



Bazy Danych

Wprowadzenie do zagadnień baz danych

Piotr Macioł
WIMiP, KSiM,
pmaciol@agh.edu.pl
Konsultacje: środa, godz. 10:00-11:30
B5, pok. 606



Definicja bazy danych:

- **Baza danych** jest zbiorem struktur danych służących do organizowania i przechowywania danych. W takim modelu danych musi istnieć zbiór reguł określających wykorzystanie takich struktur danych w aplikacji.
- **Baza danych** to zintegrowana grupa ogólnie dostępnych zbiorów danych. Podstawową formą organizacji danych w bazie jest relacja.
- **Baza danych** – zbiór danych reprezentujących dany obszar analizy. W bazie danych mamy do czynienia z faktami, które mają miejsce. Dane powinny być trwałe.

KSiM, WIMiP, AGH

2



System Zarządzania Bazą Danych (SZBD) - Database Management System (DBMS)

- Baza danych: Zbiór powiązanych ze sobą danych
- Baza danych jest zarządzana przez tzw. **system zarządzania bazą danych**, w skrócie SZBD.
- BD + SZBD = **system bazy danych**
- Z systemem bazy danych współpracują programy użytkowników, zwane **aplikacjami**. Zadaniem tych programów jest przetwarzanie danych, tj. wstawianie nowych danych, modyfikowanie danych już istniejących, usuwanie danych nieaktualnych, wyszukiwanie danych.
- Wszystkie omówione wyżej komponenty (tj. baza danych, SZBD i aplikacje) wchodzi w skład tzw. systemu informatycznego.

KSiM, WIMiP, AGH

3



System Zarządzania Bazą Danych (SZBD) - Database Management System (DBMS)

- SZBD to pakiety programowe służące do zarządzania danymi. Podstawowymi elementami tych systemów są:
 - » język zapytań (SQL)
 - » generator raportów
 - » język manipulacji danymi
 - » język definicji danych
- SZBD umożliwiają oddzielenie aplikacji od bazy.
- SZBD dostarcza wygodnego i łatwego do używania środowiska

KSiM, WIMiP, AGH

4



Geneza baz danych:

- Dawno, dawno temu... typowe aplikacje bazodanowe budowane były bezpośrednio w oparciu o system plików
- Wady bezpośredniego stosowania systemu plików w zagadnieniach bazodanowych:
 - » **Redundancja i niespójność danych**
 - Wiele formatów danych, duplikacja danych w różnych plikach
 - » **Trudności z dostępem do danych**
 - Potrzeba pisania nowych aplikacji dla nowych zadań
 - » **Rozczłonkowanie danych** — **wiele formatów** i plików
 - » **Problem integralności**
 - Warunki integralności (np. account balance > 0) stają się częścią kodu aplikacji
 - Trudności z nałożeniem nowych warunków integralności lub modyfikacją już istniejących

KSiM, WIMiP, AGH

5



Geneza baz danych:

- Wady bezpośredniego stosowania systemu plików w zagadnieniach bazodanowych (c.d.):
 - » **Granulacja modyfikacji danych**
 - Błędy mogą pozostawić bazę w stanie nie zakończonej modyfikacji
 - Np. poprawny transfer pomiędzy rachunkami to dokonanie zmian salda obu rachunków
 - » **Współbieżny dostęp wielu użytkowników**
 - Dostęp współbieżny potrzebny dla efektywności
 - Niekontrolowany dostęp współbieżny może prowadzić do niespójności
 - Np. dwóch użytkowników odczytuje i modyfikuje saldo jednocześnie
 - » **Problemy związane z bezpieczeństwem**
- SZBD oferują rozwiązanie wszystkich tych problemów

KSiM, WIMiP, AGH

6



Geneza baz danych:

- W latach 60 i 70 dominował model hierarchiczny (IMS) i sieciowy.
- W roku 1970 naukowiec z firmy IBM - E.F.Codd opublikował pracę **A relational model for large shared data banks**, w której po raz pierwszy przedstawił założenia modelu relacyjnego.
- 1979 powstaje komercyjna wersja relacyjnej bazy danych – Oracle
- W 1983 IBM przedstawił DB2 pierwszą relacyjną bazę danych dla dużych komputerów.
- W tym samym czasie pojawił się inny konkurent na rynku relacyjnych baz - firma Relational Database Systems, która wkrótce zmieniła nazwę na Informix.
- W tym czasie istniały już relacyjne bazy danych dla komputerów osobistych dBase II (Ashton Tate).

KSIM, WIMiP, AGH

7



Aktualne zagadnienia

- Architektura klient – serwer i wielowarstwowa
- Dane **multimedialne**
 - » problemy z porównywaniem wartości
 - » podejście obiektowo-relacyjne
 - » zarządzanie obiektami o dużych rozmiarach
- **Integracja danych**
 - » konieczność zarządzania rozproszonym (logicznie i fizycznie) środowiskiem
 - » hurtownie danych (opóźnienie w aktualizacji)

KSIM, WIMiP, AGH

8



Typy baz danych:

- **operacyjne bazy danych** – znajdują zastosowanie w codziennym funkcjonowaniu organizacji, instytucji i firm. Baza taka przechowuje dane dynamiczne, czyli takie, które ulegają ciągłym zmianom i odzwierciedlają aktualny stan obiektu.

OLTP – OnLine Transaction Processing
(bieżące przetwarzanie transakcji)

- **analityczne bazy danych** – wykorzystywane są do przechowywania danych historycznych i informacji związanych z pewnymi wydarzeniami. Przechowywane tutaj dane są statyczne, bardzo rzadko ulegają zmianom i zawsze odzwierciedlają stan obiektów z pewnego ustalonego momentu.

OLAP – OnLine Analytical Processing
(oprogramowanie do analiz wielowymiarowych)

KSIM, WIMiP, AGH

9



Cele systemów bazodanowych:

- **Elastyczny (efektywny) dostęp do danych** (przetwarzanie) - łatwe selekcjonowanie i prezentacja danych
- **Integralność danych** – dane są prawidłowe, spójne i aktualne
- **Bezpieczeństwo danych** – ochrona przed nieupoważnionym dostępem i uszkodzeniem
- **Obniżenie redundancji**
- **Niezależność od aplikacji** – fizyczna i logiczna organizacja danych jest oddzielona od aplikacji
- **Współdzielenie danych** (współbieżny dostęp) – umożliwienie różnym użytkownikom korzystania z tych samych (nie nadmiarowych) danych
- **Standaryzacja opisu** – jednolite definicje danych dotyczące ich nazw i opisu
- **Metadane** - dane o danych, strukturach dostępu, użytkownikach i ich prawach

KSIM, WIMiP, AGH

10



Spójność bazy danych

- Poprawność danych z punktu widzenia przyjętych kryteriów
 - » wierne odzwierciedlenie danych rzeczywistych
 - » wszystkie dane w bazie, na które nałożono pewne ograniczenia integralnościowe muszą te ograniczenia spełniać
- Odporność na anomalie będące wynikiem współbieżności dostępu do baz danych
- Odporność na błędy, awarie i inne anormalne sytuacje wynikające z zawodności środowiska sprzętowo-programowego
- Odporność na błędy użytkowników

KSIM, WIMiP, AGH

11



Funkcje:

- **Funkcje baz danych:**
 - » **aktualizujące** – zamieniają jeden stan w drugi (np. dodaj nowy produkt, zmień stawkę wynagrodzenia).
 - » **zapytań** – sprawdzanie stanu bazy danych (kwerendy pytające).
- **Funkcje SZBD:**
 - » **Zarządzanie plikami**
 - » **Wyszukiwanie informacji**
 - » **Zarządzanie bazą**
 - tworzenie i monitorowanie użytkowników
 - ograniczanie dostępu do plików
 - monitorowanie działania bazy danych

KSIM, WIMiP, AGH

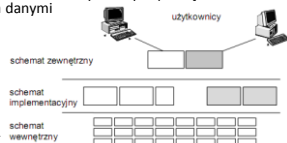
12



Architektura logiczna:

– Ujęcie ANSI/SPARC:

- » **Konceptualny** - poziom na którym zdefiniowany jest model świata w kategoriach pojęciowych użytkownika (interfejs użytkownika do bazy danych)
- » **Zewnętrzny** (implementacyjny) - języki wysokiego poziomu, umożliwiające definiowanie i dostęp do danych bez konieczności znajomości reprezentacji tych danych; odwzorowuje schemat wewnętrzny w struktury modelu danych wykorzystywanego w bazie danych
- » **Wewnętrzny** - służy do niezawodnego i trwałego składowania danych na nośnikach, na poziom ten składają się różnego rodzaju pamięci zewnętrznej, a także sposoby organizowania danych w tych pamięciach i techniki efektywnego zarządzania danymi



KSIM, WIMiP, AGH



Architektura logiczna:

– Model poziomów abstrakcji:

- » **Poziom widoków** – opis danych jest podobny do poziomu logicznego (pozbawiony takich szczegółów jak specyfikacje typów danych); zakres semantyczny każdego z widoków jest ściśle zorientowany (np. na konkretnego użytkownika) i stanowi zwykle jedynie małą część zakresu znaczeniowego całej bazy.
- » **Poziom logiczny** opisuje dane i relacje pomiędzy nimi zgodnie z ich semantyką w świecie rzeczywistym.
- » **Poziom fizyczny** opisuje, jak dane są składowane na odpowiednich nośnikach i jak realizowany jest elementarny dostęp do nich.

KSIM, WIMiP, AGH

14



Niezależność danych:

Konsekwentnie skonstruowana trzypoziomowa architektura bazy danych daje w efekcie niezależność schematu danych na wyższym poziomie od danych na niższym.

- **Logical Data Independence** – możliwość modyfikacji schematu logicznego bez konieczności przepisywania aplikacji.
- **Physical Data Independence** – możliwość modyfikacji schematu fizycznego bez zmiany schematu logicznego

W ogólności, interfejsy pomiędzy różnymi poziomami i komponentami powinny być dobrze zdefiniowane, tak aby zmiany w pewnych elementach nie wpływały istotnie na inne.

KSIM, WIMiP, AGH

15



Schematy i instancje:

– Schemat – struktura bazy danych, część intensjonalna bazy danych

- » **Schemat fizyczny**: projekt poziomu fizycznego bazy
 - Np. baza danych zawiera informację o zbiorze klientów i rachunków oraz relacjach pomiędzy nimi
- » **Schemat logiczny**: projekt poziomu logicznego bazy

– Instancja – aktualna (w danym momencie czasu) zawartość bazy danych, część ekstensjonalna bazy danych

KSIM, WIMiP, AGH

16



Części intensjonalna - przykład

- **Schemat**: MałyWirtualnyDziekanacik
- **Tabele**:
 - » studenci
 - » rozkład_zajec
 - » pracownicy
- **Związki**:
 - » rozkład zajęć pokazuje obciążenie studentów i pracowników
- **Atrybuty**:
 - » studenci mają nazwiska i imiona oraz numer albumu

KSIM, WIMiP, AGH

17



Części ekstensjonalna - przykład

- **Zawartość**: Bazy danych
- **Studenci**:
 - » Fabryczusz Kosonosy 146/2013
 - » Okowita Ambrozjowa 123/2013
- **Pracownicy**:
 - » Piotr Macioł
 - » Atanazy Angonilewicz
- **Zaplanowano**:
 - » Fabryczusz Kosonosy ma zajęcia z baz danych w dniu 6.03.2013 z dr inż. Piotrem Maciołem

KSIM, WIMiP, AGH

18



Formalizmy:

- **Formalizm reprezentacji** (*Patrick H. Winston*) to zbiór składniowych i semantycznych konwencji, które umożliwiają opisywanie rzeczy
- W terminologii baz danych idea formalizmu reprezentacji odpowiada pojęciu **modelu danych**

KSIM, WIMiP, AGH

20



Modele danych:

- **Model danych** – zbiór narzędzi formalnych opisujący: syntaktykę danych, semantykę danych, relacje pomiędzy danymi oraz ograniczenia danych.
- Modele dla poziomów widoków i logicznego oparte na koncepcji **obiektu**
 - » Entity-Relationship model
 - » Model obiektowy
- Modele dla poziomów widoków i logicznego oparte na koncepcji **rekordu**
 - » Model relacyjny
 - » Modele: sieciowy i hierarchiczny
- Fizyczne modele danych

KSIM, WIMiP, AGH

20



Przykładowa baza danych

Pracownicy									
idPrac	Nazwisko	Etat	Szef	DataZaj	Płaca	Premia	idZesp		
7343	Kowalski	referent	7648	17.12.90	1800		30		
7341	Nowak	asystent	7343	20.03.91	1500	300	30		
7342	Wojcik	asystent	7343	22.03.91	1500	600	30		
7644	Chojnacki	inżynier	7603	2.04.91	2500	100	20		
7645	Kim	inżynier	7603	28.05.91	3500		10		
7600	Dziuba	dyrektor	77.10.89		5000		10		
7650	Buda	referent	7644	23.05.95	2100		30		
7603	Majewski	inżynier	7603	28.10.91	2700		30		

Etaty			Zespoły		
Nazwa	PłacaMin	PłacaMax	idZesp	Nazwa	Adres
dyrektor	4000	9999	10	Administracyjny	Pracowni
inżynier	2000	4000	20	Specjalistyczny	Pracowni
referent	1500	2500	30	Reklamowy	Gniezno
asystent	1100	1800	40	Badawczy	Oborniki
inżynier	2500	4500			

– **Strukturami danych modelu** są w tym przypadku trzy relacje: Pracownicy, Zespoły, Etaty.

– Pierwsza z nich przechowuje dane o pracownikach, druga - o zespołach, w których ci pracownicy są zatrudnieni, a trzecia - zawiera katalog widełek płacowych.

KSIM, WIMiP, AGH

21



Interakcja z bazą danych

- **Język SQL** (projektanci aplikacji, projektanci baz danych i administratorzy baz danych)
 - » jedyny sposób interakcji z bazą danych
 - » język deklaratywny
 - specyfikujemy tylko co chcemy otrzymać, nie w jaki sposób
 - » ustandaryzowany
 - producenci systemów komercyjnych i niekomercyjnych starają się implementować ten standard

```
SELECT nazwisko, p_laca
FROM pracownicy
WHERE idZesp=30
AND etat= 'kierownik'
```

KSIM, WIMiP, AGH

22



Interakcja z bazą danych

- Aplikacje (użytkownicy końcowi)
 - » formularze
 - elektroniczne formularze z polami, listami, elementami wyboru
 - umożliwiają wstawianie, modyfikowanie, usuwanie, wyszukiwanie danych
 - » raporty
 - umożliwiają prezentowanie zawartości bazy danych:
 - teksty
 - wykresy
 - grafiki

KSIM, WIMiP, AGH

23



Problem modelowania rzeczywistości...

KSIM, WIMiP, AGH

24



Ontologia

- **Ontologia lub metafizyka** – podstawowy obok epistemologii (zajmującej się relacjami między poznawaniem, poznaniami a rzeczywistością) dział filozofii starający się badać strukturę rzeczywistości i zajmujący się problematyką związaną z pojęciami bytu, istoty, istnienia i jego sposobów, przedmiotu i jego własności, przyczynowości, czasu, przestrzeni, konieczności i możliwości.
- Termin ontologia pojawił się w kontekście informatycznym już w roku 1967 w badaniach dotyczących modelowania danych. W dobie zalewu informacją dostępną w Internecie i koniecznością jej przetwarzania zyskał szersze zainteresowanie. Ontologia zajmuje się odkrywaniem i opisywaniem „tego co jest”, pewnym fragmentem rzeczywistości, mniej lub bardziej dokładnie określonym.
- Termin **taksonomia** – początkowo zdefiniowany przez De Candolle'a (1813) jako nauka o teorii klasyfikacji organizmów żywych – jest bliskoznaczny do terminu systematyka, który w pierwotnej definicji odnosił się do zastosowań praktycznych tej klasyfikacji. W 1969 Ernst Mayr zdefiniował taksonomię jako teorię i praktykę klasyfikacji.

KSIM, WMIMP, AGH

25



Ontologia

- Aby wyraźniej podkreślić cechy charakterystyczne ontologii, należy przedstawić kilka postulatów dotyczących cech ontologii (nie samej jej konstrukcji):
 - » nie stanowi listy, katalogu czy taksonomii obiektów, stwarza natomiast **formalne przesłanki**, wedle których takowe mogą być budowane
 - » jest oderwana od teorii poznania (epistemologii), **powiązana jest z obiektem**, a nie jego subiektywnym odbiorem
 - » musi uchwycić rzeczywistość na różnych poziomach atomizacji, jak również relacje pomiędzy tymi warstwami
 - » uznanie braku możliwości stworzenia jednej ogólnej ontologii, istnienie wielu ontologii
 - » w przeciwieństwie do nauki relacje między obiektami nie są ujęte funkcyjnie (**zależności nie są ilościowe**)
 - » nauka rozpoczyna proces od mierzenia i predykcji, ontologia zaś od budowania taksonomii

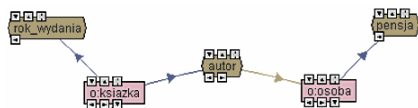
KSIM, WMIMP, AGH

26



Ontologie są...

- ...narzędziem, które umożliwia w taki sposób opisać **model rzeczywistości**, aby był on zrozumiały i przetwarzalny także dla komputera.
- ...pojęciowym modelem wiedzy, który poprzez **kategoryzację i hierarchizację** pojęć z danej dziedziny pozwala stworzyć „sieć znaczeniową”.



KSIM, WMIMP, AGH

27



Ontologie...

- Można zatem powiedzieć w pewnym uproszczeniu, że **ontologia dla repozytoriów wiedzy jest tym, czym diagram encji dla bazy danych**.
- Ontologie tworząc niejako wspólny, sformalizowany język pozwalają **zintegrować wiele różnorodnych, rozproszonych źródeł wiedzy** z danej dziedziny

Ontologia - formalny opis pojęć występujących w danej dziedzinie („słownik”, „encyklopedia”).

Geneza

- Sieci semantyczne (semantic networks ale nie Semantic Web!), Quillian 1967
- Ramy (frames), Minsky 1975

KSIM, WMIMP, AGH

28



Ontologie pozwalają...

- Ontologie pozwalają rozwiązać problemy różnic w zakresach :
 - » **systemowym** – różnice sprzętowe i różnice w zakresie stosowanych platform (np. kodowanie znaków);
 - » **syntaktycznym** – różne formaty danych (bazy danych, pliki tekstowe etc);
 - » **strukturalnym** – różne schematy danych (np. klasyfikacje według różnych norm, modele bazodanowe);
 - » **semantycznym** – różnice w znaczeniu pojęć (synonimy, polisemia, jedno pojęcie może być użyte w innych znaczeniach lub innych zakresach)
 - jak dotąd największe wyzwanie dla narzędzi informatycznych , często uniemożliwiający jednoznaczna identyfikację danego pojęcia, a w związku z tym odnalezienie informacji z nim związanej.

KSIM, WMIMP, AGH

29



Trójkąt znaczeniowy



- W 1923 Ogden i Richards wprowadzili trójkąt znaczeniowy (trójkąt semiotyczny)
- Istnieje bezpośrednie połączenie pomiędzy formą a znaczeniem oraz pomiędzy znaczeniem a referentem.
- Związek pomiędzy formą a referentem jest jedynie pośredni, ponieważ ta sama forma może odnosić się do wielu referentów.
- Na przykład: polskie słowo **zamek** może odnosić się do zamka w drzwiach, zamka przy spodniach, lub wystawnego budynku. Może oznaczać ono również część karabinu. Za każdym razem kiedy chcemy użyć słowa **zamek** mamy na myśli jedno z tych znaczeń. **Referent jest zatem konkretnym obiektem** czy też ideą do której odnosi się dana forma. Istotnym elementem jest **kontekst** w którym dana forma występuje.
- Samo słowo **zamek** przywołuje jedynie dwa **znaczenia**, które są centralne (**prototypowe**) dla tej formy. Zamek - coś co zamyka i zamek - okazała budowla. Można powiedzieć, że te dwa znaczenia są znaczeniami centralnymi tej formy. A przecież istnieją różnego rodzaju zamki np: z piasku, zamek u paska, nawet w agrafrce możemy mówić o jakimś zapięciu czy też zamku. Wszystkie pozostałe znaczenia będą jedynie peryferyjne.

KSIM, WMIMP, AGH

30



Kategorie

- Jeżeli mamy do czynienia ze zbiorem wielu podobnych obiektów to dążymy do **ujednorodnienia** ich opisu co sprowadza się do przypisania wielu w istocie różnym ale **podobnym obiektom** tych samych cech (różniących się być może wartościami)
- W ten sposób możemy stworzyć pewien **był abstrakcyjny** (który sam w sobie nie jest obiektem) i nazwać go **kategorią**
- Kategorii przypiszemy pewien **zestaw cech**, którymi opisywać będziemy wszystkie obiekty do niej należące

KSIM, WIMMUP, AGH

31



Kategoryzacja

- Umiejętność zaliczania obiektów do pewnych klas (kategorii, pojęć)
 - » *Kot Mruczek ma cztery łapy, sierść, ogon itd. ale takie same cechy mają kot Pimpus, Gienek itd.*
- Wystarczy przyjąć, że istnieje pojęcie (kategoria) kot charakteryzująca się tym, że obiekty do niej należące mają cztery łapy, sierść i ogon itd.

KSIM, WIMMUP, AGH

32



Hierarchizacja

- **Koty** rodzą się żywe i piją mleko matki ale to samo dotyczy np. **psów**, które kotami nie są
- Możemy wprowadzić pojęcie **ssaki**, które obejmuje zarówno koty jak i psy
- Wystarczy wówczas powiedzieć że Mruczek jest kotem by wiedzieć o nim wszystko to co dotyczy kotów i ssaków

KSIM, WIMMUP, AGH

33



Model danych

- Dana - symboliczna reprezentacja pewnego faktu opisującego rzeczywistość
- Dane poszerzone o **semantykę** to informacje
- Obiekt to niepodzielna jednostka danych, która opisywana być może krotką
 - <nawa obiektu, cecha obiektu, wartość cechy>

Krotka, zwana czasem n-tką, to uogólnienie pary (dwójki), trójki, czwórki, itd. na dowolną liczbę elementów. Krotka n-elementowa to uporządkowany skończony zbiór elementów (lista pewnych obiektów) przy czym kolejne elementy krotki nie muszą należeć do tego samego zbioru.

KSIM, WIMMUP, AGH

34



Związki

- **Związek** to trwała lub tymczasowa zależność występująca pomiędzy obiektami i możliwa do zapisania w modelu danych
- W niektórych przypadkach używane jest pojęcie **asocjacji** wskazujące na możliwość kojarzenia ze sobą pewnych obiektów lub pojęcie **relacji**
- Reguły opisujące konsekwencje wynikające z charakteru związku nazywać będziemy **więzami**

KSIM, WIMMUP, AGH

35



Zasady modelowania danych

- Każdy obiekt opisywany jest przez zestaw **cech (atrybutów)** z jednoznacznie określoną dziedziną, użytecznych z punktu widzenia wartości informacyjnej ontologii
- Każdy obiekt w tej samej **kategorii** musi być opisany zgodnie z tym samym **wzorcem** (choć w niektórych przypadkach nie wszystkie atrybuty obiektu muszą być podane)

KSIM, WIMMUP, AGH

36



Zasady modelowania danych

- Musi istnieć możliwość **rozróżnienia obiektów**, które w rzeczywistości są różne
- Musi istnieć możliwość **przedstawienia związków** jakie zachodzą lub mogą zajść pomiędzy obiektami
- Musi istnieć możliwość wprowadzenia do modelu dodatkowych **ograniczeń wartości** cech wynikających ze związków danych a nie tylko z dziedziny wartości atrybutów

KSIM, WIMiP, AGH

37



Abstrakcja

- Z abstrakcją mamy do czynienia, gdy wychodząc od rzeczy jednostkowych, konkretnych i indywidualnych dochodzimy, przez **proces uogólniania** i poszukiwania cech stałych i wspólnych, do pojęcia tak ogólnego, że w swej ogólności wręcz absurdalnego, gdyż nie posiadającego żadnej konkretnej cechy
- Mówimy o dwóch sposobach stosowania abstrakcji. Pierwszy z nich to **uogólnienie** a drugi to **agregacja**

KSIM, WIMiP, AGH

38



Uogólnienie

- Uogólnienie polega na próbie **klasyfikacji i generalizacji** opisów pewnych obiektów (wystąpień) i tworzeniu z nich bardziej ogólnych zbiorów cech (opisów)
- Uogólnianie cech obiektów prowadzące do stworzenia **kategorii** jest także abstrakcją
- W modelowaniu danych abstrakcję tą nazywamy nie **uogólnieniem** ale **klasyfikacją**

KSIM, WIMiP, AGH

39



Przykładowa kategoria

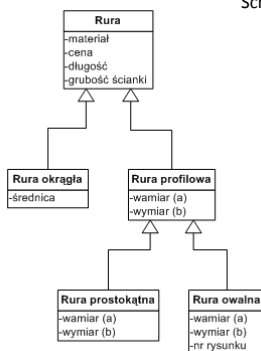
kształt rury	średnica [mm]	grubość ścianki [mm]	wymiar b [mm]	długość odcinka [m]	nr rysunku profilu
okrągła	20	1.5			
eliptyczna	40	2.0	30	4000	209/34 e
okrągła	30	2.0		6000	
prostokątna	40	1.8	20		

KSIM, WIMiP, AGH

40



Schemat uogólnienia



KSIM, WIMiP, AGH

41



Agregacja

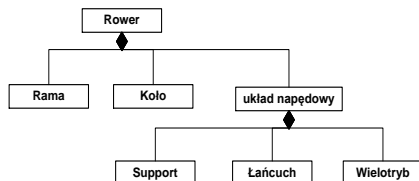
- **Agregacja** polega na traktowaniu obiektu lub kategorii (typu) jako zbioru składowych obiektów lub kategorii
- **Agregacja** to nie tylko wskazanie z jakich atrybutów składa się opis obiektu ale także jakie obiekty tworzą obiekt o bardziej złożonej strukturze
- Agregacja najprościej mówiąc oznacza zawieranie.
 - » lampka zawiera żarówkę,
 - » komputer zawiera procesor,
 - » jabłko zawiera robaka.
- Szczególnym przypadkiem agregacji jest **kompozycja**. Od agregacji różni się tym, że klasa posiada obiekty (składa się z obiektów), które bez tej klasy istnieć by nie mogły. Przykłady takich związków to:
 - » blok zawiera mieszkania (mieszkania poza blokiem nie istnieją),
 - » komputer zawiera procesor,
 - » łazienka zawiera wannę.

KSIM, WIMiP, AGH

42



Przykład agregacji



KSIM, WIMiP, AGH

43



Uogólnienie i agregacja

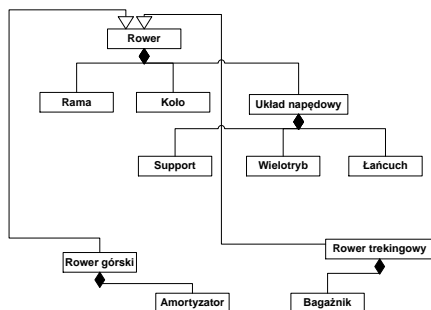
- Oba sposoby abstrakcji wykorzystujemy często łącznie
- Jest to konieczne w wielu przypadkach gdy dokonujemy abstrakcji na poziomie nie tylko obiektów ale także kategorii
- Na różnym poziomie abstrakcji różny będzie charakter związków i więzi

KSIM, WIMiP, AGH

44



Przykładowy opis obiektu



KSIM, WIMiP, AGH

45



Baza danych jako maszyna abstrakcyjna

- Maszyna abstrakcyjna to model cech systemu abstrahujący od szczegółów implementacyjnych
- Główne składniki systemu baz danych:
 - » reprezentacja a więc abstrakcyjny model danych
 - » implementacja czyli ciąg powiązanych ze sobą wystąpień kolekcji danych,
 - » zasady tworzenia i operacji na danych

KSIM, WIMiP, AGH

46



Obszar analizy

- Baza danych a właściwie jej schemat powstaje zawsze w wyniku analizy pewnego obszaru rzeczywistości
- W rezultacie powstaje model tego aspektu charakterystyczny dla danej organizacji, np.:
 - » zarządzanie personelem
 - » planowanie produkcji
 - » gospodarka materiałowa

KSIM, WIMiP, AGH

47



Zawartość bazy

- Baza danych opisuje pewien zbiór faktów
- Dane **opisują fakty i są symbolami** lub zbiorami symboli,
 - » np.: „Zygmunt Freud”
- Przypisanie danym interpretacji tworzy **informację** (rejestrację faktów), a więc dane wyposażone w semantykę,
 - » np.: „Zygmunt Freud to nazwisko pracownika”
- Obok faktów rejestrujemy **pozytywne asercje** a więc prawdziwe w danej chwili skojarzenia danych
 - » np.: „Zygmunt Freud jest ustawiczem półfabrykatów w piecach wypalowych”

KSIM, WIMiP, AGH

48



Trwałość

- Istotną cechą bazy danych jest trwałość danych
- Zarejestrowane fakty, ich związki a także sam schemat bazy są używane wielokrotnie
- W bazie danych mogą być także wykorzystywane dane nietrwałe tworzone na potrzeby użytkownika w czasie jednej sesji

KSIM, WMiP, AGH

49



Model danych i kolekcja danych

- *Schemat* to zbiór definicji w pewnym modelu danych
- W odniesieniu do baz danych pojęcie schemat jest traktowane jako identyczne z pojęciem **części intensjonalnej**
- Łączny zbiór danych zgodnych ze schematem nazywany jest **częścią ekstensjonalną** bazy danych
- Podział na część intensjonalną i ekstensjonalną dotyczy wszystkich poziomów abstrakcji modelu

KSIM, WMiP, AGH

50



Część intensjonalna i ekstensjonalna - problemy

- W życiu codziennym rzadko stosujemy abstrakcję w taki sposób w jaki wykorzystuje się ją do modelowania danych
- W sposób naturalny opisujemy pewne zbiory bądź to wyliczając i charakteryzując poszczególne ich elementy bądź wskazując na jedną lub kilka cech, które wyróżniają te elementy od elementów pochodzących z innych zbiorów

KSIM, WMiP, AGH

51



Przykład

Imię	Nazwisko	Tytuł
Bonifacy	Jagmin	Historia Literatury
Juliusz	Verne	Historia Literatury
Hans	Schmidt	BHP

KSIM, WMiP, AGH

52

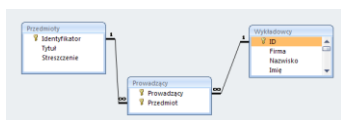


Schemat z uwzględnieniem związku

Prowadzący	Przedmiot
Jagmin	1
Verne	2
Schmidt	1

ID	Nazwisko	Imię
1	Jagmin	Bonifacy
2	Verne	Juliusz

Identyfikator	Tytuł	Streszczenie
1	Bazy Danych	
2	Zaawansowane Techniki Bazydanych	



KSIM, WMiP, AGH

53



Integralność

- Baza danych przechodzi przez ciąg zmian stanów
 - W zbiorze możliwych przyszłych stanów jedno są poprawne a inne nie
 - W schemacie należy określić zasady integralności, np.: fakt:
 - **Zaplanowano**
 - » Fabryczysz Kosonosy ma zajęcia z baz danych w dniu 6.03.2013 z dr inż. Atanazym Angonilewiczem
- jest niepoprawny bo Angonilewicz nie prowadzi baz danych

KSIM, WMiP, AGH

54



Replikacje

- W bazie danych należy zminimalizować powtarzanie się danych
- Przykładowo asercje:
 - » Krzysztof Regulski dn. 6.03.2013 prowadzi wykłady z baz danych w godzinach ...
 - » Fabrycjusz Kosonosy ma zajęcia z baz danych w dniu 6.03.2013
 - » Krzysztof Regulski dn. 6.03.2013 prowadzi wykłady z baz danych dla studentów ...

KSIEM, WIMiP, AGH

55



Więzy integralności

- Więzy integralności wskazują jakie asercje są poprawne a jakie nie
- Więzy statyczne to ograniczenia określone na stanie bazy danych, np.:
 - » nie można zaplanować zajęć z przedmiotu, którego nie ma w tabeli *przedmioty*
- Więzy przejść to reguły wiążące ze sobą stany bazy danych, np.:
 - » obciążenie studenta nie może przekroczyć 168 godzin tygodniowo

KSIEM, WIMiP, AGH

56



Funkcje i więzy

- Funkcje aktualizujące nie mogą być wykonane jeśli naruszają więzy integralności
- Funkcje aktualizujące mogą wywołać działania narzucone przez więzy integralności, np.: w pewnych przypadkach usunięcie studenta może spowodować usunięcie z bazy wszystkich ocen jakie były do niego przypisane
- Czy usunięcie pracownika powinno wiązać się z usunięciem prowadzonych przez niego przedmiotów?

KSIEM, WIMiP, AGH

57



Formalizmy

- Formalizm reprezentacji to zbiór składniowych i semantycznych konwencji, które umożliwiają opisywanie rzeczy
- W terminologii baz danych idea formalizmu reprezentacji odpowiada pojęciu modelu danych

KSIEM, WIMiP, AGH

58



Podstawowe formalizmy w bazach danych

- **encja – relacja – klasa – tabela**: zbiór podobnych obiektów opisanych w jednolity sposób
- **krotka – obiekt (instancja klasy) – rekord**: zestaw wartości atrybutów opisujących jeden obiekt identyfikowany przez wyróżnione atrybuty lub nazwę
- **więź – asocjacja**: związek pomiędzy dwoma encjami (klasami) pokazujący jakie rekordy (obiekty) z jednej encji odpowiadają rekordom z drugiej i jaki jest charakter tej odpowiedniości

KSIEM, WIMiP, AGH

59



Przykład

Encja (klasa): WydziałyProdukcyjne

KodWydziału	NazwaWydziału	CharakterPracy
1	Odlwienia	ciągły
2	Obróbka	2 zmiany

Encja (klasa): Agregaty

KodWydziału	KodAgregatu	NazwaAgregatu
1	1	piec tyglowy
1	2	formierka
2	3	tokarka

KSIEM, WIMiP, AGH

60