

**ZESZYTY NAUKOWE NR 72
WYŻSZEJ SZKOŁY MORSKIEJ
SZCZECIN 2003**

WYDZIAŁ INŻYNIERYJNO-EKONOMICZNY TRANSPORTU

Grzegorz Hołowiński

**Problemy temporalności w bazach danych
na przykładzie raportowania
w ramach systemu informacyjnego zarządzania**

Przedstawiono problem temporalności danych w systemach baz danych. Zaproponowano sposób realizacji temporalnego składowania danych w relacyjnej bazie danych na potrzeby raportowania w ramach systemów informacyjnych zarządzania, który może być zastosowany w przedsiębiorstwach i organizacjach, np. urzędzie miejskim, zarządzie portu, w celu dostarczania syntetycznych danych w różnych układach informacyjnych oraz umożliwienia monitorowania osiągniętych wyników.

**Problems of Temporariness in Data Bases
on the Example of Reporting
in the Information Management System**

The present publication presents the problem of data temporariness in data base systems. A way of temporary data storage in a relational database has been suggested, for the needs of reporting in informational management system, which may be applied in various enterprises, like the municipal office or the harbour master's office, in order to provide synthetic data in informational systems and make it possible to monitor the results obtained.

Wprowadzenie

Opracowanie i wdrażanie systemów integrujących rozproszone zasoby informacyjne organizacji stanowi jeden z najbardziej aktualnych problemów praktycznych zastosowań informatyki [1]. W przypadku systemów przeznaczonych dla decydentów istotne znaczenie ma dostęp do danych o aktualnym stanie przedsiębiorstwa czy organizacji, np. administracji publicznej czy administracji morskiej.

Dla potrzeb podejmowania decyzji istotne jest ujęcie syntetyczne danych dotyczących całej firmy lub dziedziny, konkretnego tematu. W tym celu konieczne jest zebranie danych analitycznych z różnych źródeł, często rozproszonych, a nawet niedostępnych w formie elektronicznej. Czas przygotowania i udostępnienia danych jest zazwyczaj inny dla każdego ze źródeł danych. Stąd szczególne znaczenie ma problem temporalności pozyskanych w ten sposób danych.

W dalszej części artykułu przedstawiono metody określania stanów systemu informacyjnego, zagadnienie znaczenia czasu w systemach informacyjnych zarządzania oraz zaproponowano rozwiązanie problemu temporalności danych w bazie relacyjnej dla potrzeb raportowania w ramach systemu MIS (*Management Information System*).

1. Reprezentacja czasu w bazach danych

W połowie lat siedemdziesiątych pojawiły się pierwsze rozwiązania związane z czasem w relacyjnych bazach danych [2]. Dziesięć lat później rozpoczęto badania nad temporalnymi bazami obiektowymi a później (od roku 1996) nad temporalnymi bazami danych czasu rzeczywistego [2].

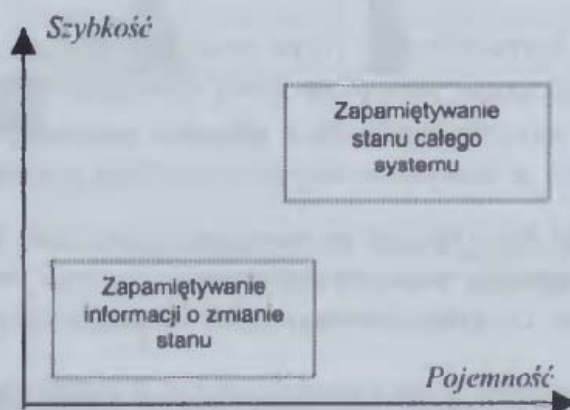
Konieczność podjęcia takich problemów była następstwem identyfikacji potrzeb związanych z uwzględnieniem wartości zmiennych w czasie [3], [4], [5], [6].

Tsotras i Hart [7] dzielą metody określania historii stanów na dwie podstawowe grupy bazujące na: 1) zapamiętywaniu stanu całego systemu i 2) zapamiętywaniu tylko informacji o zmianach stanu jego wartości.

W pierwszej z nich zmiana stanu wartości powoduje stworzenie kopii reprezentacji całego systemu, a następnie jej modyfikację poprzez uaktualnienie zmienionej wartości. Od tej pory zmodyfikowana kopia stanu systemu staje się aktualną jego reprezentacją, a dotychczasowa – historyczną. W metodach należących do drugiej grupy, modyfikacja konkretnej wartości, sprowadza się do zapisania w systemie informacji o tej zmianie.

Metody należące do pierwszej grupy są stosunkowo szybkie (rys. 1), gdyż poszukiwanie wartości sprowadza się do znalezienia wersji całej reprezentacji systemu. Ceną, jaką należy za to zapłacić jest duża objętość pamięci. Natomiast

metody należące do drugiej grupy wymagają mniej pamięci przy dłuższym czasie ustalania wartości w danym momencie czasu, co jest kosztem analizowania historii zmian wartości.



Rys. 1. Porównanie szybkości odnajdywania wartości oraz pojemności systemu w zależności od metody określania historii stanów

Fig. 1. A comparison of the speed of value findings and the system's capacity depending on the method of determining history of states

Dobór metody zależy od struktury danych, zakładanej liczby zmiennych wartości, częstotliwości zmian od bardzo częstych (np. mierzonych w sekundach lub częściach sekundy) do stosunkowo rzadkich (np. uaktualnianych raz do roku).

Przykładem praktycznej realizacji temporalnego zarządzania danymi jest system finansowy TRANS [8], [9], w którym każdy rejestrowany fakt (zdarzenie) jest określany znacznikiem czasu i opisem formalnym na potrzeby księgowania.

Bazy danych, ze względu na ujęcie czasu można podzielić [10] na:

- **bazy migawkowe**, przechowujące informacje tylko i wyłącznie o aktualnym stanie systemu;
- **bazy historyczne**, pozwalające na pamiętanie historii wartości zmiennych względem czasu ich występowania w rzeczywistości;
- **bazy transakcyjne**, które nie uwzględniają aspektu zmienności wartości w świecie rzeczywistym (jak to ma miejsce w bazach historycznych), ale rejestrują zmiany wartości w bazie danych – umożliwia to określenie stanów przeszłych w odniesieniu do stanów w bazie a nie w rzeczywistości;
- **bazy bitemporalne**, uwzględniające aspekty czasu zarówno w odniesieniu do rzeczywistości jak i do systemu bazy danych.

Dla zapisu wymiaru czasowego poszczególnych zmiennych w systemie stosuje się najczęściej [10]:

- znacznik czasowy, określający czas wystąpienia zdarzenia (np. data, godzina);
- interwał temporalny do oznaczenia przedziału czasu, w którym wartość należy traktować jako prawdziwą w świecie rzeczywistym;
- element temporalny będący zbiorem interwałów wszystkich przedziałów czasu, w którym występuje określona wartość zmiennej.

W przypadku baz danych w systemach zarządzania informacją ważne jest uwzględnianie aspektów czasu w odniesieniu do rzeczywistości oraz do systemu bazy danych, m.in. do kontrolowania systemu pod względem jakości danych.

2. Czas w systemach informacyjnych zarządzania

System informacyjny zarządzania [11], [12] dostarcza informacji w postaci syntetycznej na potrzeby podejmowania decyzji. W praktyce źródłem jego danych są systemy transakcyjne, rozproszone zasoby internetowe (np. aktualne kursy walut, wskaźniki giełdowe) a także dodatkowe informacje niemożliwe do pozyskania w sposób zautomatyzowany – wprowadzane ręcznie [13], [14].

W przypadku systemu raportowania w ramach MIS w dużym przedsiębiorstwie czy administracji, jak np. zarząd morskiego portu, urząd miejski, urząd wojewódzki, występują problemy z pozyskaniem na czas odpowiednich danych źródłowych. Związane jest to z: dużą liczbą źródeł danych dla systemu MIS, ich czasem transakcyjnym, sposobami zasilania MIS (on-line, tryb wsadowy, itp.). Ponadto poszczególne wartości mają różny interwał temporalny. Dla wymienionych przykładów systemów MIS przyjmuje się zazwyczaj zestawianie danych w ujęciu rocznym, półrocznym, kwartalnym, a także miesięcznym.

Rysunek 2 przedstawia trzy zmienne, z których każda aktualizowana jest, co miesiąc lub kwartał. Ze względu na wymienione wcześniej czynniki występują opóźnienia w zasilaniu systemu danymi. Na rysunku 2 zaznaczono je kolorem czarnym. W związku z tym próba odczytania wartości zmiennych w systemie może się nie powieść, gdyż nie będzie możliwe znalezienie w systemie wszystkich wartości na konkretny dzień, np. 12 stycznia uaktualnione będą dane A i B, podczas, gdy C (uaktualniana kwartalnie) będzie miała wartość z poprzedniego okresu – 30 września (wartość w rzeczywistości, choć zapis w bazie nastąpił dopiero w połowie października). W takim przypadku konieczne jest podjęcie odpowiedniego działania, zgodnego z przyjętą strategią publikowania informacji w systemie, np. informowania użytkownika o braku aktualnej (na 12 stycznia) i podaniu ostatniej wartości zmiennej C.

Nazwa danej: okres aktualizowania	Październik	Listopad	Grudzień	Styczeń
A: Miesiąc				
B: Miesiąc				
C: Kwartał				

Stan wartości z bazy danych np. 12 stycznia nie zawiera kompletu informacji (dana C uaktualniana kwartalnie nie została jeszcze aktualizowana)

Rys. 2. Uaktualnianie poszczególnych wartości systemu w czasie (kolorem czarnym zaznaczono opóźnienia w uaktualnianiu wartości)

Fig. 2. Updating particular vales of the system in time (delays in updating have been marked in black)

Działania takie wymagają opracowania sposobu reprezentacji czasu na potrzeby raportowania w ramach MIS.

3. Dane temporalne w bazie relacyjnej na przykładzie raportowania w systemie informacyjnym zarządzania

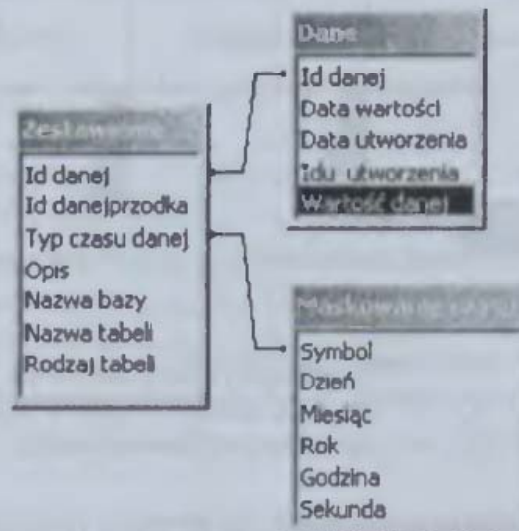
Zakładając, że moduł raportowania systemu MIS będzie składował dane w relacyjnej bazie danych, można zaproponować ogólną implementację jego struktury danych w sposób przedstawiony na rysunku 3.

Zakładając, że dane będą zapisywane w bazie relacyjnej, proponowany zapis składa się z dwóch części:

- opisu zmiennej reprezentującej fakt (np. wielkość przeladunków w tonach w kwartale, liczba ludności miasta w roku),
- znaczników czasu odpowiadających aspektom: rzeczywistemu i zapisu danej w bazie danych.

Taki układ separuje opis zmiennej w postaci (zestawienie danych):

- **id danej** unikalny identyfikator zmiennej w bazie;
- **id danej przodka**, pozwalający na hierarchizowanie zmiennych w postaci struktury informacyjnej drzewa;
- **typu czasu danej** związany z przygotowaniem znacznika czasu (aspekt rzeczywisty danej);
- **nazwy bazy, nazwy tabeli i rodzaju tabeli** pozwalające określić, w której bazie danych i tabeli znajduje się dana wartość.



Rys. 3. Struktura logiczna opisu danej (faktu) w relacyjnej bazie danych

Fig. 3. Logical structure of data (fact) description in a relational data base

Zapewnienie bitemporalności wartości w bazie wymusza przechowywanie (1) daty i czasu zapisania wartości w bazie oraz (2) zachowania aspektu rzeczywistości w postaci reprezentującej częstotliwość uaktualniania wartości rzeczywistych (roczna, półroczna, kwartalna, miesięczna, itp.). Dążąc do znormalizowania zapisu daty wartości w aspekcie rzeczywistości w bazie danych, proponuje się maskowanie **daty wartości** zgodnie z tabelą 1.

Przykład. Dla daty 04 listopada 2002 r. maskowanie względem dnia (M) spowoduje zapis 2002-11-30, względem kwartału (K) 2002-12-31, względem roku (R) 2002-12-31, podczas gdy zapis względem dnia (D) pozostanie bez zmian (2002-11-04). Dzięki takiemu zapisowi wszystkie dane dotyczące kwartału będą miały jednakową datę (wartości w rzeczywistości), tj. 2002-12-31, co ułatwi ich wyszukiwanie w systemie.

Tak maskowana data ułatwia przeszukiwanie tabel bazy danych w celu znalezienia odpowiednich wartości.

Informacja na temat normalizowania daty przechowywana jest w oddzielnej tabeli (*maskowanie czasu* z rys. 3.) w postaci:

- symbol maski (na który wskazuje pole *typ czasu danej* w zestawieniu danych);
- dzień, miesiąc, rok, godzina, minuta opisujące maskę zgodnie z tabelą 1.

Wartości danych przechowywane są natomiast w oddzielnych tabelach o strukturze (dane z rys. 3.):

- *id danej* identyfikator łączący daną z jej opisem, w zestawieniu danych;
- *data i czas wartości* data znormalizowana zgodnie z maskowaniem czasu – aspekt rzeczywistości;
- *data utworzenia* data zapisu w bazie przez użytkownika o identyfikatorze *idu użytkownika*;
- *wartość danej* konkretna wartość liczbowa, tekstowa, itp.

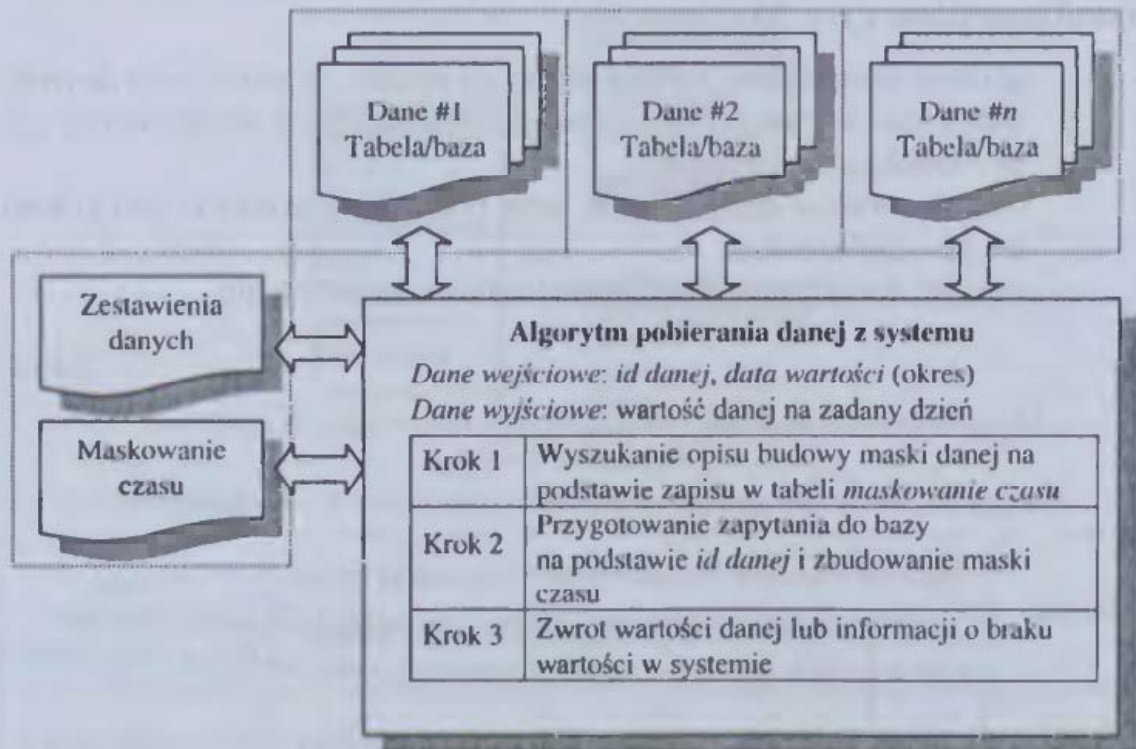
Tabela 1

Maskowanie znacznika czasu z uwzględnieniem poszczególnych składowych (+) lub ich pomijaniem (-)

Masking the time marker considering particular components (+) or skipping them (-)

Symbol	Znaczące pozycje w formacie czasu rrrr.mm.dd gg:mm:ss						Uwagi dotyczące zapisu daty w bazie danych
	Dzień dd	Miesiąc mm	Rok rrrr	Godzina gg	Minuty mm	Sekundy ss	
D (dzień)	+	+	+	-	-	-	
M (miesiąc)	-	+	+	-	-	-	<i>dd</i> = ostatni dzień miesiąca <i>mm</i>
K (kwartał)	-	+	+	-	-	-	<i>mm</i> = kwartał x 3 <i>dd</i> = ostatni dzień miesiąca <i>mm</i>
R (rok)	-	-	+	-	-	-	<i>dd</i> = 31 <i>mm</i> = 12
G (godzina)	-	-	-	+	+	+	

Na rysunku 4 przedstawiono schemat modułu bazy danych dla opisanych powyżej struktur. Proponowane rozwiązanie pozwala na realizację systemu w postaci jednej lub więcej baz danych. W najprostszym przypadku część odpowiadająca za opisy i maskowanie oraz tabele z danymi mogą znajdować się w jednej bazie danych. Jednakże uwzględnienie w *zestawieniach danych* dodatkowych pól (*nazwa bazy, nazwa tabeli, rodzaj tabeli*) umożliwia odniesienie się do innych baz danych, co pozwala na budowanie systemu rozproszonego.



Rys. 4. Schemat i algorytm wyszukiwania wartości danych w systemie z użyciem rozproszonych danych w tabelach i bazach opisanych przez zestawienia danych i o dacie normalizowanej zgodnie z zapisami w tabeli *maskowanie czasu*

Fig. 4. Diagram and algorithm of searching out data values in a system using dispersed data in Tables and data described by data comparisons and the standardized datum in accordance with the table of time masking

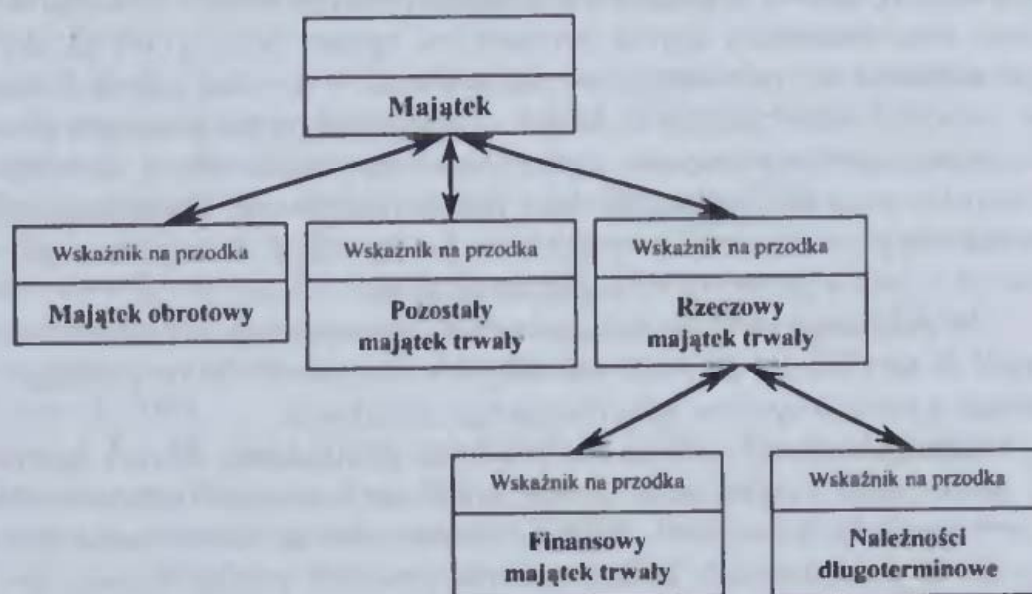
Proponowane rozwiązanie umożliwi uwzględnienie aspektu czasu w bazie modułu raportowania dla systemu MIS.

Wyszukiwanie konkretnych danych z systemu (np. liczby ton węgla przeladowanych w porcie w III kwartale, lub liczby pozwoleń na budowę wydanych przez urząd miejski) sprowadza się do realizacji algorytmu z rysunku 4.

Należy zauważyć, że zaproponowana realizacja zadania jest przeznaczona do implementacji w systemach generujących raporty informacyjne dla decydentów. Proponowane podejście nie jest efektywne (wg kryterium czasu dostępu i czasu przetwarzania) dla dużej liczby danych, jak to ma miejsce w systemach transakcyjnych. Jego przeznaczeniem jest system dostarczania informacji zbiorczych, w różnych układach informacyjnych i umożliwiających mierzenie rezultatów osiąganych przez przedsiębiorstwo – w ramach systemu MIS [11].

Tabele (*dane* z rysunku 4) zawierają wyłącznie dane jednego typu: liczbowe, znakowe, pola typu memo, itp. Wartości tekstowe trzymane są w jednej tabeli, a liczbowe w innej. Co więcej proponowane struktury pozwalają na sepa-

rowanie wybranych danych (wg identyfikatora zmiennej *id danej*) w oddzielnych tabelach, np. wszystkie wielkości przeładunków mogą znajdować się w jednej tabeli. Przyjęta metoda pozwala na szybkie przesuwanie zbiorów danych pomiędzy bazami i tabelami (modyfikację struktury informacyjnej) poprzez uaktualnienie wskazań na tabele w strukturze *zestawienia danych* oraz przeniesienie konkretnych danych w nowe miejsce. Pozwala to na takie administrowanie systemem, aby zapewnić jego możliwie wysoką wydajność poprzez rozmieszczanie baz danych na serwerach o odpowiedniej mocy obliczeniowej i połączonych z użytkownikami szybkimi łączami.



Rys. 5. Przykład zastosowania informacyjnej struktury drzewiastej w proponowanym rozwiązaniu reprezentacji danych w systemie

Fig. 5. Example of applying informational tree structure in the suggested solution of data representation in the system

Ponadto proponowane rozwiązanie umożliwia łączenie poszczególnych danych w struktury drzewiaste (rys. 5.), które są jednym ze sposobów wiązania danych na potrzeby raportowania. Na przykład majątek firmy można podzielić na majątek obrotowy, majątek rzeczowy trwały, pozostały majątek trwały. Z kolei rzeczowy majątek trwały na finansowy majątek trwały i należności długoterminowe (rys. 5.). Rodzaje majątku można dzielić dalej, jednakże z punktu widzenia proponowanego rozwiązania jest wskazanie możliwości grupowania danych w większe struktury, np. z wykorzystaniem tzw. struktury drzewa. Ma to szczególne znaczenie dla ułatwienia procesu projektowania systemów raportowania, umożliwia stosowanie mechanizmów rekurencyjnych oraz przygotowywanie modułów raportowania z wykorzystaniem technologii internetowych.

Możliwe jest modyfikowanie zaproponowanego rozwiązania, aby uzyskać powiązania wielowymiarowe danych, na podobieństwo tzw. hiperłączy w dokumentach hipertekstowych wykorzystywanych w Internecie.

Zakończenie

Zagadnienie temporalności w bazach danych związane jest ze zjawiskiem czasu. Czas, określający odniesienie zdarzeń względem teraźniejszości, przeszłości i przyszłości może być mierzony dzięki zachodzącym zmianom. W odniesieniu do baz danych, zmiany te wiążą się z przechowywanymi danymi i zmianami ich wartości. Stan teraźniejszy danych powinien być opisany precyzyjnie, tak aby dane te jak najdokładniej odzwierciedlały rzeczywistość – np. stan mienia komunalnego w mieście. Uaktualnianie tych danych, z jednoczesnym zachowaniem poprzednich wartości, umożliwia tworzenie zapisu historii danych. Co więcej, dane historyczne i aktualne pozwalają na zastosowanie metod modelowania i prognozowania, które umożliwią przewidywanie przyszłych stanów, np. zmian demograficznych w dzielnicach miasta w perspektywie najbliższych 20 lat.

W publikacji zwrócono uwagę na zagadnienie temporalności w bazach danych na przykładzie gromadzenia danych z różnych źródeł na potrzeby raportowania w ramach systemu informacyjnego zarządzania.

Zaproponowano rozwiązanie problemu gromadzenia danych temporalnych z możliwością rozpraszania zasobu informacyjnego w zależności od potrzeb i wymagań użytkowników. Zaproponowana metoda składowania danych pozwala na przedstawianie zasobu informacyjnego w postaci drzewa, co w naturalny sposób rekomenduje interfejs użytkownika w postaci przeglądarki internetowej (struktura drzewa odpowiada naturalnemu grupowaniu informacji).

Dalsze prace będą miały za zadanie przeprowadzenie badań nad efektywnością proponowanego rozwiązania w różnych relacyjnych systemach zarządzania bazami danych, m.in. MySQL, PostgreSQL, Oracle, MS SQL.

Następnym etapem będzie przeprowadzenie badań związanych z wykorzystaniem metod obiektowych w projektowaniu systemów MIS.

Literatura

1. Ochman J., *Integracja w systemach informatycznych zarządzania*. Wyd. PWE, Warszawa, 1992.
2. Sondgrass T., Ozsoyoglu G., *Temporal and Real-Time Database: A Survey*, IEEE Transactions on Knowledge and Data Engineering, No. 4, Vol. 7, 1995.
3. Maiocchi R., Percini B., *Temporal Data Management Systems: A Comparative View*, IEEE Transactions on Knowledge and Data Engineering, No. 4, Vol. 3, 1991.

4. Kania K., Gołuchowski J., *Zagadnienia czasu w klasycznych systemach baz danych*, Informatyka, Nr 7, 1996.
5. Kania K., Gołuchowski J., *Reprezentacja czasu w modelu konceptualnym SI*, Informatyka, Nr 8, 1996.
6. Kania K., Kędzierski S., Gołuchowski J., *Zależności temporalne w modelowaniu i analizie procesów gospodarczych*, Informatyka, Nr 3, 1998.
7. Tsotar V., Hart G., *Efficient Management on Time-Evolving Databases*, IEEE Transactions on Knowledge and Data Engineering, No. 4, Vol. 7, 1995.
8. Budziński R., *System rachunkowości transakcyjnej*, Polska Akademia Nauk, 1998.
9. Budziński R., *Komputerowy system przetwarzania danych ekonomiczno-finansowych w przedsiębiorstwie*, Polska Akademia Nauk, 2001.
10. Jankowski J., *Problemy reprezentacji czasu w systemach baz danych*, Zeszyty Naukowe PS Nr 557, str. 101-112, Szczecin, 2000.
11. Alter S., *Information Systems: A Management Perspective*. Wyd. The Benjamin/Cummings Publishing Company, 1996.
12. Szydłowski J.I., *Wykorzystanie narzędzi CASE w procesie projektowania systemów informacyjnych*, Zeszyty Naukowe PS Nr 557, str. 7-19, Szczecin, 2000.
13. Kimball R., *The Data Warehouse Lifecycle Toolkit. Expert Methods for Designing and Deploying Data Warehouses*. Wyd. J. Willey & Sons, 1998.
14. Tiwana A., *Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System*. Wyd. Prentice Hall, 1999.

Wpłynęło do redakcji w listopadzie 2002 r.

Recenzent

Prof. dr hab. inż. Ryszard Budziński

Adres Autora

dr inż. Grzegorz Holowiński

Wyższa Szkoła Morska w Szczecinie

Wydział Inżynieryjno-Ekonomiczny Transportu

Instytut Zarządzania Transportem

70-507 Szczecin

ul. Henryka Pobożnego 11

gholowinski@wsm.szczecin.pl