

**Zakres egzaminu dyplomowego dla studentów studiów stacjonarnych II stopnia.  
Specjalność - Inteligentne Systemy Informatyczne**

1. Modelowanie a metamodelowanie
2. Własności i zakres zastosowań języków UML i LOTOS.
3. Problemy transformacji i spójności modeli.
4. Walidacja i weryfikacja modeli.
5. Różnice między wyszukiwaniem informacji a wyszukiwaniem danych.
6. Działanie systemu informacyjnego w sieci komputerowej.
7. Technologie multimedialne stosowane w systemach informacyjnych.
8. Efektywność systemów informacyjnych.
9. Zadania projektowania sieci komputerowej.
10. Klasyfikacja ruchu teleinformatycznego.
11. Zarządzanie zasobami sieci komputerowej.
12. Metody naprawiania błędów w systemach teleinformatycznych. \
13. Koncepcje dostarczania jakości usług w sieciach teleinformatycznych.
14. Pojęcie systemu decyzyjnego oraz komputerowego systemu wspomaganie decyzji.  
todo krzysiek
15. Czynności techniki systemów.
16. Problemy decyzyjne dla kompleksu operacji. -todo Krzysiek
17. Podstawowe problemy, metody i algorytmy optymalizacji dyskretnej.
18. Podstawowe metody „obliczeń miękkich (inteligentnych)”.
19. Podejmowanie decyzji w warunkach niepewności.
20. Metody i algorytmy rozpoznawania.
21. Postulaty metodologii nauk.
22. Współczesne metody naukometrii
23. Organizacja systemów rozproszonych.
24. Ocena systemów równoległych.
25. Uczenie indukcyjne: metody, zastosowania.
26. Uczenie ze wzmocnieniem.
27. Modele obrazów: wektory cech, niezmienniki, relacje przestrzenne
28. Klasyfikacja, opisywanie, interpretacja, rozumienie.
29. Obliczenia miękkie a obliczenia tradycyjne (algorytmiczne).
30. Jakie techniki wchodzi w skład obliczeń miękkich. Charakterystyka każdej z nich.
31. Podstawy formalnego opisu języka naturalnego: założenia i stosowane metody.
32. Współczesna technologia językowa: narzędzia, zasoby językowe i ich zastosowania.
33. Proces przetwarzania języka naturalnego: typowe etapy, cele, stosowane metody

1. Modelowanie a metamodelowanie.

W najbardziej ogólnym ujęciu metamodelowanie (zapisywane również jako meta-modelowanie) określa analizę, konstrukcję i rozwój ram, reguł, ograniczeń, modeli i teorii, które są użyteczne przy procesie modelowania w pre-definiowanej grupie problemów. Sam termin składa się z dwóch członów: meta i modelowanie.

Meta-modelowanie wraz meta-modelami występuje w wielu dziedzinach wiedzy, np. w meta-nauce, meta-filozofii, meta-teorii czy też w teorii systemów. Termin ten występuje również w matematyce oraz często spotykany jest w informatyce.

W informatyce i dziedzinach jej pokrewnych meta-modelowanie oznacza konstruowanie zbioru "koncepcji" (obiektów, terminów, itp.) w zakresie pewnej dziedziny. Uznając model

za abstrakcję pewnego zjawiska ze świata rzeczywistego, to meta-model jest abstrakcją ukazującą właściwości owego modelu. Przykładem takiej zależności może być program komputerowy napisany w pewnym języku programowania oraz gramatyka owego języka programowania.

Meta-modele w informatyce najczęściej stosowane są jako:

- schematy logiczne dla danych semantycznych, które muszą być wymieniane bądź przechowywane
- język wspierający określoną metodę bądź proces
- język służący do ogólnego wyrażania semantyki pewnych informacji

Modelowanie to budowanie modeli.

Modelowanie można określić jako tworzenie opisu obiektu/zjawiska rzeczywistego lub abstrakcyjnego, wykonywane w założonym celu. Jego efektem jest model, który powinien posiadać określone własności. Z tego względu nie ma modeli złych, są tylko takie, które nie realizują założonego celu. Model może natomiast być niepoprawny, tzn. nie spełniać definiujących go własności. Co do modelowania, to zawsze posiada ono samo z siebie pewną wartość jako działanie prowadzące do usystematyzowania i ustrukturalizowania informacji.

Model - w nauce reprezentacja (abstrakcyjna, uproszczona wizja) pewnego rzeczywistego lub wyobrażanego bytu. Model zależy od przyjętej perspektywy modelowania. Perspektywa wynika z celu modelowania.

Dwie role modelu: reprezentacja tego co istnieje, reprezentacja tego co ma powstać.

Modele są reprezentowane w postaci graficznej i tekstowej. Mogą przedstawiać zarówno aspekt statyczny, jak i dynamiczny modelowanego bytu.

## 2. Własności i zakres zastosowań języków UML i LOTOS.

Własności UML:

- Graficzny język modelowania (specyfikowania, projektowania, dokumentowania)
- Jest standardem powszechnie akceptowanym, wykorzystywany do odniesień dla pojęć obiektowości
- Reprezentuje model w postaci zbioru diagramów, pełniących określone role
- Koncentruje się na danych i operacjach ich przetwarzania
- Reprezentuje zazwyczaj zbiory bytów, a nie pojedyncze byty

Zastosowanie UML:

Głównym przeznaczeniem UML jest budowa systemów informatycznych. Służy do modelowania dziedziny problemu - w przypadku stosowania go do analizy oraz do modelowania rzeczywistości, która ma dopiero powstać - tworzy się w nim głównie modele systemów informatycznych.

Własności LOTOS:

- Jest językiem interaktywnym, opisującym interakcje systemu z otoczeniem za pomocą komunikacji poprzez bramki (porty)
- Analityka nie interesuje wnętrze systemu (czarna skrzynka), a jedynie zmiany jego stanów
- Umożliwia modelowanie procesów w systemach zagnieżdżonych, gdzie jeden system zawiera w sobie kilka pomniejszych procesów

- Powstało kilka wersji języka: bazowa (bez uwzględnienia przesyłanych danych), pełna (uwzględnienie przesyłanych danych) oraz rozszerzona (upływ czasu)

Zastosowanie LOTOS:

Wykorzystany w tworzeniu systemów informatycznych, jako narzędzie pozwalające modelować systemy z wykorzystaniem algebry procesów – wyrażenie wymagań funkcjonalnych systemu w postaci sekwencji akcji.

### 3. Problemy transformacji i spójności modeli.

Transformacja modeli to opisanie modelowanego bytu (systemu) za pomocą innego modelu. Jej celem może być przekształcenie modelu do postaci mniej lub bardziej abstrakcyjnej, ograniczenie stopnia szczegółowości opisu (formalnego i nieformalnego) lub opisanie procesu / systemu / bytu z innej perspektywy celem wyszczególnienia pewnych procesów i problemów niewidocznych w opisie z wykorzystaniem innego modelu. Główny problem, jaki rodzi transformacja to możliwość utraty danych opisujących byt, która może być powodem zniekształcenia opisu (lub nawet jego zafałszowania). Transformacja może też prowadzić do niekontrolowanego rozrostu opisu / modelu, co również może być działaniem niepożądanym.

Oceną, czy model po transformacji odpowiada modelowi sprzed transformacji jest ocena ich spójności, czyli na ile różne modele opisują ten sam system / byt.

Przykładem transformacji modeli może być przekształcenie diagramu UML na proces LOTOS lub opis z użyciem sieci Petriego.

### 4. Walidacja i weryfikacja modeli.

Walidacja jest to proces wyznaczania stopnia, w jakim model jest wiernym odzwierciedleniem rzeczywistego systemu z przyjętego punktu widzenia. Ma na celu określenie, czy symulacja daje wiarygodne wyniki, w założonym stopniu zgodne z odpowiedziami rzeczywistego systemu na takie same dane wejściowe. O ile dzięki weryfikacji projektant uzyskuje informacje o zgodności systemu symulacyjnego z jego założeniami, o tyle walidacja weryfikuje zgodność jego wizji z realnym światem. Obie te fazy wzajemnie się uzupełniają i jako takie czasami przedstawiane są wspólnie jako faza oceny adekwatności modelu.

### 5. Różnice między wyszukiwaniem informacji a wyszukiwaniem danych.

Różnice między wyszukiwaniem informacji, a wyszukiwaniem danych wynikają bezpośrednio z różnic między informacjami, a danymi.

Dane (ang. data) — to litery, słowa, teksty, liczby, znaki, symbole, kombinacje liter, liczb, symboli i znaków. Dane przetwarza się po to by otrzymać informację, wiadomość. Informacja jest produktem finalnym przetwarzania danych.

Istotną cechą danych jest brak uporządkowania; jest to zbiór nieuporządkowany. Ale jednocześnie dana jest wiarygodna i pewna (przy odpowiednim przechowywaniu). Dane w każdej chwili mogą być weryfikowane pod względem poprawności i aktualizowane. Dane pełnią także rolę nośników przepływu informacji.

Informacja jest zbiorem uporządkowanym wg określonego kryterium i poddanym interpretacji.

Cechą każdej informacji (jako wyniku interpretacji) jest jej niepewność i ograniczona w czasie trwałość (wiarygodność) [4]. W popularnym pojmowaniu informacji zwracamy uwagę na ilość informacji (dużo, mało). Z informacją wiążemy także pojęcie jakości, na którą składa się wierność przekazu, wiarygodność, szybkość i sposób archiwizowania.

Z powyższych rozważań wynika, że dane i informacje różnią się własnościami. Okazuje się, że z "morza danych" możemy uzyskać niewiele informacji i odwrotnie — z niewielkiej ilości danych można otrzymać znaczące informacje. Tak więc ilość informacji nie jest zależna do ilości danych.

#### 6. Działanie systemu informacyjnego w sieci komputerowej.

Rozwój technologii telekomunikacyjnych i komputerowych w ostatniej dekadzie, zwłaszcza niezwykle ekspansja WWW – stworzyły nowe możliwości rozpowszechniania i wymiany informacji, a tym samym przyczyniły się do powstania „nowoczesnych technik informacyjnych”. Sam System WWW, po wprowadzeniu standardów i metod pozwalających na opis i organizację multimedialnych danych, zawartych na stronach WWW i opracowaniu systemów wspomagających procesy wyszukiwawcze - ma szansę stać się globalnym systemem informacyjnym, efektywnie wykorzystywanym przez użytkowników końcowych do wyszukiwania informacji o wysokiej trafności.

Jedną z cech powstającego społeczeństwa informacyjnego jest rozwój pojęcia telematyka, a więc niejako połączenia telekomunikacji i informatyki. Inaczej mówiąc są to sieciowe cyfrowe multimedia. Obecnie ważniejsze jest dotarcie do informacji niż jej gromadzenie, użytkownik ma możliwość bezpośredniego dostępu do informacji bez instytucji pośredniczących. Czas i przestrzeń dzięki połączeniom sieciowym nie odgrywają już takiej roli, jak niegdyś. Ważna jest także sprawa aktualności danych. W przypadku dokumentów drukowanych istnieje niebezpieczeństwo ich dezaktualizacji już w momencie opublikowania. Dokumenty elektroniczne w sieciach dostępne są w trybie czasu rzeczywistego. Poza tym istnieje możliwość aktualizowania danych na bieżąco.

7. Efektywność opisuje stopień wykorzystania zasobów sprzętowych i programowych stanowiących podstawę działania systemu informacyjnego.
8. Zła organizacja i niska efektywność systemu skutkują wzrostem kosztów działalności organizacji, spadkiem jakości obsługi, utrudnionym dostępem do informacji na różnych Technologie multimedialne stosowane w systemach informacyjnych.
9. Efektywność systemów informacyjnych.

szczeblach, ich znaczną niekompletnością, wydłużonym czasem generowania.

Jedną ze strategii zwiększenia efektywności pracy zespołów projektujących systemy informacyjne jest wielokrotne używanie raz opracowanych fragmentów projektu lub modułów programowych. Ponowne użycie wcześniej opracowanego fragmentu projektu albo modułu programu ma wiele zalet, wśród których na plan pierwszy wysuwają się: obniżenie kosztów, przyspieszenie realizacji projektu oraz minimalizacja błędów

#### 10. Zadania projektowania sieci komputerowej.

Warunki, które powinien spełniać dobry projekt sieci komputerowej:

1. Realizacja oczekiwań zlecniodawcy,
2. Fachowa dokumentacja,

3. Możliwość rekonfiguracji i rozbudowy sieci,
4. Łatwość rekonfiguracji w przypadku awarii,
5. Niezależność uszkodzeń w różnych segmentach sieci,
6. Bezpieczeństwo danych i serwerów.

#### Zasady:

1. Nie wolno osiągać granic możliwości sieci/sprzętu.
2. Maksymalna długość kabla dla 5 kategorii: 3m do komputera, 90m kabla poziomego, 6m kabla krosującego.
3. Zgodnie z ISO: 10 metrów kwadratowych na miejsce pracy (nie oznacza to 1 gniazdka na 10m<sup>2</sup>, ale co najmniej 1 gniazdko na 10m<sup>2</sup>).
4. Każda kondygnacja musi być wyposażona w minimum 1 punkt dystrybucyjny. W przypadku pomieszczeń o powierzchni większej od 1000m<sup>2</sup> lub w przypadku, kiedy okablowanie poziome przekracza 90 m należy wprowadzić dodatkowy punkt dystrybucyjny.
5. Powierzchnia punktu dystrybucyjnego PD (SPD) w zależności od obsługiwanej powierzchni (S):
  - S 1000 m<sup>2</sup> -> SPD min. 3.0x3.4m,
  - S 800m<sup>2</sup> -> SPD min. 3.0x2.8m,
  - S 500m<sup>2</sup> -> SPD min. 3.0x2.3m.
6. Maksymalna dozwolona ścieżka sygnału obejmuje 5 segmentów kabla połączonych 4 hubami. W takim przypadku 2 z tych segmentów mogą być użyte wyłącznie jako połączenia między hubami. Jeśli się nie da inaczej, należy podzielić sieć na domeny kolizyjne (podsieci) i wprowadzić switche.
7. Maksymalnie 1024 urządzenia na podsieć.
8. Maksymalna całkowita odległość w podsieci 500m.

#### 11. Klasyfikacja ruchu teleinformatycznego.

Ruchem telekomunikacyjnym nazywamy przepływ zgłoszeń, połączeń i wiadomości. W przypadku sieci pakietowych, w zależności od warstwy modelu OSI, na którym poziomie dokonujemy obserwacji ruchu, definicję tę można rozszerzyć na przepływ innych jednostek transmisji danych (pakietów – np. pakietów TCP, datagramów – np. datagramów IP, ramek – np. ramek Ethernet, komórek – np. komórek ATM (Asynchronous Transfer Mode), itd.). Każda aplikacja przesyłająca dane generuje w sieci pewien ruch. W przypadku sieci jednousługowej (np. sieci telefonii analogowej), ruch generowany przez poszczególnych użytkowników ma charakter homogeniczny.

W przypadku sieci wielousługowej (np. Internet) ruch telekomunikacyjny jest ruchem heterogenicznym. W sieci Internet składa się na niego m.in.:

- ruch generowany podczas transmisji danych masowych (np. z wykorzystaniem usługi ftp przesyłania plików) – charakteryzuje się on długim czasem trwania połączenia transportowego, związanym z przesyłaniem długich wiadomości (plików zawierających dane masowe), odstępy pomiędzy połączeniami transportowymi realizowanymi przez tego samego użytkownika są stosunkowo długie;
- ruch generowany podczas przesyłania poczty elektronicznej – charakteryzuje się on stosunkowo krótkimi wiadomościami tekstowymi (procentowy udział długich wiadomości, zawierających załączniki w postaci plików z danymi czy skompresowanych obrazów jest niewielki), przesyłanymi w bardzo dużych odstępach czasu;
- ruch generowany przez usługę zdalnego terminala (np. telnet), charakteryzujący się krótkimi wiadomościami, przesyłanymi w relatywnie krótkich odstępach czasu;
- ruch generowany przez usługę WWW charakteryzuje się krótkimi wiadomościami (tekst,

niewielkie obrazy – np. logo firmy, małe elementy graficzne), przesyłanymi w bardzo krótkich odstępach czasu (związanymi z przesyłaniem zawartości strony WWW), po których następują relatywnie długie okresy braku aktywności (związane z przeglądaniem zawartości stron przez użytkownika); podobnie, jak w przypadku poczty elektronicznej, procentowy udział długich wiadomości jest niewielki;

- ruch generowany przez aplikacje realizujące transmisję informacji multimedialnej w czasie rzeczywistym – charakteryzuje się przesyłaniem bardzo krótkich (np. transmisja głosu) lub bardzo długich wiadomości (np. transmisja ramek wideo), lub ich złożenia (np. równoczesna transmisja głosu, obrazu, wiadomości tekstowych); cechą charakterystyczną źródła ruchu multimedialnego jest generowanie wiadomości w stałych odstępach czasu (np. generowanie ramek wideo co 40 ms lub co ok. 33 ms).

## 12. Zarządzanie zasobami sieci komputerowej.

Typy sieci ze względu na zarządzanie zasobami:

Sieci równorzędne (każdy-z-każdym)

P2P (od ang. peer-to-peer – równy z równym) – model komunikacji w sieci komputerowej, który gwarantuje obydwu stronom równorzędne prawa (w przeciwieństwie do modelu klient-serwer). W sieciach P2P każdy komputer może jednocześnie pełnić zarówno funkcję klienta, jak i serwera. Każdy węzeł sieci (czyli komputer użytkownika) odgrywa rolę serwera przyjmując połączenia od innych użytkowników danej sieci, jak i klienta, łącząc się i pobierając dane z innych maszyn działających w tej samej sieci. Wymiana danych jest zawsze prowadzona bez pośrednictwa centralnego serwera.

Sieci oparte na serwerach (klient-serwer)

W sieciach klient-serwer zasoby często udostępniane gromadzone są w komputerach odrębnej warstwy, zwanych serwerami. Serwery zwykle nie mają użytkowników bezpośrednich. Są one raczej komputerami wielodostępnymi, które regulują udostępnianie swoich zasobów szerokiej rzeszy klientów. W sieciach tego typu zdjęty jest z klientów ciężar funkcjonowania jako serwery wobec innych klientów.

Narzędzia do zarządzania zasobami: procesory, pamięć, dyski, połączenia sieciowe itp.

Narzędzia pozwalające na równoważenie obciążenia w środowisku heterogenicznym systemów komputerowych: algorytmy rozdziału zasobów, języki i sposoby (narzędzia) specyfikacji zasobów, mapowanie zasobów do aplikacji. Funkcjonalność zarządzania zasobami jest ograniczona i charakteryzuje się takimi wadami jak:

1. niewykorzystanie zasobów obliczeniowych w czasie oczekiwania na wszystkie żądane zasoby;
2. konieczność upewnienia się, czy komponenty aplikacji nie rozpoczną działania zanim moduł przydzielający zasoby określi czy żądanie się powiedzie.

## 13. Metody naprawiania błędów w systemach teleinformatycznych.

Do detekcji i korekcji pojedynczych błędów transmisji stosuje się blokowe sekwencje znaków kontrolnych. Powszechnie stosowaną korekcją jest sekwencja BCC (Block Check Character) przedstawiająca znak lub sekwencję znaków generowaną przez algorytm kontrolny przed wysłaniem wiadomości w łączy transmisji danych. Urządzenie odbiorcze porównuje odtworzoną sekwencję kontrolną z sekwencją odebraną, aby stwierdzić, czy wystąpiły błędy transmisji.

Przy korekcji CRC blok informacyjny traktuje się jako wielomian, który w nadajniku dzieli się modulo 2 przez wielomian CRC, zwykle szesnastego stopnia (CCITT zaleca kilka, popularnym jest  $x^{16}+x^{12}+x^5+1$ ). Otrzymana reszta tworzy 16-bitową sekwencję kontrolną FCS (Frame Check Sequence) transmitowaną na końcu bloku. W odbiorniku odebrany blok informacyjny również dzieli się przez taki sam wielomian. Przez porównanie otrzymanej reszty z dzielenia z odebraną sekwencją kontrolną można stwierdzić wystąpienie błędu transmisji. Brak zgodności sekwencji wymusza przesłanie odpowiedniej informacji kanałem sprzężenia powrotnego i retransmisję błędnych bloków.

#### 14. Koncepcje dostarczania jakości usług w sieciach teleinformatycznych.

- kształtowanie i ograniczanie przepustowości
- zapewnienie sprawiedliwego dostępu do zasobów
- nadawanie odpowiednich priorytetów poszczególnym pakietom wędrującym przez sieć
- zarządzanie opóźnieniami w przesyłaniu danych
- zarządzanie buforowaniem nadmiarowych pakietów: DRR, WFQ, WRR
- określenie charakterystyki gubienia pakietów
- unikanie przeciążeń: Connection Admission Control (CAC),

#### 15. Pojęcie systemu decyzyjnego oraz komputerowego systemu wspomaganie decyzji.

- **Decyzja** jest świadomym (niełosowym) wyborem jednego z możliwych w danej sytuacji wariantów działania.
- **Sytuacja decyzyjna** charakteryzuje się istnieniem co najmniej dwóch możliwych wariantów działania różniących się między sobą stopniem korzyści.
- **Problem decyzyjny** to porównanie stanu oczekiwanego ze stanem rzeczywistym, pomiar odchyłań, oraz przymus wyboru jednego z wariantów działania.
- **Decydent**– podmiot podejmujący decyzję.
- **Przedmiot decyzji**– rzeczywistość (proces, system, obiekt) DP, której decyzja dotyczy; obiekt podejmowania decyzji - obiekt decyzyjny.
- **System decyzyjny** to model (odwzorowanie) pewnego wycinka rzeczywistości, którego celem jest rozwiązywanie występujących w nim problemów decyzyjnych.

**Systemy Wspomaganie Decyzji** są zorganizowanym zbiorem ludzi, procedur, baz danych i urządzeń wykorzystywanych w celu wspomaganie podejmowania decyzji na wszystkich etapach tego procesu, poczynając od rozpoznania czyli zdefiniowania problemu i zaklasyfikowania go do określonej grupy standardowej, następnie poprzez wybór odpowiednich danych stworzenie i analizę modelu informacyjnego opisującego rzeczywistość, dalej pomagając w generowaniu wariantów dopuszczalnych rozwiązań oraz w wyborze najlepszego rozwiązania.

Jakie decyzje potrzebują teorii?

Metody teorii decyzji wykorzystuje się wszędzie tam, gdzie podjęcie decyzji jest z pewnych powodów trudne. Przykładowo przyczynami mogą być:

- *duża liczba możliwych wariantów* – np. wybór najlepszego kandydata na dane stanowisko
- *skomplikowana sytuacja decyzyjna* – np. opracowanie takich tras i rozkładów jazdy

- autobusów, aby zapewnić wysoki poziom obsługi przy jak najniższym koszcie
- *możliwość wysokich korzyści lub dużych strat* – np. wybór sposobu ulokowania oszczędności
- *skomplikowany proces decyzyjny* – np. podejmowanie grupowych decyzji w dużych organizacjach
- *waga problemu decyzyjnego* – np. ustalenie okręgów wyborczych w wyborach prezydenckich

## 16. Czynności techniki systemów.

Teoria i technika systemów zajmuje się wspólnymi problemami, metodami i technikami dotyczącymi opisu, własności i sposobów rozwiązywania zadań, których przedmiotem są systemy o różnej naturze. W szczególności t.s. zajmuje się:

- kreowaniem modeli i modelowaniem,
- identyfikacją i rozpoznawaniem,
- analizą i projektowaniem,
- sterowaniem (kierowaniem, zarządzaniem).

**Identyfikacja systemów** lub **procesów** to termin opisujący zespół metod i narzędzi i [algorytmy](#), które mają na celu zbudować dynamiczny model [systemu](#) lub [procesu](#) na podstawie danych pomiarowych zebranych z wejścia i wyjścia. Model taki może opisywać:

- właściwości wejściowo-wyjściowe systemu - jeżeli jest tworzony w oparciu o sekwencje sygnałów wejściowych i towarzyszące im sekwencje sygnałów wyjściowych,
- przebieg wyjścia systemu o wejściach pomiarowo niedostępnych - jeżeli jest tworzony jedynie w oparciu o mierzoną sekwencję sygnału wyjściowego.

**Analiza** - Przeciwną metodą do identyfikacji jest modelowanie analityczne. Polega ono na tym, że system dzielony jest na podsystemy, których właściwości oraz prawa fizyczne nimi rządzące dają się opisać modelami matematycznymi. Metoda ta jest zależna od skali problemu, może być bardzo czasochłonna i prowadzić do uzyskania modeli matematycznych zbyt skomplikowanych, by nadawały się do dalszego wykorzystania.

**Sterowanie** - Stworzony model pozwala na syntezę [układu regulacji](#) poprzez wprowadzenie [regulatora sterującego](#) danym [obiektom](#) lub procesem tak, by ten zachowywał się w pożądanym sposób.

Wydaje mi się, że niektóre pojęcia może wprowadzać wikipedia dla hasła: system

## 17. Problemy decyzyjne dla kompleksu operacji.

Kompleks operacji jest obiektem złożonym, którego elementami są operacje, powiązane ze sobą na zasadzie kolejności czasowych, to znaczy rozpoczęcie wykonywania niektórych operacji może się rozpocząć po zakończeniu wykonywania innych operacji. Graficznym sposobem przedstawienia kompleksu operacji może być graf. Operacje wykonuje się po to, aby osiągnąć pewien określony cel, wyrażony zadaniem. Oczywiście do wykonania operacji



potrzebny jest realizator (zasób), który musi zostać przydzielony do danej operacji.

Podstawowym problemem decyzyjnym jest problem alokacji, który polega na rozdziale, czyli alokacji zasobów i zadań do operacji w kompleksie operacji. Jako kryterium alokacji zwykle przyjmuje się koszt lub czas wykonania kompleksu operacji.

Kolejnym problemem decyzyjnym jest problem szeregowania zadań. Problem ten jest szeroko znany i omawiany. Kluczowymi pojęciami występującymi w problematyce szeregowania są: zadanie, które wystąpiło już w poprzednim punkcie dotyczącym alokacji, oraz realizator rozumiany tu jako podmiot wykonujący zadanie i mogący mieć różną naturę i interpretację. Problem szeregowania można ogólnie określić jako wyznaczenie takiego dopuszczalnego przyporządkowania elementów jednego zbioru elementom drugiego zbioru, które jest najlepsze ze względu na przyjęte kryterium szeregowania.

Kolejny problem występuje w sytuacji w której nie mamy pewności co do zbiorów zasobów, zadań, operacji, czy parametrów tych elementów. Mamy do czynienia z tak zwanym problemem probabilistycznym (stochastycznym), w którym informacja o pewnych wielkościach jest określona z wykorzystaniem rozkładów prawdopodobieństwa i ma charakter stochastyczny.

Problem w sytuacji gdy mamy do czynienia z ruchomymi realizatorami. Występuje to w sytuacji gdy np. pracownicy na liniach produkcyjnych wykonują operacje na różnych liniach produkcyjnych w ramach jednego zadania. Wtedy do czasu realizacji zadania, dochodzi czas potrzebny do "przejścia" realizatora od jednego stanowiska do innego.

Bibliografia (sam, najjaśniejszy Józefczyk)

<http://www.dbc.wroc.pl/Content/1168/jozefczyk.pdf>

## 18. Podstawowe problemy, metody i algorytmy optymalizacji dyskretnej.

### OPTYMALIZACJA DYSKRETNA

Większość deterministycznych problemów planowania i sterowania w dyskretnych systemach wytwarzania jest formułowana jako zagadnienia optymalizacji, w których wszystkie zmienne decyzyjne (bądź ich część) przyjmują wartości dyskretne, całkowitoliczbowe lub binarne. Zadania takie często nazywa się problemami optymalizacji dyskretnej lub dyskretno-ciągłej, należą do klasy problemów wyjątkowo kłopotliwych z obliczeniowego punktu widzenia. Zagadnienia te sprowadzają się do zadania minimalizacji funkcji celu  $K(x)$  na zbiorze rozwiązań dopuszczalnych  $X$ , określonym przez zestaw warunków ograniczających. Głównymi powodami tych kłopotów są:

- częsty brak „klasycznych”, analitycznych własności (różniczkowalność, liniowość, itp.),
- wielo-ekstremalność ze znaczną liczbą ekstremów lokalnych.
- NP.-trudność większości problemów pochodzących z praktyki
- przekleństwo wielowymiarowości

Wielokrotnie, w celu uniknięcia kłopotów, próbuje się zamiast rozwiązywać problem dokładnie, wyznaczyć pewne jego rozwiązanie przybliżone. Dokładność tego przybliżenia posiada tendencję przeciwną do czasu obliczeń, tzn. uzyskanie dokładniejszego rozwiązania wymaga dłuższego czasu trwania algorytmu, przy czym ta ostatnia zależność

posiada charakter silnie nieliniowy. Jest to czynnikiem powstania wielu rodzajów zarówno modeli jak i metod rozwiązywania, zwykle dedykowanych dla wąskich klas zagadnień. Często dla tego samego problemu NP-trudnego występuje w literaturze kilka, kilkanaście różnych algorytmów o istotnie różnych cechach numerycznych.

Wprowadzenie dyskretnych zmiennych decyzyjnych umożliwia niekiedy względnie łatwe rozwiązanie zadań, w których zbiór rozwiązań dyskretnych nie jest zbiorem wypukłym lub też jest zbiorem niespójnym.

Metody rozwiązywania zadań optymalizacji dyskretnych można podzielić na dwie grupy:

- metody płaszczyzn tnących,
- metody heurystyczne.

Pierwsze dwie metody są metodami dokładnymi, metody heurystyczne dostarczają rozwiązań przybliżonych.

### METODA PODZIAŁU I OGRANICZEŃ (branch & bound)

Do rozwiązywania dyskretnych zadań decyzyjnych stosuje się tzw. metodę podziału i ograniczeń. Idea metody polega na tym, że tzw. przegląd zupełny (pełny) zbioru ograniczeń  $D$  zastępujemy przeglądem ukierunkowanym. Pozwala to ocenić pośrednio pewne podzbiory rozwiązań i ewentualnie je odrzucić lub czasowo pominąć, bez utraty rozwiązania optymalnego, co znacznie przyspiesza uzyskanie rozwiązania

### Metoda AHP

Analitycal Hierarchy Process „szkoła amerykańska” – Thomas L. Saaty

Działania w metodzie AHP da się ująć w trzech etapach.

I. Budowa macierzy porównań parami dla  $n$  obiektów osobno w ramach każdego kryterium (macierze  $(1) A$ ,  $(2) A$ , ...,  $(K) A$ ) oraz dla samych kryteriów (macierz  $(0) A$ ). Porównania te prowadzi do powstania  $K+1$  macierzy porównań parami ( $(0) A$ ,  $(1) A$ ,  $(2) A$ , ...,  $(K) A$ ). Ważnym uzupełnieniem etapu I jest badanie spójności ocen decydenta.

II. Wyznaczanie rankingów indywidualnych dla każdej z macierzy etapu I.

III. Wyznaczanie rankingów wielokryterialnych dla  $n$  obiektów.

### Metody przybliżone

Metoda przybliżona wyznacza pewne rozwiązanie bliskie rozwiązaniu dokładnemu. Metod przybliżonych jest zdecydowanie więcej niż dokładnych, zwykle są one problemowo-zorientowane. Zasadniczo, jakość metody przybliżonej jest oceniana z dwóch punktów widzenia: złożoność obliczeniowa algorytmu oraz dokładność przybliżenia. Dalsza charakterystyka bierze pod uwagę, między innymi, gwarancje zbieżności do rozwiązania optymalnego, szybkość tej zbieżności. Jakość wszystkich wymienionych ocen zależy od metody, problemu oraz konkretnych danych liczbowych podanych do algorytmu.

W ostatnich latach nastąpił burzliwy rozwój metod przybliżonych o dobrych i bardzo dobrych własnościach numerycznych potwierdzonych eksperymentalnie. Znaczna część tych metod czerpie swoje inspiracje z Natury i ma związek z dziedziną Sztucznej Inteligencji oraz Uczenia Maszynowego.

Rodzaje metod przybliżonych (heurystyk):

- algorytm zachłanny
- iteracyjne wspinanie się
- przeszukiwanie tabu

## 19. Podstawowe metody „obliczeń miękkich (inteligentnych)”.

Patrz odp. 30.

## 20. Podejmowanie decyzji w warunkach niepewności.

**Niepewność** jako pojęcie teorii decyzji oznacza sytuację, w której określone decyzje mogą spowodować różne skutki, w zależności od tego, który z możliwych stanów rzeczy zajdzie, przy czym nie są znane prawdopodobieństwa wystąpienia poszczególnych z nich.

Typy niepewności: (Slajdy Kwaśnickiej)

1. Wejściowe fakty są niepewne lub mają przypisane prawdopodobieństwa
2. Reguły, nawet kiedy fakty są absolutnie pewne, generują nowe nowe fakty z pewnym stopniem ufności; np., IF organism is a gram positive coccus growing in chains (100% certainty) THEN organism is a streptococcus (with 70% certainty)
3. Kombinacje przypadków 1 i 2

### **Rachunek prawdopodobieństwa jest formalnym, poprawnym mechanizmem wnioskowania**

Lecz:

Stosowanie rachunku prawdopodobieństwa wymaga od użytkownika dostarczenia pewnej liczby prawdopodobieństw warunkowych Nawet niezależne fakty początkowe nie propagują niezależności w procesie wnioskowania (patrz przykład)

Wnioskowanie z założeniem niezależności produkuje rygorystyczne (złe) wyjścia

### **Certainty Factor (CF) – Czynniki Pewności**

$$CF(h) = MB(h) - MD(h)$$

(MB – Measure of Belief),

(MD – Measure of Disbelief)

0 MB 1; 0 MD 1; -1 CF 1

- Ważne cechy CF:
  - Przesuwa zaufanie produkowanych hipotez asymptotycznie do pewności kumulując miary otrzymywane z kolejnych reguł produkujących rozważane hipotezy.
  - CF jest symetryczną miarą, tzn., jest niezależny od uporządkowania zapalanych reguł
- Lecz:
  - $CF=0$  – niemożliwe jest rozróżnienie czy MB i MD są prawie takie same (konflikt) czy obie miary bliskie zeru (hipoteza nie może być potwierdzona lub zanegowana)

### **Logika rozmyta** - wszystkim znana

Formalnie, **decyzjami podejmowanymi w warunkach niepewności** nazywamy taką klasę problemów decyzyjnych, w której dla przynajmniej jednej decyzji nie jest znany rozkład prawdopodobieństwa konsekwencji.

*Przykład:* Mamy pomysł na nowy produkt i chcemy zdecydować, czy otworzyć firmę zajmującą się produkcją i sprzedażą tego produktu. Nie jesteśmy w stanie określić

prawdopodobieństwa sukcesu naszej firmy, jednak pomimo tego decydujemy się zaryzykować. Podjęliśmy decyzję w warunkach **niepewności**.  
W praktyce prawie zawsze w wypadku niepewności określamy prawdopodobieństwo subiektywne zajścia danej konsekwencji.

Ze względu na posiadane informacje, możemy podzielić problemy decyzyjne na trzy grupy:

- **decyzja podejmowana w warunkach pewności** – każda decyzja pociąga za sobą określone, znane konsekwencje
- **decyzja podejmowana w warunkach ryzyka** – każda decyzja pociąga za sobą więcej niż jedną konsekwencję, znamy zbiór możliwych konsekwencji i prawdopodobieństwa ich wystąpienia
- **decyzja podejmowana w warunkach niepewności** – nie znamy prawdopodobieństw wystąpienia konsekwencji danej decyzji.

Metody podejmowania decyzji w warunkach niepewności:

- zbiory przybliżone
- sieci bayesowskie

**Teoria zbiorów przybliżonych** – Zbiór przybliżony (ang. *rough set*) to obiekt matematyczny zbudowany w oparciu o logikę trójwartościową. W swym pierwotnym ujęciu zbiór przybliżony to para klasycznych zbiorów: przybliżenie dolne i przybliżenie górne. Istnieje również odmiana zbioru przybliżonego, definiowana przez parę przybliżeń będących zbiorami rozmytymi (ang. *fuzzy set*). Dany element może należeć do obydwu przybliżeń, do żadnego lub tylko do przybliżenia górnego. Ten ostatni przypadek jest o tyle ciekawy, że pozwala na modelowanie niepewności.

sieci bayesowskie:

Sieć Bayesowska to acykliczny graf (DAG, Directed Acyclic Graph), składający się z: zbioru wierzchołków odpowiadających zmiennych zbioru skierowanych krawędzi łączących pary węzłów – intuicyjne znaczenie połączenia od węzła A do B oznacza, że A bezpośrednio wpływa na B

- graf nie zawiera cykli
- Węzły, od których dochodzą krawędzie do danego węzła to węzły rodzicielskie
- Każdy węzeł zawiera tabelę prawdopodobieństw warunkowych, określających wpływ węzłów 'rodzicielskich' na dany węzeł Cechy BN
- łatwiej ekspertom określić bezpośrednie zależności warunkowe w dziedzinie, niż podawać aktualne prawdopodobieństwa
- po zbudowaniu topologii BN, należy określić prawdopodobieństwa warunkowe dla węzłów, które są ze sobą bezpośrednio połączone; te prawdopodobieństwa są wykorzystywane do obliczania każdego innego prawdopodobieństwa

## 21. Metody i algorytmy rozpoznawania.

**Rozpoznawanie obiektów** – przypisanie do klasy na podstawie cech (algorytm rozpoznawania)

**schemat:** obiekt opisany cechami → pomiar cech → przetwarzanie cech → klasyfikacja

## Rozpoznawanie w oparciu o ciąg uczący – rozpoznawanie z nauczycielem

### Algorytmy odległościowe:

- **Algorytm NM** (najbliższej średniej) – obliczamy wartość średnią dla każdej klasy. Obliczamy odległość  $x$  od każdego elementu średniego, bierzemy min odległość i przypisujemy do klasy odpowiadającej tej wartości średniej
- **Algorytm KNN** (k najbliższych sąsiadów) – liczymy odległość  $x$  od każdego elementu ciągu uczącego, sortujemy w porządku niemalejącym, bierzemy  $k$  pierwszych pozycji i wybieramy najczęściej występującą klasę

### Algorytmy probabilistyczne

- **Naiwny Bayes**

### Drzewa klasyfikacyjne

algorytm (ogólny) - 2 kroki:

- 1) reguła podziału wierzchołków
- 2) reguła uznania wierzchołka za końcowy oraz wiążąca się z tym reguła przypisania wierzchołka

### Algorytm ID.3 (alg. budowy drzewa klasyfikacyjnego)

- 1) Dla każdej cechy liczymy entropię warunkową (w tym celu wyliczamy prawdopodobieństwa, że przykład przybierze daną wartość cechy oraz prawdopodobieństwa warunkowe takie że dla danej wartości cechy przykład będzie należał do klasy oraz dla danej wartości cechy przykład nie będzie należał do klasy)
- 2) Wybieramy cechę o najmniejszej wartości entropii i umieszczamy jako wierzchołek. Od wierzchołka odchodzą krawędzie etykietowane wartościami cechy (teraz będziemy brać pod uwagę tylko przykłady zawierające wartość cechy z tej krawędzi) itd.

## 22. Postulaty metodologii nauk.

Metodologia nauk – nauka zajmująca się metodami stosowanymi przy formułowaniu twierdzeń i teorii naukowych. Metodologia nauk analizuje nie tylko procedury badawcze, lecz także jej wytwory: pojęcia, hipotezy, twierdzenia.

Metodologia nauk, w aspekcie pragmatycznym – nauka o metodach działalności naukowej i stosowanych w nauce procedurach badawczych; w aspekcie teoretycznym – nauka o elementach i strukturze systemów naukowych

•Metodologia nauk nie prowadzi badań empirycznych „w terenie”, nie obserwuje rzeczywistych, konkretnych naukowców przy pracy (A)

•Metodologia nauk pyta o procedury, schematy postępowania akceptowane w nauce w ogóle i w poszczególnych dyscyplinach naukowych (B)

–jako źródła poznania, wiedzy

–jako sposoby uzasadniania stwierdzeń, wiedzy

•Pytanie – skoro metodologia nie zajmuje się (A), to skąd właściwie wie jakie są (B)?

–Nie wie, postępuje apriorycznie, jest działem logiki, zajmuje się metodami poprawnego myślenia w ogóle, m.in. bada wnioskowanie od zdania do zdania, mówi raczej o tym jakie powinny być właściwe metody prowadzące do wiedzy uzasadnionej (Bocheński, 1992, s. 20-25)

–Odnosi się do nauki jako mega-faktu, rzeczywiście istniejącego (Bocheński, 1992, s. 139-142)

Postulaty:

- falsyfikowalność - jakkolwiek żadne doświadczenie nie może wykazać prawdziwości teorii, to istnieją doświadczenia, które mogą wykazać jej fałszywość
- sprawdzalność - wiedza naukowa poddaje się sprawdzeniu, kontroli; powinno być powiedziane w jaki sposób doszliśmy do takich, a nie innych stwierdzeń, wniosków, przy użyciu jakich metod, na podstawie jakich danych, a także jak inni mogą sprawdzić nasze wyniki
- prostota - wiedza naukowa powinna być zaprezentowana w sposób możliwie najprostszy, dostępny dla jak najszerzej grupy odbiorców (???)
- bezstronność - przeciwstawienie tendencyjności, obiektywna sprawdzalność, niezależność od podmiotu sprawdzającego

### 23. Współczesne metody naukometrii

Naukometria (naukozn. dziedzina naukoznawstwa) - zajmuje się badaniem rozwoju nauki jako procesu informacyjnego; stosuje metody statystyczno-ilościowe (liczba publikacji, przyznanych stopni nauk. i nagród, placówek nauk.), pozwalające na określenie aktualnego stanu danej dyscypliny nauk. i prognozowanie perspektyw jej rozwoju.

Dużym wkładem w rozwój współczesnej naukometrii wykazało się czasopismo Journal Citation Reports zawierające informacje o czasopismach naukowych a także wskaźniki ich jakości. Czasopismo to, jak również wspomniane wskaźniki są częścią tzw. Science Citation Index powstałego w 1960 roku, posiadającego informacje nt. cytowań artykułów z największych czasopism naukowych.

Współcześnie do oceny jakości czasopism używa się następujących wskaźników:

- Impact factor - stosunek liczby cytowań do artykułów z czasopisma z ostatnich dwóch lat do liczby artykułów w tych latach, czyli średnia ilość cytowań do artykułów w ostatnim czasie
- Immediacy Index - stosunek liczby cytowań do artykułów z czasopisma z aktualnego roku do liczby artykułów opublikowanych w tym roku
- Half-Life cytujących - mediana wieku artykułów, które były cytowane przez artykuły w czasopiśmie (w danym roku)
- Half-Life cytowanych - mediana wieku artykułów z czasopisma, które były cytowane przez artykuły w innych czasopismach (w danym roku)
- liczba cytowań
- liczba artykułów

PS. Poniżej zamieściłem 3 różne miary stosowane w naukometrii: co prawda tekstu trochę jest, ale chodzi o to żeby każdą z tych miar zrozumieć. Wydaje mi się, że ich zrozumienie jest wystarczające żeby mówić i mówić

**Impact factor (IF)** – w tłumaczeniu „Miara oddziaływania” – to wskaźnik prestiżu i siły oddziaływania czasopism naukowych, ustalany przez Instytut Filadelfijski (Institute of Scientific Information, obecnie części koncernu wydawniczego Thomson), na podstawie prowadzonego przez ten instytut indeksu cytowań publikacji naukowych. Zazwyczaj jest tak, że jeśli dana publikacja wnosi coś istotnego do nauki to jest też często cytowana przez autorów innych publikacji. Stąd, liczba cytowań danej publikacji jest dobrą miarą jej wartości. Rozciągając ten tok rozumowania na czasopisma, można logicznie uznać, że średnia liczba cytowań wszystkich artykułów, które się w danym czasopiśmie ukazały jest dobrą miarą prestiżu i siły oddziaływania tego czasopisma.

IF jest ustalane wg wzoru:

$$IF = B/C$$

gdzie

**B** – to łączna lista cytowań które nastąpiły w danym roku kalendarzowym.

**C** – to liczba cytowalnych publikacji (zwykle nie są brane pod uwagę takie publikacje jak np. listy do redakcji), które ukazały się w danym czasopiśmie, w ciągu ostatnich dwóch lat.

Krytycy tego wskaźnika twierdzą, że:

- Jest on zanadto „mechaniczny” i często pokazuje nie tyle wartość naukową czasopism (i publikacji) lecz raczej aktualnie panujące mody i trendy w nauce. Czasopismo specjalizujące się w dziedzinie, która jest aktualnie modna siłą rzeczy publikuje artykuły, które opisują też „modne” badania. W modnych dziedzinach panuje większy ruch niż w niemodnych, co się przekłada bezpośrednio na liczbę cytowań.
- IF jest tworzony przez instytucję, która sama jest nastawiona na przynoszenie dochodu i ma ona w pewnym sensie „władzę” kreowania trendów w nauce – choćby poprzez arbitralne decyzje dopisywania lub wykreślania czasopism ze swojej listy. Istnieje więc ryzyko, że trendy te są kreowane w taki sposób, aby obracało się to na korzyść samego Instytutu – na zasadzie samospełniających się proroctw.
- Bardzo często podnoszony jest argument, że IF preferuje badania, które są modne w USA. Istotnie ponad 50% czasopism na liście Instytutu Filadelfijskiego jest wydawana w USA, zaś ponad 80% czasopism na tej liście to czasopisma anglojęzyczne. Argument ten jest szczególnie istotny w naukach społecznych i ekonomicznych, których większa część odbywa się w innych obszarach językowych.

**PIF - Wydaje się, że Przewidywalny Impact Factor (predicted impact factor)** to „miara” przeznaczona dla czasopism, która już została zaindeksowana przez Thomson Reuters, ale nie mają jeszcze wyznaczonego wskaźnika Impact Factor. W wyjaśnieniu (FAQ) do nowych zasad jest napisane:

Kto będzie obliczał index cytowań PIF (Przewidywalny Impact Factor)? PIF będzie obliczany automatycznie przez program po zasięgnięciu liczby cytowań czasopism naukowych umieszczonych w bazach Thomson Reuters Scientific.

ŹRÓDŁO: [FAQ](#)

Z tego można wyczytać (?), że czasopismo, które nie jest w tej bazie, nie może mieć obliczonego PIF. Ale nie dyskwalifikuje to czasopisma: może otrzymać punktację na „wykazie B”.

**Index h** - <http://www.ebib.info/2008/92/a.php?rek>

Według definicji wydawcy SCI Expanded, *indeks h* to liczba naturalna, określająca, ile spośród wyszukanych dokumentów według zadanego kryterium było tyle samo lub więcej razy cytowanych. Podany przykład  $h=20$  mówi, że w badanym zbiorze jest 20 dokumentów - co najmniej 20 razy cytowanych.

#### 24. Organizacja systemów rozproszonych.

System rozproszony to układ wielu jednostek widzianych przez użytkownika jako jedna jednostka (cechują się przezroczystością).

Ze względu na sposób połączenia poszczególnych jednostek wyróżnia się rozproszone systemy:

- połączone luźno (wielokomputery) - np. klastry obliczeniowe, składające się z jednostek heterogenicznych (różnego typu, z różnymi systemami operacyjnymi, itp.), posiadających własną pamięć, połączone ze sobą w sieć. W zależności od rodzaju rozwiązywanego problemu stosuje się układy połączeń: pełne, gwiazda, magistrala, pierścień, siatka, hiperkostka i inne.
- połączone ściśle (wieloprocесory) - np. rdzenie procesora lub rdzenie karty graficznej, składające się z jednostek tego samego typu, połączonych w sposób ściśle ustalony i niezmienny. Dane przesyłane są znacznie szybciej, często za pośrednictwem wspólnej pamięci.

Ze względu na architekturę rozróżnia się systemy typu:

- SISD - Single Instruction Single Data - naraz wykonywana jest jedna instrukcja dla jednego strumienia danych (standardowe CPU)
- SIMD - Single Instruction Multiple Data - jedna instrukcja wykonywana jest dla wielu danych (karty graficzne, procesor wektorowy CELL)
- MISD - Multiple Instruction Single Data - wiele instrukcji wykonywanych dla jednego strumienia danych, układ nie występujący w praktyce
- MIMD - Multiple Instruction Multiple Data - wiele strumieni instrukcji przetwarza wiele strumieni danych (komputery równoległe i rozproszone)

#### 25. Ocena systemów równoległych.

Dobry system równoległy/rozproszony powinien być skalowalny - jego osiągi nie powinny spadać podczas zwiększania rozmiaru problemu obliczeniowego, a powinny rosnać w miarę rozbudowywania warstwy sprzętowej. Do oceny tych osiągnięć stosuje się miary zwane przyspieszeniem i efektywnością.

**Przyspieszenie** jest stosunkiem czasu wykonania programu na jednym procesorze do czasu wykonania go na wielu procesorach ( $S_p = T_1 / T_p$ ).

**Efektywność** to iloraz przyspieszenia i liczby procesorów tego samego typu, za pomocą których to przyspieszenie to osiągnięto ( $E = S_p / p$ ). Wskazuje ona poziom wykorzystania dostępnych procesorów (przyspieszenie na procesor).



Gene **Amdahl** sprecyzował sposób obliczania maksymalnego oczekiwanego przyspieszenia systemu po przyspieszeniu jego części.

Metoda ta, znana jako prawo Amdahla znalazła zastosowanie w systemach równoległych, umożliwiając szacowanie maksymalnego przyspieszenia programu składającego się z części sekwencyjnej i równoległej. Narzuca ono górną granicę przyspieszenia dla programu składającego się z części równoległej i sekwencyjnej (niemożliwej do zrównoleglenia) ograniczonego właśnie częścią sekwencyjną.

**Gustafson** argumentował nieprawdziwość prawa Amdahla tym, że dla większej ilości procesorów zwykle zwiększamy rozmiar problemu, co z kolei powoduje zwiększenie ilości obliczeń możliwych do zrównoleglenia. W ten sposób próbował znieść "mentalną blokadę" narzuconą wg niego przez prawo Amdahla.

Powyższe miary ze względu na założenie, że użyte procesory są tego samego typu (podczas wykonywania programów zarówno sekwencyjnych i równoległych), nie mogą zostać zastosowane w obliczeniach na kartach graficznych - jednostki tam dostępne uniemożliwiają uruchomienie na nich wersji sekwencyjnej.

## 26. Uczenie indukcyjne: metody, zastosowania.

Uczenie indukcyjne prowadzi do znajdowania hipotez najlepiej wyjaśniających i uogólniających obserwowane fakty. W przypadku uczenia indukcyjnego fakty te często nazywane są przykładami.

Wyróżniamy następujące rodzaje indukcyjnego uczenia się, które jednocześnie warunkują jego zastosowania:

- uczenie się pojęć - pojęcia służą klasyfikacji obiektów do kategorii, np. pojęcie krzesła umożliwia nam wskazanie go pośród innych mebli (klasyfikacja)
- tworzenie pojęć - uczeń na podstawie zaobserwowanych przykładów, grupuje je samodzielnie w kategorie zgodnie z pewnymi kryteriami podobieństwa zależnymi od algorytmu (klasteryzacja)
- uczenie się aproksymacji funkcji - gdzie zbiorem wartości podlegających uczeniu nie jest zbiór kategorii jak w uczeniu się pojęć, a zbiór liczb rzeczywistych, przykłady trenujące złożone są z argumentu funkcji (wektora) i jej wartości dla tego argumentu (regresja?)

Tryby uczenia:

- inkrementacyjny - przykłady dostarczane pojedynczo, udoskonalenie hipotezy po każdym z nich
- epokowy - cykl uczenia podzielony na epoki, w którym hipotezę udoskonala się po przetworzeniu danej liczby przykładów
- wsadowy - uczeń otrzymuje wszystkie przykłady, po przetworzeniu których podaje gotową hipotezę, nie można odczytać hipotez w trakcie procesu uczenia. W przypadku zmiany zbioru przykładów uczenie przeprowadza się od nowa.
- korekcyjny - odmiana inkrementacyjnego, w którym uczeń najpierw przedstawia swoją odpowiedź na podany przykład, a następnie otrzymuje informację korygującą, powstałą na podstawie różnicy wartości otrzymanej od oczekiwanej

Przykładem zastosowania jest klasyfikacja obrazów na podstawie wyodrębnionych z nich wektorów cech, a przykładem implementacji - jednokierunkowe warstwowe sieci neuronowe (perceptrony).

## 27. Uczenie ze wzmocnieniem.

Nie jest uczeniem indukcyjnym czyli generalizacją dużej liczby przykładów

Nie jest to inna forma uczenia się pojęć

Nie jest to uczenie z nauczycielem (choć tu niektórzy uważają inaczej)

Jest to uczenie się umiejętności (tzw. wiedzy proceduralnej)

Źródłem informacji trenującej jest krytyk, nie będący nauczycielem. W uczeniu nadzorowanym nauczyciel podaje informację trenującą w formie gotowej odpowiedzi (np. klasy obiektu). Z kolei w przypadku uczenia ze wzmocnieniem uczeń najpierw daje odpowiedź na podstawie dotychczasowej strategii, a następnie krytyk ją ocenia.

Krytyk jest częścią środowiska, w odróżnieniu od ucznia, który środowiska nie zna, nie kontroluje i nie jest go pewnym. Uczeń poznaje środowisko jedynie za pośrednictwem wykonywanych przez siebie akcji.

Proces uczenia podzielony jest na kroki wykonywane w dyskretnych przedziałach czasowych, a każdy z kroków składa się z następujących etapów:

- obserwacja aktualnego stanu  $x(t)$
- wybór akcji na podstawie strategii ( $a = PI(x(t))$ ), strategia dla podanego stanu zwraca akcję
- wykonanie akcji, a następnie zaobserwowanie otrzymanej nagrody/wzmocnienia  $r(t)$
- obserwacja stanu  $x(t+1)$  po wykonaniu akcji
- zmiana strategii na podstawie  $\langle x(t), a(t), r(t), x(t+1) \rangle$

Zmiana strategii powinna powodować naukę celowego zachowania. Cel wyznaczany jest otrzymywanymi nagrodami, a dążenie do niego wiąże się z maksymalizacją kryterium sumy nagród (ważonej). Maksymalizacja ta może być długo lub krótkoterminowa, odpowiednio gdy uczeń pamięta wszystkie poprzednie nagrody lub tylko kilka ostatnich.

Poprawianie strategii może następować po każdym kroku (tryb uczenia inkrementacyjny) lub po epoce uczenia (tryb epokowy).

Formalnym modelem środowiska jest proces decyzyjny Markowa (procesy stochastyczne w środowisku). Rozwiązaniem problemu uczenia ze wzmocnieniem jest poznanie prawdopodobieństw i wyznaczenie optymalnej strategii na tej podstawie - wartościowanie strategii. Używa się w tym celu metod:

- programowania dynamicznego - metoda analityczna, wymaga pełnej wiedzy o procesie, daje w wyniku strategię optymalną, rozwiązanie za pomocą równań Bellmana
- różnic czasowych TD - metody numeryczne, szybsze i prostsze, wynik przybliżony, algorytmy AHC, Q-learning, SARSA.

## 28. Modele obrazów: wektory cech, niezmienniki, relacje przestrzenne.

Wektor cech obrazu to zestaw liczbowych miar opisujących dany obraz (przykład).

Ekstrakcja cech stanowi syntezę danych wizualnych, mająca na celu wydobyć informacji istotnej dla użytkownika

Cechy niskopoziomowe:

- Wyznaczane są bezpośrednio z danych rastrowych.

- Sa cechami ogólnymi, charakteryzują dowolne obrazy.
- Nie mają bezpośredniego przełożenia na semantykę.

Cechy wysokopoziomowe:

- Sa zaprojektowane do wydobywania specjalizowanej informacji.
- Sa cechami dostosowanymi do konkretnej dziedziny obrazów.
- Mają bezpośrednie przełożenie na semantykę.
- Sa wyznaczane przez złożone metody analizy obrazu.

Globalne:

- Jeden obszar zainteresowania obejmujący cały obraz.
- Charakteryzują cały obraz, pomijając szczegóły.

Semi-globalne, semi-lokalne:

- Reprezentują duże i istotne części obrazu.
- Zazwyczaj jest kilka do kilkunastu regionów.
- Mogą odzwierciedlać obiekty znajdujące się na obrazie.
- Mogą dzielić obraz sztywny, ustalony sposób.

Lokalne:

- Reprezentują bardzo małe fragmenty obrazu.
- Wiele cech lokalnych składa się na jeden obiekt.
- Bardzo często nakładają się na siebie (nie są rozłączne).
- Często mają ściśle określone formy geometryczne (np. elipsy).
- Zazwyczaj jest ich kilkaset do kilku tysięcy.

Podział cech ze względu na syntezowane dane

Kolor:

- Rozkład koloru ogólnie charakteryzuje obraz.
- Różne modele barw niosą różną informację.
- Wszystkie dziedziny, w których kolor stanowi istotny nośnik informacji.

Tekstura:

- Charakteryzuje ziarnistość obrazu.
- Charakteryzuje powtarzalność wzorców.
- Może być powiązana ze znaczeniem obrazu.

Kształt:

- Charakteryzuje geometrie regionów zainteresowania.
- Ma znaczenie tylko gdy regiony zainteresowania odzwierciedlają obiekty widoczne na obrazie.

Niezmienne - stałe cechy postrzeganych obiektów

- dzięki ich detekcji możliwe jest rozpoznawanie przedmiotu w sytuacji, kiedy jest on niekompletny (częściowo zakryty, zdeformowany);
- wysoka tolerancja na takie zmiany w rozpoznawaniu przedmiotu wskazuje na dużą wrażliwość systemu percepcyjnego w wykrywaniu niezmienników.

*Relacje przestrzenne* określają względne pozycje obiektów na obrazie (*na lewo, pod, z przodu*).

## 29. Klasyfikacja, opisywanie, interpretacja, rozumienie.

“najniższym i najmniej skomplikowanym poziomem jest klasyfikacja, gdzie przypisujesz jedną klasę do obrazu, potem opisywanie, gdzie wskazujesz obiekty poprzez przypisanie im etykiet, potem interpretacja, gdzie nakreślasz jaki jest stosunek opisanych obiektów względem siebie, a na końcu rozumienie, gdzie na podstawie zinterpretowanej sytuacji wyciągasz wiedzę”

© Karol Radziszewski

### (poniższe informacje ze slajdów paradowskiego)

Rozpoznawanie obrazu (klasyfikacja) jest to przypisanie do obrazu, na podstawie jego cech, jednej klasy ze ściśle określonego zbioru klas

Luka semantyczna jest brakiem zgodności pomiędzy informacją, którą można wydobyć z danych wizualnych a interpretacją użytkownika tych danych w zadanej sytuacji.

Opisywanie obrazu jest to przypisanie do obrazu, na podstawie jego cech, podzbioru klas ze ściśle określonego zbioru klas (słownika).

- Z założenia działa na danych o dużych wolumenach.
- Nie wymaga danych wysokiej jakości:
  - opis obrazu może być niekompletny (brakuje pewnych klas),
  - opis obrazu może być niespójny (na skutek opisywania przez wiele osób).
- Nie wymaga wcześniejszej, ręcznej pracy nad:
  - segmentacją obrazów,
  - rozpoznawaniem poszczególnych segmentów.
- Ma bardzo duży aspekt praktyczny.
- Zazwyczaj daje gorsze wyniki, niż rozpoznawanie ze zbiorem uczącym.
- Bardzo duże znaczenie odgrywa również kontekst, to on pozwala na podjęcie ostatecznej decyzji z jakim obiektem na zdjęciu mamy do czynienia. Pozwala na rozróżnienie obiektów, które bez jego analizy były by nierozróżnialne

## 30. Obliczenia miękkie a obliczenia tradycyjne (algorytmiczne).

Obliczenia miękkie w przeciwieństwie do tradycyjnych nie mają na celu ustalenia dokładnego rozwiązania problemu a jedynie jego oszacowanie z pewną dokładnością. Powoduje to, że obliczenia np. z obszaru problemów NP-trudnych mogą zostać oszacowane z zadowalającą dokładnością w krótkim czasie, w przeciwieństwie do analitycznych metod o wykładniczym czasie działania.

Ze względu na działania biorące pod uwagę niepewność, nieprecyzyjność, przybliżenia i częściową zgodność z prawdą, obliczenia miękkie przypominają procesy zachodzące w naturze i z tych procesów uczeni korzystają w modelowaniu nowych rozwiązań z tej dziedziny.

Tyle, jeśli chodzi o różnice, po techniki zapraszam do pkt 30.

31. Jakie techniki wchodzą w skład obliczeń miękkich. Charakterystyka każdej z nich.

**- Alg. bazujące na naturze:**

- **immunologiczne** – stanowią odpowiednik realizacji procesu adaptacji i dywersyfikacji naturalnego systemu immunologicznego. Ich zadaniem jest, poprzez sterowanie populacją przeciwciał, doprowadzenie do otrzymania rozwiązania. Algorytmy immunologiczne można podzielić na populacyjne (selekcja klonalna, selekcja negatywna i sieciowe (sieć idiotypowa). Algorytm *selekcji klonalnej* składa się z dwóch etapów: ekspansji klonalnej oraz hipermutacji. Pierwszy odpowiedzialny jest za wyselekcjonowanie najlepiej dopasowanych przeciwciał i ich sklonowanie. Hipermutacja natomiast realizuje dojrzewanie przeciwciał w celu jeszcze lepszego ich dopasowania.

Algorytm *selekcji negatywnej* stosuje się w celu wyeliminowania przeciwciał, które rozpoznają własne struktury jako obce. Algorytm ten znajduje zastosowanie najczęściej w problemach znajdowania anomalii.

Model *sieci idiotypowej* proponuje system immunologiczny charakteryzujący się dynamicznym działaniem nawet w przypadku braku antygenów ciał obcych. Model ten różni się od selekcji, czyniąc limfocyty zdolne do rozpoznawania siebie nawzajem. Znajdują one zastosowanie głównie w problemach rozpoznawania obrazów, analizie danych, maszynowym uczeniu i problemach optymalizacji.

- **rojowe** (mrówkowe, pszczele, świetlikowe, kukułcze) - wywodzące się z algorytmu optymalizacji kolonii cząstek - particle swarm optimization PSO. W zależności od stworzenia, na którego obserwacjach bazuje algorytm, przeszukiwanie dziedziny i wybór najlepszych rozwiązań następuje w różny sposób, wszystkie mają jednak jedną bazę i w zapisie formalnym różnią się nieznacznie.

- **oparte o teorię chaosu** - bazujące na deterministycznych układach, tak wrażliwych na drobne zmiany warunków początkowych, że podczas obserwacji uchodzą za działające losowo (niedeterministycznie)

- **oparte o transformatę falkową** - będącą rozwinięciem transformaty Fouriera dla sygnałów niestacjonarnych, czyli zmieniających widmo w czasie. Transformata falkowa pozwala na wyznaczenie momentu występowania poszczególnych składowych harmonicznych.

- **Logika rozmyta typu 2** - wprowadzająca stany pośrednie w logice, umożliwiające określenie stopnia przynależności obiektu do zbioru.

Dzieli się na przedziałowe i uogólnione:

- przedziałowe: współczynniki przynależności są przedziałami ostrymi, umożliwiają modelowanie niepewności w przeciwieństwie do logiki rozmytej typu 1, a operacje and i or są proste i szybkie,

- uogólnione: współczynniki przynależności nie są ostre, należą do zbioru rozmytego. Podejście na razie raczej teoretyczne ze względu na słabo poznane zasady matematyczne i złożoność obliczeniową (zamiast operacji and i or - meet i join, bardzo wymagające obliczeniowo).

Przykładowo:

- typ-1: Karol jest w 0.72 wysoki,

- typ-2 przedziałowy: Karol jest w [0.62-0.82] wysoki, rozkład przynależności równomierny

- typ-2 uogólniony: Karol jest w  $N(0.72, 0.1)$  wysoki, rozkład przynależności normalny (mam nadzieję że nie zakręciłem) czy jakkolwiek inny niż równomierny

- **Techniki agentowe** - system oparty o autonomiczne byty zwane agentami, rozproszone podejście do rozwiązania problemu.

Agent:

- oddziałuje na środowisko,
- komunikuje się z innymi agentami,
- działa realizując wyznaczone cele,
- ma dostęp i dysponuje zasobami,
- posiada jakiś zbiór umiejętności,
- posiada ograniczoną percepcję,
- posiada wiedzę nt środowiska lub ją gromadzi
- czasami może się rozmnażać (pewnie jak spotka ładną agentkę :) )

- **Zbiory przybliżone** - rough sets, dla zbiorów o nieregularnych zakresach definiujemy przybliżenie górne i dolne o zakresach regularnych. Dzięki temu możemy określić nieostre pojęcie w ścisły sposób. Przynależność sprawdza się na podstawie klas równoważności  $R$ , zwanych atomami. Obiekty należące do tej samej klasy równoważności są nierozróżnialne.

- aproksymacja dolna: składa się z obiektów, które z całkowitą pewnością należą do zbioru  $X$

- aproksymacja górna: składa się z obiektów, które MOGĄ należeć do zbioru  $X$
- obszar brzegowy: różnica między aproksymacją dolną i górną
- zbiór dokładny (crisp): obszar brzegowy nie zawiera żadnych obiektów
- zbiór przybliżony (rough): obszar brzegowy zawiera jakieś obiekty

- **Hybrydy** - rozwiązania powstałe poprzez połączenie powyższych z innymi rozwiązaniami, np. rozmyte sieci neuronowe, transformata falkowa połączona z sieciami lub też połączenie między sobą itd.

- prócz tego: **sieci neuronowe, algorytmy ewolucyjne, heurystyki, sieci bayesowskie...**

## 32. Podstawy formalnego opisu języka naturalnego: założenia i stosowane metody.

Formalizacja w opisie języka naturalnego

- morfologia:
  - gramatyki formalne poziomu znaków i morfemów
- składnia:
  - gramatyki formalne
- semantyka:
  - opis znaczeń leksykalnych (słów) za pomocą grafu relacji
  - logika jako podstawa reprezentacji znaczenia
- pragmatyka
  - modele oparte na logice i formalnej reprezentacji wiedzy
  - modele wieloagentowe

formalne:

- Gramatyka regularna
- Automat skończony
- Transduktor

## Reprezentacja znaczeń leksykalnych

### Postulaty znaczeniowe — aksjomaty logiczne

#### Reprezentacja dekompozycyjna

- wyrażenia odwołujące się do znaczeń atomowych
- zalety: precyzja, duża siła ekspresji
- wady: pracochłonna, problem spójności

#### Leksykalna sieć semantyczna

- opisująca leksykalne relacje znaczeniowe pomiędzy poszczególnymi jednostkami leksykalnymi
  - jednostka leksykalna — jedno lub wielowyrazowy leksem, reprezentowany jako para ⟨lemat, sens⟩, np. zamek 1,
- zalety: efektywność budowania, generalizacja
- wady: opis nie wprost, ograniczona siła ekspresji

33. Współczesna technologia językowa: narzędzia, zasoby językowe i ich zastosowania.

#### Zasoby językowe:

- uporządkowane zasoby danych językowych, najczęściej anotowanych (opatrzonej metaopisem) lub bazy wiedzy opisujące język naturalny
- prowadzone są prace nad standaryzacją zasobów językowych
- przykłady anotowanych danych językowych
  - korpus tekstu opisanego morfologicznie i ujednoliconego
  - korpus tekstu z oznaczeniem sensów poszczególnych wystąpień wyrazów
- przykłady baz wiedzy:
  - gramatyka formalna opisująca podzbiór języka
  - leksykon semantyczny zawierający sformalizowane opisy znaczeń słów
  - model językowy opisujący prawdopodobieństwo występowania ciągów wyrazów

#### Narzędzia językowe

specjalistyczne programy do przetwarzania danych językowych

przykłady:

- analizatory morfologiczne, tagery, parsery
- konwertery, programy tłumaczące

- a. WordNet - Słowność
- b. Takipi
- c. Odgadywacz

Korpus to dowolny zbiór tekstów, w którym czegoś szukamy. O korpusach w tym znaczeniu mówią najczęściej językoznawcy, ale także archiwiści, historycy i informatycy.

Korpus tekstów polskich to fragment słownikowej kuchni, czyli autentyczny materiał językowy, na którego podstawie opisujemy znaczenia słów i konstrukcji. Zrównoważenie korpusu jest równie ważne jak jego wielkość. Drobne fragmenty tekstów korpusu są zamieszczane w postaci pojedynczych zdań w słownikach jako przykłady ilustrujące znaczenia.

Dzięki próbkę naszego redakcyjnego korpusu w sieci teraz i Państwo mogą zobaczyć, w jakich kontekstach używa się danego słowa w różnych typach tekstów. Zachęcamy do sprawdzenia pod tym kątem adekwatności opisów w naszych słownikach!

Wydawnictwo Naukowe PWN przygotowało i udostępniło **sieciową wersję Korpusu Języka Polskiego PWN wielkości 40 milionów słów**. Korpus składa się z fragmentów 386 różnych książek, 977 numerów 185 różnych gazet i czasopism, 84 nagranych rozmów, 207 stron internetowych oraz kilkuset ulotek reklamowych.

Pełna wersja sieciowa korpusu jest dostępna odpłatnie, a bezpłatnie **wersja demonstracyjna wielkości ponad 7,5 miliona słów**.

#### 34. Proces przetwarzania języka naturalnego: typowe etapy, cele, stosowane metody

- a. 1 - wyodrębnienie tekstu z dokumentu,
- b. 2 - segmentacja (tokenizacja),
  - Gramatyka regularna
  - Automat skończony
  - Transduktor
- c. 3 - analiza morfologiczna,
  - Steming
  - Lematyzacja
  - Odgadywacz
- d. 4 - ujednoznacznianie morfo-syntaktyczne,
  - Statystyczne,
  - indukcyjne,
  - ręczne,
  - regułowe,
  - oparte na pamięci
- e. 5 - ujednoznacznianie sensu słów
  - TAKIPI
- f. 6 - płytki parsing,
  - chunking
  - Przetwarzanie częściowe
- g. 7 - rozpoznawanie wyrażeń wielowyrazowych,
- h. 8 - rozpoznawanie związków w tekście,
- i. 9 - głęboki parsing,
- j. 10 - głęboka analiza semantyczna,
- k. 11 - analiza pragmatyczna

#### Cele :

- wydobywanie wiedzy z tekstu,
- wyszukiwanie informacji,
- systemy automatycznego odpowiadania na pytania,
- definiuje zbiór znaczeń
- rozumienie języka naturalnego.