

ESTIMATING VEHICLES POLLUTANT EMISSION CHARACTERISTICS USING DATABASES

Krzysztof Brzozowski

Leonard Grinke

University of Bielsko-Biala
Willowa 2 str., 43-309 Bielsko-Biala, Poland
tel.: +48 33 8279221, fax: +48 33 8279000
e-mail: kbrzozowski@ath.bielsko.pl
lgrinke@ath.bielsko.pl

Abstract

The schema of databases system which is dedicated to store pollutant emissions data from motor vehicles is presented in the paper. This databases system could be used to: building emission models, estimating car emissions, making statistical analysis etc. The system is divided to three levels accordingly the data characteristics: raw and reconciled data level, integrated (aggregated) data level, models data level. Each level shares dictionary database. Each level has own procedures, methods, and tools to evaluate data stored in it. The paper includes representative examples to illustrate above mentioned features. There are represented general schema of a system of databases to taking into account the lexical base, measuring database, integrated database, supplementary bases, and database of models.

Examples of opportunity application technology of databases for propriety ecological vehicles appointing are presented. Implementation of data placed in the paper will make possible: integration of data deriving from different sources, the dispersion data with contemporary access for them from any location, acquisition of independence of applications using data from manners and methods of their storage, the access for embedded tools in the System of The Data Base Management for the advanced data analysis, format uniformities and data representations.

Keywords: relational databases, pollutants emission, SQL, Data Analysis

ZASTOSOWANIE BAZ DANYCH DO WYZNACZANIA WŁASNOŚCI EKOLOGICZNYCH POJAZDÓW W ZAKRESIE EMISJI ZWIĄZKÓW SZKODLIWYCH SPALIN

Streszczenie

W pracy przedstawiono system baz danych przeznaczonych do pamiętania danych związanych z emisją związków szkodliwych pojazdów. Dane zgromadzone w bazach danych mogą zostać wykorzystane do: tworzenia modeli emisji, modelowania stężeń, porównywania wyników uzyskanych z pomiarów eksperymentalnych i modeli emisji. System, zgodnie z właściwościami danych został podzielony na trzy poziomy: poziom danych źródłowych, poziom danych zagregowanych i poziom danych modeli. Każdy poziom korzysta z bazy słownikowej zawierającej opisy danych, norm, jednostek itp. Baza słownikowa stanowi repozytorium całego systemu. Przedstawiono ogólny schemat systemu baz danych uwzględniając bazę słownikową, bazę danych pomiarowych, bazę danych zintegrowanych, bazy dodatkowe oraz bazę danych modeli.

W artykule zilustrowano przykładami możliwości zastosowania technologii baz danych do wyznaczania właściwości ekologicznych pojazdów. Wykorzystanie danych zamieszczonych w artykule umożliwi: integrację danych pochodzących z różnych źródeł, rozproszenie danych z jednoczesnym dostępem do nich z dowolnego miejsca, uzyskanie niezależności aplikacji wykorzystujących dane od sposobów i metod ich przechowywania, dostęp do wbudowanych w System Zarządzania Bazą Danych narzędzi do zaawansowanej analizy danych jednolitości formatów i reprezentacji danych.

Słowa kluczowe: relacyjna baza danych, emisja związków szkodliwych, SQL, analiza danych

1. Wstęp

Celem tworzenia systemu baz danych jest konieczność przechowywania danych, opisów metod pomiarowych oraz modeli i dokumentów związanych z własnościami ekologicznymi pojazdów w zakresie emisji związków szkodliwych. Otrzymane z baz danych wyniki i raporty mogą służyć do:

- modelowania i szacowania emisji w oparciu o zgromadzone dane,
- odszukiwania dokumentów związanych z badaniami lub modelowaniem emisji,
- weryfikacji tworzonych modeli emisji,
- rozbudowy modeli emisji.

Brak pełnych i powszechnie dostępnych danych pomiarowych wymusza konieczność stosowania i gromadzenia wszystkich dostępnych danych dotyczących emisji z pojazdów samochodowych. Ogólną koncepcję takiego systemu przedstawiono w pracy [1]. Stwierdzenie „wszystkich dostępnych” wymusza ciągły proces uaktualniania i modyfikacji bazy danych oraz wbudowanie mechanizmów pamiętania danych w różnych formatach i postaciach. Nie jest możliwe utworzenie systemu baz danych realizującego od początku pełną funkcjonalność. Uwaga ta dotyczy przede wszystkim zintegrowanych baz danych, których rozbudowa opiera się głównie o metody heurystyczne. Na tym poziomie budowa konkretnego modelu może wymusić definicję innej metody integracji danych. Ważne jest również, aby procesy tworzenia nowych modeli i ich weryfikacji korzystały ze źródeł danych o dużej wiarygodności i ustalonym formacie [2]. Proces pozyskiwania danych z różnych źródeł, ich czyszczenie i weryfikacja a następnie wprowadzenie do systemu jest podstawowym czynnikiem decydującym o ich jakości.

Pamiętanie danych pomiarowych w bazie danych ma następujące zalety [3]:

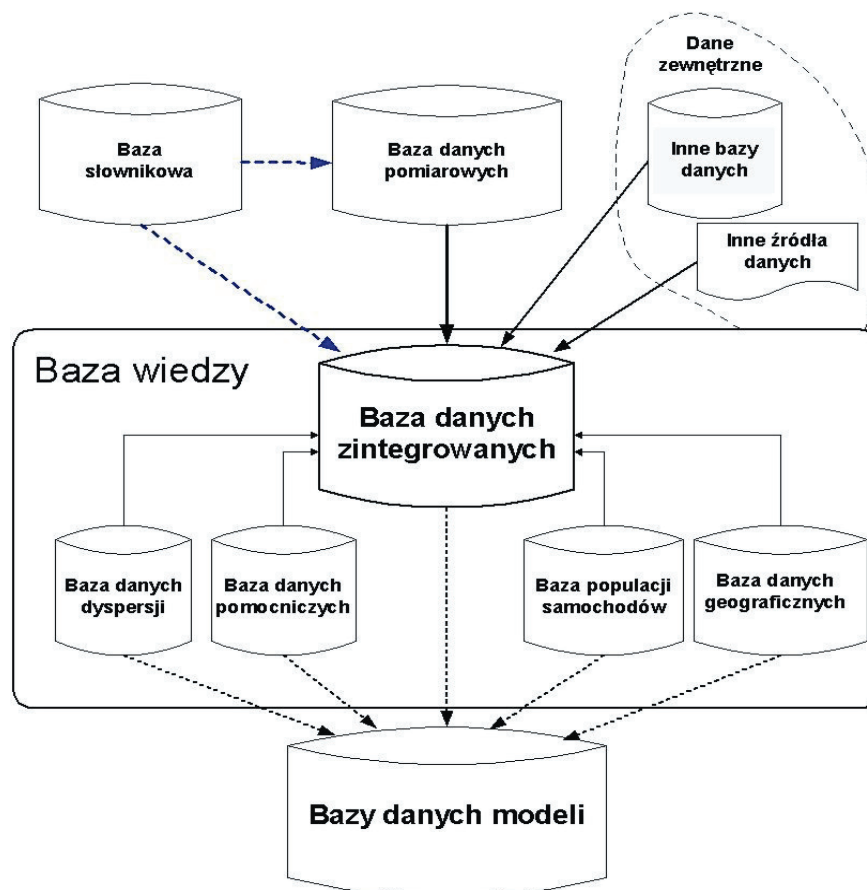
- wymusza stosowanie jednolitych postaci i formatów zapisu danych,
- dane są zgromadzone w jednym miejscu i są łatwo dostępne,
- można wykorzystać wbudowane w System Zarządzania Bazą Danych mechanizmy manipulacji danymi.

Na Rys. 1. przedstawiono ogólny schemat systemu baz danych. Podział na poszczególne bazy danych wynika z:

- konieczności uzyskania autonomiczności i niezależności poszczególnych elementów,
- heterogeniczności źródeł danych,
- potrzeby pamiętania oprócz danych pomiarowych, dokumentów w różnych formatach i postaciach.

Podstawowymi bazami danych są:

- **Baza słownikowa** - przechowuje definicje, normy, współczynniki, wzory, dokumenty. Dane te są niezbędne w obsłudze danych pozostałych baz.
- **Baza danych pomiarowych** - przechowuje dane pomiarowe pochodzące z różnych źródeł, w dalszej części pracy zostanie przedstawiona jej podstawowa struktura oraz przykłady ilustrujące jej wykorzystanie.
- **Baza danych zintegrowanych** - przechowuje dane scalone, najczęściej w postaci struktur wielowymiarowych, stanowi podstawę zaawansowanej analizy danych.
- **Bazy dodatkowe np. danych geograficznych, populacji samochodów itp.**, przechowują semantyczne jednorodnie dane niezbędne do tworzenia modeli wykorzystywanych w pracach związanych z bilansowaniem emisji.
- **Baza(y) danych modeli** - przechowuje dane uzyskane z wykorzystaniem modeli emisji, służy głównie jako podstawa do pamiętania opisów modeli i ich weryfikacji.



Rys. 1. Ogólny schemat systemu baz danych

Fig. 1. Databases system - general scheme

2. Dane pomiarowe i proste analizy

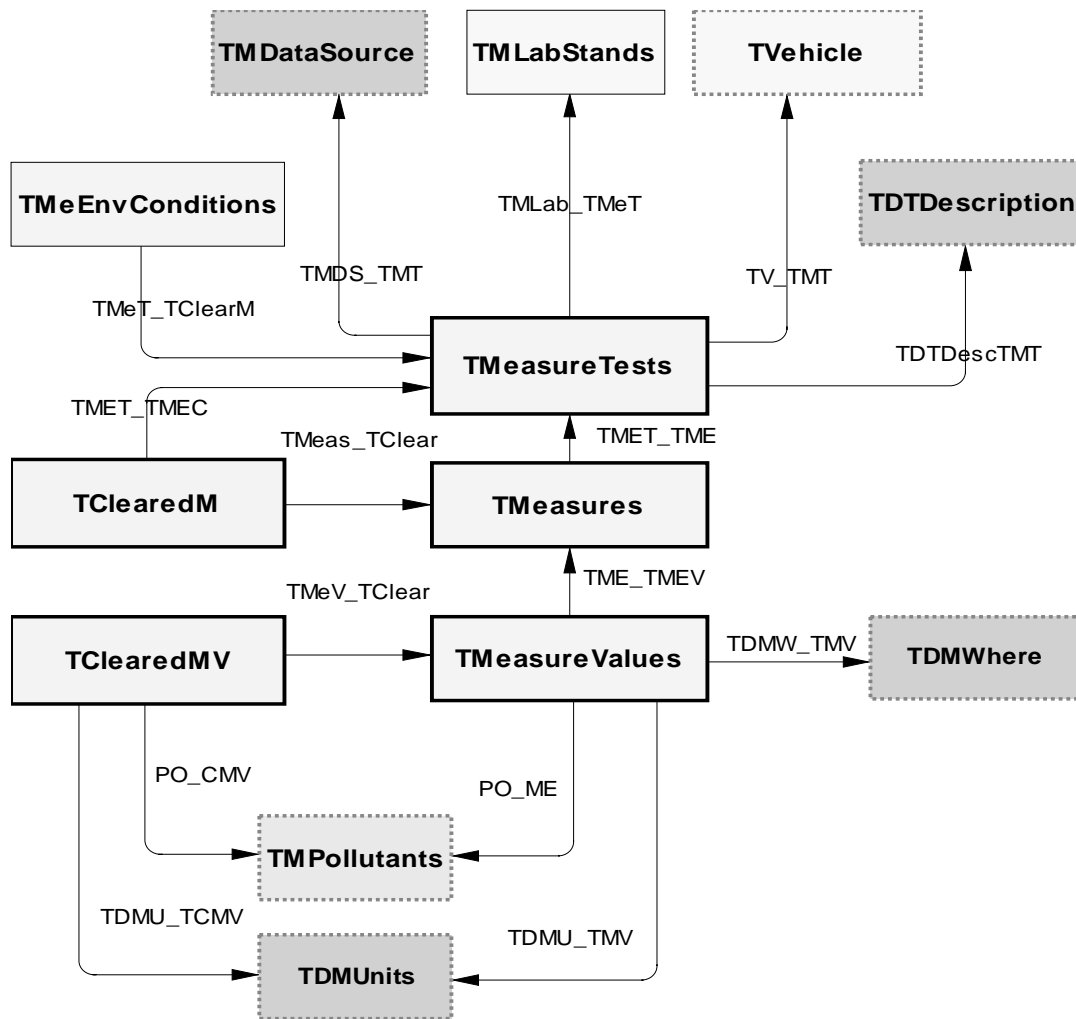
Na Rys. 2 przedstawiono część schematu bazy pomiarowej. Podstawowymi elementami (tabelami) są: *TMeasureTests* (przechowuje opis eksperymentu pomiarowego), *TMeasures* (przechowuje dyskretne wartości czasu i przełożenia skrzyni biegów), *TMeasureValues* (przechowuje wyniki pomiarów). Tabele *TClearedM* i *TClearedMV* przechowują dane pomiarowe po normalizacji. Dane pomiarowe mogą mieć wiele postaci znormalizowanych.

Tabele *TDMWhere*, *TMPollutants* i *TDMUnits* pamiętają kolejno: miejsce poboru próbki (np. przed reaktorem katalitycznym), typ czynnika (np. CO₂ wraz z opisem), jednostki pomiarowe (np. mg, ppm). Pozostałe tabele służą do pamiętania warunków w jakich pomiary zostały dokonane.

W przypadku korzystania z wyników pomiarów otrzymanych z wykorzystaniem różnej aparatury pomiarowej, opartych o różne testy jezdne lub metodologie, dane pomiarowe należy sprowadzić do wspólnej postaci. Zadaniem procesu normalizacji danych jest oczyszczenie i uzupełnienie danych pomiarowych i przekształcenie do zdefiniowanej postaci.

Otrzymane pomiary powinny podlegać normalizacji celem:

- usunięcia pomiarów niewiarygodnych,
- usunięcia wartości wynikłych z fluktuacji aparatury pomiarowej,
- sprowadzenia do zdefiniowanej postaci,
- określenia błędów pomiarowych.



Rys. 2. Schemat logiczny bazy danych pomiarowych

Fig. 2. Measurement data database logical scheme

Na Rys. 3 został pokazany przykład wyrażenia SQL (język baz danych) obliczającego podstawowe statystyki dla natężenia emisji CO₂ (μg/s) w przedziale prędkości od 40 do 45 km/h (prędkość chwilowa), dla próbek pobranych za reaktorem katalitycznym.

```

SELECT  ROUND(AVG(MEValue),3) AS Srednia_CO2, -- percents
          ROUND(MAX(MEValue),3) AS Maksymalna_CO2,
          ROUND(MIN(MEValue),3) AS Minimalna_CO2,
          ROUND(STDEV(MEValue),3) AS OdchylStd_CO2,
          ROUND(VAR(MEValue),3) AS Wariancja_CO2
FROM    TMeasureValues AS MV JOIN TMeasures AS M
ON      (MV.METID=M.METID AND MV.MEID=M.MEID)
WHERE   DPCode = 'CO2' AND MV.DMWID = 2 AND DMUID = 2
          AND MESpeedInt BETWEEN 40 AND 45
  
```

Rys. 3. Przykład kodu SQL obliczającego podstawowe statystyki dla emisji CO₂

Fig. 3. Example of SQL query expression for CO₂ emission calculation

Tab. 1. Wynik zapytania z przykładu na Rys. 3

Tab. 1. Query result from example on Fig. 3

	Srednia_CO2	Maksymalna_CO2	Minimalna_CO2	OdchylStd_CO2	Wariancja_CO2
1	9,72	16,255	0,318	5,625	31,644

Wynik wykonania tego wyrażenia został przedstawiony w tabeli Tab. 1. Samo wyrażenie jest bardzo zwarte, proste oraz podane w sposób deklaracyjny. Wykonanie tego typu obliczeń z wykorzystaniem algorytmicznych języków programowania (np. Pascal, C, C++) wymaga bardziej rozbudowanego kodu a korzystanie z wielu plików przechowujących dane znacznie wydłuża czas obliczeń. W przypadku zapytań kierowanych do systemu baz danych optymalizacją ich wykonania zajmuje się sam System Zarządzania Bazą Danych. Dane do których kierowane są zapytania muszą istnieć, ich wprowadzenie z różnych, często heterogenicznych źródeł jest poważnym problemem.

Bardziej zaawansowane zapytanie zostało przedstawione na Rys. 4 (ze względu na wielkość został pokazany tylko początkowy fragment kodu), które wylicza średnią emisję NO_x (ppm) w zależności od rodzaju przeprowadzonego testu jezdnego (FTP, NEDC) w różnych przedziałach prędkości pojazdu. Próbkę pomiarową została pobrana za reaktorem katalitycznym. Wynik obliczeń przedstawiono w tabeli Tab. 2.

Tab. 2. Wynik zapytania z przykładu na Rys. 4

Tab. 2. Query result from example on Fig. 4

	TestType	SpeedRange	NOx_AVG_ppm
1	FTP	0	72,684
2	FTP	1-20	119,778
3	FTP	21-40	127,083
4	FTP	41-60	105,551
5	FTP	61-80	164,937
6	FTP	80 >	131,425
7	NEDC	0	50,312
8	NEDC	1-20	123,455
9	NEDC	21-40	87,108
10	NEDC	41-60	76,856
11	NEDC	61-80	96,721
12	NEDC	80 >	98,509

```

SELECT 'TestType' = CASE TestID
  WHEN 1 THEN 'FTP'
  WHEN 2 THEN 'NEDC' END, '0' AS SpeedRange,
  ROUND(AVG(MEValue),3) AS NOx_AVG_ppm
FROM TMeasureValues AS MV JOIN TMeasures AS M
ON (MV.METID=M.METID AND MV.MEID=M.MEID)
JOIN TMeasureTests AS MT ON M.METID=MT.METID

```

```

WHERE  DPCode = 'NOx' AND MESpeedInt =0 AND MV.DMWID = 2
        AND DMUID = 2
GROUP BY TestID
UNION
SELECT 'TestType' = CASE TestID
        WHEN 1 THEN 'FTP'
        WHEN 2 THEN 'NEDC'END, '1-20' AS SpeedRange,
        ROUND(AVG(MEValue),3) AS NOx_AVG_ppm
FROM   TMeasureValues AS MV JOIN TMeasures AS M
        ON (MV.METID=M.METID AND MV.MEID=M.MEID)
        JOIN TMeasureTests AS MT ON M.METID=MT.METID
WHERE  (DPCode = 'NOx') AND (MESpeedInt >0 AND MESpeed <= 20)
        AND MV.DMWID = 2 AND DMUID = 2
GROUP BY TestID
UNION

SELECT 'TestType' = CASE TestID

        WHEN 1 THEN 'FTP'
        WHEN 2 THEN 'NEDC'END,
        '1-20' AS SpeedRange,
        ...
SELECT 'TestType' = CASE TestID
        WHEN 1 THEN 'FTP'
        WHEN 2 THEN 'NEDC'END,
        '80 >' AS SpeedRange,
        ROUND(AVG(MEValue),3)
        AS NOx_AVG_ppm
FROM   TMeasureValues AS MV
        JOIN TMeasures AS M
        ON (MV.METID=M.METID
        AND MV.MEID=M.MEID)
        JOIN TMeasureTests AS MT
        ON M.METID=MT.METID
WHERE  (DPCode = 'NOx') AND (MESpeedInt >80)
        AND MV.DMWID = 2 AND DMUID = 2
GROUP BY TestID

```

Rys. 4. Kod SQL obliczenia średniej emisji NO_x zależności od typu testu i prędkości

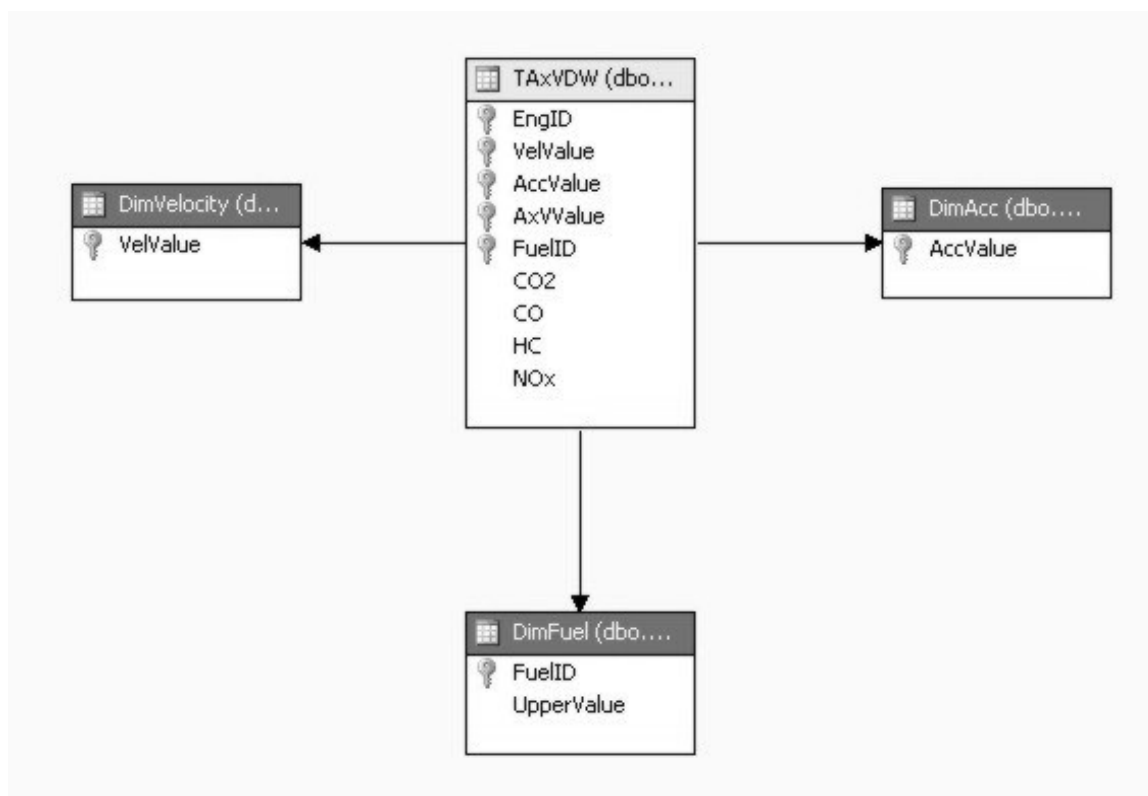
Fig. 4. SQL query expression for average NO_x emission grouped by test type and speed range

3. Zaawansowana analiza danych

Metody analizy danych podane w poprzednim paragrafie ograniczają się do prostych metod obliczeniowych opartych o proste wyrażenia. Nie umożliwiają natomiast stosowania zaawansowanych metod statystycznych oraz metod związanych z uczeniem maszynowym. Współczesne Systemy Zarządzania Bazami Danych zawierają moduły pozwalające na taką analizę. Podstawą takiej analizy są najczęściej wielowymiarowe struktury danych.

Na Rys. 5. przedstawiono schemat logiczny takiej struktury, zwanej kostką wielowymiarową. W tym przypadku pamiętane są wartości danych pomiarowych dotyczących emisji takich związków szkodliwych jak CO₂, CO, HC i NO_x w zależności od prędkości (*DimVelocity*), przyspieszenia (*DimAcc*) i zużycia paliwa (*DimFuel*). Dane do przygotowania kostki mogą zostać pobrane z podanej w poprzednim paragrafie bazy danych pomiarowych lub z dowolnego innego źródła. W tym konkretnym przypadku dane zostały pobrane z zewnętrznego źródła danych [4]. Wstępna obróbka polegała na podziale prędkości na poszczególne zakresy i obliczenia średnich emisji w tych zakresach dla wszystkich uwzględnianych składników spalin [5].

Do przechowywania przetworzonych danych do dalszej analizy została utworzona baza danych zintegrowanych (Rys. 1.)



Rys. 5. Schemat kostki wielowymiarowej

Fig. 5. Multidimensional cube scheme

W tabeli Tab.3. przedstawiono zależności natężenia emisji CO od przyspieszenia, zużycia paliwa i średniej prędkości. Wymiar prędkości został zwinięty (wartość „Dim Velocity - Multiple Items”) a wartość przyspieszenia została przeskalowana wartością 10. Wartość w kolumnie „Dim Acc” równa np. 2 to przyspieszenie o wartości 0,2 m/s².

Pewne kombinacje zużycia paliwa i przyspieszeń nie występują. Zmierzone wartości zużycia paliwa są najczęściej zapisywane w formacie liczb rzeczywistych i zawierają się w przedziałach od 0 do pewnej maksymalnej wartości. Grupowane danych względem wartości ciągłych utrudnia analizę wyników i stanowi je mało użytecznymi. W tym przypadku wymiar reprezentujący zużycie paliwa (*Dim Fuel*) został podzielony na przedziały. Prędkość również została poddana zamianie z wartości ciągłych na dyskretne. Podział został dokonany co 5 km/h, a w podanym przykładzie wyniki analizy ograniczono dla prędkości w zakresie od 40 do 60 km/h.

Zaawansowana analiza danych najczęściej umożliwia korzystanie z następujących klas algorytmów:

- klasyfikacji danych,
- obliczania regresji,
- segmentacji,
- analizy sekwencji.

Wraz z narzędziami do analizy danych dostarczane są najczęściej narzędzia do tworzenia raportów i wizualizacji otrzymanych wyników.

Tab. 3. Tab. zależności natężenia emisji CO od przyspieszenia i zużycia paliwa

Tab.3. CO emission dependency of fuel and acceleration factors

Dim Velocity ▾ (Multiple Items)		Dim Fuel ▾							
		2	3	4	5	6	7	8	9
Dim Acc ▾	CO	CO	CO	CO	CO	CO	CO	CO	CO
-12		36.143			383.567				
-11	2.932	46.289			527.9				
-10	9.919	58.279			407.032	294.314			
-9		67.279	9.491			732.061			
-8		68.854	14.379			603.797	176.23		
-7		64.558	24.29			297.782	331.862	193.541	
-6		29.841	58.912	5.477		152.42	504.228	207.025	
-5			84.561	16.626			316.694	364.436	197.75
-4			68.509	27.37		11.663	326.277	180.694	394.157
-3				97.233		18.195		522.397	203.549
-2				5.085		63.507	55.049	349.823	187.617
-1						27.526	44.793	36.217	385.179
0						10.785	24.21	40.777	412.888
1							11.073	27.616	55.381
2								11.361	31.045
3								11.595	6.387
4									11.813
5									6.095

4. Wymiana danych

Osobnymi i bardzo istotnymi zagadnieniami wykorzystania baz danych są:

- wymiana danych pomiędzy różnymi systemami,
- reprezentacja danych do dalszego wykorzystania w oparciu o silne gramatyki narzucające jednolitą ich postać (standaryzacja).

Jednolita postać zapisu danych jest szczególnie istotna w przypadku danych, które nie przyjmują postaci cyfrowej np. tekst, wzory matematyczne, schematy i symbole. Wykorzystanie tych elementów do wyznaczania własności ekologicznych pojazdów np. wzory korekcyjne związane ze zużyciem pojazdu, opisy norm itp. są niezbędne. Najczęściej są one dostępne w postaci raportów organizacji europejskich. Powszechnie używaną technologią do reprezentacji tego typu danych jest XML. Współczesne systemy bazodanowe tą technologię bardzo mocno wspierają.

Na Rys. 6. pokazano fragment dokumentu XML zawierającego klasyfikacje pojazdów zgodnie z COPERT III [6]. Dokument ten powstał jako wynik zapytania do tabel przechowujących klasyfikacje pojazdów w bazie słownikowej (Rys. 1.). Postać ta ma następujące zalety:

- jest czytelna,
- jest łatwa w obsłudze przez ogólnodostępne programy komputerowe,
- jest powszechnie uznanym standardem,

- jest obsługiwana przez większość systemów baz danych,
- istnieje możliwość określenia silnej gramatyki np. w postaci XML Schema.

Zastosowanie gramatyk do sprawdzania poprawności struktur dokumentów i zawartych w nich danych powoduje znaczne obniżenie kosztów obsługi danych i zwiększenie ich jakości.

Należy oczekiwać, że w niedalekiej przyszłości większość dokumentów tworzonych przez oprogramowanie biurowe będzie miała postać dokumentów XML. Obecnie pakiet OpenOffice realizuje tą ideę, również firma Microsoft zapowiada wprowadzenie takiej możliwości do pakietu Office.

```

    <Vehicle_class>
    <TSector Sector="Passenger Cars">
    <TSubsector Subsector="Gasoline &lt;1,4">
    <TTech Tech="PRE ECE" />
    ...
    <TTech Tech="Euro IV - 98/69/EC Stage2005" />
    <TTech Tech="Euro V (post 2005)" />
    </TSubsector>
    <TSubsector Subsector="Gasoline 1,4 - 2,0">
    ...
    <TSubsector Subsector="Diesel &lt;2,0">
    </TSector>
    <TSector Sector="Heavy Duty Vehicles">
    <TSubsector Subsector="Gasoline &gt;3,5">
    <TTech Tech="Conventional" />
    ...
    </TSubsector>
    <TSubsector Subsector="Diesel 3,5 - 7,5">
    <TTech Tech="Conventional" />
    <TTech Tech="Euro I - 91/542/EEC Stage I" />
    <TTech Tech="Euro II - 91/542/EEC Stage II" />
    <TTech Tech="Euro III - 2000 Standards" />
    <TTech Tech="Euro IV - 2005 Standards" />
    <TTech Tech="Euro V - 2008 Standards" />
    </TSubsector>
    </TSubsector>
    ...
    </TSector>
    </Vehicle_class>

```

Rys. 6. Dokument XML zawierający klasyfikacje pojazdów samochodowych

Fig. 6. XML document consists vehicles classification

5. Wnioski

W artykule zilustrowano przykładami możliwości zastosowania technologii baz danych do wyznaczania własności ekologicznych pojazdów. Wykorzystanie tego typu środowiska umożliwia:

- integrację danych pochodzących z różnych źródeł,

- rozproszenie danych z jednoczesnym dostępem do nich z dowolnego miejsca,
- uzyskanie niezależności aplikacji wykorzystujących dane od sposobów i metod ich przechowywania,
- dostęp do wbudowanych w System Zarządzania Bazą Danych narzędzi do zaawansowanej analizy danych,
- jednolitości formatów i reprezentacji danych.

Literatura

- [1] Grinke, L., *System baz danych do bilansowania emisji spalin*, X Konferencja Naukowa, Badania Symulacyjne w Technice Samochodowej, Kazimierz Dolny 2005. Polska Akademia Nauk Oddział w Krakowie, Konstrukcja, badania, eksploatacja, technologia pojazdów samochodowych i silników spalinowych, Zeszyt Nr 29-30, Kraków 2005, pp. 155-162.
- [2] Chłopek, Z., *Modelowanie procesów emisji spalin w warunkach eksploatacji trakcyjnej silników spalinowych*, Prace Naukowe Politechniki Warszawskiej z.173. Warszawa, 1999.
- [3] Molina, H., Ullman, J. Widom, J., *Database Systems. The complete book*, Prentice Hall 2002.
- [4] Sturm, P., Boulter, P., Do Haan, P., Joumard, R., Hausberger, S., Hickman, J., Keller, M., Niederle, W., Ntziachristos, L., Reiter, C., Samaras, Z., Schinagl, G., Schweizer, T., Pischinger, R., *Instantaneous emission data and their use in estimating passenger car emissions*, Graz, VKM-THD Report, vol. 74, Verlag der Tech. Univ. Graz, 1998.
- [5] Brzozowski, K., *Wybrane aspekty modelowania natężenia emisji w warunkach dynamicznych*, Czasopismo Techniczne Mechanika z.6-M/2004 s. 125-132. Kraków, 2004.
- [6] Ntziachristos, L., Samaras, Z., *Methodologies for Estimating Air Pollutant Emissions from Transport. Average Hot Emission Factors For Passenger Cars and Light Duty Trucks*, Thessaloniki. 1998. COPERT III. Copenhagen. 2000.