

Bazy danych



Andrzej Łachwa, UJ, 2015

andrzej.lachwa@uj.edu.pl

4/14

Własności SZBD:

- możliwość bezpiecznego przechowywania przez długi czas danych mierzonych w tera- i petabajtach,
- istnienie mechanizmów „masowego” wprowadzania danych,
- zapewnienie bezpieczeństwa danych,
- efektywny dostęp do danych (m. in. możliwość tworzenia zapytań o dane oraz aktualizowania danych za pomocą odpowiedniego języka operacji na danych),
- sterowanie współbieżnością,
- trójwarstwowa struktura (jądro, interfejs i narzędzia).
- możliwość tworzenia nowej bazy danych, tzn. określenia jej schematu i więzów integralności za pomocą jakiegoś języka definiowania danych.

Zalety podejścia bazodanowego:

- niezależność między danymi a programami,
- ukrywanie szczegółów dotyczących sposobu przechowywania danych w pamięci fizycznej,
- przechowywanie metadanych,
- dostarczanie wielu widoków (perspektyw) danych dla różnych grup użytkowników,
- zapewnienie integralności danych,
- współdzielenie danych i współbieżne transakcje wielu użytkowników,
- wysoka niezawodność i bezpieczeństwo
- ...

Tabele Codda

Teoretyczne podstawy relacyjnych baz danych zostały wprowadzone przez Edwarda Codda w 1970 roku. Zaproponował on przechowywanie danych w tabelach.

Tabela jest tutaj czymś podobnym do matematycznej relacji, a zarazem do zwykłej tabeli złożonej z nagłówka, kolumn i wierszy.

W nagłówku występuje nazwa tabeli oraz nazwy kolumn.

W polach wierszy występują wartości atrybutów.

W tabeli Codda nie jest ustalony ani porządek kolumn, ani porządek wierszy!

Tabela pusta zawiera tylko nagłówek!

Kryteria Codd

Informacje są reprezentowane (na poziomie logicznym) w tabelach.

Dane są (na poziomie logicznym) dostępne przez podanie nazwy tabeli, wartości klucza głównego i nazwy kolumny.

Wartości *null* są traktowane w jednolity sposób jako „brakujące informacje”. Nie mogą być traktowane jako puste łańcuchy znaków, puste miejsca, czy zera.

Metadane (dane dotyczące bazy danych) są umieszczone w bazie danych dokładnie tak, jak zwykłe dane.

Język obsługi danych ma możliwość definiowania danych, perspektyw i więzów integralności, przeprowadzenia autoryzacji, obsługi transakcji i manipulacji danymi.

Perspektywy reagują na zmiany swoich tabel bazowych. Odwrotnie, zmiana w perspektywie powoduje automatycznie zmianę w tabeli bazowej.

Istnieją pojedyncze operacje pozwalające na wyszukanie, wstawienie, uaktualnienie i usunięcie danych.

Operacje użytkownika są logicznie oddzielone od fizycznych danych i metod dostępu do danych.

Operacje użytkownika pozwalają na zmianę struktury bazy bez konieczności tworzenia od nowa bazy czy aplikacji ją obsługującej.

Więzy integralności są umieszczone i dostępne w metadanych, a nie w programie obsługi bazy danych.

Język manipulacji danymi powinien działać bez względu na to, jak i gdzie są rozmieszczone fizyczne dane oraz nie powinien wymagać żadnych zmian, gdy fizyczne dane są centralizowane lub rozpraszane.

Operacje na pojedynczych wierszach tabel przeprowadzane w systemie podlegają tym samym zasadom i więzom, co operacje na zbiorach danych.

Definicja

Relacyjna baza danych jest zbiorem danych tworzących tabele (w sensie Codda) i spełniających określone warunki integralności danych.

Tabela Codda reprezentuje relację. Kolumny i wiersze tabel odpowiadają atrybutom i krotkom relacji.

Kolumny i wiersze nie są uporządkowane!

Nagłówek tabeli odpowiada schematowi relacji.

Niech dany będzie zbiór atrybutów $\{A_1, A_2, \dots, A_n\}$, a z każdym atrybutem A_i związana będzie dziedina D_i wartości tego atrybutu. Dowolny podzbiór iloczynu kartezjańskiego $D_1 \times D_2 \times \dots \times D_n$ nazywamy relacją rozpiętą na schemacie $F(A_1, A_2, \dots, A_n)$.

Schematem relacyjnej bazy danych nazywamy zbiór pustych tabel oraz więzów integralności danych, które można do tych tabel wprowadzić.

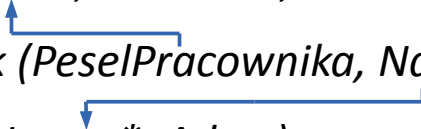
Tabela wypełniona danymi będzie kojarzona z relacją opartą na danym schemacie.

Konwencja 1

Osoba (Pesel, Nazwisko, DataUr)*

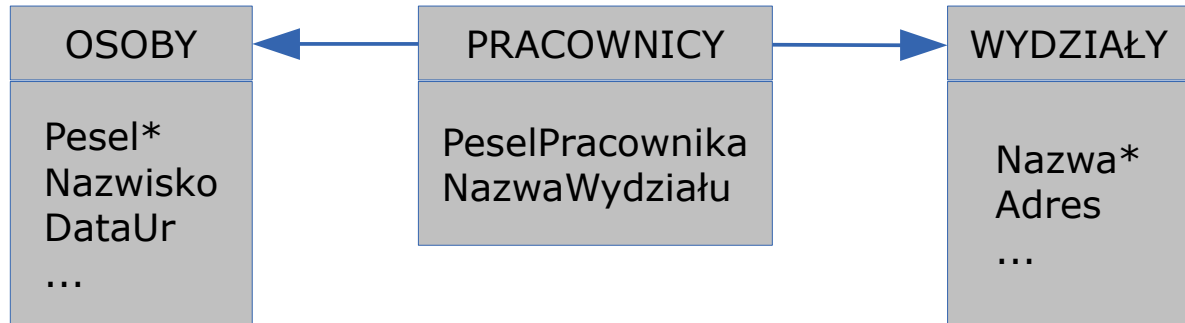
Pracownik (PeselPracownika, NazwaWydziału)

Wydział (Nazwa, Adres)*

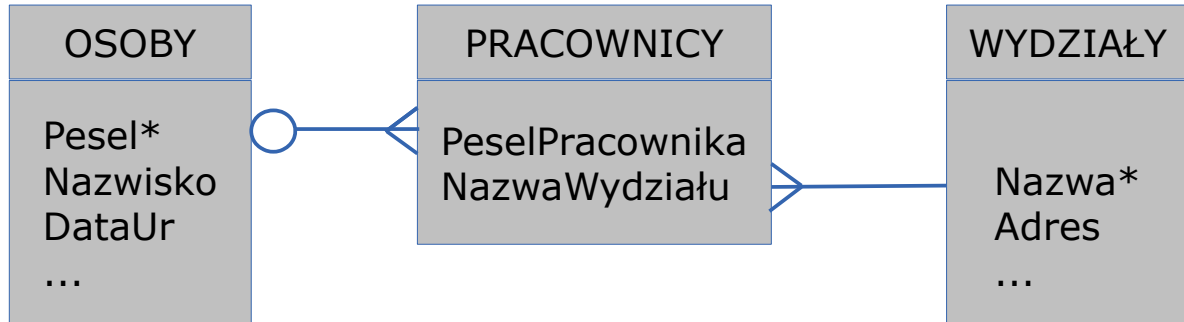


Symbolem * oznacza się atrybuty tworzące klucz główny tabeli (wartości klucza głównego służą do identyfikowania wierszy tabeli). Strzałki pokazują związek referencyjny między wartościami zwany kluczem obcym.

Konwencja 2



Konwencja 3



Więzy

WIĘZY DOMENOWE

WIĘZY KLUCZY PODSTAWOWYCH

WIĘZY INTEGRALNOŚCI REFERENCYJNEJ

(WIĘZY KLUCZA OBCEGO)

ZALEŻNOŚCI FUNKCYJNE

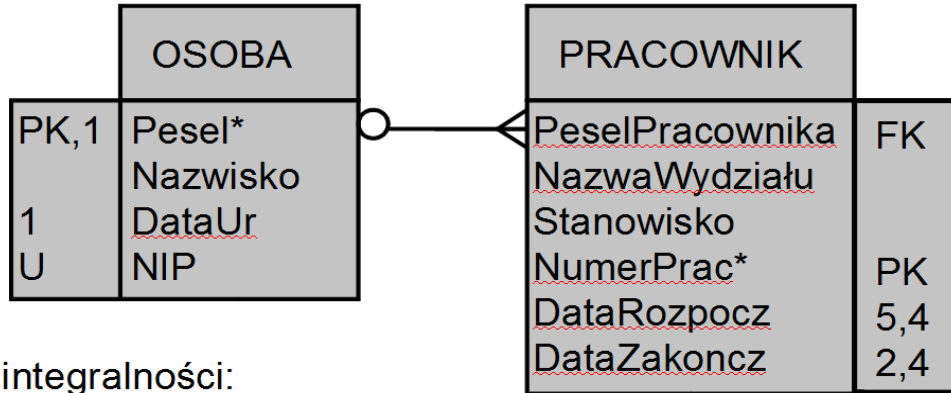
ZALEŻNOŚCI WIELOWARTOŚCIOWE

ZALEŻNOŚCI ZŁACZENIOWE

WARTOŚCI WYMAGANE

WIĘZY OGÓLNE

Konwencja 4



Więzy integralności:

PK — klucz główny (Primary Key)

FK — klucz obcy (Foreign Key)

U — wartość unikatowa (klucz alternatywny)

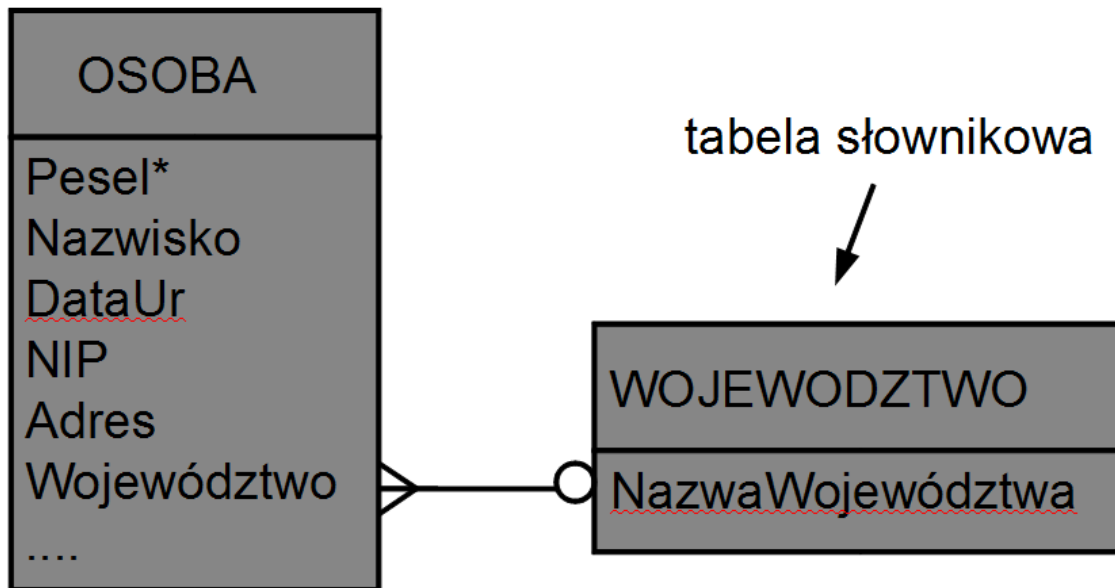
1 — Pesel ma być zgodny z datą urodzenia

2 — data zakończenia < od aktualnej lub NULL

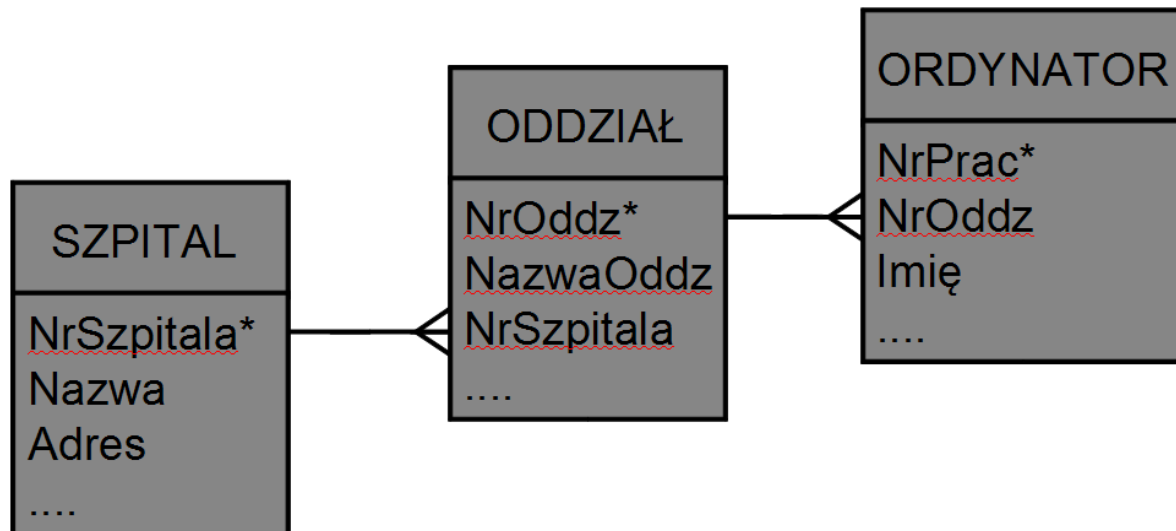
4 — data rozpoczęcia pracy < zakończenia

5 — data rozpoczęcia pracy < DataUr - "18 lat"

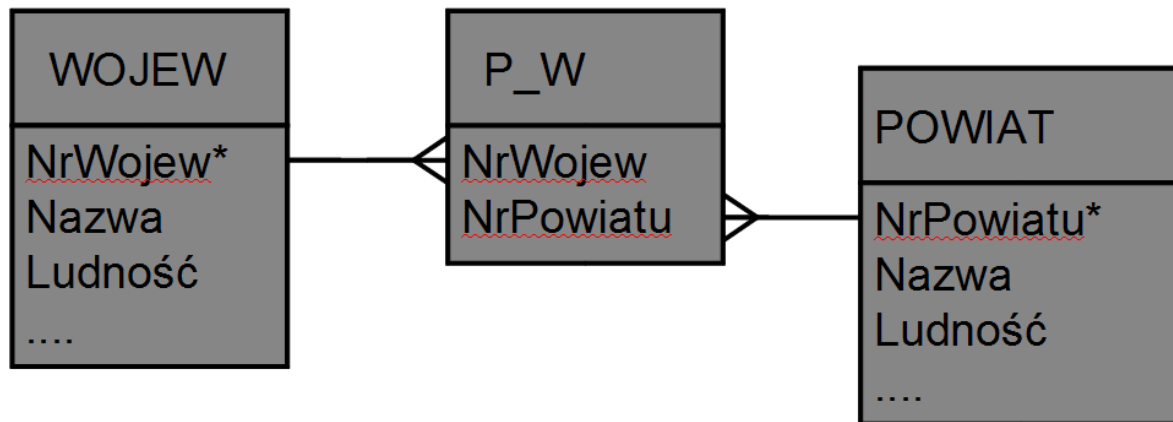
Wybrane rozwiązania



Modelowanie hierarchii

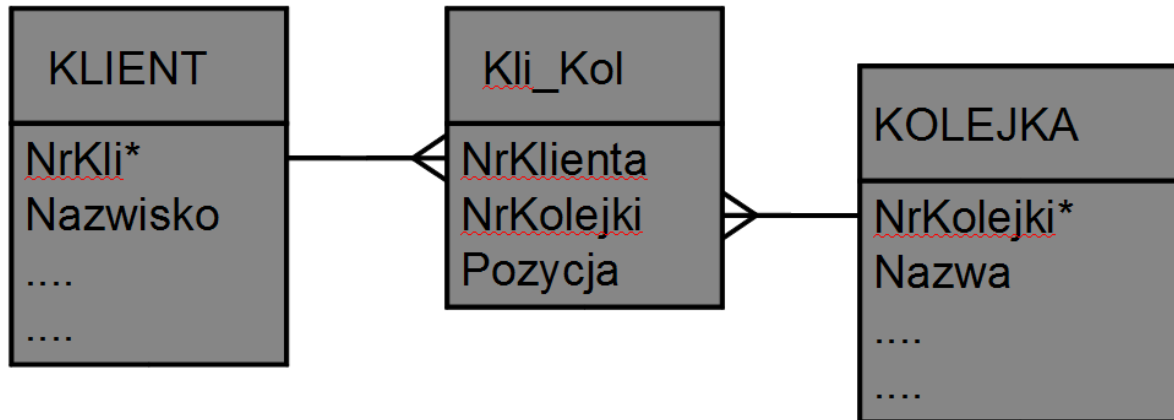


Modelowanie zbiorów (województwo jako zbiór powiatów)

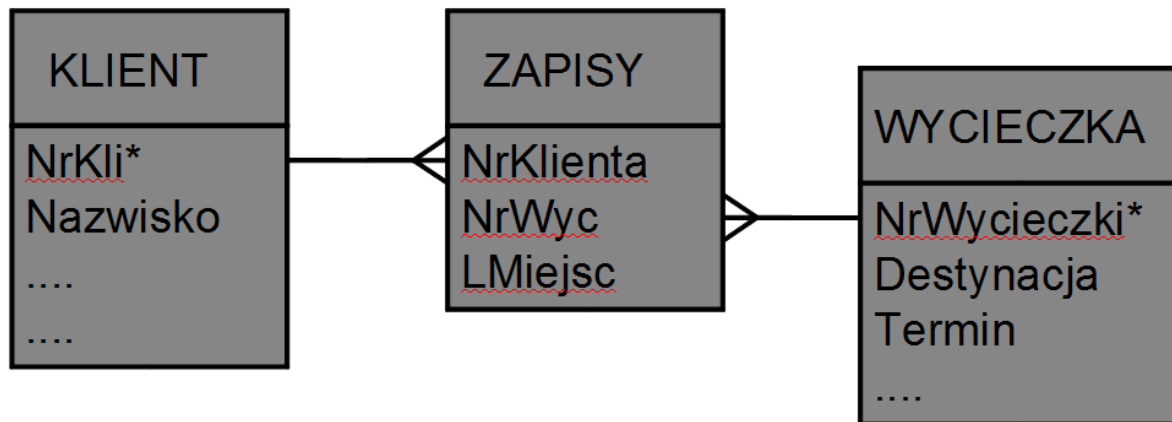


Modelowanie list

(kolejka jako lista klientów)



Modelowanie wielozbioru (zamówienia jako torby)



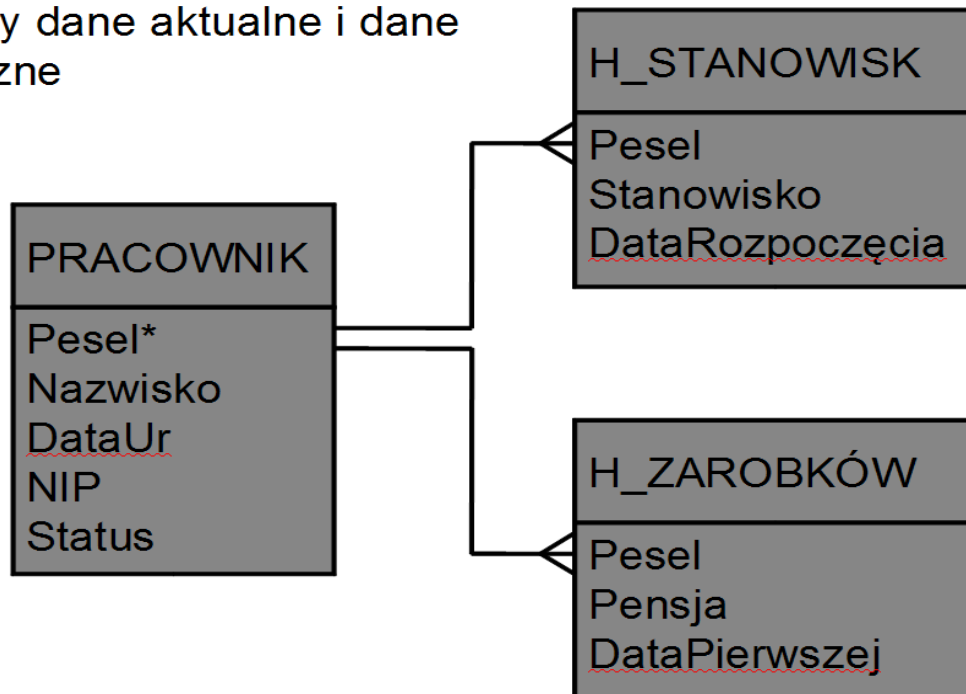
Dane aktualne czy historyczne?

OSOBA
Pesel*
Nazwisko
<u>DataUr</u>
NIP
Stanowisko
Pensja

To są aktualne dane
pracownika



Tu mamy dane aktualne i dane historyczne



Część DDL języka MySQL

zob.

<http://dev.mysql.com/doc/refman/5.6/en/sql-syntax-data-definition.html>

SQL 2 (1992), SQL 3 (1999-2002)

Liczbowe typy danych:

- INTEGER (INT), SMALLINT, BIGINT, [TINYINT]
- DECIMAL(p,s), NUMERIC(p,s)
- FLOAT(p), REAL, DOUBLE PRECISION
- [BIT, BYTE]

W różnych implementacjach różne zakresy. Należy również uważać na zaokrąglanie, obcinanie i konwersje.

Cztery operacje arytmetyczne są rozszerzane o różne funkcje matematyczne, np. kwadrat, pierwiastek kwadratowy, modulo, logarytm, funkcje trygonometryczne etc.

Przekształcanie liczb na liczebniki

Procedura byłaby zbyt skomplikowana. W języku angielskim najlepszym kompromisem między wielkością kodu procedury i wielkością tabeli jest przygotowanie tabeli tłumaczenia liczb mniejszych od 1000 na liczebniki, a potem kolejne obliczanie: liczby mniejszej od tysiąca, liczby tysięcy, milionów, miliardów ... i na koniec wykonanie konkatenacji odpowiednich liczebników i słów *thousand, million, billion* ... (ewentualnie z końcówką *s*).

W języku polskim sprawa znacznie się komplikuje. Prosta procedura dla angielskiego produkowałaby takie niegramatyczne wyrażenia, jak dla przykładu:

pięć miliony trzysta dwa tysiący sto jedenaście

Koncepcja elementu danych jako pojedynczej (atomowej) wartości nie zawsze ma sens!

Przykład 1: PUNKTY NA PŁASZCZYŹNIE

Położenie punktu na płaszczyźnie możemy określić parą liczb w odniesieniu do ustalonego układu współrzędnych. Modelowanie tego w jednej kolumnie (jako wartości atrybutu POŁOŻENIE) byłoby niewygodne. Lepsze jest modelowanie w dwóch kolumnach: ODCIĘTA, RZĘDNA.

Przykład 2: PUNKTY NA ZIEMI

Położenie punktu na Ziemi można określić przy pomocy współrzędnych geograficznych: wielkości kątowych liczonych względem równika (szerokość geograficzna, ang. *Latitude*, *Lat*) i południka zerowego (długość geograficzna, ang. *Longitude*, *Lng*).

Np. wejście do tego budynku to

$50^{\circ} 01' 46.0'' N 19^{\circ} 54' 21.9'' E$ (format DMS)

Lat: 50.029451 Lng: 19.906082 (format WGS84)

N: 50° 01.7670600', E: 19° 54.3649200' (format GPS)

Modelowanie formatu DMS można wykonać w 8 kolumnach, a formatu WGS84 w 2 kolumnach.

Przykład procedury

Aby obliczyć odległość między dwoma punktami na Ziemi można użyć wzoru ze strony

http://pl.wikibooks.org/wiki/Astronomiczne_podstawy_geografii/Odleg%C5%82o%C5%9Bci

Potrzebne funkcje (pierwiastek, kwadrat, cos) są dostępne w SQL w większości serwerów. Sprawdzić poprawność na GoogleMaps.

Przykład 3: ADRESY IP

Przechowywanie adresu IP można zrealizować na trzy sposoby: jako łańcuch, jako liczbę całkowitą lub jako cztery oktety. Ten ostatni sposób jest wygodny i czytelny.

Przykład 4: WALUTY

Waluty można przechowywać z rozbiciem na wartość i kod waluty. Trzeba ustalić liczbę miejsc po przecinku branych pod uwagę w trakcie obliczeń. Przy transakcjach międzynarodowych trzeba również uwzględniać tabelę kursów walut i dokonywać przeliczeń.

Zob. <http://www.nbp.pl/home.aspx?f=/statystyka/kursy.html>

Przykład 5: PESEL

Procedura, zob. <https://msw.gov.pl/pl/sprawy-obywatelskie/centralne-rejestry-pan/32,PESEL.html>

Sprawdzanie poprawności:

<http://www.kalkulatory.gofin.pl/Sprawdzanie-pesel-weryfikacja-pesel,12.html>