

RELACYJNA BAZA DANYCH W ARKUSZU KALKULACYJNYM

Wanda Jochemczyk Katarzyna Olędzka
Ośrodek Edukacji Informatycznej i Zastosowań Komputerów
02-026 Warszawa, ul. Raszyńska 8/10
wanda.jochemczyk@oeiizk.waw.pl
katarzyna.oliedzka@oeiizk.waw.pl

Abstract. A table with data in a spreadsheet can be a database. However, it is rarely known how to create relationships between tables in Excel. We will show how to create a relational database in an Excel spreadsheet using the Power Query add-in and perform basic database operations.

1. Wstęp

W XXI wieku można zaobserwować wzrost zapotrzebowania na umiejętność analizowania zbiorów danych. Rozwój technologii, a przede wszystkim internetu spowodował możliwość gromadzenia ogromnej liczby danych. Stąd potrzeba specjalistów z zakresu przetwarzania danych, którzy łączą umiejętności informatyczne, matematyczne z kreatywnym rozwiązywaniem problemów, ponadto efektywnie posługują się różnymi narzędziami informatycznymi. Równie cenna jest umiejętność samodzielnego zdobywania wiedzy na temat metod i sposobów radzenia sobie w sytuacjach problemowych.

Chociaż funkcjonalność poszczególnych baz danych się różni, można wyodrębnić pewne cechy charakterystyczne. Każdy model bazy danych powinien definiować następujące elementy: strukturę danych, operacje, jakie mogą być wykonywane na danych i ograniczenia nakładane na dane, tzw. więzy integralności. W relacyjnej bazie dane są zorganizowane w tabele, które są ze sobą wzajemnie powiązane. Do podstawowych operacji na danych zaliczamy:

- wprowadzanie danych (*create*) – można jest wpisywać ręcznie lub importować ze źródła elektronicznego,
- przeglądania danych (*read*) – zarówno dotyczące pojedynczego obiektu, jak i zestawu,

- uaktualnianie danych (*update*), które powinno być stosunkowo łatwe i bezpieczne,
- usuwanie (*remove*) z zachowaniem zasady archiwizowania.

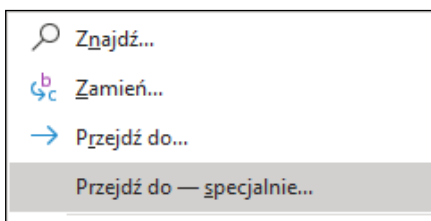
Postaramy się pokazać, że wszystkie te operacje create-read-update-remove można wykonywać w arkuszu kalkulacyjnym, który choć nie jest dedykowanym narzędziem bazodanowym, ma wiele przydatnych funkcji. Dlatego warto kształtować u uczniów umiejętność analizy danych za pomocą arkusza kalkulacyjnego.

2. Przygotowanie i import danych

Nie zawsze dane, które będziemy przetwarzać są uporządkowane. Gdy są zbierane ręcznie lub pozyskiwane z pomiarów, trzeba je uporządkować i ujednolicić. Często należy:

- usunąć zbędne spacje:
funkcja =USUŃ.ZBĘDNE.ODSTĘPY(),
- ujednolicić wielkości liter:
funkcje =Litery.małe() lub =Litery.wielkie(),
- usunąć formatowanie,
- dokonać podziału na kolumny:
za pomocą opcji **Dane** → **Tekst jako kolumny**,
- zamienić liczby zapisane jako tekst na format liczbowy,
- usunąć duplikaty – wybieramy **Dane** → **Usuń duplikaty**.

Przydatna może też umiejętność automatycznego wstawiania wartości w pustych komórkach. Można to zrobić na wiele sposobów, na przykład wybrać opcję **Znajdź i zaznacz | Przejdź do – specjalnie**, **Puste** i wpisać wprowadzaną nową zawartość, a następnie wybrać **Ctrl+Enter**.



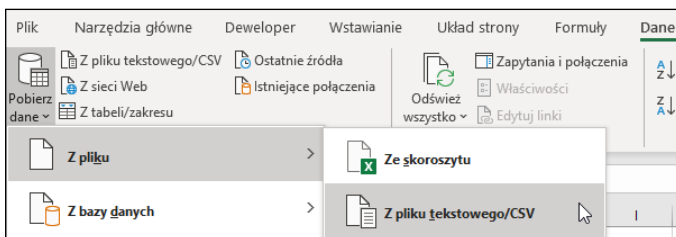
Rysunek 1 Wybór komórek do automatycznego wstawienia wartości

Dane importowane do arkusza kalkulacyjnego mogą mieć różny format. Wśród nich wyróżniamy dane w postaci tabelarycznej – do skopiowania przez schowek,

w pliku tekstowym lub csv, w formacie JSON oraz w innych np. *.ods. Warto również nadmienić o możliwości przygotowania formularzy online do zbierania danych. Są bardzo przydatne w praktyce.

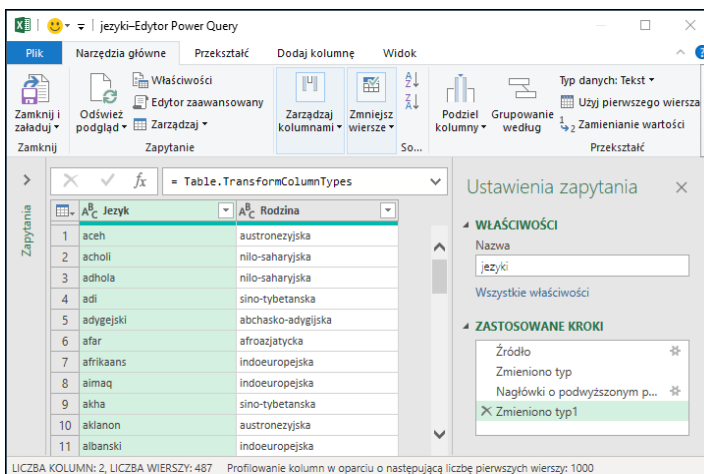
Dodatek Power Query został przygotowywany dla arkusza kalkulacyjnego Excel począwszy od wersji 2010, w nowych wersjach jest wbudowany. Aby zaimportować dane z pliku tekstowego do Power Query należy:

- na karcie **Dane**, w sekcji **Pobieranie i przekształcanie danych** wybrać **Pobierz dane** → **Z pliku** → **Z pliku tekstowego/CSV**



Rysunek 2 Import z pliku tekstowego

- wskazać odpowiedni plik na dysku, wybrać **Importuj**
- sprawdzić, czy został wybrany odpowiedni ogranicznik, czy kodowanie polskich znaków jest prawidłowe, następnie wybrać przycisk **Przekształć dane**
- wybrać przycisk **Użyj pierwszego wiersza jako nagłówków**



Rysunek 3 Dane po zaimportowaniu

- po zamknięciu dodatku Power Query i zapisaniu zmian, znajdziemy się w arkuszu (w nowym skoroszybie)
- aby wrócić do Power Query trzeba na karcie **Dane**, w sekcji **Pobieranie i przekształcanie danych** wybrać **Pobierz dane** → **Uruchom edytor dodatku Power Query...**

Import powtarzamy dla wszystkich tabel. Po przygotowaniu danych, jesteśmy gotowi do implementowania zapytań. Warto zauważyć, że dane nie zostały wstawione do arkusza, ale tylko z nim połączone. Zmiana danych źródłowych i odświeżenie widoku, sprawi, że arkusz kalkulacyjny będzie przetwarzać nowe dane.

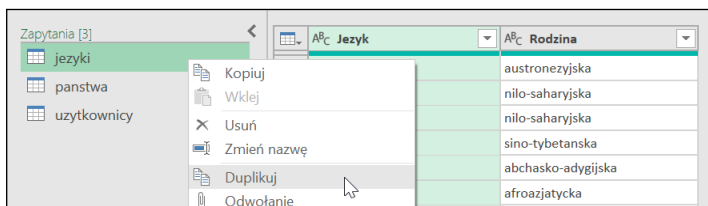
3. Przeglądanie danych w różnej formie

Arkusz kalkulacyjny zawiera kilka funkcjonalności, które ułatwiają przeglądanie danych. Jedną z nich jest blokada fragmentu arkusza. Stosujemy ją, aby podczas przewijania zachować w polu widzenia te kolumny lub wiersze, które chcemy aktualnie wyświetlać na ekranie. Zupełnie inną funkcjonalnością jest korzystanie z formularzy. Za ich pomocą można wyświetlać oraz wprowadzać dane w arkuszu bez konieczności przewijania w poziomie. Jest to bardzo wygodne, ponieważ cały rekord jest dobrze widoczny. Można też ustawić filtrowanie z wyświetlaniem wierszy spełniającym określone kryteria oraz sortowaniem według zadanego porządku.

Bardziej zaawansowanym narzędziem jest definiowanie zapytań w dodatku Power Query. Pokażemy to na przykładzie zdania maturalnego roku 2020[3].

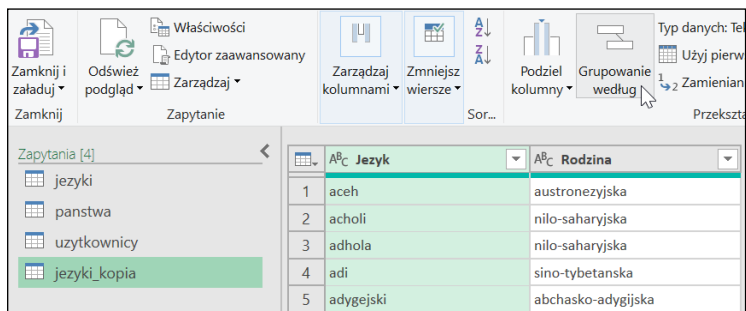
Jedno z poleceń w zadaniu maturalnym polegało na utworzeniu zestawienia, które dla każdej rodziny językowej podaje, ile języków do niej należy. Otrzymane dane powinny być posortowane nierosnąco według liczby języków.

Potrzebne dane znajdują się w tabeli **języki**. Musimy je pogrupować i posortować. W zakładce zapytania klikamy prawy przyciskiem myszy w tabelę **języki** i wybieramy **Duplikuj**. Od teraz będziemy modyfikować kopię tego zapytania, by w każdej chwili mieć możliwość skorzystania z danych ze źródłowej tabeli.



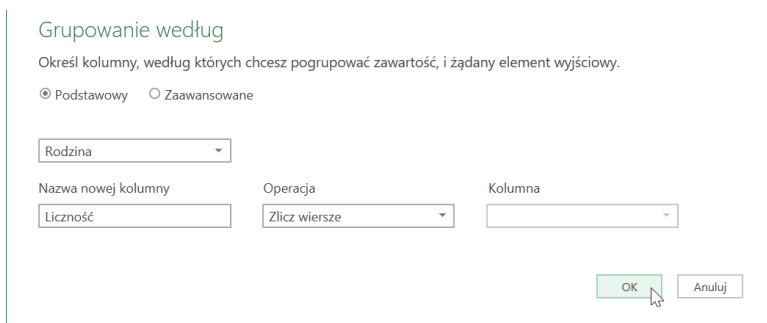
Rysunek 4 Duplikowanie zapytania

W kolejnym kroku grupujemy dane. W tym celu wybieramy opcję **Grupuj dane**.



Rysunek 5 Ustawianie grupowania

Musimy wskazać, według której kolumny chcemy grupować dane i rodzaj operacji, jaka ma być wykonywana. W tym przypadku zliczamy wiersze w kolumnie **Rodzina** o takich samych wartościach.



Rysunek 6 Kryteria grupowania

Następnie strzałką przy polu **Liczność** wybieramy odpowiedni porządek sortowania. Zapytanie jest gotowe.

	AB:C Rodzina	1.2 Liczność
1	nigero-kongijska	137
2	austronezyjska	65
3	indoeuropejska	63
4	sino-tybetanska	43
5	nilo-saharyjska	30
6	afroazjatycka	28
7	dajaska	23
8	austroazjatycka	20
9	drawidyjska	15
10	turecka	15

Rysunek 7 Dane po zgrupowaniu

Na koniec wybieramy opcję **Zamknij i załaduj**, by dane zostały wyświetlone w arkuszu kalkulacyjnym.

	A	B
1	Rodzina	Liczność
2	nigero-kongijska	137
3	austronezyjska	65
4	indoeuropejska	63
5	sino-tybetyjska	43
6	nilo-saharyjska	30
7	afroazjatycka	28
8	dajjska	23
9	austroazjatycka	20
10	drawidyjska	15
11	turecka	15

Rysunek 8 Dane po załadowaniu do skoroszytu

Kolejnym zestawieniem w zadaniu maturalnym było podanie wszystkich języków, którymi posługują się użytkownicy na co najmniej czterech kontynentach.

Będziemy korzystać z tabel **uzytkownicy** i **panstwa**, dlatego wykonujemy ich kopie. W tabeli **uzytkownicy** usuwamy zbędne kolumny, zostawiamy dwie pierwsze **Panstwo** i **Jezyk**. Kolejnym krokiem będzie połączenie ich ze sobą. Wybieramy **Połącz** → **Scal zapytania**. W oknie scalania wybieramy nazwy łączonych tabel i zaznaczamy pole **Panstwo**, które je łączy.

Scalanie ×

Wybierz tabelę i pasujące kolumny, aby utworzyć scaloną tabelę.
uzytkownicy_kopia 📄

Panstwo	Jezyk
Chiny	mandaryński
Indie	hindi
USA	angielski
Brazylia	portugalski
Bangladesz	bengalski

panstwa_kopia 📄

Panstwo	Kontynent	Populacja
Afganistan	Azja	32,5
Algieria	Afryka	39,7
Argentyna	Ameryka Południowa	43,4
Bangladesz	Azja	161
Birma	Azja	51,4


Rodzaj sprzężenia

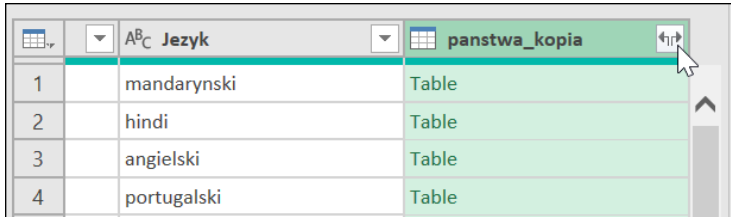
 Użyj dopasowywania rozmytego w celu wykonania scalenia

▶ Opcje dopasowywania rozmytego

Zaznaczenie jest zgodne z 656 z 656 wierszy z pierwszej tabeli.

Rysunek 9 Tworzenie powiązania między tabelami

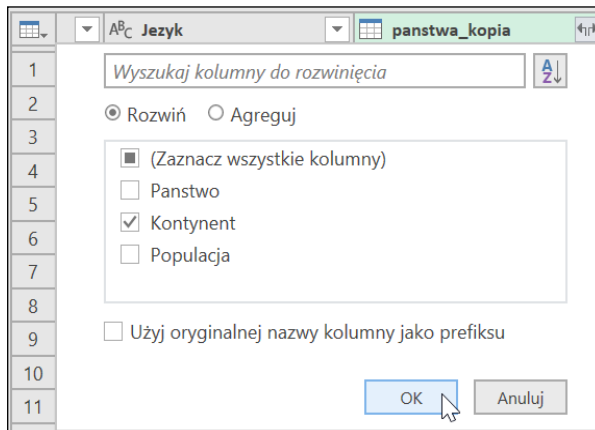
Kolejnym krokiem będzie rozwinięcie tabeli, wybieramy kolumnę z obiektem tabelarycznym i wskazujemy przycisk .



	ABC Jezyk	panstwa_kopia
1	mandarynski	Table
2	hindi	Table
3	angielski	Table
4	portugalski	Table

Rysunek 10 Dołączanie pól

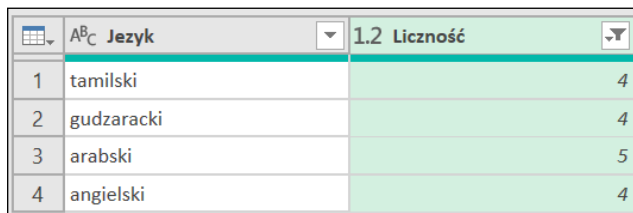
Wskazujemy pola, które powinny być wyświetlane, tutaj będzie to **Kontynent**.



	ABC Jezyk	panstwa_kopia
1	Wyszukaj kolumny do rozwinięcia	
2	<input checked="" type="radio"/> Rozwiń <input type="radio"/> Agreguj	
3	<input checked="" type="checkbox"/> (Zaznacz wszystkie kolumny)	
4	<input type="checkbox"/> Panstwo	
5	<input checked="" type="checkbox"/> Kontynent	
6	<input type="checkbox"/> Populacja	
7	<input type="checkbox"/> Użyj oryginalnej nazwy kolumny jako prefiksu	
8		
9		
10	OK	
11	Anuluj	

Rysunek 11 Wybór dołączanych pól

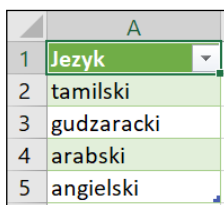
Usuujemy kolumnę **Panstwo**, a następnie duplikaty wierszy. W tym celu zaznaczamy obie kolumny i wybieramy **Zmniejsz wiersze** → **Usuń wiersze** → **Usuń duplikaty**. Grupujemy wiersze po języku zliczając kontynenty. Odfiltrowujemy te, które występują co najmniej 4 razy.



	ABC Jezyk	1.2 Liczność
1	tamilski	4
2	gudzaracki	4
3	arabski	5
4	angielski	4

Rysunek 12 Wybór dołączanych pól

Usuujemy pole **Liczność** i ładujemy tabelę do arkusza. Lista tych czterech języków jest odpowiedzią na postawione pytanie.



	A
1	Jezyk
2	tamilski
3	gudzaracki
4	arabski
5	angielski

Rysunek 13 Odpowiedź na zadanie

Podsumowując, w arkuszu kalkulacyjnym Excel z wykorzystaniem dodatku Power Query można tworzyć zapytania, które zawierają dane z kilku tabel powiązanych ze sobą. Możliwe jest sortowanie, filtrowanie i grupowanie danych, co jest podstawą do tworzenia raportów.

Do raportów można skorzystać z tabel przestawnych, które również umożliwiają zmianę sposobu prezentowania danych. Możemy za ich pomocą analizować dużą liczbę danych, wykorzystując możliwości związane z przestawianiem, grupowaniem, a także podsumowywaniem danych, łączeniem poszczególnych grup danych zgodnie z potrzebami oraz opracowywaniem wykresów prezentujących wyniki.

4. Modyfikowanie i usuwanie danych

Warto jeszcze raz podkreślić, że standardowo importowanie danych w arkuszu kalkulacyjnym w rzeczywistości tworzy powiązanie z plikami źródłowymi. Usuwanie i modyfikowanie danych w plikach źródłowych, powoduje zmianę w wyświetlanych zapytaniach, a w konsekwencji w raportach. Można co prawda przechowywać dane tylko w arkuszu, ale trzeba to wyraźnie zaznaczyć przy imporcie.

Powiązanie danych z plikami źródłowymi ma ogromną zaletę przy tworzeniu systemów informatycznych. Raz zaprojektowane zapytanie i raport będą służyły opracowaniu różnych danych. Przygotowujemy w ten sposób narzędzie, które może być pomocne dla stale aktualizowanych danych. Wadą tego rozwiązania jest konieczność redefiniowania odnośników do plików przy zmianie ich lokalizacji.

5. Zakończenie

Argumentem przemawiającym za wyborem arkusza kalkulacyjnego do rozwiązywania zadań bazodanowych jest jego powszechność oraz bogactwo narzędzi do obróbki danych. Można w nim wykonać podstawowe operacje bazodanowe łącząc dane z różnych źródeł. Możliwość tworzenia powiązań między tabelami jest istotną

cechą przy wyborze narzędzia do implementacji bazy danych. Dodatkowo przechowywanie danych w surowych plikach powiększa obszar zastosowań.

Trzeba jednak zdawać sobie sprawę, że arkusz kalkulacyjny nie został zaprojektowany głównie z myślą o bazach danych, istnieje w tym celu dedykowane oprogramowanie. Bardziej złożone bazy wymagają innych narzędzi. Nie jest też on oparty o język SQL. Arkusza kalkulacyjnego można używać w zastosowaniach codziennych, jak i do rozwiązywania zdań maturalnych.

Literatura

1. Jochemczyk W., *Relacyjna baza danych w arkuszu kalkulacyjnym, W cyfrowej szkole*, OEliZK, Warszawa nr 2/2020.
2. Olędzka K., *Maturalne zadanie bazodanowe a arkusz kalkulacyjny, W cyfrowej szkole*, OEliZK, Warszawa nr 3/2020.
3. Centralna Komisja Egzaminacyjna,
<https://cke.gov.pl/egzamin-maturalny/egzamin-w-nowej-formule/arkusze/2020-2/>,
ostatni dostęp 04.09.2020 roku.