
Spis treści

Przedmowa	25
O autorach	33
I. Wprowadzenie do baz danych	35
1. Bazy danych i ich użytkownicy	37
1.1. Wprowadzenie	38
1.2. Przykład	41
1.3. Właściwości rozwiązań opartych na bazach danych	44
1.3.1. Samoopisująca natura systemów baz danych	45
1.3.2. Oddzielenie programów od danych oraz abstrakcja danych	46
1.3.3. Obsługa wielu perspektyw dla tych samych danych	48
1.3.4. Współdzielenie danych oraz wielodostępne przetwarzanie transakcji	49
1.4. Aktorzy na scenie	50
1.4.1. Administratorzy bazy danych	50
1.4.2. Projektanci bazy danych	50
1.4.3. Użytkownicy końcowi	51
1.4.4. Analitycy systemowi i programiści aplikacji (inżynierowie oprogramowania)	52
1.5. Pracownicy poza sceną	52
1.6. Zalety stosowania rozwiązań opartych na systemach zarządzania bazami danych	53
1.6.1. Kontrola nadmiarowości	53
1.6.2. Ograniczanie możliwości uzyskania nieautoryzowanego dostępu	55
1.6.3. Zapewnianie miejsca trwałego przechowywania dla obiektów stosowanych w programach	56
1.6.4. Zapewnianie struktur przechowywania dla efektywnego przetwarzania zapytań	56
1.6.5. Zapewnianie możliwości tworzenia kopii bezpieczeństwa i odzyskiwania danych	57
1.6.6. Zapewnianie interfejsów dla wielu użytkowników	57
1.6.7. Reprezentowanie skomplikowanych relacji pomiędzy danymi	57
1.6.8. Wymuszanie więzów integralności	58

1.6.9. Zezwalanie na wnioskowanie i podejmowanie działań w oparciu o zdefiniowane reguły	59
1.6.10. Dodatkowe własności wynikające ze stosowania rozwiązań opartych na bazach danych	59
1.7. Krótka historia praktycznych zastosowań baz danych	60
1.7.1. Wczesne zastosowania baz danych oparte na systemach hierarchicznych i sieciowych	60
1.7.2. Zapewnianie elastyczności w rozwiązaniach opartych na relacyjnych bazach danych	61
1.7.3. Aplikacje obiektowe i konieczność wprowadzenia bardziej skomplikowanych baz danych	62
1.7.4. Wykorzystywana w handlu elektronicznym wymiana danych za pośrednictwem internetu z użyciem XML-a	62
1.7.5. Rozszerzanie możliwości współczesnych systemów baz danych z myślą o nowych zastosowaniach	63
1.7.6. Powstanie systemów przechowywania big data i baz NOSQL	64
1.8. Kiedy nie należy używać systemów zarządzania bazami danych	64
1.9. Podsumowanie	65
Pytania powtórkowe	66
Ćwiczenia	66
Wybrane publikacje	67

2. Architektura systemów baz danych i związane z nimi pojęcia **69**

2.1. Modele danych, schematy i egzemplarze	70
2.1.1. Kategorie modeli danych	71
2.1.2. Schematy, egzemplarze i stany baz danych	72
2.2. Trójwarstwowa architektura i niezależność danych	74
2.2.1. Architektura trójwarstwowa	74
2.2.2. Niezależność danych	76
2.3. Języki i interfejsy baz danych	77
2.3.1. Języki systemów zarządzania bazami danych	77
2.3.2. Interfejsy systemów zarządzania bazami danych	80
2.4. Środowisko systemu bazy danych	82
2.4.1. Moduły składające się na system zarządzania bazą danych	82
2.4.2. Narzędzia systemu bazy danych	85
2.4.3. Narzędzia, środowiska aplikacji oraz mechanizmy komunikacji	86
2.5. Architektury systemów zarządzania bazami danych — scentralizowane i typu klient-serwer	87
2.5.1. Scentralizowane architektury systemów zarządzania bazami danych	87
2.5.2. Podstawowe architektury typu klient-serwer	87
2.5.3. Dwuwarstwowe architektury typu klient-serwer dla systemów zarządzania bazami danych	90
2.5.4. Trójwarstwowe i n-warstwowe architektury typu klient-serwer dla aplikacji internetowych	90
2.6. Klasyfikacja systemów zarządzania bazami danych	92
2.7. Podsumowanie	95
Pytania powtórkowe	96
Ćwiczenia	97
Wybrane publikacje	97

II. Konceptyjne modelowanie danych i projektowanie baz danych

99

3. Modelowanie danych zgodnie z modelem związków encji	101
3.1. Stosowanie wysokopoziomowych, konceptyjnych modeli danych podczas projektowania bazy danych	103
3.2. Przykładowa aplikacja bazy danych	105
3.3. Typy encji, zbiory encji, atrybuty i klucze	106
3.3.1. Encje i atrybuty	107
3.3.2. Typy encji, zbiory encji, klucze i zbiory wartości	109
3.3.3. Początkowy projekt konceptyjny bazy danych FIRMA	113
3.4. Typy związków, zbiory związków, role i ograniczenia strukturalne	114
3.4.1. Typy, zbiory i egzemplarze związków	115
3.4.2. Stopień związku, nazwy ról oraz związki rekurencyjne	115
3.4.3. Ograniczenia dla typów związków	118
3.4.4. Atrybuty typów związków	121
3.5. Słabe typy encji	122
3.6. Udoskonalanie projektu ER dla bazy danych FIRMA	123
3.7. Diagramy ER, konwencje nazewnictwa oraz zagadnienia związane z projektowaniem	124
3.7.1. Podsumowanie notacji diagramów związków encji (ER)	124
3.7.2. Prawdłowe nazewnictwo konstrukcji schematu	125
3.7.3. Decyzje projektowe związane z tworzeniem schematu konceptyjnego ER	127
3.7.4. Notacje alternatywne względem tradycyjnych diagramów związków encji (ER)	128
3.8. Przykładowa inna notacja: diagramy klas UML	129
3.9. Typy związków stopnia wyższego niż drugi	132
3.9.1. Wybór pomiędzy związkami binarnymi a trójskładnikowymi (lub wyższych stopni)	132
3.9.2. Ograniczenia związków trójskładnikowych (i wyższych stopni)	136
3.10. Inny przykład — baza danych UNIWERSYTET	137
3.11. Podsumowanie	139
Pytania powtórkowe	140
Ćwiczenia	141
Ćwiczenia laboratoryjne	147
Wybrane publikacje	149
4. Rozszerzony model związków encji	151
4.1. Podklasy, nadklasy i dziedziczenie	152
4.2. Specjalizacja i generalizacja	154
4.2.1. Specjalizacja	154
4.2.2. Generalizacja	155
4.3. Ograniczenia i właściwości hierarchii specjalizacji i generalizacji	157
4.3.1. Ograniczenia dotyczące specjalizacji i generalizacji	157
4.3.2. Hierarchie i kraty specjalizacji i generalizacji	160
4.3.3. Stosowanie procesów specjalizacji i generalizacji podczas udoskonalania schematów konceptyjnych	163
4.4. Modelowanie typów UNII w oparciu o kategorie	164
4.5. Przykład schematu EER dla bazy danych UNIWERSYTET oraz formalne definicje dla modelu EER	166
4.5.1. Inny przykład bazy danych UNIWERSYTET	166
4.5.2. Wybory projektowe związane ze specjalizacją i generalizacją	169
4.5.3. Formalne definicje pojęć stosowanych w modelu EER	170

4.6. Przykładowa inna notacja: reprezentowanie specjalizacji-generalizacji na diagramach klas języka UML	171
4.7. Abstrakcja danych, reprezentacja wiedzy oraz zagadnienia związane z ontologią	173
4.7.1. Klasyfikacja i tworzenie egzemplarzy	174
4.7.2. Identyfikacja	175
4.7.3. Specjalizacja i generalizacja	176
4.7.4. Agregacja i asocjacja	176
4.7.5. Ontologia i sieć semantyczna	178
4.8. Podsumowanie	179
Pytania powtórkowe	180
Ćwiczenia	181
Ćwiczenia laboratoryjne	188
Wybrane publikacje	190

III. Relacyjny model danych i SQL

193

5. Relacyjny model danych i ograniczenia relacyjnych baz danych	195
5.1. Pojęcia z modelu relacyjnego	196
5.1.1. Dziedziny, atrybuty, krotki i relacje	197
5.1.2. Właściwości relacji	199
5.1.3. Notacja modelu relacyjnego	203
5.2. Ograniczenia modelu relacyjnego i schematy relacyjnych baz danych	203
5.2.1. Ograniczenia dziedziny	204
5.2.2. Ograniczenia klucza i ograniczenia wartości pustych	205
5.2.3. Relacyjne bazy danych i schematy relacyjnych baz danych	206
5.2.4. Integralność encji, integralność odwołań i klucze obce	209
5.2.5. Pozostałe typy ograniczeń	210
5.3. Operacje aktualizacji, transakcje i obsługa naruszeń więzów integralności	212
5.3.1. Operacja wstawiania	212
5.3.2. Operacja usuwania	214
5.3.3. Operacja aktualizacji	215
5.3.4. Transakcje	216
5.4. Podsumowanie	216
Pytania powtórkowe	217
Ćwiczenia	218
Wybrane publikacje	222
6. Podstawy języka SQL	223
6.1. Definicje danych i typy danych języka SQL	225
6.1.1. Stosowane w języku SQL pojęcia schematu i katalogu	226
6.1.2. Polecenie CREATE TABLE języka SQL	226
6.1.3. Typy danych atrybutów oraz dziedziny wartości w standardzie SQL	229
6.2. Określanie ograniczeń w języku SQL	231
6.2.1. Definiowanie ograniczeń i wartości domyślnych dla atrybutów	232
6.2.2. Definiowanie ograniczeń klucza i więzów integralności odwołań	233
6.2.3. Nadawanie nazw definiowanym ograniczeniom	235
6.2.4. Stosowanie klauzuli CHECK do określania ograniczeń dla krotek	235
6.3. Podstawowe zapytania języka SQL	236
6.3.1. Struktura podstawowych zapytań języka SQL: SELECT-FROM-WHERE	236
6.3.2. Niejednoznaczne nazwy atrybutów, mechanizm nazw zastępczych (aliasów) oraz zmienne krotek	239
6.3.3. Nieokreślona klauzula WHERE i zastosowania symbolu gwiazdki	241

6.3.4. Tabele i zbiory w języku SQL	242
6.3.5. Dopasowywanie podciągów znaków do wzorca oraz operacje arytmetyczne	244
6.3.6. Sortowanie wyników zapytań	246
6.3.7. Omówienie i podsumowanie prostych zapytań języka SQL	246
6.4. Dostępne w języku SQL polecenia INSERT, DELETE i UPDATE	247
6.4.1. Polecenie INSERT	247
6.4.2. Polecenie DELETE	249
6.4.3. Polecenie UPDATE	250
6.5. Dodatkowe własności języka SQL	250
6.6. Podsumowanie	252
Pytania powtórkowe	252
Ćwiczenia	253
Wybrane publikacje	255

7. Jeszcze o języku SQL — złożone zapytania, wyzwalacze, perspektywy i modyfikowanie schematów 257

7.1. Bardziej skomplikowane zapytania języka SQL pobierające dane	257
7.1.1. Operacje porównania z wartością pustą (NULL) oraz logika trójwartościowa	258
7.1.2. Zapytania zagnieżdżone, krotki oraz porównywanie zbiorów i wielozbiorów	260
7.1.3. Zagnieżdżone zapytania skorelowane	262
7.1.4. Dostępne w języku SQL funkcje EXISTS i UNIQUE	263
7.1.5. Jawne deklarowanie zbiorów i zmienianie nazw atrybutów w języku SQL	265
7.1.6. Tabele połączone w języku SQL	266
7.1.7. Funkcje agregujące w języku SQL	268
7.1.8. Grupowanie: klauzule GROUP BY i HAVING	270
7.1.9. Inne konstrukcje języka SQL: WITH i CASE	273
7.1.10. Zapytania rekurencyjne w języku SQL	274
7.1.11. Omówienie i podsumowanie zapytań języka SQL	275
7.2. Definiowanie ograniczeń w postaci asercji i działań w postaci wyzwalaczy	277
7.2.1. Definiowanie ogólnych ograniczeń w postaci asercji w języku SQL	277
7.2.2. Wprowadzenie do wyzwalaczy w języku SQL	278
7.3. Perspektywy (tabele wirtualne) w języku SQL	280
7.3.1. Pojęcie perspektywy w języku SQL	280
7.3.2. Definiowanie perspektyw w języku SQL	280
7.3.3. Implementacja perspektyw i mechanizm ich aktualizowania	282
7.3.4. Perspektywy jako mechanizm uwierzytelniania	284
7.4. Dostępne w języku SQL polecenia zmiany schematu	285
7.4.1. Polecenie DROP	285
7.4.2. Polecenie ALTER	286
7.5. Podsumowanie	287
Pytania powtórkowe	289
Ćwiczenia	289
Wybrane publikacje	291

8. Algebra relacyjna i rachunek relacji 293

8.1. Relacyjne operacje unarne: selekcja i projekcja	295
8.1.1. Operacja selekcji	295
8.1.2. Operacja projekcji	297
8.1.3. Sekwencje operacji i operacja ZMIANA NAZWY	299
8.2. Operacje algebry relacyjnej pochodzące z teorii zbiorów	301
8.2.1. Operacje sumy, części wspólnej i różnicy	301
8.2.2. Operacja iloczynu (produktu) kartezjańskiego	303

8.3.	Binarne operacje na relacjach: złączenie i dzielenie	305
8.3.1.	Operacja złączenia	305
8.3.2.	Odmiany operacji złączenia: operacje równo-złączenia i złączenia naturalnego	307
8.3.3.	Kompletny zbiór operacji algebry relacyjnej	309
8.3.4.	Operacja dzielenia	310
8.3.5.	Notacja drzew zapytań	313
8.4.	Dodatkowe operacje relacyjne	314
8.4.1.	Uogólniona projekcja	314
8.4.2.	Funkcje agregujące i mechanizm grupowania	314
8.4.3.	Rekurencyjne operacje domknięcia	316
8.4.4.	Operacje złączenia zewnętrznego	317
8.4.5.	Operacja sumy zewnętrznej	319
8.5.	Przykłady zapytań w algebrze relacyjnej	320
8.6.	Relacyjny rachunek krotek	323
8.6.1.	Zmienne krotek i relacje zakresowe	324
8.6.2.	Wyrażenia i wzory w relacyjnym rachunku krotek	325
8.6.3.	Kwantyfikatory uniwersalne i egzystencjalne	327
8.6.4.	Przykładowe zapytania w relacyjnym rachunku krotek	328
8.6.5.	Notacja używana dla grafów zapytań	329
8.6.6.	Wzajemne przekształcanie kwantyfikatorów uniwersalnych i egzystencjalnych	330
8.6.7.	Stosowanie kwantyfikatorów uniwersalnych w zapytaniach	331
8.6.8.	Bezpieczne wyrażenia	333
8.7.	Relacyjny rachunek dziedzin	334
8.8.	Podsumowanie	336
	Pytania powtórkowe	338
	Ćwiczenia	338
	Ćwiczenia laboratoryjne	343
	Wybrane publikacje	346

9.	Projektowanie relacyjnych baz danych przez odwzorowywanie modelu ER i EER w model relacyjny	347
9.1.	Projektowanie relacyjnych baz danych w oparciu o odwzorowywanie modelu ER w model relacyjny	348
9.1.1.	Algorytm odwzorowujący model ER w model relacyjny	348
9.1.2.	Omówienie i podsumowanie odwzorowania konstrukcji modelu ER w odpowiednie konstrukcje modelu relacyjnego	355
9.2.	Odwzorowania konstrukcji modelu EER w relacje	357
9.2.1.	Odwzorowywanie specjalizacji i generalizacji	357
9.2.2.	Odwzorowywanie współdzielonych podklas (konstrukcji dziedziczenia wielokrotnego)	360
9.2.3.	Odwzorowywanie kategorii (typów unii)	361
9.3.	Podsumowanie	361
	Pytania powtórkowe	362
	Ćwiczenia	362
	Ćwiczenia laboratoryjne	364
	Wybrane publikacje	365

IV. Techniki programowania baz danych**367**

10. Wprowadzenie do technik programowania w języku SQL	369
10.1. Przegląd technik i zagadnień z obszaru programowania baz danych	370
10.1.1. Strategie programowania baz danych	371
10.1.2. Niezgodność impedancji	372
10.1.3. Typowa sekwencja operacji składających się na interakcję w programowaniu baz danych	373
10.2. Osadzony język SQL, dynamiczny język SQL oraz język SQLJ	374
10.2.1. Wyszukiwanie pojedynczych krotek za pomocą poleceń osadzonego języka SQL	375
10.2.2. Przetwarzanie wyników zapytań za pomocą kursorów	379
10.2.3. Określanie zapytań w czasie wykonywania programu — stosowanie dynamicznego języka SQL	381
10.2.4. SQLJ: osadzanie poleceń języka SQL w języku Java	383
10.2.5. Używanie iteratorów do przetwarzania wyników zapytań w standardzie SQLJ	385
10.3. Programowanie baz danych z wywołaniami funkcji i bibliotekami klas: SQL/CLI oraz JDBC	388
10.3.1. Programowanie baz danych z wykorzystaniem interfejsu SQL/CLI oraz języka C w roli nadrzędnego języka programowania	389
10.3.2. JDBC: biblioteka klas języka SQL służąca do programowania w języku Java	393
10.4. Procedury składowane w bazie danych i technika SQL/PSM	398
10.4.1. Procedury i funkcje składowane w bazie danych	399
10.4.2. SQL/PSM: Rozszerzenie standardu SQL o możliwość określania trwale składowanych modułów	400
10.5. Porównanie trzech opisanych podejść	402
10.6. Podsumowanie	403
Pytania powtórkowe	403
Ćwiczenia	404
Wybrane publikacje	404
11. Programowanie internetowych baz danych z użyciem języka PHP	405
11.1. Prosty przykład zastosowania PHP	406
11.2. Przegląd podstawowych mechanizmów języka PHP	408
11.2.1. Zmienne, typy danych i konstrukcje programistyczne języka PHP	409
11.2.2. Tablice w PHP	410
11.2.3. Funkcje w języku PHP	412
11.2.4. Zmienne i formularze serwera PHP	414
11.3. Przegląd programowania baz danych za pomocą PHP	415
11.3.1. Nawiązywanie połączenia z bazą danych	415
11.3.2. Pobieranie danych z formularzy i wstawianie rekordów	417
11.3.3. Zapytania pobierające dane z tabel bazy	418
11.4. Krótki przegląd technologii programowania internetowych baz danych w Javie	419
11.5. Podsumowanie	420
Pytania powtórkowe	420
Ćwiczenia	421
Wybrane publikacje	421

V. Podejścia obiektowe, obiektowo-relacyjne i XML: zagadnienia, modele, języki i standardy

423

12. Bazy obiektowe i obiektowo-relacyjne	425
12.1. Przegląd pojęć obiektowych	427
12.1.1. Wprowadzenie do pojęć i cech obiektowych	427
12.1.2. Tożsamość obiektów i porównanie obiektów z literałami	429
12.1.3. Złożone struktury typów obiektów i literałów	430
12.1.4. Enkapsulacja operacji i trwałość obiektów	432
12.1.5. Hierarchia typów i dziedziczenie	436
12.1.6. Inne pojęcia obiektowe	438
12.1.7. Podsumowanie zagadnień dotyczących obiektowych baz danych	440
12.2. Rozszerzenia obiektowe w standardzie SQL	440
12.2.1. Typy definiowane przez użytkownika za pomocą polecenia CREATE TYPE i obiekty złożone	443
12.2.2. Identyfikatory obiektów oparte na odwołaniach	444
12.2.3. Tworzenie tabel z wykorzystaniem UDT	444
12.2.4. Enkapsulacja operacji	445
12.2.5. Dziedziczenie i przeciążanie funkcji	446
12.2.6. Określanie związków za pomocą odwołań	446
12.3. Model obiektowy ODMG i język definiowania obiektów ODL	447
12.3.1. Przegląd modelu obiektowego ODMG	448
12.3.2. Dziedziczenie w modelu obiektowym ODMG	453
12.3.3. Wbudowane interfejsy i klasy w modelu obiektowym	454
12.3.4. Obiekty atomowe (definiowane przez użytkownika)	456
12.3.5. Ekstensje, klucze i obiekty-fabryki	458
12.3.6. Język definicji obiektów ODL	460
12.4. Projektowanie koncepcyjne obiektowej bazy danych	465
12.4.1. Różnice pomiędzy koncepcyjnym projektowaniem obiektowych i relacyjnych baz danych	465
12.4.2. Odwzorowywanie schematu EER na schemat obiektowy	466
12.5. Obiektowy język zapytań OQL	468
12.5.1. Proste zapytania OQL, punkty wejścia bazy danych i zmienne iterujące	469
12.5.2. Wyniki zapytań i wyrażenia ścieżkowe	470
12.5.3. Inne cechy OQL	472
12.6. Przegląd wiązania z językiem C++ w standardzie ODMG	476
12.7. Podsumowanie	478
Pytania powtórkowe	479
Ćwiczenia	480
Wybrane publikacje	481
13. XML — rozszerzalny język znaczników	483
13.1. Dane strukturalne, półstrukturalne i niestukturalne	484
13.2. Hierarchiczny (drzewiasty) model danych w dokumentach XML	488
13.3. Dokumenty XML, DTD i schematy	491
13.3.1. Dobrze uformowane i prawidłowe dokumenty XML oraz XML DTD	491
13.3.2. Schematy XML	494
13.4. Zapisywanie dokumentów XML w bazach i ich pobieranie	499
13.5. Języki związane ze standardem XML	500
13.5.1. XPath, czyli określanie ścieżek w dokumentach XML	500
13.5.2. XQuery: definiowanie zapytań w XML	502
13.5.3. Inne języki i protokoły związane ze standardem XML	504

13.6. Pobieranie dokumentów XML z relacyjnych baz danych	505
13.6.1. Tworzenie hierarchicznych perspektyw w formacie XML dla danych płaskich lub zapisanych w grafie	505
13.6.2. Przerywanie cykli w celu zamiany grafów w drzewa	509
13.6.3. Dodatkowe kroki związane z tworzeniem dokumentu XML na podstawie bazy danych	510
13.7. XML/SQL: funkcje języka SQL generujące dane w formacie XML	511
13.8. Podsumowanie	512
Pytania powtórkowe	513
Ćwiczenia	513
Wybrane publikacje	514

VI. Teoria projektowania baz danych i normalizacja **515**

14. Podstawy zależności funkcyjnych i normalizacji w relacyjnych bazach danych	517
14.1. Nieformalne wskazówki dotyczące projektowania schematów relacji	519
14.1.1. Wymuszanie jednoznacznej semantyki atrybutów relacji	519
14.1.2. Nadmiarowe informacje w krotkach oraz anomalie aktualizacji	521
14.1.3. Wartości null w krotkach	525
14.1.4. Generowanie fałszywych krotek	526
14.1.5. Podsumowanie i omówienie wskazówek projektowych	528
14.2. Zależności funkcyjne	528
14.2.1. Definicja zależności funkcyjnej	529
14.3. Postaci normalne oparte na kluczach głównych	532
14.3.1. Normalizacja relacji	532
14.3.2. Praktyczne zastosowania postaci normalnych	533
14.3.3. Definicje kluczy i atrybutów należących do kluczy	534
14.3.4. Pierwsza postać normalna	535
14.3.5. Druga postać normalna	539
14.3.6. Trzecia postać normalna	540
14.4. Definicje ogólne drugiej i trzeciej postaci normalnej	542
14.4.1. Definicja ogólna drugiej postaci normalnej	542
14.4.2. Definicja ogólna trzeciej postaci normalnej	544
14.4.3. Interpretacja definicji ogólnej trzeciej postaci normalnej	544
14.5. Postać normalna Boyce'a-Codda	545
14.5.1. Dekompozycja relacji niebędących w BCNF	547
14.6. Zależności wielowartościowe i czwarta postać normalna	549
14.6.1. Formalna definicja zależności wielowartościowej	549
14.7. Zależności złączeniowe i piąta postać normalna	552
14.8. Podsumowanie	553
Pytania powtórkowe	555
Ćwiczenia	556
Ćwiczenia laboratoryjne	560
Wybrane publikacje	560
15. Algorytmy projektowania relacyjnych baz danych i dodatkowe zależności	563
15.1. Inne zagadnienia z obszaru zależności funkcyjnych: reguły wnioskowania, równoważności i pokrycie minimalne	564
15.1.1. Reguły wnioskowania dla zależności funkcyjnych	565
15.1.2. Równoważność zbiorów zależności funkcyjnych	569
15.1.3. Zbiory minimalne zależności funkcyjnych	570

15.2. Właściwości dekompozycji relacyjnych	573
15.2.1. Dekompozycja relacji i niewystarczalność postaci normalnych	573
15.2.2. Właściwość zachowania zależności dekompozycji	574
15.2.3. Właściwość złączenia bezstratnego (nieaddytywnego) dekompozycji	575
15.2.4. Testowanie dekompozycji binarnych pod względem występowania właściwości złączenia nieaddytywnego	577
15.2.5. Kolejne dekompozycje o złączeniach nieaddytywnych	578
15.3. Algorytmy projektowania schematów relacyjnych baz danych	579
15.3.1. Dekompozycja na schematy w trzeciej postaci normalnej z zachowaniem zależności i właściwością złączenia nieaddytywnego (bezstratnego)	579
15.3.2. Dekompozycja ze złączeniem nieaddytywnym na schematy w postaci normalnej Boyce'a-Codda	582
15.4. Problemy związane z wartościami pustymi i krotkami zawieszonymi oraz inne projekty relacyjne	584
15.4.1. Problemy związane z wartościami pustymi i krotkami zawieszonymi	584
15.4.2. Omówienie algorytmów normalizacyjnych i innych projektów relacyjnych	586
15.5. Dalsze omówienie zależności wielowartościowych i 4NF	588
15.5.1. Reguły wnioskowania dla zależności funkcyjnych i wielowartościowych	588
15.5.2. Jeszcze o czwartej postaci normalnej	589
15.5.3. Dekompozycja ze złączeniem nieaddytywnym na relacje w czwartej postaci normalnej	591
15.6. Inne zależności i postaci normalne	592
15.6.1. Zależności złączeniowe i piąta postać normalna	592
15.6.2. Zależności zawierania	592
15.6.3. Zależności funkcyjne oparte na funkcjach i procedurach arytmetycznych	593
15.6.4. Postać normalna klucza dziedziny	594
15.7. Podsumowanie	595
Pytania powtórkowe	595
Ćwiczenia	596
Ćwiczenia laboratoryjne	598
Wybrane publikacje	598

VII. Struktury plikowe, funkcje mieszające, indeksowanie i projekty fizyczne baz danych **601**

16. Składowanie danych na dysku, podstawowe struktury plikowe, funkcje mieszające i nowoczesne struktury składowania	603
16.1. Wprowadzenie	604
16.1.1. Hierarchie pamięciowe i urządzenia składowania danych	605
16.1.2. Przechowywanie baz danych	607
16.2. Drugorzędne urządzenia pamięciowe	609
16.2.1. Opis sprzętowy napędów dyskowych	609
16.2.2. Zwiększanie wydajności dostępu do danych na dysku	615
16.2.3. Pamięć masowa SSD	616
16.2.4. Taśmowe urządzenia pamięciowe	617
16.3. Buforowanie bloków	619
16.3.1. Zarządzanie buforem	620
16.3.2. Strategie zastępowania danych w buforze	622
16.4. Rozmieszczanie rekordów plików na dysku	623
16.4.1. Rekordy i typy rekordów	623
16.4.2. Pliki oraz rekordy o stałej i zmiennej długości	624

16.4.3. Rozmieszczenie rekordów w blokach i rekordy segmentowane oraz niesegmentowane	626
16.4.4. Alokowanie bloków pliku na dysku	627
16.4.5. Nagłówki plików	627
16.5. Operacje wykonywane na plikach	627
16.6. Pliki nieuporządkowanych rekordów (pliki stertowe)	631
16.7. Pliki uporządkowanych rekordów (pliki posortowane)	632
16.8. Techniki mieszania	636
16.8.1. Mieszanie wewnętrzne	636
16.8.2. Mieszanie zewnętrzne dla plików na dysku	639
16.8.3. Techniki mieszania umożliwiające dynamiczne rozszerzanie plików	642
16.9. Inne podstawowe metody organizacji plików	647
16.9.1. Pliki rekordów mieszanych	647
16.9.2. B-drzewa i inne struktury danych służące jako podstawowe metody organizacji	648
16.10. Zapewnianie równoległego dostępu do dysku przy użyciu architektury RAID	648
16.10.1. Zwiększanie niezawodności przy użyciu architektury RAID	650
16.10.2. Poprawianie wydajności przy użyciu architektury RAID	651
16.10.3. Metody organizacji i poziomy architektury RAID	651
16.11. Nowoczesne architektury składowania danych	653
16.11.1. Sieci obszarów składowania danych	653
16.11.2. Technologia NAS	654
16.11.3. iSCSI i inne sieciowe protokoły składowania danych	655
16.11.4. Technologia Automated Storage Tiering	656
16.11.5. Obiektowa pamięć masowa	656
16.12. Podsumowanie	657
Pytania powtórkowe	658
Ćwiczenia	660
Wybrane publikacje	663
17. Struktury indeksowe dla plików i fizyczne projekty baz danych	665
17.1. Rodzaje jednopoziomowych indeksów uporządkowanych	666
17.1.1. Indeksy główne	667
17.1.2. Indeksy klastrowania	670
17.1.3. Indeksy drugorzędne	673
17.1.4. Podsumowanie	677
17.2. Indeksy wielopoziomowe	677
17.3. Dynamiczne indeksy wielopoziomowe z użyciem B-drzew i B+-drzew	681
17.3.1. Drzewa wyszukiwania i B-drzewa	682
17.3.2. B+-drzewa	687
17.4. Indeksy na wielu kluczach	696
17.4.1. Indeks uporządkowany na wielu atrybutach	697
17.4.2. Mieszanie partycjonowane	697
17.4.3. Pliki matrycowe	697
17.5. Inne rodzaje indeksów	699
17.5.1. Indeksy oparte na mieszaniu	699
17.5.2. Indeksy bitmapowe	699
17.5.3. Indeksowanie oparte na funkcji	702
17.6. Ogólne zagadnienia związane z indeksami	703
17.6.1. Indeksy logiczne a fizyczne	703
17.6.2. Tworzenie indeksu	704
17.6.3. Dostrajanie indeksów	705
17.6.4. Dodatkowe kwestie związane ze składowaniem relacji i indeksów	706

17.7. Fizyczne projektowanie baz danych w przypadku baz relacyjnych	708
17.7.1. Czynniki wpływające na fizyczny projekt bazy danych	708
17.7.2. Decyzje dotyczące fizycznego projektu bazy danych	710
17.8. Podsumowanie	711
Pytania powtórkowe	713
Ćwiczenia	714
Wybrane publikacje	716

VIII. Przetwarzanie i optymalizacja zapytań

719

18. Strategie przetwarzania zapytań	721
18.1. Translacja zapytań języka SQL do postaci wyrażeń algebry relacji i innych operacji	724
18.1.1. Dodatkowe operatory połączeń częściowych i antyzłączeń	725
18.2. Algorytmy sortowania zewnętrznego	727
18.3. Algorytmy operacji selekcji	729
18.3.1. Możliwości implementacji operacji SELECT	729
18.3.2. Metody wyszukiwania dla selekcji na podstawie warunku koniunktywnego	731
18.3.3. Metody wyszukiwania dla selekcji na podstawie alternatywy logicznej	733
18.3.4. Szacowanie selektywności warunku	733
18.4. Implementacja operacji JOIN	734
18.4.1. Metody implementacji połączeń	735
18.4.2. Wpływ dostępnej przestrzeni bufora i pliku używanego w pętli zewnętrznej na wydajność operacji połączenia w pętli zagnieżdżonej	738
18.4.3. Wpływ współczynnika selekcji połączenia na wydajność tej operacji	739
18.4.4. Ogólna postać partycjonowanego połączenia mieszającego	741
18.4.5. Hybrydowe łączenie mieszające	742
18.5. Algorytmy operacji projekcji i teoriomnogościowych	743
18.5.1. Stosowanie antyzłączeń w operacji SET DIFFERENCE (EXCEPT lub MINUS w języku SQL)	744
18.6. Implementacja operacji agregujących oraz połączeń różnego rodzaju	745
18.6.1. Implementacja operacji agregujących	745
18.6.2. Implementacja różnego rodzaju połączeń	746
18.7. Łączenie operacji poprzez mechanizm potokowy	748
18.7.1. Iteratory używane do implementowania operacji fizycznych	749
18.8. Algorytmy równoległego przetwarzania zapytań	750
18.8.1. Równoległość na poziomie operatorów	751
18.8.2. Równoległość w jednym zapytaniu	754
18.8.3. Równoległość w wielu zapytaniach	754
18.9. Podsumowanie	755
Pytania powtórkowe	755
Ćwiczenia	756
Wybrane publikacje	756
19. Optymalizacja zapytań	757
19.1. Drzewa zapytań i heurystyki optymalizacji zapytań	758
19.1.1. Notacja drzew zapytań i grafów zapytań	758
19.1.2. Heurystyczna optymalizacja drzew zapytań	760
19.2. Wybór planów wykonania zapytań	767
19.2.1. Różne sposoby wykonywania zapytań	767
19.2.2. Optymalizacja podzapytań zagnieżdżonych	768
19.2.3. Scalanie podzapytań (perspektyw)	770
19.2.4. Perspektywy zmaterializowane	772

19.3. Wykorzystanie selektywności w optymalizacji kosztowej	775
19.3.1. Składowe koszty wykonywania zapytań	776
19.3.2. Informacje z katalogu używane w funkcjach kosztu	777
19.3.3. Histogramy	778
19.4. Funkcje kosztu dla operacji SELECT	778
19.4.1. Przykład optymalizacji selekcji na podstawie wzorów szacowania kosztów	781
19.5. Przykłady funkcji kosztu dla operacji JOIN	783
19.5.1. Selektywność i liczność złączeń częściowych i antyzłączeń	785
19.5.2. Przykład optymalizacji złączenia na podstawie funkcji kosztu	786
19.5.3. Zapytania dotyczące wielu relacji i porządkowanie złączeń	787
19.5.4. Optymalizacja fizyczna	789
19.5.5. Określanie kolejności złączeń za pomocą programowania dynamicznego	790
19.6. Przykład ilustrujący kosztową optymalizację zapytań	792
19.7. Dodatkowe zagadnienia związane z optymalizacją zapytań	794
19.7.1. Wyświetlanie planu wykonania zapytania uzyskanego przez system	794
19.7.2. Szacowanie wielkości wyników dla innych operacji	795
19.7.3. Zapis planu w pamięci podręcznej	796
19.7.4. Optymalizacja z wykorzystaniem pierwszych k wyników	796
19.8. Przykład optymalizacji zapytań w hurtowniach danych	796
19.9. Optymalizacja zapytań w bazach Oracle	799
19.9.1. Optymalizator fizyczny	799
19.9.2. Globalny optymalizator zapytań	799
19.9.3. Optymalizacja adaptacyjna	800
19.9.4. Przetwarzanie tablicowe	801
19.9.5. Wskazówki	801
19.9.6. Zarysy	802
19.9.7. Zarządzanie planami wykonywania instrukcji języka SQL	802
19.10. Semantyczna optymalizacja zapytań	803
19.11. Podsumowanie	804
Pytania powtórkowe	804
Ćwiczenia	805
Wybrane publikacje	806

IX. Przetwarzanie transakcji, sterowanie współbieżne i odtwarzanie baz danych

809

20. Wprowadzenie do problematyki i teorii przetwarzania transakcji	811
20.1. Wprowadzenie do problematyki przetwarzania transakcji	812
20.1.1. Systemy jedno- i wieloużytkownikowe	812
20.1.2. Transakcje, elementy baz danych, operacje odczytu i zapisu oraz bufora SZBD	813
20.1.3. Uzasadnienie potrzeby stosowania sterowania współbieżnego	815
20.1.4. Uzasadnienie potrzeby odtwarzania	818
20.2. Pojęcia dotyczące transakcji i systemu	819
20.2.1. Stany transakcji i dodatkowe operacje	819
20.2.2. Dziennik systemowy	821
20.2.3. Punkt zatwierdzenia transakcji	822
20.2.4. Strategie zastępowania bufora specyficzne dla SZBD	823
20.3. Pożądane właściwości transakcji	824
20.4. Charakteryzowanie harmonogramów na podstawie możliwości odtwarzania	825
20.4.1. Harmonogramy (historie) transakcji	825
20.4.2. Charakterystyka harmonogramów według możliwości odtwarzania	827



20.5. Charakterystyka harmonogramów według ich szeregowalności	829
20.5.1. Harmonogramy szeregowo, nieszeregowo oraz konfliktowo-szeregowalne	830
20.5.2. Sprawdzanie występowania szeregowalności konfliktowej harmonogramu	834
20.5.3. Wykorzystywanie szeregowalności do sterowania współbieżnego	837
20.5.4. Równoważność perspektywiczna i szeregowalność perspektywiczna	839
20.5.5. Inne rodzaje równoważności harmonogramów	840
20.6. Obsługa transakcji w języku SQL	841
20.7. Podsumowanie	843
Pytania powtórkowe	844
Ćwiczenia	845
Wybrane publikacje	846
21. Techniki sterowania współbieżnego	847
21.1. Techniki blokowania dwufazowego dla celów sterowania współbieżnego	848
21.1.1. Rodzaje blokad i systemowe tabele blokad	848
21.1.2. Gwarantowanie szeregowalności blokowania dwufazowego	853
21.1.3. Problem zakleszczenia i zagłodzenia	856
21.2. Sterowanie współbieżne w oparciu o uporządkowanie według znaczników czasu	860
21.2.1. Znaczniki czasu	860
21.2.2. Algorytm uporządkowania według znaczników czasu	860
21.3. Techniki wielowersyjnego sterowania współbieżnego	862
21.3.1. Technika wielowersyjna oparta na porządkowaniu według znaczników czasu	863
21.3.2. Wielowersyjne blokowanie dwufazowe z użyciem blokad certyfikujących	864
21.4. Sterowanie współbieżne z użyciem technik walidacyjnych (optymistycznych) i izolacji snapshotów	866
21.4.1. Walidacyjne (optymistyczne) sterowanie współbieżne	866
21.4.2. Sterowanie współbieżne oparte na izolacji snapshotów	868
21.5. Ziarnistość elementów danych i blokowanie z wieloma poziomami ziarnistości	869
21.5.1. Kwestie dotyczące poziomu ziarnistości w przypadku blokowania	869
21.5.2. Blokowanie z wieloma poziomami ziarnistości	870
21.6. Użycie blokad dla celów sterowania współbieżnego w przypadku indeksów	872
21.7. Inne kwestie związane ze sterowaniem współbieżnym	874
21.7.1. Wstawianie, usuwanie i rekordy fantomowe	875
21.7.2. Transakcje interaktywne	876
21.7.3. Zatrzaski	876
21.8. Podsumowanie	876
Pytania powtórkowe	877
Ćwiczenia	878
Wybrane publikacje	879
22. Techniki odtwarzania baz danych	881
22.1. Pojęcia związane z odtwarzaniem	882
22.1.1. Zarys problematyki odtwarzania i podział algorytmów odtwarzania na odrębne kategorie	882
22.1.2. Zapisywanie w pamięci podręcznej (buforowanie) bloków dyskowych	883
22.1.3. Rejestrowanie zapisów z wyprzedzeniem, technika zabierania oraz wymuszania	885
22.1.4. Punkty kontrolne w dzienniku systemowym oraz tworzenie przybliżonych punktów kontrolnych	887
22.1.5. Wycofywanie transakcji i wycofywanie kaskadowe	888
22.1.6. Działania transakcji niewpływające na bazy danych	890
22.2. Techniki odtwarzania NO-UNDO/REDO oparte na aktualizacjach odroczonej	890
22.3. Techniki odtwarzania oparte na aktualizacjach natychmiastowych	893

22.4. Stronicowanie z przesłanianiem	895
22.5. Algorytm odtwarzania ARIES	897
22.6. Odtwarzanie w systemach wielu baz danych	901
22.7. Tworzenie kopii bezpieczeństwa bazy danych i odtwarzanie po awariach katastroficznych	902
22.8. Podsumowanie	903
Pytania powtórkowe	904
Ćwiczenia	905
Wybrane publikacje	908

X. Rozproszone bazy danych, systemy NOSQL i big data **911**

23. Zagadnienia z obszaru rozproszonych baz danych	913
23.1. Zagadnienia z obszaru rozproszonych baz danych	914
23.1.1. Co sprawia, że baza danych jest rozproszona?	914
23.1.2. Przezroczystość	915
23.1.3. Stabilność i dostępność	916
23.1.4. Skalowalność i odporność na podział	917
23.1.5. Autonomia	917
23.1.6. Zalety rozproszonych baz danych	917
23.2. Techniki fragmentacji, replikacji i alokacji danych w projekcie rozproszonej bazy danych	918
23.2.1. Fragmentacja danych i sharding	918
23.2.2. Replikacja i alokacja danych	921
23.2.3. Przykłady fragmentacji, alokacji i replikacji danych	922
23.3. Techniki sterowania współbieżnego i odtwarzania danych w rozproszonych bazach danych	925
23.3.1. Rozproszone sterowanie współbieżne oparte na wyróżnionej kopii danych	925
23.3.2. Rozproszone sterowanie współbieżne oparte na głosowaniu	927
23.3.3. Rozproszone odtwarzanie danych	927
23.4. Przegląd zarządzania transakcjami w rozproszonych bazach danych	928
23.4.1. Protokół zatwierdzania dwufazowego	929
23.4.2. Protokół zatwierdzania trójfazowego	929
23.4.3. Obsługa zarządzania transakcjami w systemie operacyjnym	930
23.5. Przetwarzanie zapytań i optymalizacja w rozproszonych bazach danych	930
23.5.1. Przetwarzanie zapytań rozproszonych	931
23.5.2. Koszty przesyłu danych w przetwarzaniu zapytań rozproszonych	931
23.5.3. Rozproszone przetwarzanie zapytań z użyciem złączeń częściowych	933
23.5.4. Dekompozycja zapytań i aktualizacji	934
23.6. Rodzaje rozproszonych systemów baz danych	937
23.6.1. Zarządzanie federacyjnymi systemami baz danych	938
23.7. Architektury rozproszonych baz danych	940
23.7.1. Architektura równoległa a rozproszona	940
23.7.2. Ogólna architektura czystych baz rozproszonych	942
23.7.3. Architektura federacyjnych baz danych	942
23.7.4. Przegląd trójwarstwowej architektury klient-serwer	944
23.8. Zarządzanie rozproszonym katalogiem	946
23.9. Podsumowanie	947
Pytania powtórkowe	948
Ćwiczenia	949
Wybrane publikacje	951

24. Bazy danych NOSQL i systemy składowania big data	955
24.1. Wprowadzenie do systemów NOSQL	955
24.1.1. Powstanie systemów NOSQL	955
24.1.2. Cechy systemów NOSQL	957
24.1.3. Kategorie systemów NOSQL	959
24.2. Twierdzenie CAP	960
24.3. Dokumentowe systemy NOSQL i baza MongoDB	961
24.3.1. Model danych z systemu MongoDB	962
24.3.2. Operacje CRUD w systemie MongoDB	965
24.3.3. Cechy systemu rozproszonego MongoDB	965
24.4. Magazyny NOSQL z parami klucz-wartość	967
24.4.1. Przegląd systemu DynamoDB	967
24.4.2. Rozproszony magazyn danych z parami klucz-wartość — Voldemort	968
24.4.3. Przykładowe inne magazyny z parami klucz-wartość	971
24.5. Kolumnowe systemy NOSQL	972
24.5.1. Model danych i wersjonowanie w systemie HBase	972
24.5.2. Operacje CRUD w systemie HBase	974
24.5.3. Zagadnienia związane ze składowaniem danych i systemem rozproszonym w HBase	974
24.6. Grafowe bazy NOSQL i system Neo4j	975
24.6.1. Model danych w systemie Neo4j	975
24.6.2. Język zapytań Cypher w systemie Neo4j	977
24.6.3. Cechy interfejsów i systemu rozproszonego w Neo4j	979
24.7. Podsumowanie	980
Pytania powtórkowe	981
Wybrane publikacje	982
25. Technologie z obszaru big data oparte na modelu MapReduce i systemie Hadoop	983
25.1. Czym jest big data?	986
25.2. Wprowadzenie do technologii MapReduce i Hadoop	988
25.2.1. Tło historyczne	988
25.2.2. Model MapReduce	989
25.2.3. Wersje Hadoopa	993
25.3. System HDFS	993
25.3.1. Wymagania wstępne związane z systemem HDFS	994
25.3.2. Architektura systemu HDFS	994
25.3.3. Operacje wejścia-wyjścia na plikach i zarządzanie replikami w systemie HDFS	996
25.3.4. Skalowalność systemu HDFS	997
25.3.5. Ekosystem Hadoopa	998
25.4. Model MapReduce: dodatkowe szczegóły	999
25.4.1. Środowisko uruchomieniowe MapReduce	999
25.4.2. Przykład: złączenia w modelu MapReduce	1002
25.4.3. Apache Hive	1006
25.4.4. Zalety technologii Hadoop i MapReduce	1008
25.5. Hadoop 2 (nazywany też YARN)	1009
25.5.1. Uzasadnienie powstania platformy YARN	1009
25.5.2. Architektura platformy YARN	1012
25.5.3. Inne platformy w YARN	1015
25.6. Ogólne omówienie	1017
25.6.1. Hadoop i MapReduce a równoległe relacyjne SZBD	1017
25.6.2. Big data w chmurach obliczeniowych	1019

25.6.3. Problemy z lokalnością danych i optymalizacja zasobów w aplikacjach z obszaru big data działających w chmurze	1021
25.6.4. YARN jako platforma usług z obszaru danych	1022
25.6.5. Wyzwania związane z technologiami z obszaru big data	1023
25.6.6. Przyszłość	1024
25.7. Podsumowanie	1026
Pytania powtórkowe	1027
Wybrane publikacje	1028

XI. Zaawansowane modele, systemy i zastosowania baz danych

1031

26. Rozszerzone modele danych: wprowadzenie do aktywnych, czasowych, przestrzennych, multimedialnych i dedukcyjnych baz danych	1033
26.1. Wyzwalacze i inne pojęcia związane z aktywnymi bazami danych	1035
26.1.1. Uogólniony model aktywnych baz danych i wyzwalacze w Oracle	1035
26.1.2. Projektowanie i implementacja aktywnych baz danych	1039
26.1.3. Przykładowe aktywne reguły poziomu wyrażenia w systemie STARBURST	1042
26.1.4. Możliwe zastosowania aktywnych baz danych	1044
26.1.5. Wyzwalacze w SQL-99	1044
26.2. Koncepcja czasowych baz danych	1045
26.2.1. Reprezentacja czasu, kalendarze i wymiary czasu	1046
26.2.2. Wprowadzenie czasu do relacyjnych baz danych z obsługą wersji krotek	1048
26.2.3. Czas w obiektowych bazach danych z obsługą wersji atrybutów	1053
26.2.4. Konstruowanie zapytań czasowych i język TSQL2	1055
26.2.5. Szeregi czasowe	1057
26.3. Zagadnienia z obszaru przestrzennych baz danych	1058
26.3.1. Wprowadzenie do przestrzennych baz danych	1058
26.3.2. Typy i modele danych przestrzennych	1059
26.3.3. Operatory i zapytania przestrzenne	1060
26.3.4. Indeksowanie danych przestrzennych	1062
26.3.5. Eksploracja danych przestrzennych	1063
26.3.6. Zastosowania danych przestrzennych	1065
26.4. Zagadnienia z obszaru multimedialnych baz danych	1065
26.4.1. Automatyczna analiza obrazów	1067
26.4.2. Wykrywanie obiektów na obrazach	1068
26.4.3. Semantyczne opisywanie obrazów	1069
26.4.4. Analizy danych audio	1069
26.5. Wprowadzenie do dedukcyjnych baz danych	1070
26.5.1. Przegląd dedukcyjnych baz danych	1070
26.5.2. Notacja języków Prolog i Datalog	1071
26.5.3. Notacja języka Datalog	1073
26.5.4. Forma klauzulowa i klauzule Horna	1074
26.5.5. Interpretacja reguł	1075
26.5.6. Programy w języku Datalog i ich bezpieczeństwo	1077
26.5.7. Zastosowanie operacji relacyjnych	1081
26.5.8. Wykonywanie zapytań nierekurencyjnych	1081
26.6. Podsumowanie	1083
Pytania powtórkowe	1085
Ćwiczenia	1086
Wybrane publikacje	1089

27. Wprowadzenie do wyszukiwania informacji i danych w internecie	1093
27.1. Zagadnienia z obszaru wyszukiwania informacji (WI)	1093
27.1.1. Wprowadzenie do wyszukiwania informacji	1094
27.1.2. Porównanie baz danych i systemów WI	1097
27.1.3. Krótka historia WI	1098
27.1.4. Tryby interakcji w systemach WI	1099
27.1.5. Ogólny proces WI	1100
27.2. Modele wyszukiwania	1101
27.2.1. Model logiczny	1101
27.2.2. Model oparty na przestrzeni wektorowej	1102
27.2.3. Model probabilistyczny	1104
27.2.4. Model semantyczny	1106
27.3. Typy zapytań w systemach WI	1107
27.3.1. Zapytania oparte na słowach kluczowych	1107
27.3.2. Zapytania logiczne	1107
27.3.3. Zapytania do wyszukiwania fraz	1108
27.3.4. Zapytania z określaniem odległości słów	1108
27.3.5. Zapytania z symbolami wieloznacznymi	1108
27.3.6. Zapytania w języku naturalnym	1109
27.4. Wstępne przetwarzanie tekstu	1109
27.4.1. Usuwanie słów pomijanych	1109
27.4.2. Stemming	1109
27.4.3. Korzystanie z tezaury	1110
27.4.4. Inne etapy przetwarzania: cyfry, myślniki, znaki przestankowe, wielkość znaków	1111
27.4.5. Wydobywanie informacji	1112
27.5. Indeksy odwrócone	1112
27.5.1. Wprowadzenie do systemu Lucene	1114
27.6. Miary oceny adekwatności wyników wyszukiwania	1115
27.6.1. Czułość i precyzja	1116
27.6.2. Średnia precyzja	1118
27.6.3. Krzywa czułość/precyzja	1118
27.6.4. Miara F	1118
27.7. Wyszukiwanie i analizy w sieci WWW	1119
27.7.1. Analizy danych internetowych i ich związki z WI	1119
27.7.2. Analizy struktury sieci WWW	1121
27.7.3. Analizowanie struktury odsyłaczy na stronach internetowych	1122
27.7.4. Analizy treści w sieci WWW	1123
27.7.5. Podejścia analizowania treści w sieci WWW	1125
27.7.6. Analizy użytkowania witryn	1126
27.7.7. Praktyczne zastosowania analiz użytkowania witryn	1128
27.8. Trendy w wyszukiwaniu informacji	1129
27.8.1. Wyszukiwanie fasetowe	1129
27.8.2. Wyszukiwanie społecznościowe	1130
27.8.3. Wyszukiwanie informacji w dialogach	1131
27.8.4. Probabilistyczne modelowanie tematu	1131
27.8.5. Systemy odpowiadania na pytania	1132
27.9. Podsumowanie	1135
Pytania powtórkowe	1136
Wybrane publikacje	1138
28. Elementy eksploracji danych	1141
28.1. Przegląd technologii eksploracji danych	1142
28.1.1. Eksploracja danych kontra hurtownie danych	1142
28.1.2. Eksploracja danych jako część procesu odkrywania wiedzy	1142

28.1.3. Cele eksploracji danych i odkrywania wiedzy	1143
28.1.4. Rodzaje wiedzy odkrywanej w procesie eksploracji danych	1145
28.2. Reguły asocjacyjne	1146
28.2.1. Model koszyka klienta supermarketu, poziom wsparcia i poziom ufności	1146
28.2.2. Algorytm Apriori	1148
28.2.3. Algorytm próbkujący	1150
28.2.4. Drzewa częstych wzorców i algorytm ich tworzenia	1151
28.2.5. Algorytm partycjonujący	1155
28.2.6. Pozostałe typy reguł asocjacyjnych	1156
28.2.7. Dodatkowe problemy związane z regułami asocjacyjnymi	1159
28.3. Klasyfikacja	1160
28.4. Grupowanie	1163
28.5. Strategie rozwiązywania pozostałych problemów związanych z eksploracją danych	1166
28.5.1. Odkrywanie wzorców sekwencyjnych	1166
28.5.2. Odkrywanie wzorców w szeregach czasowych	1167
28.5.3. Regresja	1167
28.5.4. Sieci neuronowe	1168
28.5.5. Algorytmy genetyczne	1169
28.6. Zastosowania technik eksploracji danych	1170
28.7. Komercyjne narzędzia eksploracji danych	1170
28.7.1. Interfejs użytkownika	1171
28.7.2. Interfejs programowy aplikacji	1171
28.7.3. Kierunki przyszłego rozwoju	1171
28.8. Podsumowanie	1173
Pytania powtórkowe	1173
Ćwiczenia	1174
Wybrane publikacje	1176
29. Przegląd hurtowni danych i rozwiązań OLAP	1177
29.1. Wprowadzenie, definicje i terminologia	1177
29.2. Właściwości hurtowni danych	1179
29.3. Modelowanie danych dla hurtowni danych	1181
29.4. Budowanie hurtowni danych	1187
29.5. Typowe funkcje hurtowni danych	1191
29.6. Hurtownie danych kontra perspektywy	1192
29.7. Trudności z implementowaniem hurtowni danych	1193
29.8. Podsumowanie	1194
Pytania powtórkowe	1195
Wybrane publikacje	1195

XII. Dodatkowe zagadnienia z obszaru baz danych: bezpieczeństwo

1197

30. Bezpieczeństwo w bazach danych	1199
30.1. Wprowadzenie do bezpieczeństwa baz danych	1200
30.1.1. Rodzaje zabezpieczeń	1200
30.1.2. Środki kontroli	1201
30.1.3. Bezpieczeństwo a administrator bazy danych	1203
30.1.4. Ochrona dostępu, konta użytkowników i audyty bazy danych	1204
30.1.5. Dane wrażliwe i typy ujawnień	1205
30.1.6. Związki między bezpieczeństwem a prywatnością informacji	1206

30.2. Dyspozycyjna kontrola dostępu polegająca na nadawaniu i odbieraniu uprawnień	1207
30.2.1. Typy uprawnień dyspozycyjnych	1207
30.2.2. Określanie uprawnień przy użyciu perspektyw	1208
30.2.3. Cofanie uprawnień	1209
30.2.4. Propagacja uprawnień poprzez opcję GRANT	1209
30.2.5. Przykład	1209
30.2.6. Określanie ograniczeń propagacji uprawnień	1211
30.3. Realizacja zabezpieczeń wielopoziomowych	
za pomocą obowiązkowej kontroli dostępu i zabezpieczeń opartych na rolach	1212
30.3.1. Porównanie dyspozycyjnego i obowiązkowego modelu bezpieczeństwa	1215
30.3.2. Kontrola dostępu oparta na rolach	1215
30.3.3. Zabezpieczenia oparte na etykietach i kontrola dostępu na poziomie wierszy	1217
30.3.4. Kontrola dostępu dla danych w formacie XML	1218
30.3.5. Polityki kontroli dostępu dla aplikacji sieciowych i mobilnych	1219
30.4. Wstrzykiwanie kodu w języku SQL	1220
30.4.1. Metody wstrzykiwania kodu w języku SQL	1221
30.4.2. Zagrożenia związane ze wstrzykiwaniem kodu w języku SQL	1222
30.4.3. Techniki ochrony przed wstrzykiwaniem kodu w języku SQL	1223
30.5. Wprowadzenie do bezpieczeństwa statystycznych baz danych	1224
30.6. Wprowadzenie do kontroli przepływu	1225
30.6.1. Ukryte kanały	1226
30.7. Szyfrowanie i infrastruktura klucza publicznego	1227
30.7.1. Szyfry DES i AES	1228
30.7.2. Algorytmy z kluczem symetrycznym	1228
30.7.3. Szyfrowanie kluczem publicznym (asymetrycznym)	1228
30.7.4. Podpis cyfrowy	1230
30.7.5. Certyfikaty cyfrowe	1230
30.8. Problemy z prywatnością i jej zachowywanie	1231
30.9. Wyzwania związane z utrzymaniem bezpieczeństwa baz danych	1232
30.9.1. Jakość danych	1232
30.9.2. Prawa własności intelektualnej	1232
30.9.3. Odporność baz danych	1233
30.10. Zabezpieczenia oparte na etykietach w bazach Oracle	1233
30.10.1. Technologia wirtualnych prywatnych baz danych	1234
30.10.2. Architektura zabezpieczeń opartych na etykietach	1234
30.10.3. Współdziałanie etykiet danych i etykiet użytkowników	1235
30.11. Podsumowanie	1236
Pytania powtórkowe	1237
Ćwiczenia	1239
Wybrane publikacje	1239

Dodatki **1241**

A. Alternatywne notacje modeli związków encji	1243
B. Parametry dysków	1247
C. Omówienie języka QBE	1251
Bibliografia	1259
Skorowidz	1319