

## 15.9. Dostęp do systemów bazodanowych – język SQL

Relacyjne SZBD są wyposażone w język zapytań SQL (*Structured Query Language*). Spełnia on funkcje zarówno opisu, jak i manipulacji danymi. SQL umożliwia definiowanie, wyszukiwanie i aktualizację danych w relacyjnej bazie danych [Matosek, 2005].

Język SQL jest postrzegany przede wszystkim jako język zapytań. Istotnie, jego podstawową instrukcją jest **instrukcja wyszukiwania**, składająca się z bloku SELECT FROM WHERE, która ma postać:

```
SELECT (lista atrybutów)  
FROM (lista relacji)  
WHERE (warunek)
```

Lista atrybutów określa nazwy atrybutów, których wartości są wyszukiwane, lista relacji – nazwy relacji, które będą przetwarzane w wyniku zapytania, a warunek – krotki, których zapytanie dotyczy. Pełna składnia instrukcji wyszukiwania przedstawia się następująco:

```
SELECT (lista atrybutów)  
FROM (lista relacji)  
[ WHERE (warunek) ]  
[ GROUP BY (atrybuty grupowania) ]  
[ HAVING (warunek grupowy) ]  
[ ORDER BY (lista atrybutów) ]
```

Pierwsze dwa zdania są obligatoryjne, a pozostałe opcjonalne.

Załóżmy następujące zapytanie (wyszukiwanie kwalifikowane). Podaj numery legitymacji studentów, będących na czwartym roku i studiujących towaroznawstwo.

W języku SQL przyjmie ono następującą postać:

```
SELECT NR-LEGIT  
FROM STUDENT  
WHERE ROK = 4 AND KIER-STUD = 'towaroznawstwo'
```

W wyrażeniach logicznych następujących po WHERE korzysta się z:

- operatorów porównania: <, <=, =, >, >=;
- operatorów logicznych: AND, OR i NOT;
- nawiasów określających kolejność obliczenia wartości wyrażenia.

Do aktualizacji bazy danych w języku SQL służą operacje wprowadzania (INSERT), skreślenia (DELETE) i modyfikacji (UPDATE). Tak więc wprowadzenie kolejnej krotki do relacji STUDENT przeprowadza się przy użyciu następującej instrukcji:

```
INSERT INTO STUDENT  
VALUES (29201, 'Kinga', 'Jankowska', 1, 17, 'towaroznawstwo')
```

Istnieje możliwość wprowadzania instrukcją INSERT również krotek niepełnych. Zakłada ona przyjęcie wartości zerowych dla atrybutów niewymienionych w instrukcji INSERT:

```
INSERT INTO WYKŁAD (NAZWA, IL-GODZ) VALUES 'programowanie', 45)
```

Instrukcja UPDATE pozwala zmieniać wartość atrybutów w danej krotce, np. nazwisko wykładowcy finansów:

```
UPDATE WYKŁAD
SET WYKŁADOWCA = ‚Jabłkowski’
WHERE NAZWA = ‚finanse’
```

Instrukcja skreślenia usuwa krotki z relacji, np. krotkę o kluczu 15921 z relacji STUDENT:

```
DELETE FROM STUDENT
WHERE NR- LEG = 15921
```

Usunięcie wszystkich wierszy z tabeli można dokonać instrukcją:

```
DELETE FROM OCENA
```

Język SQL może być językiem interaktywnym dla relacyjnej bazy danych bądź osadzonym (*embedded*) w języku macierzystym (*host language*), np. w języku Java.

## 15.10. Systemy Big Data

**Big Data** jest terminem, który został spopularyzowany dzięki wszechobecnym dużym zbiorom danych, przede wszystkim znajdujących się w internecie. Najczęściej wykorzystywana definicja dotyczy ujęcia Big Data od strony źródeł danych.

**Źródła danych typu Big Data** można scharakteryzować za pomocą atrybutów określanych wspólnym kodem 3V: duży wolumen danych (*Volume*), duża szybkość przepływu i przetwarzania danych (*Velocity*) oraz duża zmienność danych (*Variety*) [Chen, Chiang, Storey, 2012].

**Volume** oznacza zbiory o bardzo dużych rozmiarach w porównaniu z tradycyjnymi bazami danych. **Velocity** oznacza, że występuje ciągły napływ danych uniemożliwiający ich zatrzymanie i przetwarzanie z wykorzystaniem tradycyjnych technik. **Variety** odnosi się do braku uporządkowania tych danych, tj. ich różnorodności i heterogeniczności. Wraz z rozwojem Big Data dodawane są kolejne atrybuty V, wśród których najpowszechniej występują: duża wartość danych (*Value*) – postrzeganie danych typu Big Data jako wartość przynosząca znaczące korzyści dla przedsiębiorstw oraz wiarygodność danych (*Veracity*) – oznaczające możliwość oceny poprawności danych typu Big Data [Morton, Runciman, Gordon, 2014].

Big Data odnosi się do wyników analiz, które uzyskuje się z dużych zbiorów danych, a które nie byłyby możliwe do osiągnięcia na podstawie małych zbiorów danych. Analizy te wykonuje się w celu uzyskania nieznanego obrazu danych lub utworzenia nowej informacji w sposób umożliwiający zmianę rynku, organizacji, powiązań między rządem a obywatelami i inne [Mayer-Schoenberger, Cukier, 2013]. Systemy Big Data mają zastosowanie wtedy, gdy dane są tak dużych rozmiarów, że nie mogą być przetwarzane z wykorzystaniem tradycyjnych technologii i narzędzi, takich jak relacyjne systemy zarządzania bazami danych [Maślankowski, 2015a].

Jeśli chcemy scharakteryzować systemy Big Data w kategorii jednostek danych, należy stwierdzić, że zbiory przetwarzane przez te narzędzia mają wielkość wielu petabajtów lub eksabajtów danych.