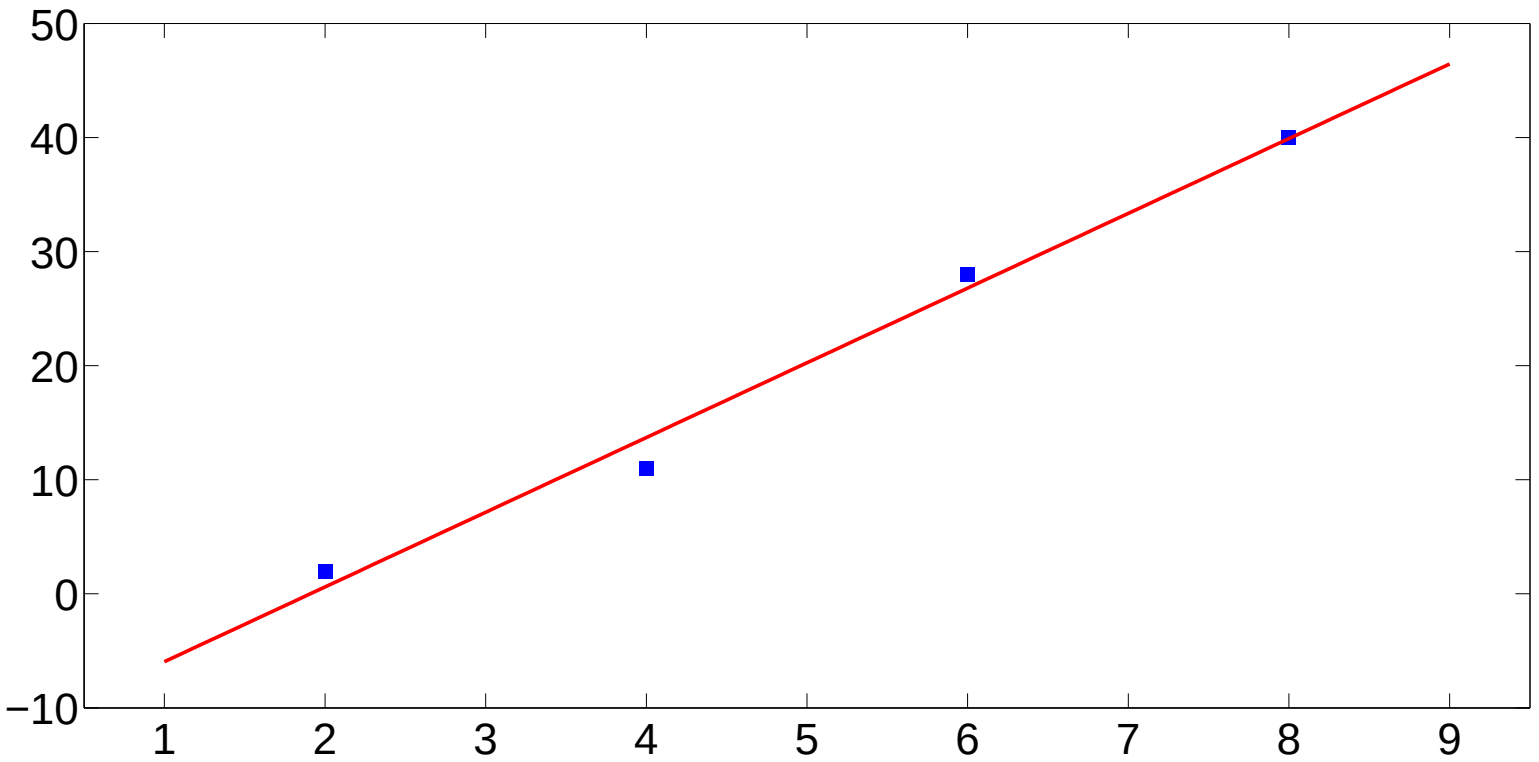
i	x_i	$f(x_i)$
0	2	2
1	4	11
2	6	28
3	8	40

W tym przypadku $n = 3$ i $m = 1$. Wykorzystując układ równań normalnych, możemy napisać

$$\begin{cases} 4a_0 + 20a_1 = 81 \\ 20a_0 + 120a_1 = 563 \end{cases}.$$

Rozwiązanie układu prowadzi do końcowego rozwiązania: $a_0 = -12,5$ i $a_1 = 6,55$. Wielomian aproksymacyjny wynosi

$$P_1(x) = -12,5 + 6,55x.$$



Interpolacja Lagrange'a

Interpolacja może być traktowana jako szczególny przypadek aproksymacji, polegający na tym, że funkcja aproksymowana i funkcja aproksymująca przyjmują te same wartości w wybranych punktach obszaru zwanych węzłami.

Funkcję interpolującą $P_n(x)$, najczęściej przedstawia się w postaci **wielomianu uogólnionego** utworzonego z odpowiednio dobranych funkcji bazowych

$$P_n(x) = a_0 \cdot \varphi_0(x) + a_1 \cdot \varphi_1(x) + a_2 \cdot \varphi_2(x) + \ldots + a_n \cdot \varphi_n(x) = \sum_{i=0}^{n} a_i \cdot \varphi_i(x).$$

Funkcje bazowe wielomianu interpolującego muszą być liniowo niezależne. Korzystając z warunku, że wartości funkcji interpolowanej i interpolującej mają takie same wartości, możliwe jest obliczenie współczynników funkcji interpolującej.

Najczęściej stosowaną i najprostszą bazą, jest baza funkcji jednomianowych

$$\varphi_n(x) = \left[x^0, x^1, x^2, \ldots, x^n\right],$$

i wtedy wielomian interpolacyjny będzie miał postać

$$P_n(x) = a_0 \cdot x^0 + a_1 \cdot x^1 + a_2 \cdot x^2 + \ldots + a_n \cdot x^n = \sum_{i=0}^{n} a_i \cdot x^i.$$

Wspomniana wcześniej procedura poszukiwania funkcji interpolującej jest w przypadku ogólnym dość złożona obliczeniowo: rozwiązywanie układu równań w celu określenia współczynników a_i; a współczynniki a_i zwykle nie mają interpretacji fizycznej.

Dlatego też funkcję interpolującą można zapisać w sposób

$$P_n(x) = f_0 \cdot N_{n,0}(x) + f_1 \cdot N_{n,1}(x) + f_2 \cdot N_{n,2}(x) + \ldots + f_n \cdot N_{n,n} = \sum_{i=0}^{n} f_i \cdot N_{n,i}(x),$$

z tak zwanymi funkcjami bazowymi Lagrange,a o postaci

$$N_{n,i}(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} = \frac{(x - x_0)(x - x_1)\ldots(x - x_{i-1})(x - x_{x+1})\ldots(x - x_n)}{(x_i - x_0)(x_i - x_1)\ldots(x_i - x_{i-1})(x_i - x_{i+1})\ldots(x_i - x_n)}.$$

Warto zauważyć, że wartości współczynników, poprzednio oznaczonych przez a_i to teraz wartości funkcji interpolowanej f_i.

Przykład

Wyprowadźmy wzór interpolacyjny Lagrange'a stopnia drugiego przybliżający funkcję $f(x) = \frac{1}{x}$ i przyjmując węzły interpolacji: $x_0 = 2$, $x_1 = 2,5$ i $x_3 = 4$.

Wielomian będzie miał postać

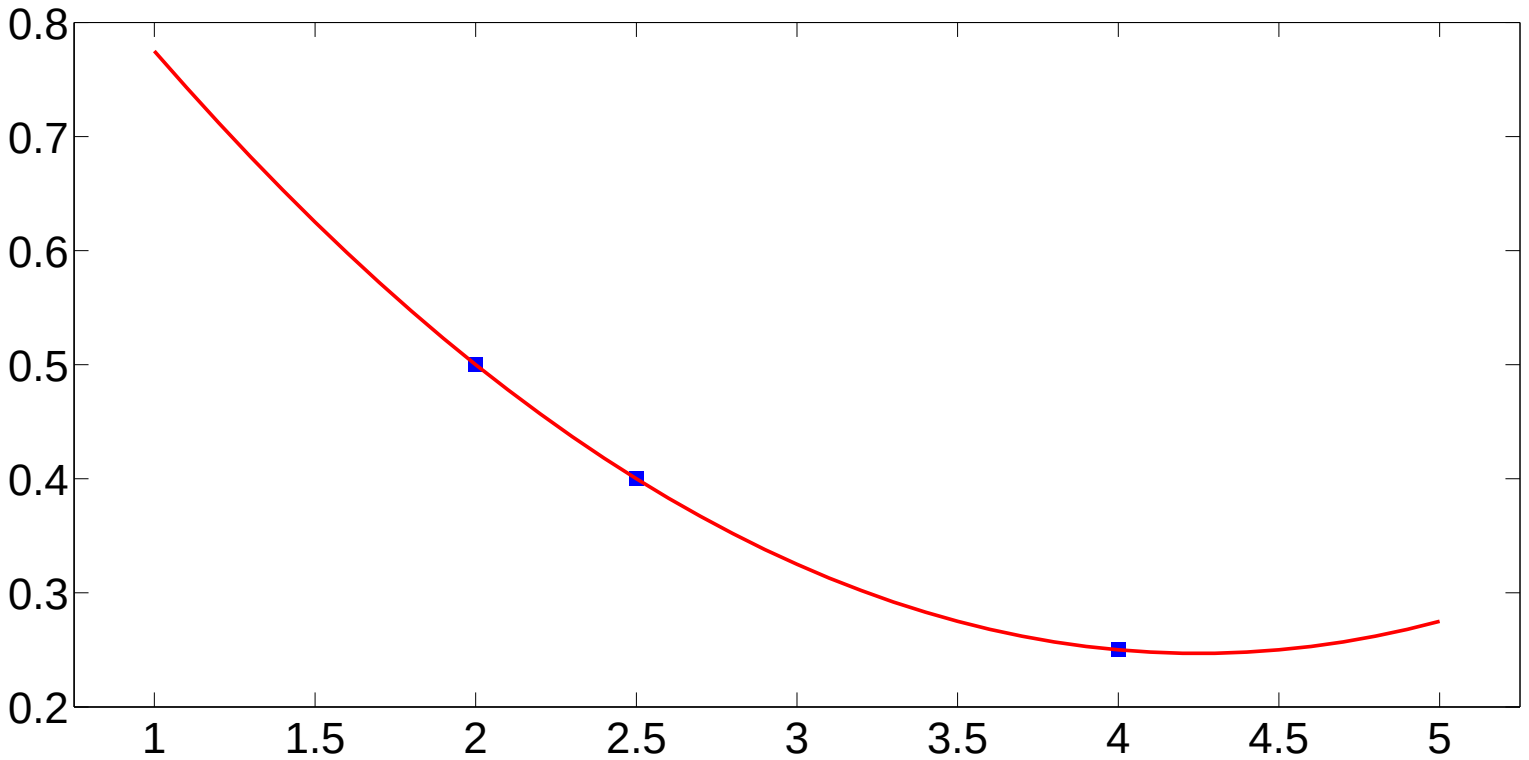
$$P_2(x) = f_0 \cdot N_{2,0}(x) + f_1 \cdot N_{2,1}(x) + f_2 \cdot N_{2,2}(x),$$

Wartości funkcji w węzłach interpolacji wynoszą odpowiednio: $f_0 = 0,5$, $f_1 = 0,4$ i $f_3 = 0,25$. Obliczmy funkcje bazowe

$$N_{2,0} = \frac{(x - 2,5)(x - 4)}{(2 - 2,5)(2 - 4)},$$
$$N_{2,1} = \frac{(x - 2)(x - 4)}{(2,5 - 2)(2,5 - 4)},$$
$$N_{2,2} = \frac{(x - 2)(x - 2,5)}{(4 - 2)(4 - 2,5)}.$$

Podstawienie funkcji bazowych i uproszczenie prowadzi do wielomianu interpolacyjnego Lagrange'a w postaci

$$P_2(x) = 0,5\frac{(x-2,5)(x-4)}{(2-2,5)(2-4)} + 0,4\frac{(x-2)(x-4)}{(2,5-2)(2,5-4)} + 0,25\frac{(x-2)(x-2,5)}{(4-2)(4-2,5)} = \\ = 1,15 - 0,425x + 0,05x^2.$$



W przypadkach elementarnych obliczenie wartości całki oznaczonej odbywa się na podstawie wzoru Newtona-Leibnitza

$$I(f) = \int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

Powyższy wzór można stosować wtedy, gdy znana jest tzw. **funkcja pierwotna** $F(x)$ spełniająca warunek

$$\frac{dF(x)}{dx} f(x).$$

Jeśli wyznaczenie funkcji pierwotnej jest bardzo trudne lub niemożliwe, albo też jeśli f(x) dane jest w postaci dyskretnej, to wartość całki można obliczyć tylko numerycznie stosując **wzory kwadraturowe** (kwadratury).

Idea postępowania przy całkowaniu numerycznym jest zawsze taka sama, a mianowicie zastępujemy funkcję podcałkową funkcją łatwą do scałkowania analitycznego (wielomian) w drodze interpolacji, a następnie funkcję interpolującą całkujemy ściśle (analitycznie).

W zależności od sposobu postępowania przy wyborze położeń węzłów interpolacji w przedziale całkowania możemy mieć do czynienia z kwadraturami z:

- **zamkniętymi** końcami (końce przedziału całkowania wchodzą do wzorów), a węzły są rozmieszczone równomiernie,
- **otwartymi** końcami (końce przedziału całkowania nie wchodzą do wzorów), a węzły są rozmieszczone w przedziale całkowania nierównomiernie.

Przykładem kwadratur pierwszego typu są kwadratury Newtona-Cotesa, a kwadratur drugiego typu kwadratury Gaussa.

Stosując funkcje bazowe interpolacji Lagrange'a

$$\int_a^b f(x)\,\mathrm{d}x \approx \int_a^b P_n(x)\,\mathrm{d}x = \int_a^b \sum_{i=0}^{n} f(x_i) N_{n,i}(x)\,\mathrm{d}x$$

wyznaczmy kilka kwadratur całkowania

- **kwadratura prostokątów** czyli interpolacja wielomianowa stopnia 0 ($n = 0$, funkcja stała)

$$\int_a^b f(x)\,\mathrm{d}x \approx \int_a^b f(x_0)\,\mathrm{d}x = (b-a) f(x_0),$$

- **kwadratura trapezów** czyli interpolacja wielomianowa stopnia 1 ($n = 1$, funkcja liniowa)

$$\int_a^b f(x)\,\mathrm{d}x \approx \int_a^b \left(f(x_0)\frac{x-x_1}{x_0-x_1} + f(x_1)\frac{x-x_0}{x_1-x_0} \right)\,\mathrm{d}x = \frac{b-a}{2}(f(x_0)+f(x_1)),$$

- **kwadratura Simpsona** czyli interpolacja wielomianowa stopnia 2 ($n = 2$, funkcja kwadratowa)

$$\int_a^b f(x)\,\mathrm{d}x \approx \int_a^b \left(f(x_0)\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + f(x_1)\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + \right.$$
$$\left. + f(x_1)\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \right) \mathrm{d}x = \frac{b-a}{3}\left(f(x_0) + 4f(x_1) + f(x_2)\right).$$

Przykład

Policzmy całkę

$$I = \int_0^{\frac{1}{2}} \sqrt{1+x}\,\mathrm{d}x = \frac{2}{3}\left(\left(\frac{3}{2}\right)^{\frac{3}{2}} - 1\right) = \frac{1}{2}\sqrt{6} - \frac{2}{3} = 0,5580782.$$

Policzmy przykład jeszcze raz korzystając ze wzorów przybliżonych:

- wzór trapezów

$$I \approx \frac{1}{2}\cdot\frac{1}{2}\left((1+0)^{\frac{1}{2}} + \left(1+\frac{1}{2}\right)^{\frac{1}{2}}\right) = 0,5561862,$$

- wzór Simpsona

$$I \approx \frac{1}{3}\cdot\frac{1}{4}\left((1+0)^{\frac{1}{2}} + 4\left(1+\frac{1}{4}\right)^{\frac{1}{2}} + \left(1+\frac{1}{2}\right)^{\frac{1}{2}}\right) = 0,5580734.$$

Wady kwadratur Newtona-Cotesa:

- duże przedziały całkowania wymagają interpolacji wysokiego stopnia co utrudnia obliczanie współczynników interpolacji,
- wielomiany interpolacyjne wysokiego stopnia z równo rozmieszczonymi węzłami mają wysoki stopień oscylacji na brzegach.

Wyjściem z tej sytuacji są **kwadratury złożone**. Kwadratury złożone otrzymujemy dzieląc przedział całkowania na pewną liczbę równych pod przedziałów i stosując dla każdego z nich tę samą kwadraturę.

Do podwyższenia dokładności wzorów kwadraturowych można zastosować propozycję **Gaussa**, polegającą na optymalizacji położenia węzłów interpolacyjnych, rezygnacji z warunku równomiernego rozmieszczenia węzłów interpolacji oraz doborze odpowiednich wartości współczynników wagowych.

We wzorach kwadraturowych Gaussa wykorzystuje się aproksymację wielomianami Legendre'a. Wartość przybliżoną całki można policzyć korzystając z zależności

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} w_i \cdot f(x_i),$$

gdzie w_i to współczynniki wagowe, a x_i to współrzędne węzłów interpolacji.

Wygodnie jest korzystać z wartości tablicowych współczynników wagowych i współrzędnych węzłów interpolacji, które można znaleźć w każdym podręczniku dotyczącym metod numerycznych.

Rozwiązywanie równań nieliniowych

Kolejnym prostym zastosowaniem metod numerycznych jest rozwiązywanie nieliniowego równania (algebraicznego, trygonometrycznego) o ogólnej postaci

$$f(x) = 0.$$

Przyjmiemy, ze znany jest **przedział izolacji**, w którym znajduje się szukany pierwiastek i jest tylko jeden (w badanym przedziale).

Pierwiastek będziemy liczyć **metodą iteracyjną**, polegającą na powtarzaniu pewnego ustalonego ciągu operacji arytmetycznych. Wymaga to przyjęcia odpowiedniego **przybliżenia początkowego** (punktu startowego).

Obliczenia uznamy za zakończone jeśli osiągniemy odpowiednie kryterium. Mogą to być kryteria związane z **liczbą iteracji** (niezależnie od znalezienia rozwiązania), z tempem zbieżności (zmiana kolejnych przybliżeń rozwiązania) lub z **residuum** (zmiana kolejnych wartości funkcji dla kolejnych przybliżeń rozwiązania).

Metoda bisekcji

Metoda bisekcji należy do najstarszych i naiprostszych metod iteracyjnych, oprócz znajdowania pierwiastków równania jest wykorzystywana przy zagadnieniach optymalizacji funkcji.

Dla wyjściowego równania $f(x) = 0$ szuka przybliżenia w przedziale $x \in (a, b)$. Stąd, aby metoda zadziałała, musi być gwarancja istnienia miejsca zerowego w tym przedziale: $f(a) \cdot f(b) < 0$.

Przy każdej iteracji oblicza się środek przedziału $x = \frac{a+b}{2}$. Następnie sprawdza się, w którym z podprzedziałów (a, x) oraz (x, b) leży miejsce zerowe i ten podprzedział podlega dalszemu dzieleniu.

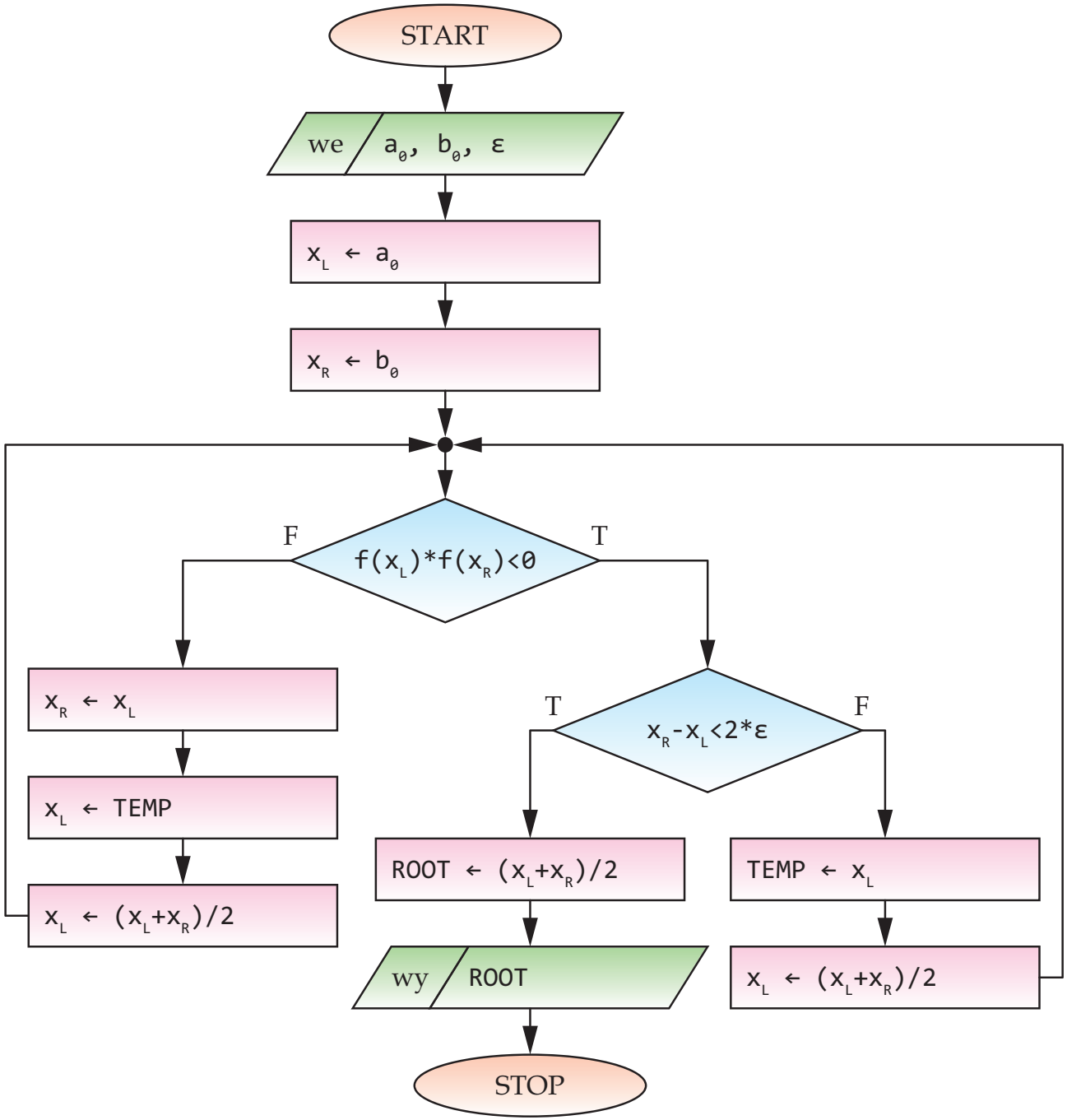
Podział przedziału (a, b) niekoniecznie musi odbywać się na dwie równe części.

Przykład

Rozwiążmy równanie $sin(x) + x^2 = 2$, przyjmując przedział początkowy $(1, 3)$.
Rozważane równanie ma pierwiastki $x_1^{sc} = -1,06155$ i $x_2^{sc} = 1,728466$.

Równanie wyjściowe zapiszemy w postaci: $f(x) = sin(x) + x^2 - 2$, $f(x) = 0$,
a końce przedziałów: $a_0 = 1$ i $b_0 = 3$.

i	$x_i = \frac{a_{i-1}+b_{i-1}}{2}$	$f(x_i) \cdot f(a_{n-1})$	a_i	b_i	$\varepsilon_i = \left\lvert \frac{x_i - x_{i-1}}{x_i} \right\rvert$	$\delta_i = \lvert f(x_i) \rvert$
1	2,000	-2,008497	1,000	2,000	0,500	1,090703
2	1,500	1,376490	1,500	2,000	0,333333	0,747495
...	...	...	...	...	...	...
11	1,727539	0,000023	1,727539	1,728516	0,000565	0,003350


START
we
a_0, b_0, ε
$x_L \leftarrow a_0$
$x_R \leftarrow b_0$
F
$f(x_L)*f(x_R)<0$
T
$x_R \leftarrow x_L$
$x_L \leftarrow$ TEMP
$x_L \leftarrow (x_L+x_R)/2$
T
$x_R-x_L<2*\varepsilon$
F
ROOT $\leftarrow (x_L+x_R)/2$
wy
ROOT
STOP
TEMP $\leftarrow x_L$
$x_L \leftarrow (x_L+x_R)/2$

Metoda stycznych (Newtona-Raphsona)

Dla pewnego otoczenia h punktu x rozwinięcie wartości wyjściowej funkcji $f(x+h)$ w szereg Taylora będzie wynosiło

$$f(x+h) = f(x) + f'(x) \cdot h + \frac{1}{2} f''(x) \cdot h^2 + \frac{1}{6} f'''(x) \cdot h^3 + \ldots \approx f(x) + f'(x) \cdot h$$

Po odrzuceniu wyrazów rozwinięcia wyższych rzędem niż pierwszy, ustalając x i podstawiając $f(x+h) = 0$ można obliczyć przyrost h jako

$$h = -\frac{f(x)}{f'(x)}.$$

Dla danej pary sąsiednich przybliżeń zachodzi: $x_{i+1} = x_i + h$, stąd otrzymujemy schemat metody

$$\begin{cases} x_0 \\ x_{i+1} = x_i - \dfrac{f(x)}{f'(x)} \end{cases}.$$

Graficznie metoda polega na budowaniu stycznych w kolejnych przybliżeniach x_i począwszy od punktu startowego oraz na szukaniu miejsc zerowych tych stycznych.

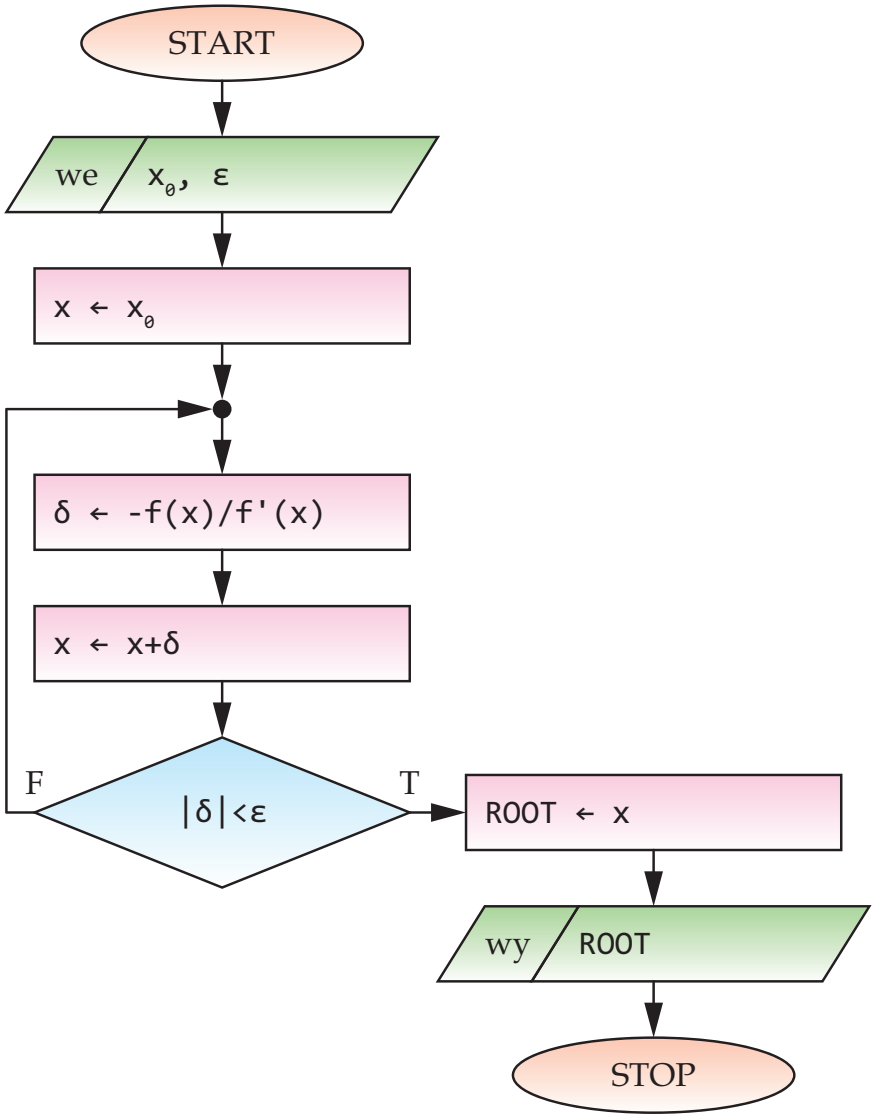
Przykład

Rozwiążmy równanie $sin(x) + x^2 = 2$, przyjmując punkt startowy $x_0 = -2$. Rozważane równanie ma pierwiastki $x_1^{sc} = -1,06155$ i $x_2^{sc} = 1,728466$.

Równanie wyjściowe zapiszemy w postaci: $f(x) = sin(x) + x^2 - 2$, $f(x) = 0$. Pochodna będzie równa: $f'(x) = cos(x) + 2x$. Schemat iteracyjny będzie miał postać:

$$\begin{cases} x_0 = -2 \\ x_{i+1} = x_i - \dfrac{sin(x_i) + x_i^2 - 1}{cos(x_i) + 2x_i} \end{cases}.$$

i	x_i	$\varepsilon_i = \left\lvert \frac{x_i - x_{i-1}}{x_i} \right\rvert$	$\delta_i = \left\lvert \frac{f(x_i)}{f(x_{i-1})} \right\rvert$
1	-1,188221	0,683189	0,116721
2	-1,064728	0,115985	0,002854
...	...	...	...
4	-1.061550	$< 10^{-6}$	$< 10^{-8}$


START
we
x_0, ε
x ← x_0
δ ← -f(x)/f'(x)
x ← x+δ
F
|δ|<ε
T
ROOT ← x
wy
ROOT
STOP

Metoda siecznych

W metodzie stycznych do schematu iteracyjnego potrzebna jest znajomość pochodnej rozpatrywanej funkcji. Aby uniknąć jej różniczkowania, można liczbową pochodną obliczać w sposób przybliżony korzystając z wartości ilorazu różnicowego

$$f'(x_i) \approx \frac{f(x_{i-1}) - f(x_i)}{x_{i-1} - x_i}$$

i wymaga dwóch punktów startowych. Prowadzi to do schematu iteracyjnego metody

$$\begin{cases} x_0, \ x_1 \\ x_{i+1} = x_i - f(x_i)\dfrac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})} \end{cases}.$$

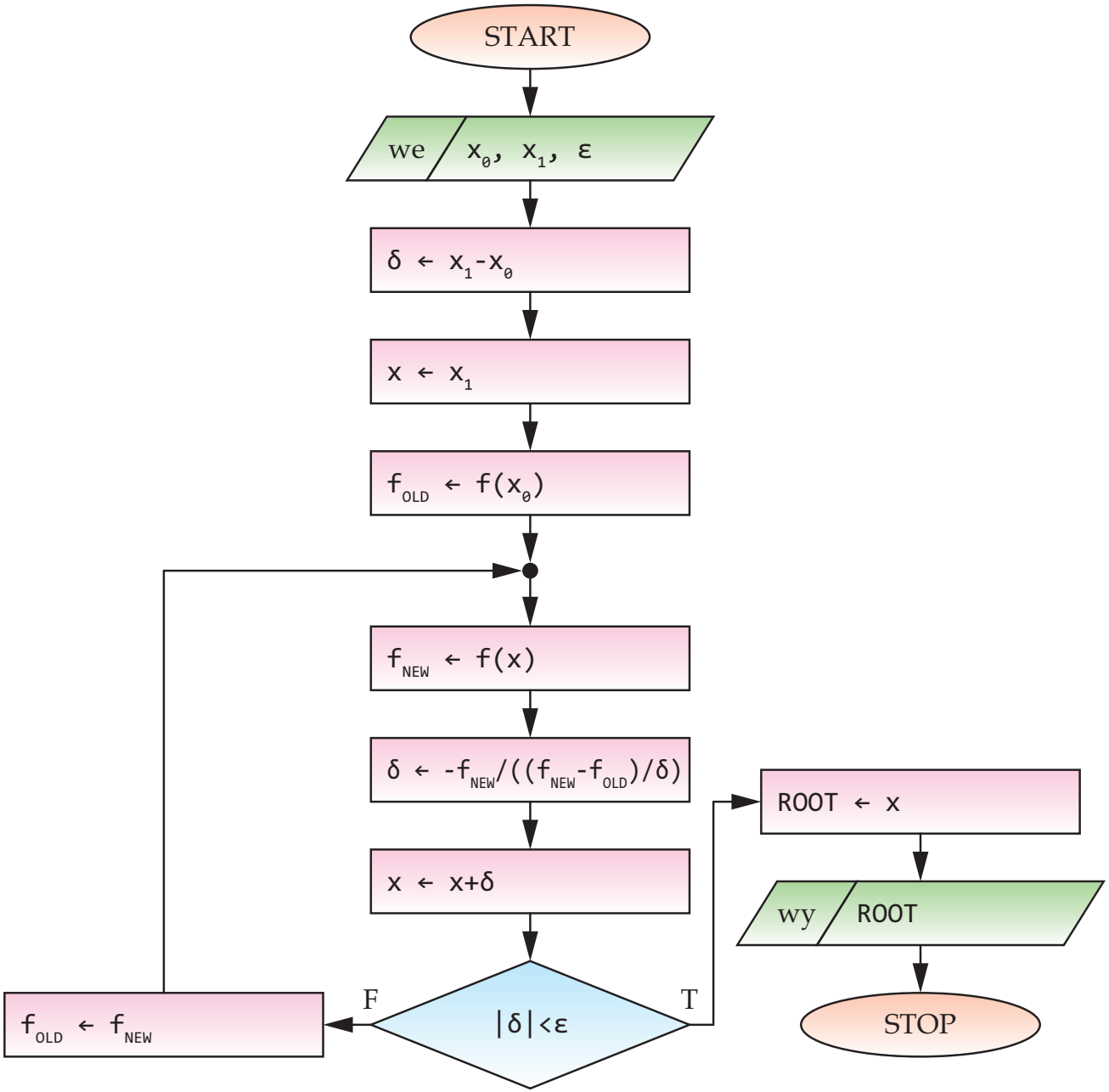
Przykład

Rozwiążmy równanie $sin(x) + x^2 = 2$, przyjmując punkty startowe $x_0 = -2$ i $x_1 = -0,5$. Rozważane równanie ma pierwiastki $x_1^{sc} = -1,06155$ i $x_2^{sc} = 1,728466$.

Równanie wyjściowe zapiszemy w postaci: $f(x) = sin(x) + x^2 - 2$, $f(x) = 0$. Schemat iteracyjny będzie miał postać:

$$\begin{cases} x_0 = -2, \ x_1 = -0,5 \\ x_{i+1} = x_i - (sin(x_i) + x_i^2 - 2)\dfrac{x_{i-1} - x_i}{sin(x_{i-1}) + x_{i-1}^2 - sin(x_n) - x_i^2} \end{cases}.$$

i	x_i	$\varepsilon_i = \left\lvert \frac{x_i - x_{i-1}}{x_i} \right\rvert$	$\delta_i = \left\lvert \frac{f(x_i)}{f(x_{i-1})} \right\rvert$
1	-0,955962	0,476967	0,092554
2	-1,078578	0,113683	0,015336
...	...	...	...
5	-1.061550	$< 10^{-6}$	$< 10^{-8}$


START
we
x_0, x_1, ε
δ ← x_1-x_0
x ← x_1
f_{OLD} ← f(x_0)
f_{NEW} ← f(x)
δ ← -f_{NEW}/((f_{NEW}-f_{OLD})/δ)
x ← x+δ
|δ|<ε
F
T
f_{OLD} ← f_{NEW}
ROOT ← x
wy
ROOT
STOP

Układ liniowych równań algebraicznych

Z liniowym (oznaczonym) układem równań algebraicznych mamy do czynienia w sytuacji, gdy wszystkie zmienne występujące w równaniach układu występują jedynie w pierwszej potędze. Można go skrótowo zapisać w następującej postaci

$$\sum_{l=1}^{n} c_{kl} \cdot x_l = r_k, \quad k = 1, 2, \ldots, m,$$

w którym oznaczono odpowiednio:

- c_{kl} – wartości współczynników stojących przy niewiadomych (elementy macierzy głównej),
- x_l – symbole niewiadomych (elementy wektora niewiadomych),
- r_l – wartości prawej strony (elementy wektora prawej strony).

Dla układu równań liniowych, warunek liniowej niezależności jest równoważny żądaniu by macierz główna układu $\mathbf{C}$, o elementach c_{kl}, miała wyznacznik różny od zera ($\det(\mathbf{C}) \neq 0$). W takiej sytuacji układ równań ma jedno i tylko jedno rozwiązanie. Podstawowe metody jego znajdowania można podzielić na dwie kategorie:

- metody eliminacyjne (np. Gaussa),
- metody iteracyjne (np. Gaussa-Seidla).

Do zalet metod eliminacyjnych należą między innymi łatwość rozwiązywania układu równań z wieloma prawymi stronami (sytuacja często występująca w praktyce), możliwość precyzyjnego oszacowania czasu obliczeń na podstawie rozmiaru zadania czy wreszcie gwarancja uzyskania wyniku po wykonaniu możliwej do określenia z góry liczby operacji.

Zaletą metod iteracyjnych jest prostota algorytmu i niewielkie zapotrzebowanie na pamięć operacyjną w trakcie obliczeń.

Metoda eliminacji Gaussa

Układ równań przekształcamy tak, aby macierz główną układu doprowadzić do postaci trójkątnej górnej (górnotrójkątnej).

Przekształcenia prowadzimy korzystając z faktu, że:

- przestawienie dwóch wierszy (równań),
- przemnożenie wszystkich elementów wiersza (współczynników równania) przez tę samą liczbę,
- dodanie do elementów jednego wiersza (współczynników równania) odpowiednio elementów innego wiersza,
- dodanie do elementów jednego wiersza odpowiednio kombinacji liniowej elementów innych wierszy,

przekształcają tablicę równoważnie, czyli w sposób nie zmieniający rozwiązania układu równań.

Obliczenia rozpoczynamy od pierwszej niewiadomej pierwszego równania

$$\begin{cases} c_{11} \cdot x_1 + c_{12} \cdot x_2 = r_1 \\ c_{21} \cdot x_1 + c_{22} \cdot x_2 = r_2 \end{cases},$$

naszym celem jest takie przekształcenie równania, aby współczynnik c_{11} miał wartość 1, a współczynnik c_{21} miał wartość 0. W tym celu podzielimy pierwsze równanie przez c_{11} (element wiodący)

$$\begin{cases} 1 \cdot x_1 + c_{12}/c_{11} \cdot x_2 = r_1/c_{11} \\ c_{21} \cdot x_1 + c_{22} \cdot x_2 = r_2 \end{cases},$$

następnie, od równania drugiego odejmujemy równanie pierwsze pomnożone przez c_{21}

$$\begin{cases} 1 \cdot x_1 + c_{12}/c_{11} \cdot x_2 = r_1/c_{11} \\ 0 \cdot x_1 + (c_{22} - c_{12}/c_{11} \cdot c_{21}) \cdot x_2 = r_2 - r_1/c_{11} \cdot c_{21} \end{cases},$$

w kolejnym kroku dzielimy drugie (kolejne) równanie przez $(c_{22} - c_{12}/c_{11} \cdot c_{21})$ (element wiodący)

$$\begin{cases} 1 \cdot x_1 + c_{12}/c_{11} \cdot x_2 = r_1/c_{11} \\ 0 \cdot x_1 + 1 \cdot x_2 = (r_2 - r_1/c_{11} \cdot c_{21})/(c_{22} - c_{12}/c_{11} \cdot c_{21}) \end{cases}.$$

W kolejnych krokach wyznaczamy wartości niewiadomych.

Przykład

Rozwiążmy metodą eliminacji Gaussa układ równań

$$\begin{cases} 2x_1 + 1x_2 - 2x_3 = 0 \\ -1x_1 - 3x_2 + 4x_3 = 3 \\ 3x_1 + 2x_2 + 1x_3 = 4 \end{cases}.$$

W pierwszym kroku, pierwsze równanie podzielimy przez 2 (element wiodący)

$$\begin{cases} 1x_1 + 0,5x_2 - 1x_3 = 0 \\ -1x_1 - 3x_2 + 4x_3 = 3 \\ 3x_1 + 2x_2 + 1x_3 = 4 \end{cases},$$

następnie, od równania 2 odejmiemy równanie 1 pomnożone przez -1

$$\begin{cases} 1x_1 + 0,5x_2 - 1x_3 = 0 \\ 0x_1 - 2,5x_2 + 3x_3 = 3 \\ 3x_1 + 2x_2 + 1x_3 = 4 \end{cases},$$

w kolejnym kroku, od równania 3 odejmiemy równanie 1 pomnożone przez 3

$$\begin{cases} 1x_1 + 0,5x_2 - 1x_3 = 0 \\ 0x_1 - 2,5x_2 + 3x_3 = 3 \\ 0x_1 + 0,5x_2 + 4x_3 = 4 \end{cases}.$$

Drugie równanie dzielimy przez 2,5 (element wiodący)

$$\begin{cases} 1x_1 + 0,5x_2 - 1x_3 = 0 \\ 0x_1 + 1x_2 - 1,2x_3 = -1,2 \\ 0x_1 + 0,5x_2 + 4x_3 = 4 \end{cases},$$

następnie, od równania 3 odejmiemy równanie 2 pomnożone przez 0,5

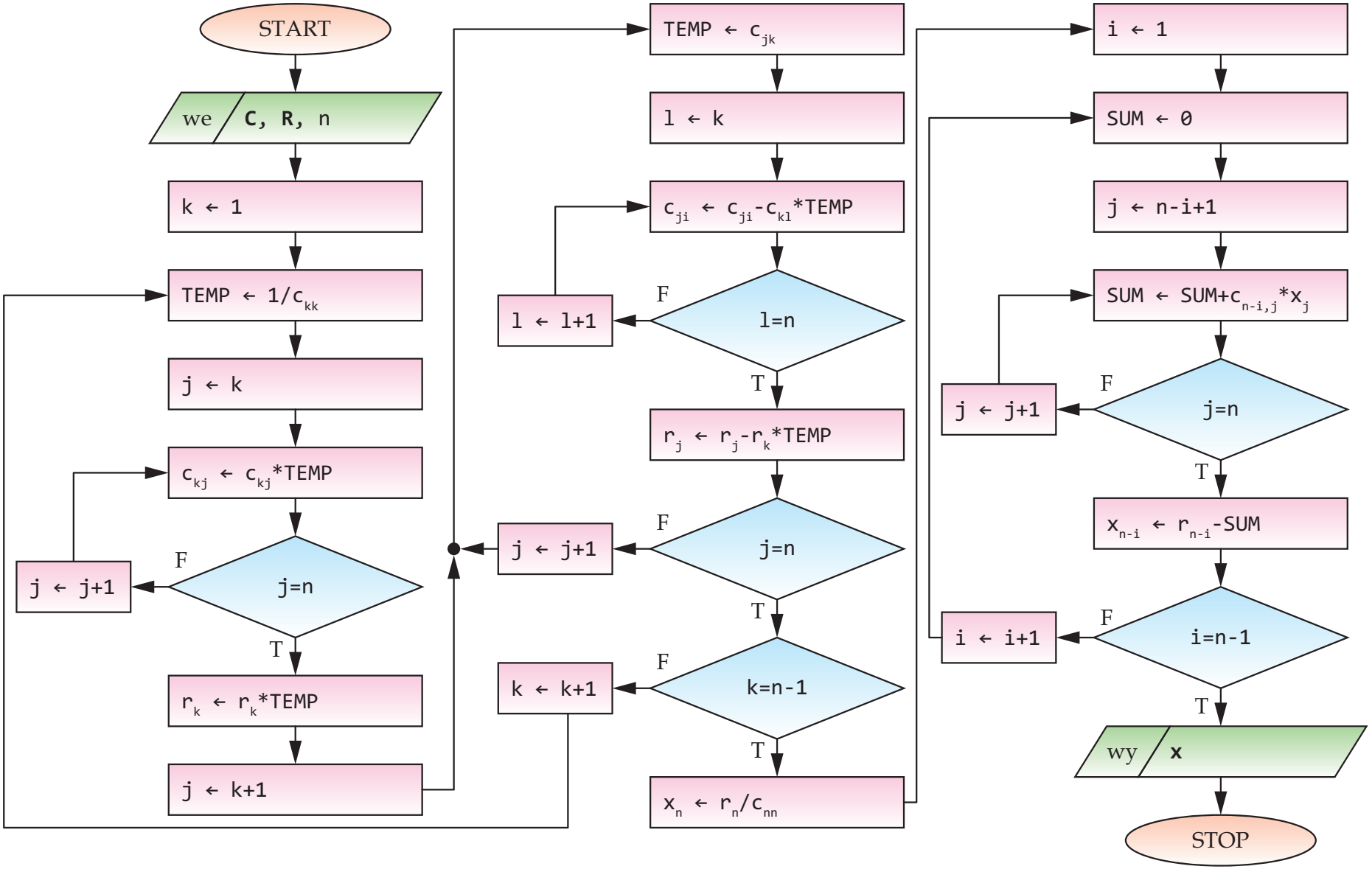
$$\begin{cases} 1x_1 + 0,5x_2 - 1x_3 = 0 \\ 0x_1 + 1x_2 - 1,2x_3 = -1,2 \\ 0x_1 + 0x_2 + 4,6x_3 = 4,6 \end{cases},$$

ostatnim krokiem będzie podzielenie równania 3 przez 4,6

$$\begin{cases} 1x_1 + 0,5x_2 - 1x_3 = 0 \\ 0x_1 + 1x_2 - 1,2x_3 = -1,2 \\ 0x_1 + 0x_2 + 1x_3 = 1 \end{cases}.$$

Z powyższego układu równań, poczynając od ostatniego będziemy mogli wyznaczyć kolejne niewiadome

$$\begin{cases} 1x_3 = 1 & \Rightarrow & x_3 = 1 \\ 1x_2 - 1,2x_3 = -1,2 & \Rightarrow & x_2 = 0 \\ 1x_1 + 0,5x_2 - 1x_3 = 0 & \Rightarrow & x_1 = 1 \end{cases}.$$


START
we C, R, n
k ← 1
TEMP ← 1/c_kk
j ← k
c_kj ← c_kj*TEMP
j=n
F
T
j ← j+1
r_k ← r_k*TEMP
j ← k+1
TEMP ← c_jk
l ← k
c_ji ← c_ji-c_kl*TEMP
l=n
F
T
l ← l+1
r_j ← r_j-r_k*TEMP
j=n
F
T
j ← j+1
k=n-1
F
T
k ← k+1
x_n ← r_n/c_nn
i ← 1
SUM ← 0
j ← n-i+1
SUM ← SUM+c_n-i,j*x_j
j=n
F
T
j ← j+1
x_n-i ← r_n-i-SUM
i=n-1
F
T
i ← i+1
wy x
STOP

Metoda Gaussa-Seidla

W tej metodzie układ równań

$$\begin{cases} c_{11} \cdot x_1 + c_{12} \cdot x_2 = r_1 \\ c_{21} \cdot x_1 + c_{22} \cdot x_2 = r_2 \end{cases},$$

przekształcamy tak, aby z kolejnych równań móc obliczyć kolejne niewiadome

$$\begin{cases} x_1 = (r_1 - c_{12} \cdot x_2)/c_{11} \\ x_2 = (r_2 - c_{21} \cdot x_1)/c_{22} \end{cases}.$$

Obliczenia rozpoczynamy od założenia początkowych wartości niewiadomych $x_{1\{0\}}$ i $x_{2\{0\}}$. W pierwszej iteracji wyznaczamy nowe przybliżenie szukanych wartości niewiadomych, wykorzystując ich ostatnie dostępne wartości

$$\begin{cases} x_{1\{1\}} = (r_1 - c_{12} \cdot x_{2\{0\}})/c_{11} \\ x_{2\{1\}} = (r_2 - c_{21} \cdot x_{1\{1\}})/c_{22} \end{cases}.$$

Po pierwszej iteracji otrzymujemy przybliżone wartości rozwiązania $x_{1\{1\}}$ i $x_{2\{1\}}$. Powtarzamy procedurę iteracji

$$\begin{cases} x_{1\{2\}} = (r_1 - c_{12} \cdot x_{2\{1\}})/c_{11} \\ x_{2\{2\}} = (r_2 - c_{21} \cdot x_{1\{2\}})/c_{22} \end{cases},$$

i kolejno

$$\begin{cases} x_{1\{i+1\}} = (r_1 - c_{12} \cdot x_{2\{i\}})/c_{11} \\ x_{2\{i+1\}} = (r_2 - c_{21} \cdot x_{1\{i+1\}})/c_{22} \end{cases},$$

aż do otrzymania wyników spełniających zakładaną dokładność, zwykle określaną jako dopuszczalną, względną wielkość zmiany wyników pomiędzy kolejnymi iteracjami.

Przykład

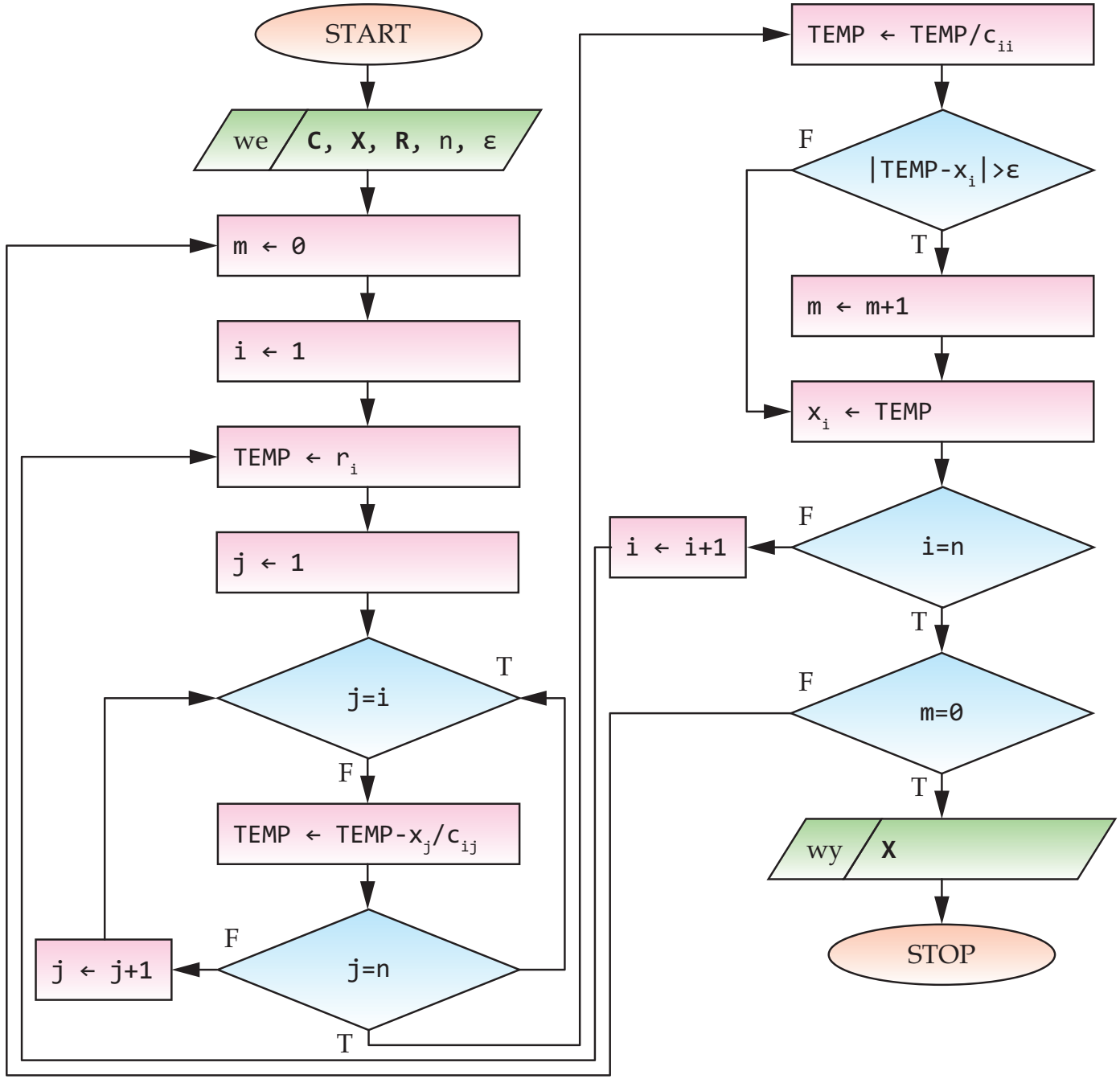
Rozwiążmy metodą iteracyjną Gaussa-Seidla układ równań,

$$\begin{cases} 3x_1 + 1x_2 - 1x_3 = 2 \\ 1x_1 + 4x_2 + 1x_3 = 12 \\ 2x_1 + 1x_2 + 2x_3 = 10 \end{cases},$$

przyjmując wartości startowe: $x_{1\{0\}} = 1$, $x_{2\{0\}} = 1$ i $x_{3\{0\}} = 1$. Przekształcony układ równań:

$$\begin{cases} x_1 = (2 - 1x_2 + 1x_3)/3 \\ x_2 = (10 - 1x_1 - 1x_3)/4 \\ x_3 = (12 - 2x_1 - 1x_2)/2 \end{cases}.$$

i	$x_{1\{i\}}$	$x_{2\{i\}}$	$x_{3\{i\}}$	$\varepsilon_1^{\{i\}} = \left\lvert \frac{x_{1\{i\}} - x_{1\{i-1\}}}{x_{1\{i\}}} \right\rvert$	$\varepsilon_2^{\{i\}} = \left\lvert \frac{x_{2\{i\}} - x_{2\{i-1\}}}{x_{2\{i\}}} \right\rvert$	$\varepsilon_3^{\{i\}} = \left\lvert \frac{x_{3\{i\}} - x_{3\{i-1\}}}{x_{3\{i\}}} \right\rvert$
1	0.6667	2.5834	3.0416	0.5	0.6129	0.6712
2	0.8194	2.0347	3.1632	0.1864	0.2696	0.0384
...	...	...	...	...	...	...
8	1.0000	2.0000	2.9999	0.0004	0.0001	0.0000


START
we C, X, R, n, ε
m ← 0
i ← 1
TEMP ← r_i
j ← 1
j=i
T
F
TEMP ← TEMP-x_j/c_ij
j=n
F
T
j ← j+1
TEMP ← TEMP/c_ii
|TEMP-x_i|>ε
F
T
m ← m+1
x_i ← TEMP
i=n
F
T
i ← i+1
m=0
F
T
wy X
STOP