

Bazy danych

dr inż. Andrzej Macioł

<http://amber.zarz.agh.edu.pl/amaciol/>

Ontologia

- Dziedzina metafizyki, która para się badaniem i wyjaśnianiem natury jak i kluczowych właściwości oraz relacji rządzących wszelakimi bytami bądź głównymi zasadami i przyczynami bytu
 - jak można wszystko poklasyfikować?,
 - jakie klasy bytów są niezbędne do opisu i wnioskowania na temat zachodzących procesów?
 - jakie klasy bytu pozwalają wnioskować o prawdzie?
 - na podstawie jakich klas bytu można wnioskować o przyszłości?

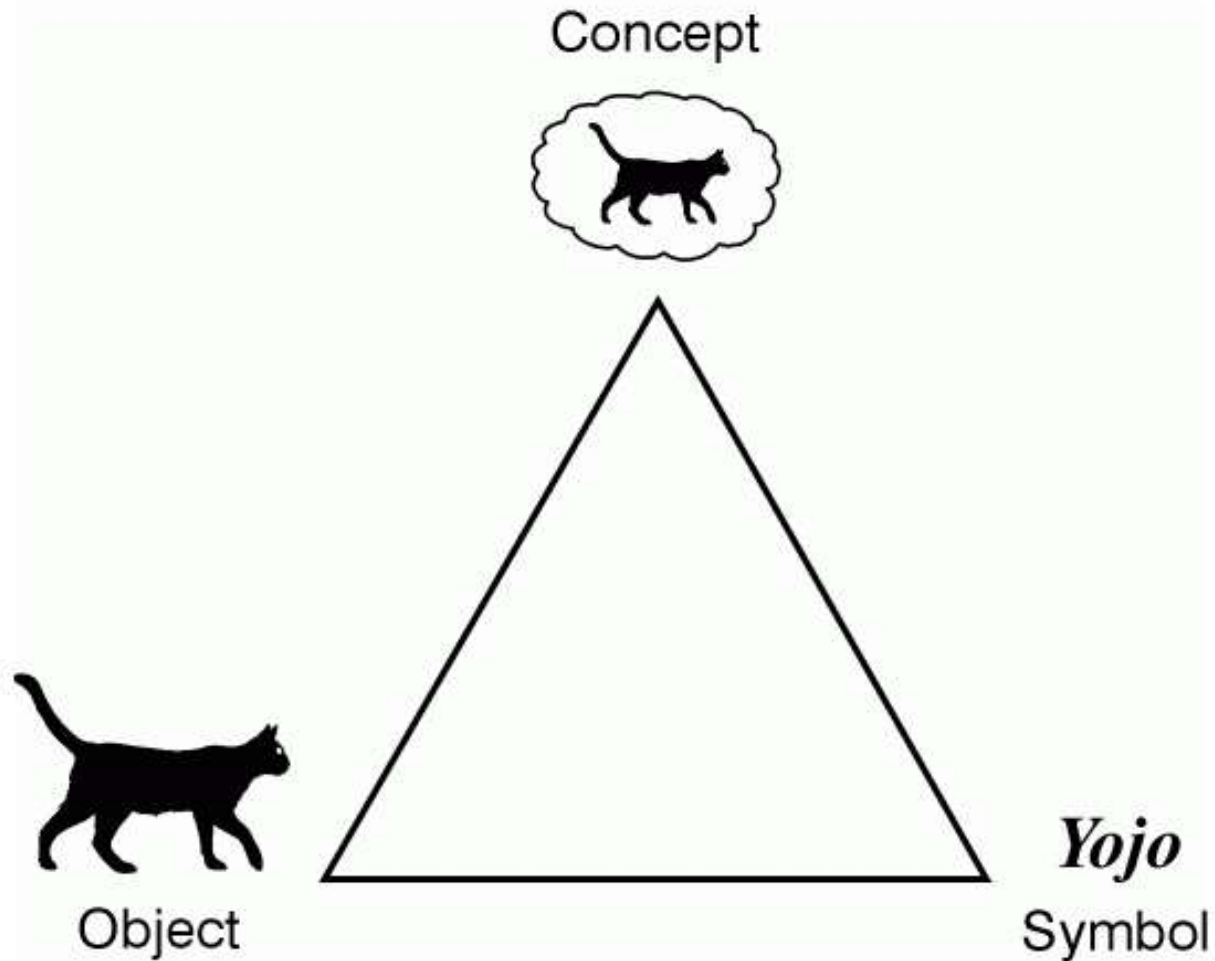
Byt

- Wszystko co jest może być bytem: kot Mruczek, Andrzej Macioł, towar w sklepie, indeks studenta itd.
- Konstruując ontologię formułujemy zestaw komunikatów, którymi możemy opisać byty, np.:
kot Mruczek ma cztery łapy

Kategoryzacja

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

Trójkąt znaczeniowy



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

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.

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 byt 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

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*

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)

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

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

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*

Uogólnienie - przykład



PHILIPS

PHILIPS PD7032T/12

Odtwarzacz DVD

nr kat.: 1149225]

★★★★★ (0)

- » Ekran: 2 x LCD, 480 x 234 piksele
- » Przekątna ekranu [cal]: 7
- » Format plików: MP3, DivX, JPEG, MPEG4
- » Pilot: TAK

więcej › + schowek

cena online ⓘ

799,-

[sprawdź dostępność ›](#)

odbierz w markecie

[sprawdź koszty i termin dostawy ›](#)

zamów do domu



GRUNDIG

GRUNDIG Music Boy 51

Radio

nr kat.: 1091965

★★★★★ (2)

- » Rodzaj radia: Cyfrowe
- » Zegar: TAK
- » System fonii: Mono
- » Zasilanie: Baterie/sieć

więcej › + schowek

cena online ⓘ

149,90

[sprawdź dostępność ›](#)

odbierz w markecie

[sprawdź koszty i termin dostawy ›](#)

zamów do domu

Klasa (kategoria) i obiekty – bez uogólnienia

DrobnySprzetAudioVideo

- RodzajProduktu
- Producent
- Cena
- Ocena
- Ekran
- PrzekatnaEkranu
- FormatPlikow
- Pilot
- RodzajRadia
- Zegar
- SystemFonii
- Zasilanie

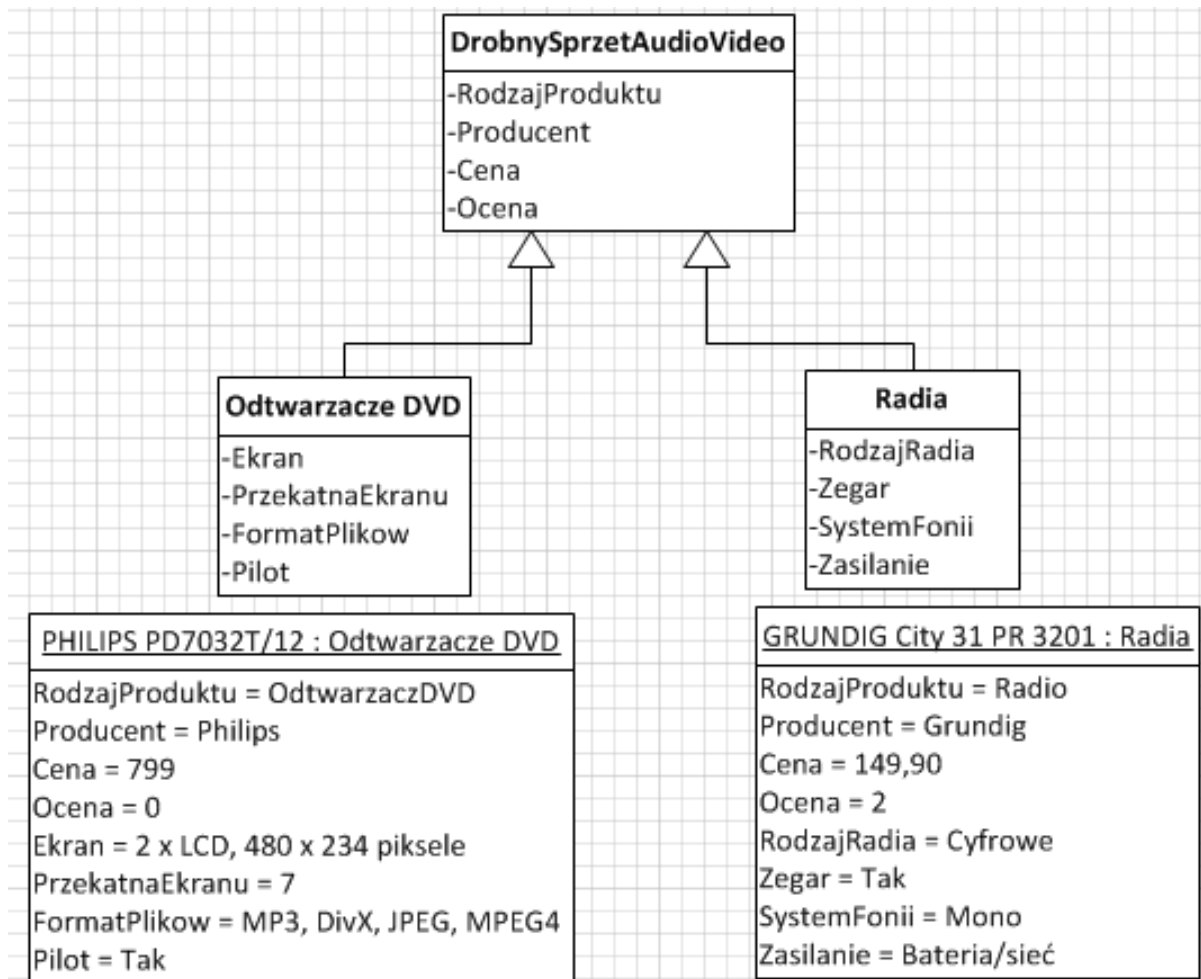
PHILIPS PD7032T/12 : DrobnySprzetAudioVideo

RodzajProduktu = Odtwarzacz DVD
Producent = Philips
Cena = 799
Ocena = 0
Ekran = 2 x LCD, 480 x 234 piksele
PrzekatnaEkranu = 7
FormatPlikow = DivX, JPEG, MPEG4, MP3
Pilot = Tak
RodzajRadia
Zegar
SystemFonii
Zasilanie

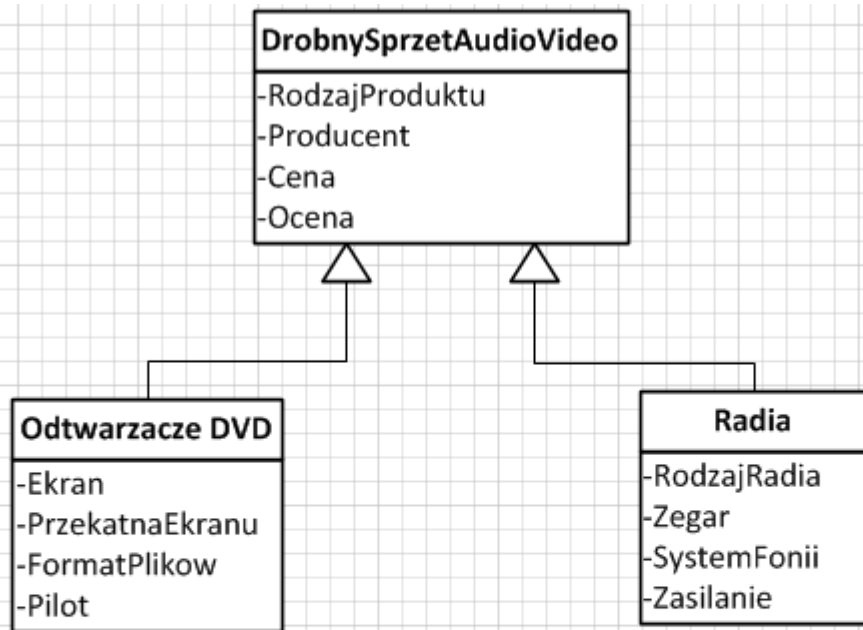
GRUNDIG City 31 PR 3201 : DrobnySprzetAudioVideo

RodzajProduktu = Radio
Producent = Grundig
Cena = 149,90
Ocena = 2
Ekran
PrzekatnaEkranu
FormatPlikow
Pilot
RodzajRadia = Analogowe
Zegar = Tak
SystemFonii = Mono
Zasilanie = Bateriajne/sieć

Klasa (kategoria) i obiekty uogólnione



Klasa (kategoria) i obiekty uogólnione



PHILIPS PD7032T/12 : DrobnySprzetAudioVideo

RodzajProduktu = OdtwarzaczDVD
Producent = Philips
Cena = 799
Ocena = 0

GRUNDIG City 31 PR 3201 : DrobnySprzetAudioVideo

RodzajProduktu = Radio
Producent = Grundig
Cena = 149,90
Ocena = 2

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

Przykładowy opis obiektu

nazwa produktu: OCR Composite 2

cena produktu: 9999 zł

producent: Giant

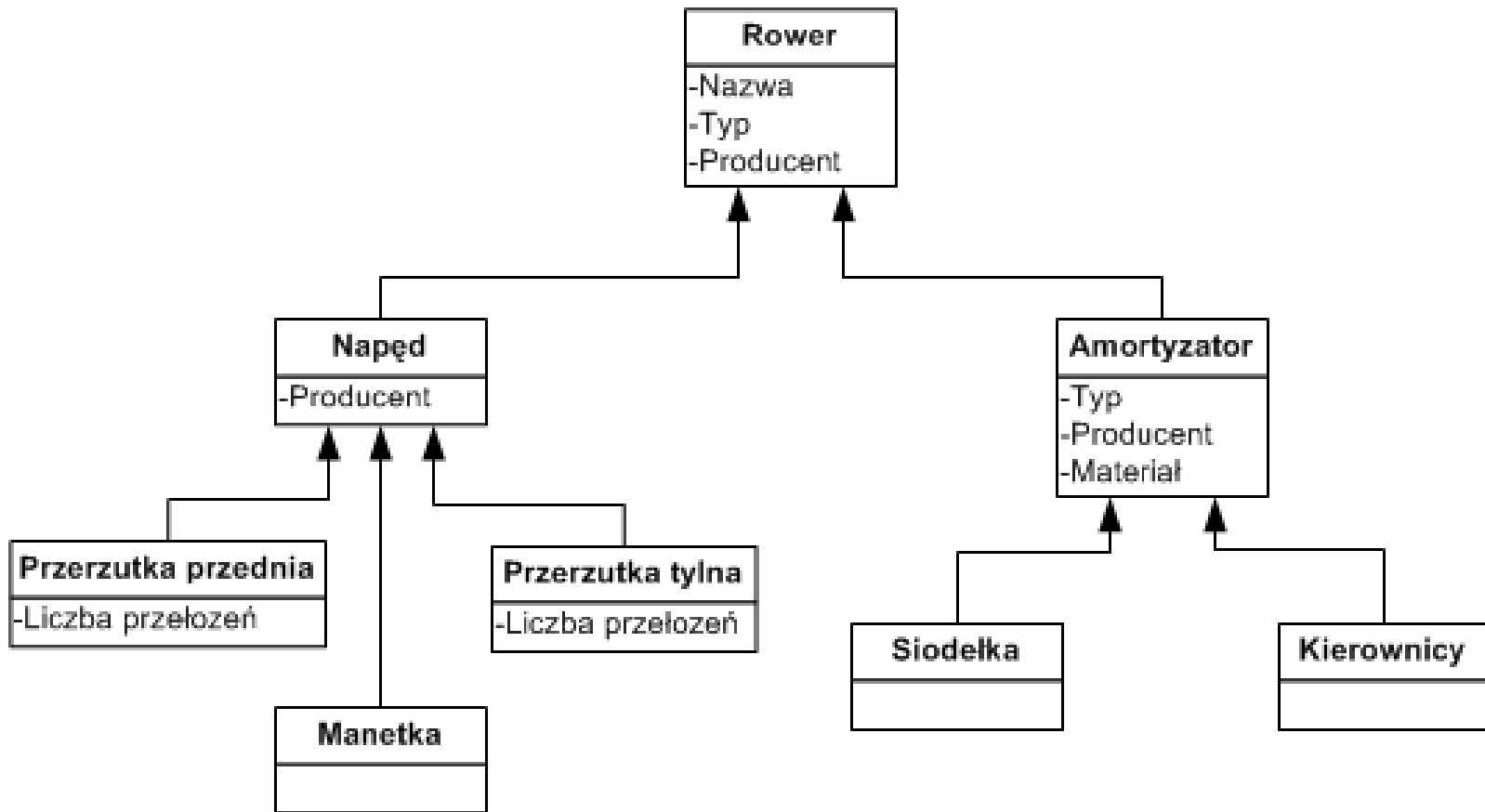
kategoria: Rowery szosowe

Rower szosowy na lekkiej ramie kompozytowej. 20-biegowy napęd, wymienny hak, profesjonalny osprzęt

Specyfikacja

- napęd: 20-biegowy:
- rama: kompozyt z włókna węglowego, Compact Road, wymienny hak
- widelec: kompozyt z włókna węglowego, aluminiowa rura sterowa 1-1/8", areo
- kierownica: Easton EA50 Road 31.8x400/420/440
- wspornik kierownicy: Easton EA50 Road 28.6-31.8-90/100/110-60

Schemat agregacji



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

Przykładowy opis obiektu

nazwa produktu: OCR 3
cena produktu: 2999 zł
producent: Giant
kategoria: Rowery szosowe

Rower szosowy na lekkiej ramie kompozytowej. 24-biegowy napęd, wymienny hak

Specyfikacja

- napęd: 24-biegowy:
- rama: aluminium AluxX 6061, Compact Road, wymienny hak
- widelec: kompozyt z włókna węglowego z aluminiową rurą sterową 1-1/8" aero
- kierownica: alu HL 26.0x400/420/440
- wspornik kierownicy: alu Mike Ahead 28.6x400/420/440

Przykładowy opis obiektu tej samej kategorii

nazwa produktu: OCR Composite 2

cena produktu: 9999 zł

producent: Giant

kategoria: Rowery szosowe

Rower szosowy na lekkiej ramie kompozytowej. 20-biegowy napęd, wymienny hak, profesjonalny osprzęt

Specyfikacja

- napęd: 20-biegowy:
- rama: kompozyt z włókna węglowego, Compact Road, wymienny hak
- widelec: kompozyt z włókna węglowego, aluminiowa rura sterowa 1-1/8", areo
- kierownica: Easton EA50 Road 31.8x400/420/440
- wspornik kierownicy: Easton EA50 Road 28.6-31.8-90/100/110-60

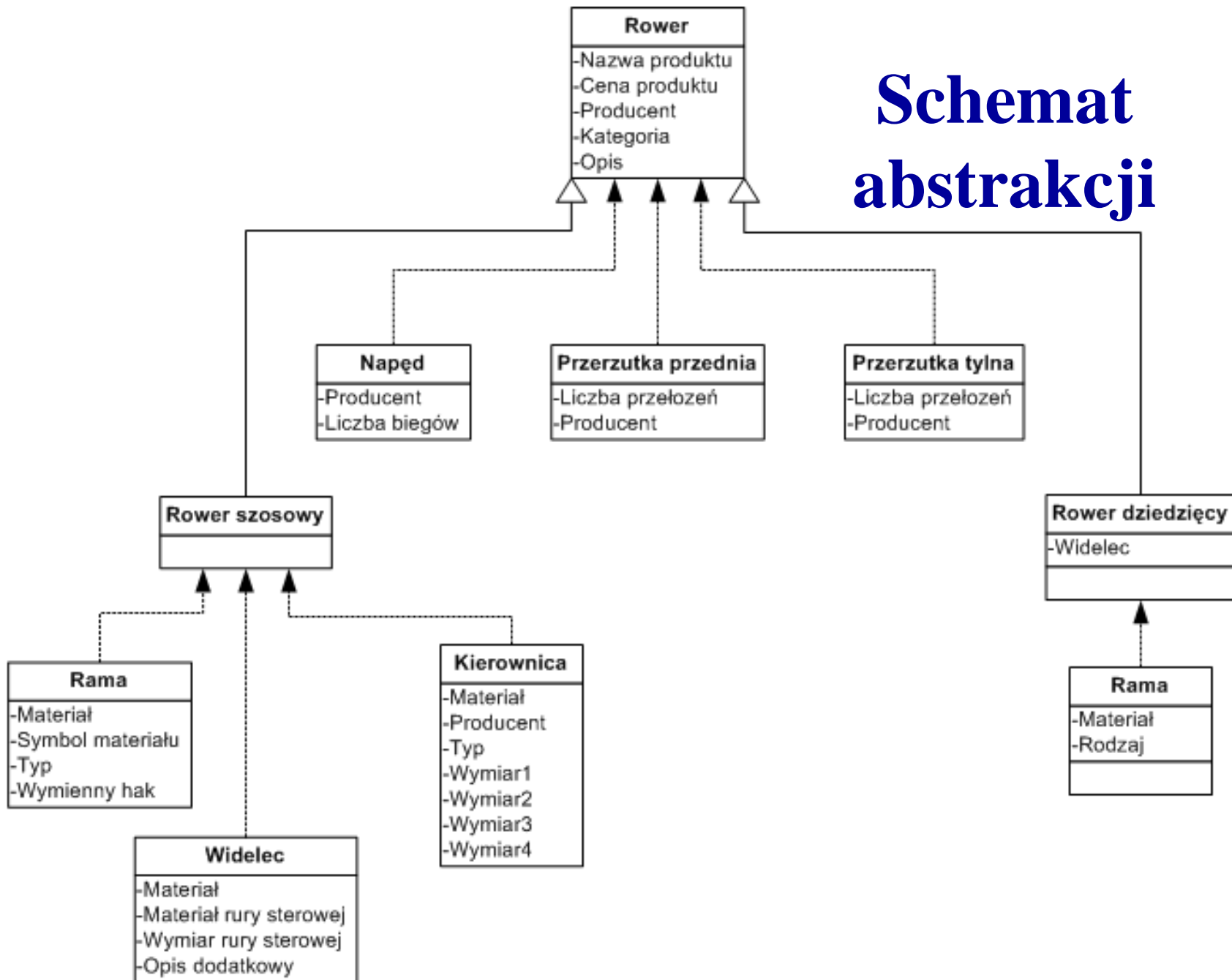
Przykładowy opis obiektu tej samej kategorii

nazwa produktu:	MIX 150
cena produktu:	669 zł
producent:	Giant
kategoria	Rowery dziecięce

Rower dziecięcy na kole 20". Wytrzymała rama aluminiowa, 5-biegowy napęd.
Specyfikacja

- napęd: 5-biegowy:
- rama: aluminiowa unisex
- widelec: Hi-Ten
- hamulec przedni: aluminiowy typu V
- hamulec tylny: aluminiowy typu V

Schemat abstrakcji



Baza danych – próba definicji

- Baza danych to zbiór danych dotyczących wyróżnionego obszaru zorganizowanych we wzajemnie powiązane pliki
- Jej istotą jest wewnętrzna struktura i organizacja, która pozwala na zaspokojenie potrzeb jednego lub wielu użytkowników bez uprzedniego sortowania w różne pożądane struktury potrzebne do przetwarzania lub bezpośredniego udzielania informacji

Baza danych – próba definicji

- Dzięki wewnętrznej organizacji stanowiącej integralną część bazy danych, korzystanie z jej zawartości jest niezależne od oprogramowania i sprzętu
- Według J. Martina baza danych to zbiór wystąpień różnych typów rekordów oraz opisów powiązań między rekordami, danymi zagregowanymi i danymi elementarnymi

Baza danych - właściwości

- współdzielenie danych, a więc możliwość spełniania potrzeb wielu użytkowników nie raz w jednym czasie
- integracja danych gwarantująca, że dane i związki między nimi nie powtarzają się jeśli nie jest to konieczne ale wszelkie zmiany w obrębie bazy nie powodują wieloznaczności

Baza danych - właściwości

- integralność danych pozwalająca na dokładne odzwierciedlenie stanu i zmian obszaru analizy ze szczególnym uwzględnieniem charakteru związków między danymi
- bezpieczeństwo danych pozwalające na wypełnienie postulatów integracji i integralności

Baza danych - właściwości

- abstrakcja danych a więc przedstawianie tylko tych informacji, które są istotne z punktu widzenia celu tworzenia bazy
- niezależność danych polegająca na oddzieleniu danych od procesów, które ich używają

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

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

Fabryka butów

Produkty

IdProduktu	NazwaProduktu	Materiał	Kolor	Rozmiar
1	But męski	skóra	czarny	44
2	But damski	płótno	biały	37

Walcownia blach

Blachy					
IdProduktu	NazwaProduktu	Material	Grubosc	Szerokosc	
1	Blacha zimnowalcowana	St3s		1	1200
2	Blacha goracowalcowana	St3s		4	1400

Rozwiązanie - 1

Produkt	
IdProduktu	NazwaProduktu
1	blacha zimnowalcowana
2	blacha goracowalcowana
3	but męski
4	but damski

Rozwiązanie - 2

Cechy		
IdCechy	Cecha	Typ
1	Materiał	słownik
2	Grubość	liczba
3	Szerokość	liczba
4	Kolor	słownik
5	Rozmiar	liczba

Rozwiązanie - 3

Słownik			
IdWartosci	Wartosc	IdCechy	
	1 St3s		1
	2 skóra		1
	3 płótno		1
	4 czarny		4
	5 biały		4

Rozwiązanie - 4

CechyProduktow				
IdProduktu	IdCechy	IdWartosci	Wartosc	
	1	1	1	0
	1	2	0	1
	1	3	0	1200
	2	1	1	0
	2	2	0	4
	2	3	0	1400
	3	1	2	0
	3	4	4	0
	3	5	0	44
	4	1	3	0
	4	4	5	0
	4	5	0	37

Wynik

OpisProduktow				
NazwaProduktu	Cecha	Typ	Sloownik.Wartosc	CechyProduktow.Wartosc
blacha zimnowalcowana	Materiał	sloownik	St3s	0
blacha zimnowalcowana	Grubość	liczba		1
blacha zimnowalcowana	Szerokość	liczba		1200
blacha goracowalcowana	Materiał	sloownik	St3s	0
blacha goracowalcowana	Grubość	liczba		4
blacha goracowalcowana	Szerokość	liczba		1400
but męski	Materiał	sloownik	skóra	0
but męski	Kolor	sloownik	czarny	0
but męski	Rozmiar	liczba		44
but damski	Materiał	sloownik	plótno	0
but damski	Kolor	sloownik	biały	0
but damski	Rozmiar	liczba		37

Modele danych

- Model konceptualny – spojrzenie na dane jako całość, model najbardziej stabilny, powinien on być podstawą, na której opierać się będzie przetwarzanie danych
- Model wewnętrzny, niskiego poziomu – opisuje sposób przechowywania danych w pamięci komputerów i przedstawia formaty rekordów czy ścieżki dostępu, modelami takimi są metody adresowania, struktury łańcuchowe i pierścieniowe

Modele danych

- Modele użytkowe – stanowią podstawę do budowy systemu informatycznego:
 - hierarchiczny
 - sieciowy
 - relacyjny
 - obiektowy

Model hierarchiczny

- Każdy element zwany rekordem może uczestniczyć w roli podrzędnej w co najwyżej jednym powiązaniu rekordów, w roli nadrzędnej w dowolnej liczbie powiązań
- Rekord podrzędny nie może istnieć bez rekordu nadrzędnego

Więzi w modelu hierarchicznym

KodPojazdu	Nazwa	Typ
1	Accent	kombi

KodPojazdu	KodCzęści	NazwaCzęści
1	1	silnik V6

KodPojazdu	KodCzęści	NazwaCzęści
1	2	szkrzynia biegów

Model sieciowy

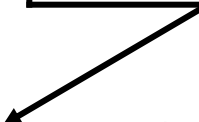
- Każdy rekord może jednocześnie uczestniczyć w wielu powiązaniach rekordów
- Rekord taki może równocześnie i wielokrotnie wystąpić w roli nadrzędnej oraz w roli podrzędnej, powiązania realizowane są przez rekordy specjalne zwane łącznikami

Więzi w modelu sieciowym

KodPojazdu	Nazwa	Typ
1	Accent	kombi

KodPojazdu	Nazwa	Typ
1	Getz	cupe

KodPojazdu	KodCzęści	NazwaCzęści
1	2	szkrzynia biegów



Modele obiektowy

- Wymaga istnienia trwałych obiektów, obiekty te pozostają zapisane w pamięci pomocniczej przed i po wykonaniu programów
- Obiekty mogą komunikować się bezpośrednio z użytkownikiem lub z innymi obiektami przez przesyłanie komunikatów, obiekty posiadają identyfikator, co umożliwia tworzenie powiązań między nimi
- Obiekt jest to abstrakcja czegoś w dziedzinie problemu, który odzwierciedla zdolność systemu do przechowywania informacji o tym lub interakcji z tym czymś