

Sebastian Łacheciński

Uniwersytet Łódzki

e-mail: sebastian.lachecinski@uni.lodz.pl

MODELOWANIE DANYCH TEMPORALNYCH W RELACYJNYM MODELU DANYCH

MODELING OF TEMPORAL DATA IN THE RELATIONAL DATA MODEL

DOI: 10.15611/ie.2017.4.08

JEL Classification: C88, L86

Streszczenie: Artykuł porusza problem modelowania zmian zachodzących w czasie w odwzorowaniu do relacyjnego modelu danych. Zagadnienie to jest szczególnie istotne wszędzie tam, gdzie należy przechowywać w bazie danych nie tylko dane zawsze aktualne, ale również dane historyczne. Prezentowane są rozwiązania uwzględniające temporalność atrybutów encji, temporalność związku encji oraz temporalność całej encji, jak również temporalność całej encji i związków do niej przynależących. W skondensowany sposób opisane zostały także zagadnienia reprezentacji czasu, rodzaje atrybutów i związków z punktu widzenia ich zmienności w czasie. Ponadto zarysowano podział temporalnych baz danych. Celem artykułu jest porównanie sposobów modelowania zmian wybranych cech encji logicznych w czasie.

Słowa kluczowe: modelowanie danych, modelowanie czasu, temporalny model danych, relacyjny model danych.

Summary: This article presents a problem of changes modeling which occurs over time, in mapping to relational data model. This issue is particularly important wherever not only always current data, but also historical data should be stored in the database. Solutions are presented that take into account the temporality of the attributes of entity, temporality of the relationships of the entity, and of entire entities, as well as temporality of entire entity and relationships belonging to entity. The issues of time representation, types of attributes and relationships from the point of view of their variability over time are also described in a condensed manner. In addition, the division of temporal databases is described. The aim of this article is a comparison of the ways of modeling changes of selected features of logical entities over time.

Keywords: data modeling, time modeling, temporal data model, relational data model.

1. Wstęp

Modelowanie zmian zachodzących w czasie jest bardzo istotnym zagadnieniem. Przede wszystkim dotyczy to przypadków, gdy zachodzi konieczność przechowania, oprócz danych, których zarówno wartość i status są zawsze i wyłączenie aktualne, także danych historycznych. Konieczność ta poparta jest zazwyczaj wymaganiami biznesowymi, które z kolei przekładają się na określoną funkcjonalność projektowanego systemu informatycznego opartego na bazie danych, z możliwością rejestrowania i przechowywania zmian temporalnych. Istota tego zagadnienia jest podkreślana w pracach wielu badaczy, m.in.: [Mach-Król, Michalik 2014; Michalik, Mach-Król 2014, Mach-Król 2014].

Zmiany temporalne mogą dotyczyć wybranych pojedynczych atrybutów lub grup atrybutów danej encji, zmian związków zachodzących pomiędzy encjami czy też zmian całej encji (ze wszystkimi jej atrybutami), a także z uwzględnieniem zmian wszystkich związków, w których dana encja bierze udział.

Przykładów zastosowań modelu temporalnego jest nieskończenie wiele, wśród których m.in. wyróżnić można gromadzenie informacji nt. historii zatrudnienia pracowników w danej firmie. Przy czym należy zwrócić uwagę, iż historia ta może dotyczyć awansu pracowników, tj. zmiany stanowisk, ale także zmian miejsca zatrudnienia, a więc historii zatrudnienia, w różnych działach, w tej samej firmie; należy również uwzględnić okresy zatrudnienia u innych pracodawców. Równie istotne może być przechowywanie historii wynagrodzeń pracownika, dzięki czemu możliwe będzie wyliczenie wynagrodzenia pracownika w poszczególnych latach jego zatrudnienia. Inny przypadek może dotyczyć składowania historii cen sprzedawanych produktów, produktów oferowanych w przeszłości, jak również dostawców produktów z uwzględnieniem czasookresu ważności kontraktu z danym dostawcą oraz gamą produktów przez niego dostarczanych w danym okresie.

2. Podstawowe ujęcia czasu oraz reprezentacja encji w temporalnym modelu danych

Atrybuty encji można podzielić ze względu na charakter ich zmian w czasie na:

- statyczne, a więc takie, których wartości nie zmieniają się w czasie,
- temporalne, czyli takie, których wartości są zmienne wraz z upływem czasu.

Interwały czasowe opisujące czasokres obowiązywania, a zatem ważność czy też aktualność konkretnych danych w modelu danych, mogą być odwzorowane w postaci jawnej i niejawnej. Elementy jawne nazywane są stemplami czasowymi lub zegarami fizycznymi. Najczęściej definiowane są jako pojedyncza wartość, np. minuta, godzina, dzień itd. Tym samym wskazują na konkretny moment w czasie bądź też przedział czasowy (w przypadku zdefiniowania dwóch stempli czasowych, w którym określone dane są aktualne). Elementy niejawne nazywane są zegarami

logicznymi. Najczęściej definiowane są jako czasookres, np. kwartał, półrocze, semestr, okres rozliczeniowy itd. [Matysiak 2017].

Nie bez znaczenia podczas tworzenia temporalnego modelu danych pozostaje również charakter upływającego czasu, a więc sposób jego zmiany. Ma on wpływ na sposób przechowywania i zapisywania atrybutów, a także związków temporalnych. Wymiar upływającego czasu determinuje także podział temporalnych baz danych, skutkujący ujęciem w swoich modelach czasu rzeczywistego, transakcyjnego czy też obu jednocześnie.

Dodatkowym argumentem świadczącym o wadze i istotności rejestracji zmian temporalnych w bazach danych, a także ich odwzorowaniu w modelu logicznym jest gros prac badawczych prowadzonych od kilkadziesiąt lat [Mahmood i in. 2010; Fuller 2007], opracowanie przez C.J. Date'a specjalnej, dodatkowej, przeznaczonej dla temporalnej postaci normalnej nazywanej szóstą postacią normalną 6NF [Date i in. 2003] oraz uwzględnienie w zapisie standardu języka SQL [Kulkarni, Michels 2012] rozwiązań przeznaczonych dla temporalnych baz danych.

2.1. Reprezentacja czasu

Czas można sklasyfikować według kilku różnych kryteriów [Hajnicz 1996; Mach-Król 2007].

Ze względu na sposób zmiany czasu:

- ciągły – w którym dla każdych dwóch dowolnych chwil czasowych istnieje także określona chwila czasowa pomiędzy nimi,
- dyskretny – w którym pomiędzy dwiema chwilami czasowymi nie istnieje żadna chwila czasu.

Ze względu na skończoność (lub brak skończoności) czasu:

- lewostronnie skończony – w którym przeszłość jest ograniczona do pewnego momentu (najczęściej teraźniejszości), sama zaś przyszłość jest nieskończona,
- prawostronnie skończony – w którym przeszłość jest nieskończona, przyszłość zaś jest ograniczona do pewnego momentu (najczęściej teraźniejszości),
- obustronnie skończony – w którym zarówno przeszłość, jak i przyszłość są ograniczone do pewnych momentów w czasie,
- nieskończony – w którym zarówno przeszłość, jak i przyszłość nie są ograniczone do żadnego momentu w czasie.

Ze względu na sposób upływu czasu [Kania 2004; Mach-Król 2014]:

- liniowy – gdy czas przebiega w sposób uporządkowany, od przeszłości ku przyszłości. Przy czym istnieje tylko jeden sposób upływu czasu. W systemach opartych na tym modelu niemożliwe jest zapisanie zdarzenia odwołującego się do przyszłości. Czas ten kończy się w punkcie oznaczającym teraźniejszość. Jest to model używany w transakcyjnych bazach danych;
- gałęziowy – nazywany bywa modelem alternatywnych przyszłości. Istnieją różne sposoby upływu, które mają swój punkt wspólny. Wyróżnić można czas rozgałęziony lewo- i prawostronnie liniowy. W przypadku modelu czasu lewostron-

nie liniowego czas jest liniowy od przeszłości do terażniejszości. Przy czym może on dzielić się na wiele linii odwzorowujących możliwe scenariusze przyszłych działań czy też wartości istotnych z punktu widzenia użytkownika. Model ten jest przydatny podczas budowy systemów planowania oraz systemów generujących scenariusze działań. Dzięki temu możliwe jest przechowywanie wielu różnych wersji planów działań. Systemy zbudowane na podstawie tego modelu czasu pozwalają na porównanie wartości bieżących z prognozowanymi;

- równoległy – w którym istnieje wiele różnych sposobów przepływu czasu, przy czym żaden z nich nie ma z pozostałymi ani jednego wspólnego punktu czasowego. Czas ten określany jest mianem czasu jednocześnie lewo- i prawostronnie liniowego czy też czasem alternatywnych rzeczywistości.

Ze względu na powtarzalność zdarzeń w czasie:

- cykliczny – stosowany tam, gdzie zachodzi powtarzalność zdarzeń, a więc gdzie bardziej istotna jest zależność pomiędzy zdarzeniami realizowanego procesu niż sam kalendarz, np. sezonowość produkcji, zmiana pór roku itd. Jest często stosowany w systemach wspomagających planowanie produkcji sezonowej czy też zarządzanie organizacją pracy w danej jednostce z uwzględnieniem sezonowości;
- spiralny – stanowi połączenie modelu liniowego i cyklicznego. Uwzględnia upływ czasu wewnątrz cyklu oraz niezależny od zmian procesu upływ czasu zewnętrznego, najczęściej będącego odwzorowaniem kalendarza. Sprawdza się w obrazowaniu przebiegów powtarzających się procesów w ujęciu historycznym.

Ze względu na strukturę czasu [Mach-Król 2014]:

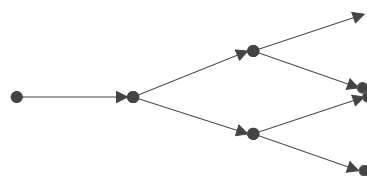
- punktowy – uwzględnia tylko punkty czasowe,
- przedziałowy – uwzględnia tylko przedziały czasowe,
- przedziałowo-punktowy – uwzględnia zarówno punkty, jak i przedziały czasowe.

Na rysunkach 1-6 przedstawiono interpretację graficzną podstawowych modeli czasu.



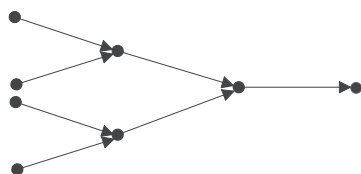
Rys. 1. Liniowy model czasu

Źródło: opracowanie własne.



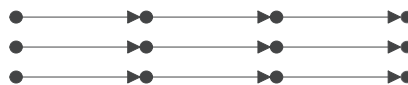
Rys. 2. Rozgałęziony lewostronny liniowy model czasu

Źródło: opracowanie własne.



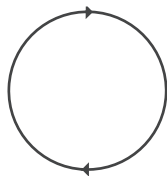
Rys. 3. Rozgałęziony prawostronny liniowy model czasu

Źródło: opracowanie własne.



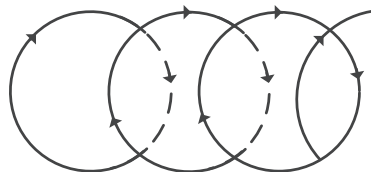
Rys. 4. Równoległy model czasu

Źródło: opracowanie własne.



Rys. 5. Cykliczny model czasu

Źródło: opracowanie własne.



Rys. 6. Spiralny model czasu

Źródło: opracowanie własne.

2.2. Rodzaje atrybutów i związków

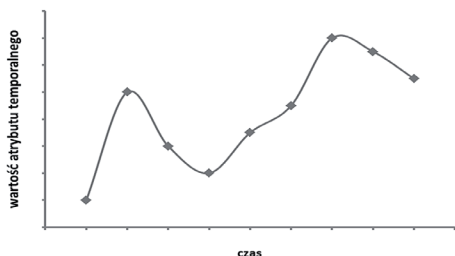
Mając na uwadze istniejącą zależność między atrybutami temporalnymi a upływem czasu, atrybuty temporalne można podzielić ze względu na sposób zmiany czasu:

- **ciągłe** – wartości tych atrybutów mogą być odczytywane w dowolnych punktach czasu. Jednakże istnieje możliwość szacowania wartości punktów w czasie, dla których nie przeprowadzono do tej pory odczytu wartości atrybutu. Jako przykład można podać m.in. odczyt wilgotności, temperatury, ciśnienia, częstotliwości itd.
- **dyskretne** – wartości tych atrybutów są prawdziwe (aktualne) tylko w konkretnym punkcie w czasie, np. liczba zamówień zrealizowanych w ciągu dnia;

oraz ze względu na strukturę upływającego czasu:

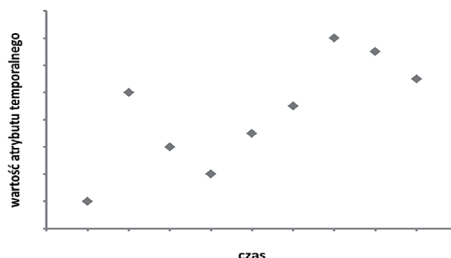
- **punktowe** – wartości tych atrybutów mogą być odczytywane w dowolnych punktach czasu. Mogą to być atrybuty zarówno o ciągłym, jak i dyskretnym sposobie zmiany czasu,
- **przedziałowe** – wartości tych atrybutów są aktualne (prawdziwe) w określonym przedziale czasu. Atrybuty te mogą mieć stale zdefiniowaną wartość lub też tylko w wybranych okresach. Jako przykład można wymienić m.in. liczbę zamówień zrealizowanych w danym okresie, liczbę dostaw z danego okresu, wielkość produkcji w danym okresie itp.
- **przedziałowo-punktowe** – wartości tych atrybutów mogą być odczytywane zarówno w wybranych okresach, jak i w konkretnych punktach czasu.

Na rysunkach 7-10 przedstawiono interpretację graficzną atrybutów temporalnych z uwzględnieniem ich charakteru zmienności. Rysunki 7-10 opracowano na podstawie [Kania 2004].



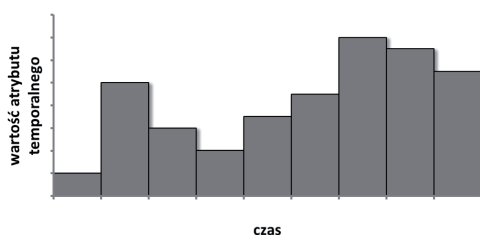
Rys. 7. Atrybut ciągły

Źródło: opracowanie własne.



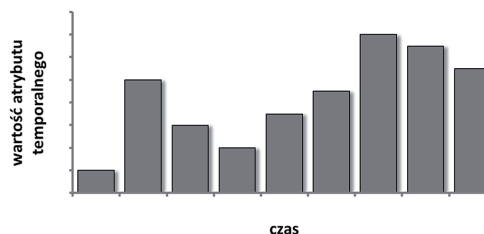
Rys. 8. Atrybut dyskretny

Źródło: opracowanie własne.



Rys. 9. Atrybut przedziałowy stale istniejący

Źródło: opracowanie własne.



Rys. 10. Atrybut przedziałowy okresowo istniejący

Źródło: opracowanie własne.

Podobny podział można wprowadzić ze względu na istniejące związki między encjami z punktu widzenia ich zmienności w czasie. Można wyróżnić związki:

- punktowe – implementujące związki między encjami w skutek zaistnienia konkretnego zdarzenia, np. zakupu określonego towaru przez klienta. Przy czym związek ten traktowany jest jako trwały. W świetle tego stwierdzenia klient nadal jest właścicielem nabytego przez siebie towaru (ewentualne zbycie przez klienta nabytego towaru nie jest w żaden sposób odnotowywane w systemie sprzedaży – z punktu widzenia sprzedażowej bazy danych to, co dalej dzieje się z zakupionym przez klienta towarem, jest zupełnie nieistotne);
- okresowe – implementujące związki między encjami, które zachodzą tylko w pewnych okresach. W odróżnieniu zatem od związków punktowych są związkami nietrwałymi. W tym przypadku rejestrowany jest moment rozpoczęcia i zakończenia trwania związku. Jako przykład można podać zakup i rejestrację pojazdu mechanicznego. Związek trwa (jest prawdziwy) aż do momentu, gdy pojawi się nowy wpis świadczący o zakończeniu danego związku, a będący odpowiedzią na wyrejestrowanie lub zbycie pojazdu.

2.3. Rodzaje temporalnych baz danych

Ze względu na wymiar wpływającego czasu temporalne bazy danych można podzielić na:

- historyczne bazy danych – ujmują czas rzeczywisty. Pozwalają odtwarzać przeszłe stany, które zaistniały w rzeczywistości. Są bazami o ciągle przyrastających danych. Dane gromadzone są z uwzględnieniem wartości początku i końca okresu, w którym przyjmowały określoną wartość. Bazy te pozwalają uzyskiwać informacje nt. stanu przeszłego i teraźniejszego, a także implementować gałęziowy model wpływu czasu;
- transakcyjne bazy danych – ujmują czas transakcyjny. Rejestrują czas, w którym dane znalazły się w bazie danych, niezależnie od czasu, w którym dane zdarzenie zaistniało w rzeczywistości. Są szczególnie przydatne tam, gdzie istotny jest fakt rejestracji określonych działań na bazie danych, tzn. kiedy to miało miejsce, kto przeprowadzał zmiany itd. Czas rzeczywisty jest mniej istotny. Pozwalają

odtworzyć historię stanów bazy danych. Czas, w którym nastąpiła zmiana stanu bazy danych, ujmowany jest jako czas transakcyjny. Rejestracja samego czasu transakcyjnego jest wystarczająca tylko wówczas, gdy czas transakcyjny jest praktycznie równy czasowi rzeczywistemu (system o natychmiastowej rejestracji zdarzeń) bądź też, gdy bardziej istotny jest fakt i moment pojawienia się danych w bazie (wystąpienia zdarzenia) od rejestrowanych wartości związanych z wystąpieniem danego zdarzenia w świecie realnym;

- bitemporalne bazy danych – uwzględniają czas rzeczywisty i transakcyjny. [Rozmus 2016]. Pozwalają uzyskać informacje nt. wartości atrybutów w danym momencie oraz dodatkowo wskazują, jaka wartość go opisywała w tym czasie. Przechowują pełną informację, co pozwala np. ustalić różnorakie rozbieżności między stanem bazy danych a wiedzą użytkowników. Pozwalają także łatwo identyfikować opóźnienia we wprowadzaniu danych. Wydawać by się mogło, iż w dobie systemów realizujących transakcje bazodanowe w trybie *online*, a więc *de facto* w czasie rzeczywistym, model ten wydaje się zbędny. Jednakże nie jest tak do końca. Wszędzie tam, gdzie konieczne jest składowanie w bazie wpisów dotyczących aktualizacji przeprowadzanych w odniesieniu do istniejących danych, niezbędna jest bitemporalna baza danych.

2.4. Postulaty szóstej postaci normalnej 6NF

C.J. Date zaproponował postulaty składające się na szóstą postać normalną 6NF [Date i in. 2003]. Zadaniem tej postaci miało być rozwiązanie anomalii mogących się pojawić podczas pracy z danymi tymczasowymi. Zgodnie z tą definicją baza danych jest w szóstej postaci normalnej 6NF wtedy i tylko wtedy, gdy nie zawiera żadnych nietrywialnych zależności złączeń. Zastosowanie uogólnionych ograniczeń i związków (wymagających zawierania się atrybutów interwału i równości dla pozostałych atrybutów) nie spowoduje powstania żadnych nietrywialnych zależności złączeń, skutkujących powstaniem niepożądanych krotek [Ben-Gan i in. 2009a; Ben-Gan i in. 2009b].

Rozwiązanie proponowane przez C.J. Date'a wprowadza dekompozycję poziomą oryginalnej encji do dwóch encji [Mahmood i in. 2010]. Jedna z nich stanowi rolę kontenera na dane częściowo tymczasowe (a więc dane, które są prawidłowe od pewnego momentu w czasie), natomiast druga na dane w pełni tymczasowe (dane, które są prawidłowe w określonym interwale czasu). Encja ulega podziałowi w taki sposób, że każdy atrybut temporalny, funkcjonalnie niezależny od pozostałych atrybutów, jest składowany w osobnej encji. Aczkolwiek w przypadku, gdy wartości kilku atrybutów zmieniają się jednocześnie, mogą być one składowane w tej samej encji logicznej.

A więc głównym celem szóstej postaci normalnej 6NF jest eliminacja wszystkich zależności złączeniowych, co z kolei wyraźnie powinno wpłynąć na poprawę obsługi zmian atrybutów temporalnych [Karwin 2012]. Postać ta jest używana

w zwinnej technice modelowania danych nazywanej modelowaniem zakotwiczeń (*anchor modeling*) [Anchor modeling 2017]. Jest to technika przystosowana do danych, które zmieniają się w czasie nie tylko pod względem samej zawartości, ale również ich struktury. Technika ta wprowadza określone konstrukcje: kotwicę, atrybut, wiązanie i węzeł. Każda z tych konstrukcji odwzorowuje określone aspekty modelowanej domeny [Rönnbäck i in. 2010]. Następnie stworzony model jest transformowany za pomocą określonych formalizmów (sformalizowanych reguł) na schemat bazy danych, którego postać powinna być zgodna z 6NF.

3. Modelowanie zmian w czasie

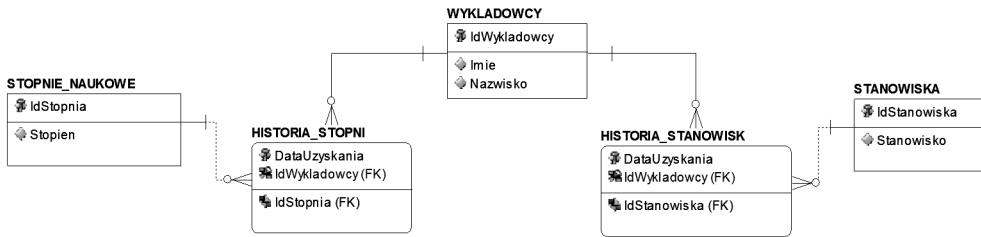
W zależności od kontekstu użycia zmiany w czasie dotyczyć mogą pojedynczych atrybutów bądź też grup atrybutów budujących daną encję, lecz także zmian związków zachodzących pomiędzy encjami, a w szczególnym przypadku także zmian całej encji (wszystkich jej atrybutów), również z uwzględnieniem zmian wszystkich związków, w których dana encja bierze udział. Każda z wymienionych zmian temporalnych ma swój własny sposób realizacji na poziomie modelu logicznego [Łacheciński 2012]. Sposób ich uwzględnienia w modelu logicznym prezentowany jest w kolejnych podrozdziałach.

3.1. Modelowanie temporalności atrybutu

Modelowanie zmian wartości atrybutu w czasie polega na wprowadzeniu dodatkowej encji dla każdego atrybutu temporalnego. Wyjątkiem może być umieszczenie grupy atrybutów temporalnych w jednej nowo powstałej encji, pod warunkiem że ich wartości zmieniają się w tym samym czasie, co zostało nadmienione w punkcie 2.4. Związek istniejący pomiędzy encją pierwotną a nowo powstałą encją temporalną jest związkiem o charakterze identyfikującym. W skład klucza podstawowego dodatkowo jest włączany atrybut przechowujący wartości typu data i czas przechowujący konkretną wartość daty – w przypadku użycia zegara fizycznego, lub też okres – w przypadku użycia zegara logicznego.

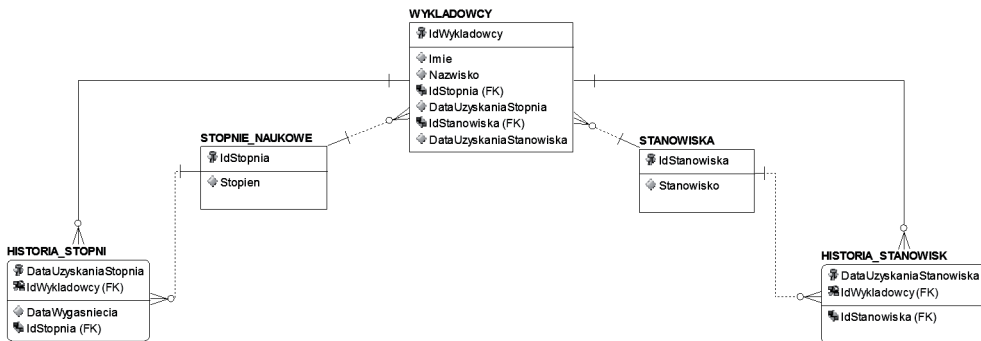
Na rysunku 11 zaprezentowano model temporalny przechowujący historię awansu pracowników naukowych uczelni z wyszczególnieniem posiadanych przez nich stopni naukowych oraz stanowisk. Wartości te zostały rozdzielone do dwóch encji temporalnych ze względu na to, iż zdobycie kolejnego stopnia naukowego nie jest jednoznaczne z jednoczesną zmianą stanowiska pracy (zazwyczaj są to zdarzenia niezależne i oddalone w czasie od siebie) [Łacheciński 2012].

Na rysunku 12 przedstawiono zmodyfikowaną postać poprzedniego rozwiązania. W encji WYKLADOWCY umieszczono dodatkowo atrybuty temporalne. Dzięki tej modyfikacji możliwe jest osobne składowanie aktualnych wartości atrybutu od ich wartości historycznych. W przypadku zaistnienia pierwszych wartości encje



Rys. 11. Zmiana atrybutu w czasie – zastosowanie zegara fizycznego

Źródło: opracowanie własne.



Rys. 12. Zmiana atrybutu w czasie – zastosowanie zegara fizycznego.

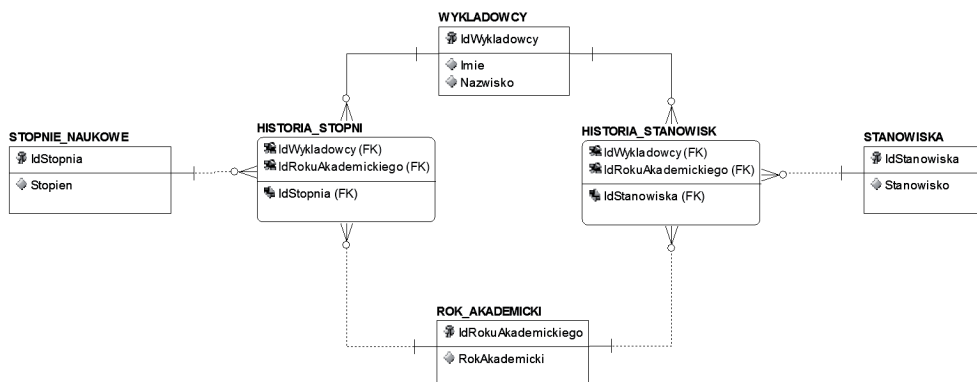
Oddzielne składowanie aktualnej wartości atrybutu

Źródło: opracowanie własne.

temporalne pozostają puste. Inicjowane są one wartościami w momencie pojawienia się nowych, aktualnych wartości [Matysiak 2017]. Oczywiście wiąże się to z dodatkową implementacją mechanizmu przenoszącego wartości atrybutów temporalnych z encji WYKLADOWCY do encji temporalnych w momencie pojawienia się nowych aktualnych wartości.

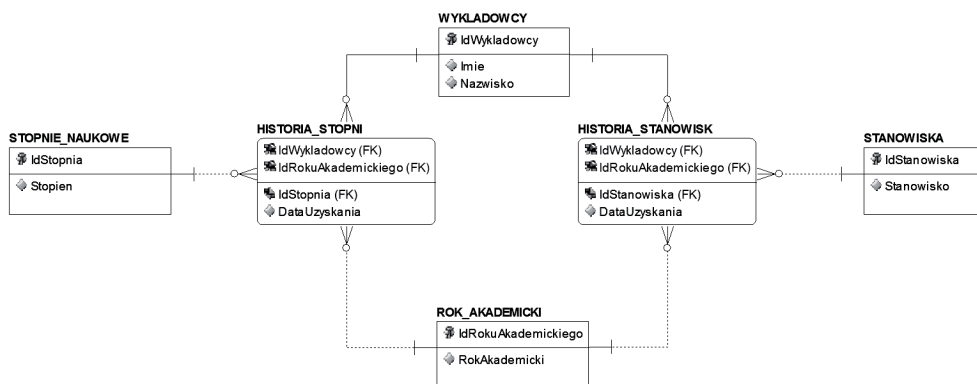
Na rysunku 13 przedstawiono zmodyfikowaną postać rozwiązania z rys. 11 z uwzględnieniem zegara logicznego. Rozwiązanie to jest użyteczne w przypadkach, gdy w tym samym przedziale czasu zmianie ulegają wartości atrybutów temporalnych w wielu krotkach.

Na rysunku 14 przedstawiono zmodyfikowaną postać rozwiązania z rys. 13. Dodatkowo oprócz wprowadzenia zegara logicznego (uwzględniającego zadany czasookres) rejestracji podlegają rzeczywiste, konkretne daty uzyskania danego stopnia naukowego czy też stanowiska, które zawierają się w danym czasookresie. Tym samym uwzględniono dodatkowe składowanie danych w bardziej szczegółowym ujęciu. Najczęściej podejście takie wynika z określonych reguł biznesowych.



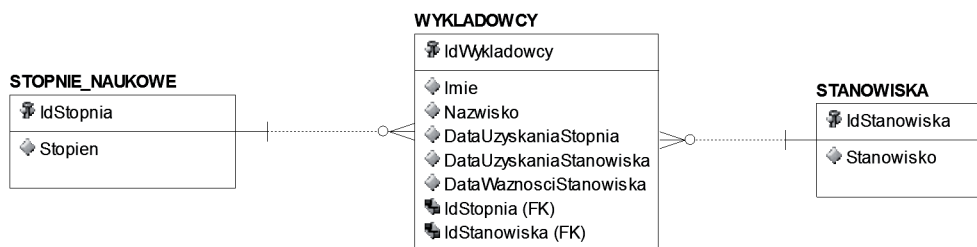
Rys. 13. Zmiana atrybutu w czasie – zastosowanie zegara logicznego

Źródło: opracowanie własne.



Rys. 14. Zmiana atrybutu w czasie – zastosowanie zegara logicznego z dodatkowym uwzględnieniem konkretnej daty uzyskania danego stopnia i stanowiska

Źródło: opracowanie własne.



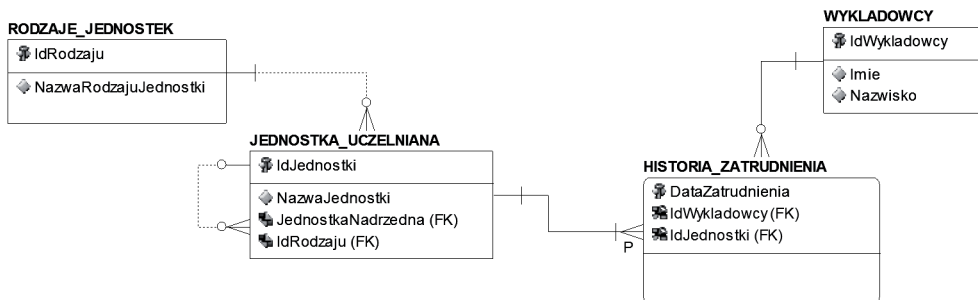
Rys. 15. Zmiana atrybutu w czasie – zastosowanie zegara fizycznego. Składowanie tylko aktualnej wartości atrybutu z uwzględnieniem czasu jego ważności (czasu obowiązywania)

Źródło: opracowanie własne.

Na rysunku 15 przedstawiony jest przypadek, w którym nieistotne są poprzednie wartości atrybutów, ważny zaś jest okres ważności tych atrybutów (o jedynie aktualnej wartości).

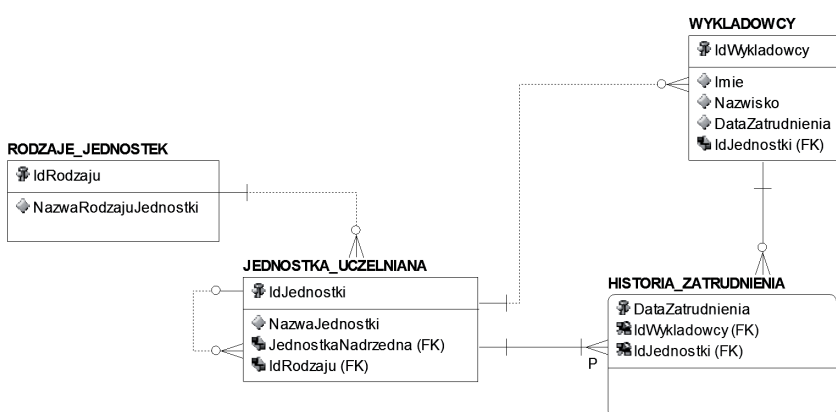
3.2. Modelowanie temporalności związku

W celu odwzorowania zmian związków zachodzących pomiędzy encjami w czasie wprowadza się do modelu danych dodatkową encję temporalną, która stanowi odpowiednik encji asocjacyjnej. Klucz podstawowy encji temporalnej modelującej zmienność związku w czasie stanowią kopie kluczy podstawowych z encji nadrzędnych uczestniczących w tym związku wraz z atrybutem przechowującym wartość typu data i czas (a więc wartość czasu, kiedy nastąpiła zmiana związku) [Łacheciński, s. 2012].



Rys. 16. Zmiana związku w czasie – zastosowanie zegara fizycznego

Źródło: opracowanie własne.



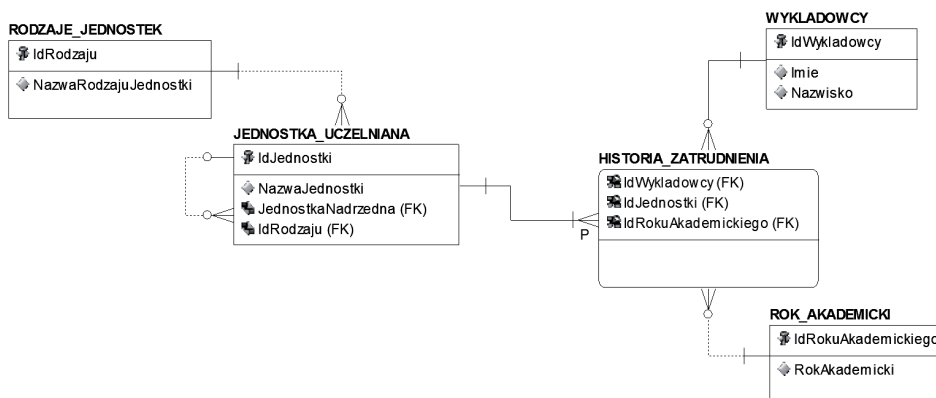
Rys. 17. Zmiana związku w czasie – zastosowanie zegara fizycznego.

Oddzielne składowanie aktualnego związku

Źródło: opracowanie własne.

Na rysunku 16 przedstawiono model temporalny przechowujący historię zatrudnienia pracowników w poszczególnych jednostkach w kolejnych latach ich zatrudnienia z uwzględnieniem zegara fizycznego.

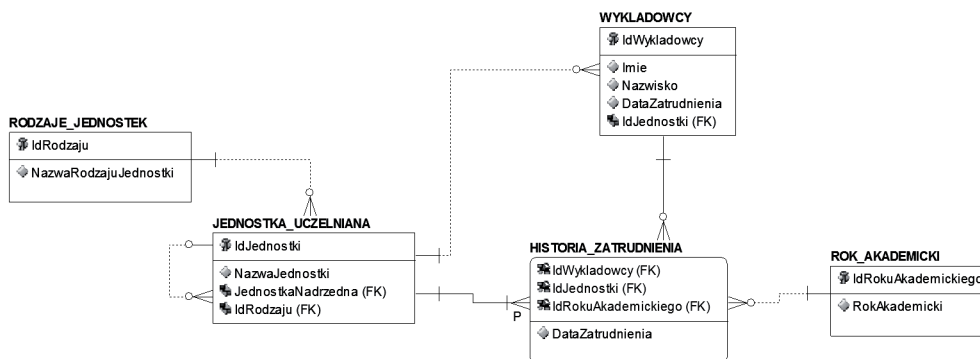
Na rysunku 17 przedstawiono zmodyfikowaną postać rozwiązania z rysunku 16. Uwzględniono w nim separowane składowanie związku aktualnego od jego poprzednich wartości. Encja temporalna pozostaje pusta aż do momentu pojawienia się drugiej, nowej aktualnej wartości. Wówczas poprzednia aktualna wartość jest przenoszona do encji temporalnej HISTORIA_ZATRUDNIENIA.



Rys. 18. Zmiana związku w czasie – zastosowanie zegara logicznego

Źródło: opracowanie własne.

Na rysunku 18 przedstawiono zmodyfikowany przykład z rys. 16. Modyfikacja dotyczy zastąpienia zegara fizycznego zegarem logicznym. Sens wprowadzenia tej zmiany jest wówczas, gdy bardziej istotna jest informacja, w jakim roku akademickim



Rys. 19. Zmiana związku w czasie – zastosowanie zegara logicznego, z uwzględnieniem osobnego składowania związku oraz dodatkowym przechowaniem rzeczywistej (pełnej) daty zatrudnienia

Źródło: opracowanie własne.

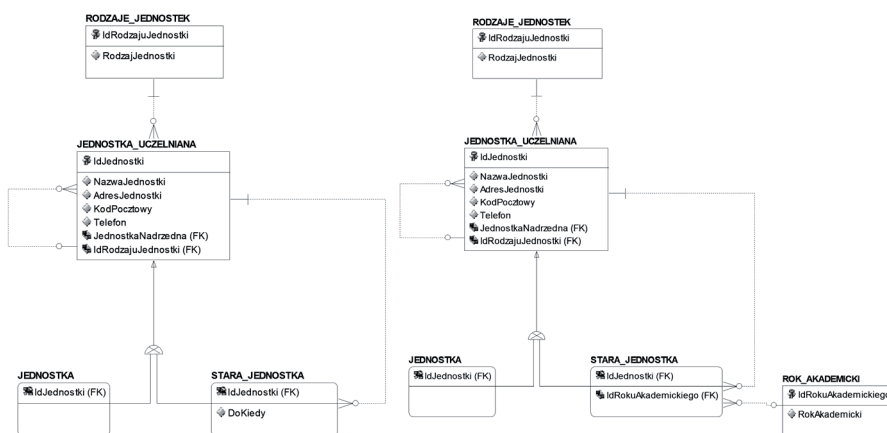
kim (okresie) nastąpiła zmiana zatrudnienia pracownika, od konkretnego dnia tego zdarzenia. Ponadto sprawdza się także w przypadkach, gdy w tym samym czasookresie istnieje wiele takich zdarzeń.

Na rysunku 19 zaprezentowano wariację rozwiązań z rys. 17 i 18. Rozwiązanie sprawdzi się tam, gdzie dodatkowo oprócz sezonowości zachodzących zmian potrzebne jest przechowanie dokładnej wartości wystąpienia zdarzenia.

3.3. Modelowanie temporalności encji

W określonych okolicznościach może zachodzić konieczność uwzględnienia zmian w czasie całej encji wraz ze wszystkimi jej atrybutami, często także z uwzględnieniem zmienności wraz z upływem czasu także związków, w których dana encja uczestniczy. Przy czym należy uwzględnić, oprócz jednoczesnej zmiany encji wraz ze związkami, możliwość niezależnej zmiany związków w czasie od zmian całej encji [Matysiak 2017].

Na rysunku 20 zaprezentowano temporalność samej encji bez uwzględnienia zmian związków. Przedstawione zostały dwa rozwiązania z użyciem zegara fizycznego i logicznego. Zamodelowane one zostały za pomocą związku kategorii encji (z uwzględnieniem podtypów-specjalizacji). Encja JEDNOSTKA_UCZELNIANA zawiera wystąpienia aktualne i historyczne encji. Zadaniem encji JEDNOSTKA jest składowanie aktualnego egzemplarza encji, wszystkie zaś przeszłe wystąpienia encji składowane są w encji STARA_JEDNOSTKA. Każde nowe wystąpienie encji JEDNOSTKA_UCZLENIANA jest rejestrowane jako wystąpienie encji podtypu JEDNOSTKA. Zamodelowany związek pomiędzy encjami STARA_JEDNOSTKA

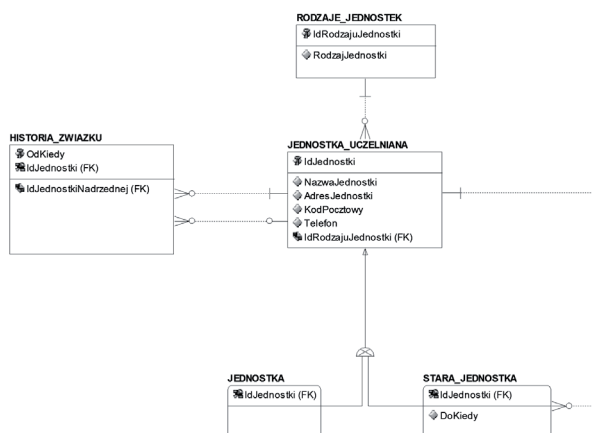


Rys. 20. Zmiana encji w czasie – zastosowanie zegara fizycznego i logicznego z uwzględnieniem osobnego składowania aktualnego wystąpienia encji

Źródło: opracowanie własne.

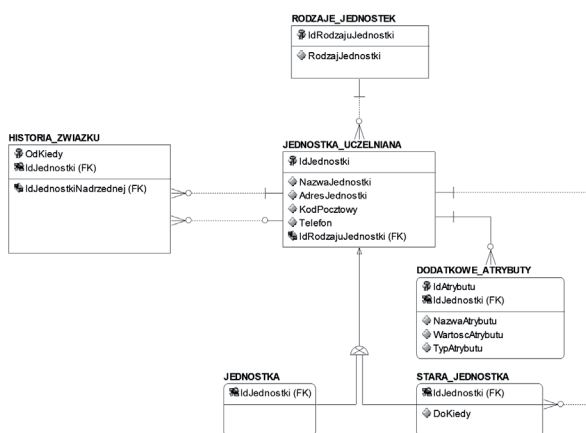
i JEDNOSTKA_UCZELNIANA odzwierciedla zależność mówiącą o tym, że starsza jednostka jest poprzednią wersją aktualnej jednostki. Należy zaznaczyć, iż prezentowane rozwiązanie dotyczy modelu logicznego. Odmienną kwestią jest sposób jego transformacji do modelu fizycznego i jego ostateczna postać.

Na rysunku 21 przedstawiono rozwiązanie zmiany w czasie całej encji z dodatkowym uwzględnieniem zmian związków. Prezentowany problem dotyczy przypadku, gdy dana jednostka wchodzi w skład zupełnie innej jednostki organizacyjnej. Związek rekurencyjny został przekształcony do związku sieciowego. Prezentowane



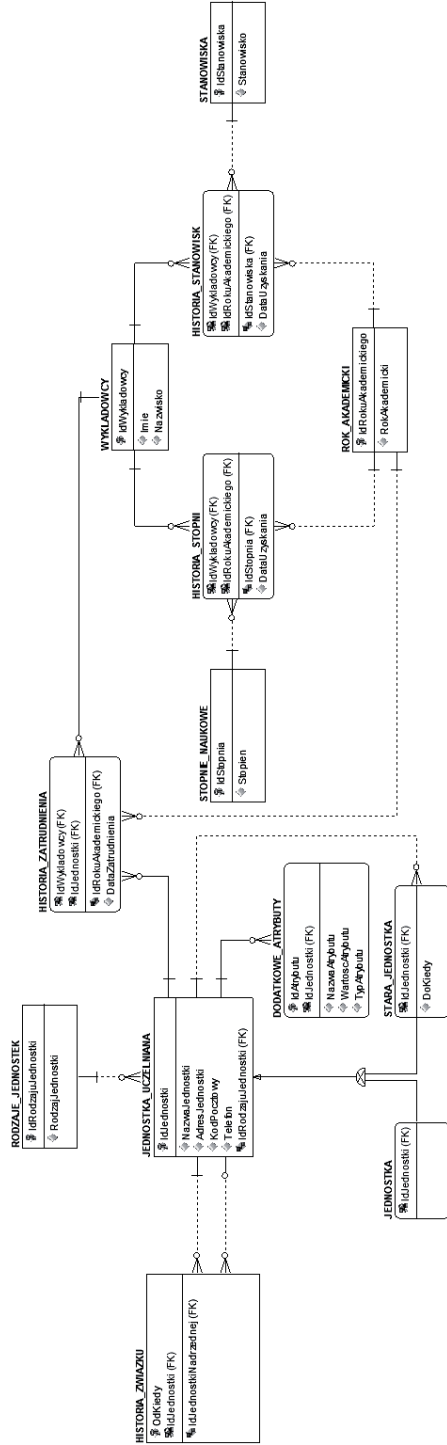
Rys. 21. Zmiana encji w czasie – zastosowanie zegara fizycznego z dodatkowym uwzględnieniem zmiany związku

Źródło: opracowanie własne.



Rys. 22. Zmiana encji w czasie – zastosowanie zegara fizycznego z dodatkowym uwzględnieniem zmiany związku oraz zmianą struktury samej encji

Źródło: opracowanie własne.



Rys. 23. Kompleksowe rozwiązanie uwzględniające temporalność atrybutów, związków, encji wraz ze związkami oraz struktury encji
 Źródło: opracowanie własne.

rozwiązanie pozwala również na niezależną zmianę samego związku bez jednoczesnej zmiany całej encji.

Na rysunku 22 dodatkowo uwzględniono możliwość zmiany w czasie, oprócz zmiany całej encji i związków, w których uczestniczy, również samej struktury encji. Rozwiązanie to zostało stworzone przez użycie wzorca encja-atrybut-wartość EAV. Funkcjonalność ta przydatna jest w sytuacji, gdy dana encja z upływem czasu wzbogaci się o nowe cechy, np. nieuwzględniony wcześniej adres e-mail i strony WWW.

Rysunek 23 stanowi rozwiązanie kompleksowe wszystkich dotychczasowych rozważań z uwzględnieniem temporalności atrybutów, związków, encji wraz ze związkami, jak również struktury samej encji.

4. Zakończenie

Znormalizowanie bazy danych do temporalnej postaci normalnej pozwala w znacznym stopniu ograniczyć redundancję danych, eliminuje anomalie związane z różnymi działaniami na danych temporalnych, a także pozwala na stosunkowo prosty odczyt historii zmian wartości poszczególnych atrybutów temporalnych. Ceną, jaką należy zapłacić za wspomniane zalety, jest dodatkowo zwiększona liczba encji, co może mieć pewien wpływ na efektywność realizacji zapytań, podobnie zresztą jak w przypadku standardowej normalizacji bazy danych.

Modele logiczne uwzględniające temporalność całych encji mogą być transformowane do modeli fizycznych na kilka różnych sposobów, podobnie jak w przypadku transformacji związku kategorii encji. Transformacja ta może sprowadzać model logiczny do pojedynczej tabeli zawierającej wszystkie pola, do dwóch osobnych tabel, gdzie w każdej z nich przechowywane są wspólne pola, bądź też do postaci, w której liczba tabel jest równa liczbie encji logicznych. Wybór rozwiązania powinien być poparty analizą konkretnego przypadku użycia. Należy wziąć pod uwagę m.in. liczbę atrybutów wspólnych, całkowitą liczbę atrybutów, sposób rozłożenia atrybutów czy też chociażby rząd wielkości egzemplarzy encji temporalnych.

Artykuł ten w całości został poświęcony zagadnieniom modelowania danych temporalnych z wykorzystaniem relacyjnego modelu danych. Aczkolwiek nie mniej istotna jest także sprawa użycia danych temporalnych, ich dodawanie do bazy, aktualizacja, przetwarzanie czy też wreszcie udostępnianie w postaci użytecznej informacji użytkownikom systemu informatycznego uwzględniającego dodatkową funkcjonalność w postaci modułu temporalnego. Co prawda uwzględnienie w trakcie modelowania normalizacji temporalnej pozwala używać atrybutów temporalnych do opisu modelowanej rzeczywistości w relacyjnym modelu danych, jednakże dane temporalne są danymi specyficznymi, a więc wymagają także specyficznego podejścia do ich obsługi w postaci dodatkowych, specjalnych operatorów i klauzul ułatwiających pracę z danymi temporalnymi [Morelos 2017]. W standardzie SQL 2011 wyraźnie polepszone obsługę danych temporalnych oraz wprowadzono wiele nowych rozwiązań, jak choćby tymczasowe klucze podstawowe czy też tymczasowe

więzy integralności referencyjnej i wiele innych [Kulkarni, Michels 2012; Zemke 2012]. Niemniej jednak dostęp do tych rozwiązań w komercyjnych SZBD w większości produktów pozostawia jeszcze wiele do życzenia. Jednocześnie należy zauważyć, że kilku producentów zaimplementowało w swoich flagowych produktach wybrane rozwiązania opracowane i zawarte w standardzie języka SQL 2011. Oczywiście różnie się to przedstawia u poszczególnych producentów. W różnym zakresie i stopniu zostały zaimplementowane wybrane funkcjonalności, nie zawsze w sposób w pełni zgodny z postacią ujętą w standardzie SQL. Do grona producentów, którzy oferują w swoich produktach obsługę danych temporalnych, zaliczyć można: IBM DB2 od wersji 10 [Saracco i in. 2012], ORACLE od wersji 10g/11g, choć obsługa danych temporalnych w zgodności ze standardem SQL 2011 dotyczy wersji 12c [Oracle 2017; Salvisberg 2014], Teradata [Snodgrass 2010; Teradata 2017], MS SQL Server od wersji 2016 [SQL Server 2017].

Wymienione zagadnienie stanowi kolejny etap badawczy, wyniki badań zaś będą zaprezentowane w kolejnych publikacjach. Tym samym artykuł ten stanowi początek prac, a zarazem punkt wyjścia do dalszych poszukiwań i prac badawczych. Jego kontynuacją będzie prezentacja wybranych części standardu SQL pod kątem obsługi danych temporalnych, również z uwzględnieniem rozwiązań temporalnych baz danych co do składni Temporal SQL, oraz implementacji tego standardu w wybranych komercyjnych rozwiązaniach dostępnych na rynku.

Literatura

- Anchor modeling, <http://www.anchor modeling.com/> (1.12.2017).
- Ben-Gan I. i in., 2009b, *Microsoft SQL Server 2008 od środka: Zapytania w języku T-SQL*, APN Promise, Warszawa.
- Ben-Gan I. i in., 2009a, *Microsoft SQL Server 2008 od środka: Programowanie w języku T-SQL*, APN Promise, Warszawa.
- Date C.J., Darwen H., Lorentzos N., 2003, *Temporal Data and the Relational Model*, Morgan Kaufman.
- Fuller A., 2007, *Database Design: A Point in Time Architecture*, <https://www.red-gate.com/simple-talk/sql/database-administration/database-design-a-point-in-time-architecture/> (1.12.2017).
- Hajnicz E., 1996, *Reprezentacja logiczna wiedzy zmieniającej się w czasie*, Akademicka Oficyna Wydawnicza PLJ, Warszawa.
- Kania K., 2004, *Temporalne bazy danych w systemach informatycznych zarządzania*, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice.
- Karwin B., 2012, *Antywzorce języka SQL. Jak uniknąć pułapek podczas programowania baz danych*, Helion, Gliwice.
- Kulkarni K., Michels J.-E., 2012, *Temporal features in SQL:2011*, <https://sigmodrecord.org/publications/sigmodRecord/1209/pdfs/07.industry.kulkarni.pdf> (1.12.2017).
- Łacheciński S., 2012, *Tworzenie relacyjnych modeli danych*, Polskie Stowarzyszenie Zarządzania Wiedzą-Studia i Materiały nr 62, Bydgoszcz, s. 42-55.
- Mach-Król M., 2007, *Temporalna analiza otoczenia przedsiębiorstwa: techniki i narzędzia inteligentne*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, Seria: Monografie i Opracowania nr 100, Wrocław.

- Mach-Król M., 2014, *Ontologia czasu nieliniowego dla opisu rzeczywistości ekonomicznej*, Informatyka Ekonomiczna 1(31), Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- Mach-Król M., Michalik K., 2014, *Selected Aspects of Temporal Knowledge Engineering*, Ganzha M., Maciaszek L. A., Paprzycki M. (red.), [in:] Proc. Federated Conference on Computer Science and Information Systems – FedCSIS'2014.
- Mahmood N., Burney A., Ahsan K., 2010, *A Logical Temporal Relational Data Models*, <https://www.ijcsi.org/papers/7-1-1-1-9.pdf> (1.12.2017).
- Matysiak M., *Modelowanie wersji danych*, http://www.ploug.org.pl/?page_id=1072 (1.12.2017).
- Michalik K., Mach-Król M., 2014, *Logika temporalna w koncepcji baz wiedzy systemu Logos do reprezentacji wiedzy w domenie prawniczej*, Roczniki Kolegium Analiz Ekonomicznych SGH, nr 33, Warszawa.
- Morelos A., Holmes A., *Temporal Databases: Why you should care and how to get started*, <https://www.datasciencecentral.com/profiles/blogs/temporal-databases-why-you-should-care-and-how-to-get-started> (1.12.2017).
- Oracle: <http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/12c/r1/ilm/temporal/temporal.html> (1.12.2017).
- Rönnbäck L., Regardt O., Bergholtz M., Johannesson P., Wohed P., 2010, *Agile Information Modeling in Evolving Data Environments*, <http://www.anchor modeling.com/wp-content/uploads/2011/05/Anchor-Modeling.pdf> (1.12.2017).
- Rozmus S., 2016, *Projektowanie baz danych z pełną historią zmian danych – model bitemporalnej bazy danych i operacje zapisu*, Biuletyn WAT, vol. LXV, nr 1.
- Salvisberg P., *Multi-temporal Features in Oracle 12c*, <https://www.salvis.com/blog/2014/01/04/multi-temporal-database-features-in-oracle-12c/>.
- Saracco M., Nicola M., Gandhi L., 2012, *A matter of time: Temporal data management in DB210*, <https://www.ibm.com/developerworks/data/library/techarticle/dm-1204db2temporaldata/dm-1204db2temporaldata-pdf.pdf> (1.12.2017).
- Snodgrass R.T., 2010, *A Case Study of Temporal Data*, Teradata Corporation, <http://www2.cs.arizona.edu/people/rts/pubs/Teradatacasestudy.pdf> (1.12.2017).
- SQL Server: <https://docs.microsoft.com/en-us/sql/relational-databases/tables/temporal-tables> (1.12.2017).
- Teradata: https://www.info.teradata.com/HTMLPubs/DB_TTU_16_00/index.html#page/SQL_Reference%2FB035-1182-160K%2Fyxa1472240621730.html%23 (1.12.2017).
- Zemke F., 2012, *What's new in SQL:2011*, SIGMOD Record, vol. 41, no. 1, March pp. 67-73, <https://sigmodrecord.org/publications/sigmodRecord/1203/pdfs/10.industry.zemke.pdf> (1.12.2017).