Maciej Kiedrowicz[1], Jarosław Koszela[2],
Paulina Szczepańczyk-Wysocka[3]

Wykrywanie przestępstw finansowych i przeciwdziałanie im z wykorzystaniem sieciowych baz danych – system IAFEC

1. Wstęp

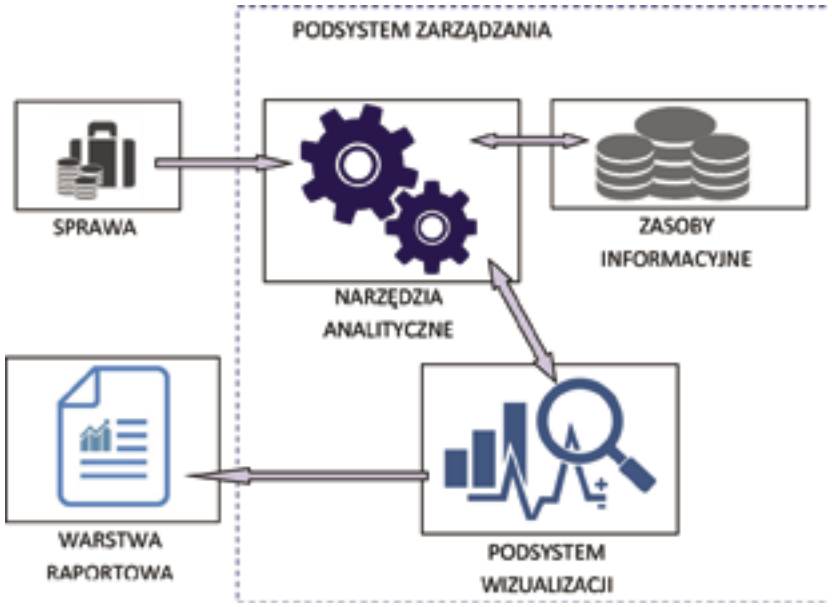
Każdego roku państwo odnotowuje coraz wyższe straty w budżecie wynikające z przestępstw finansowych, a w szczególności z „prania brudnych pieniędzy". Funkcjonujący w Polsce model wykrywania i przeciwdziałania takiej działalności ma charakter rozproszony, gdyż opiera się na działaniach wielu niezależnych organów i instytucji. Ponadto rozwój technologii utrudnia skuteczne wykrywanie przestępczości finansowej, a stosowane dotychczas metody opierają się na analizie manualnej. Szansę na zwiększenie skuteczności wykrywania prania pieniędzy stanowi automatyzacja procesów pozyskiwania danych oraz ich analizy. Przedstawiony problem stał się podstawą podjęcia prac koncepcyjnych nad narzędziem wspomagającym pracę analityków, a w konsekwencji – opracowania systemu IAFEC (ang. *Information Analysis of Financial and Economic Crime*), przeznaczonego do analizy danych pod kątem wykrywania przestępstw finansowych i przeciwdziałania im.

2. Koncepcja systemu IAFEC

Działanie systemu oparte jest na mechanizmie spraw, które są głównym obiektem w systemie. Do nich dołączane są wszystkie dane, a także przeprowadzane są analizy. W ramach pracy nad sprawą analityk pobiera dane z systemów

[1] Wojskowa Akademia Techniczna w Warszawie, Wydział Cybernetyki.
[2] Wojskowa Akademia Techniczna w Warszawie, Wydział Cybernetyki.
[3] Wojskowa Akademia Techniczna w Warszawie, Wydział Cybernetyki.

zewnętrznych, a następnie dokonuje ich analizy za pomocą zaimplementowanych w systemie specjalistycznych narzędzi, których raport interpretuje. System umożliwia też skierowanie do zasobów informacyjnych zapytania o dodatkowe dane, a także powtórzenie całego procesu. Gdy dane i analizy są kompletne, następuje ich agregacja oraz utworzenie raportu końcowego. Dalsze postępowanie prowadzone jest poza systemem IAFEC.



Rysunek nr 1. Koncepcja systemu IAFEC

Źródło: opracowanie własne

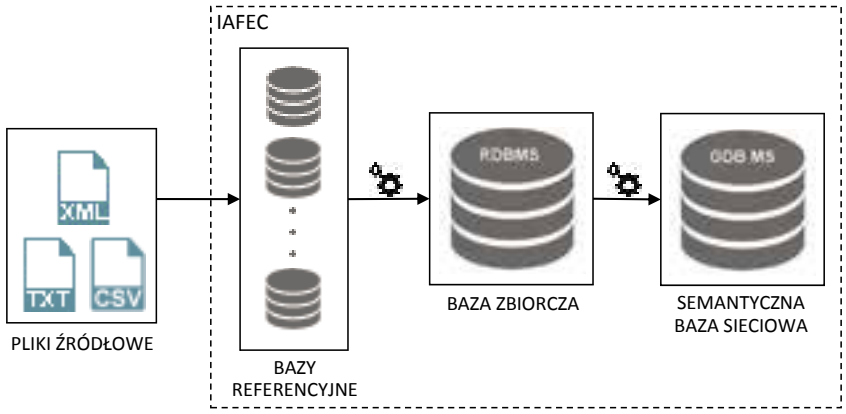
Architektura systemu IAFEC oparta jest na koncepcji SOA (ang. *Service-Oriented Architecture*) i bazuje na niezależnych komponentach, dzięki czemu system jest elastyczny i może być rozbudowywany i dostosowany do aktualnych potrzeb.

Moduł główny systemu składa się z wielu podmodułów, m.in. podsystemu zarządzania, wizualizacji, narzędzi analitycznych oraz zasobów informacyjnych. Dzięki zastosowanym programowym mechanizmom refleksji, moduł główny wykrywa dostępne komponenty i udostępnia je. Taka budowa umożliwia w szczególności rozbudowywanie systemu o kolejne narzędzia analityczne, stanowiące zbiór programów implementujących różnorodne algorytmy i sposoby analizy. Każde z narzędzi musi implementować interfejsy systemu, umożliwiające ich obsługę z panelu głównego, a także obsługiwać systemowe zasoby informacyjne (istnieje możliwość rozbudowy modelu danych, lecz bez jego modyfikacji).

3. Zasoby informacyjne systemu IAFEC

Zasoby informacyjne składają się z baz referencyjnych, ujednoliconej bazy zbiorczej oraz semantycznej bazy sieciowej, powiązanych procesami ETL (ang. *Extract*, *Transform and Load*). Dane do zasobów ładowane są za pomocą modułu importu poprzez bazy referencyjne.

Pozyskiwanie danych jest niezbędne do przeprowadzenia analizy oraz wykrywania i przeciwdziałania przestępczości finansowej, w związku z czym stanowi jedno z podstawowych zadań systemu IAFEC. Dane gromadzone w systemie to przede wszystkim informacje o osobach fizycznych i prawnych, a także o ich zobowiązaniach czy środkach finansowych, m.in. o udziałach i papierach wartościowych, kredytach, podatkach, przepływach na rachunkach bankowych. Dane te mogą zostać pozyskane z wielu źródeł informacyjnych, m.in. z systemów obsługiwanych przez organy administracji publicznej, rejestrów gromadzących dane zgodnie z obowiązującymi przepisami, a także pochodzić z akt operacyjnych sprawy. W zależności od możliwości, uwzględnia się dwie metody pozyskiwania danych – w trybie online oraz w trybie offline.



Rysunek nr 2. Schemat działania modułu importu danych

Źródło: opracowanie własne

Jeśli możliwy jest bezpośredni dostęp do danych, stosuje się metodę online pozyskiwania danych poprzez interfejsy do systemów zewnętrznych. Analityk wysyła żądania do zasobów zewnętrznych, inicjując tym samym proces ETL. Żądanie jest przetwarzane, następnie dane są wybierane z systemu źródłowego, transformowane i przesyłane do systemu docelowego w akceptowalnej przez

niego formie. Istnieje możliwość uruchamiania ładowania danych zarówno w trybie na żądanie, jak i zgodnie z ustalonym harmonogramem.

W przypadku ograniczeń w dostępie do zasobów źródłowych możliwe jest pozyskanie danych poprzez pliki wynikowe (np. XML, CSV). Proces ten wymaga uczestnictwa osób odpowiedzialnych za zasoby źródłowe, co powodować może błędy oraz opóźnienia w prowadzonych sprawach.

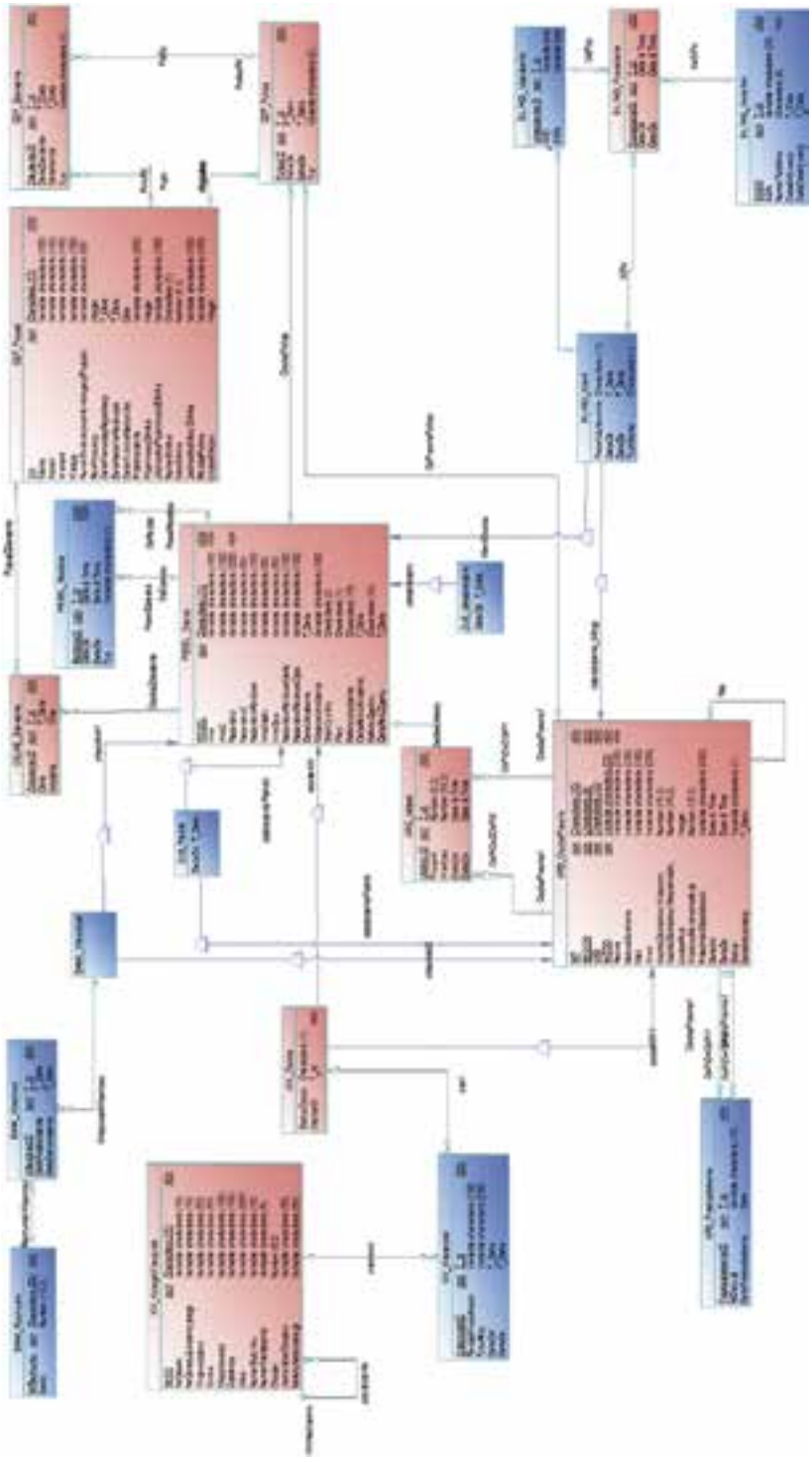
4. Relacyjny model danych systemu IAFEC

Jak przedstawiono na rysunku nr 2, pliki źródłowe, po uzyskaniu ich z systemów zewnętrznych (zarówno w trybie online, jak i offline), są importowane do systemu IAFEC, a następnie dane z nich ładowane są do baz referencyjnych, które odzwierciedlają modele danych stosowane w systemach źródłowych.

Bazy referencyjne (inicjalne) stanowią: PESEL (dane osób oraz związków pokrewieństwa i powinowactwa), KRS/EDG/REGON (dane dotyczące prowadzonych działalności), ZUS (dane płatników i osób ubezpieczonych), CEPIK (dane pojazdów i kierowców), CELNE (dane dotyczące przekraczania granic kraju), KW (dane ksiąg wieczystych), BILINGI (rejestr połączeń), GIIF (rejestr transakcji bankowych). Po załadowaniu dane są transformowane (ujednolicane), a następnie przenoszone w sposób przyrostowy do relacyjnej bazy zbiorczej. Na jej podstawie następuje kolejna transformacja danych do modelu sieciowego (tworzony jest graf zależności i powiązań) oraz załadowanie danych do semantycznej sieciowej bazy danych.

Uogólniony model konceptualny danych przedstawiono na rysunku nr 3.

Przedstawiony model jest modelem uproszczonym. Implementacja modelu relacyjnego danych, wymaganych przez zastosowane metody analizy, jest uszczegółowiona, a także rozszerzona o struktury wspomagające narzędzia analityczne oraz warstwę raportową systemu.



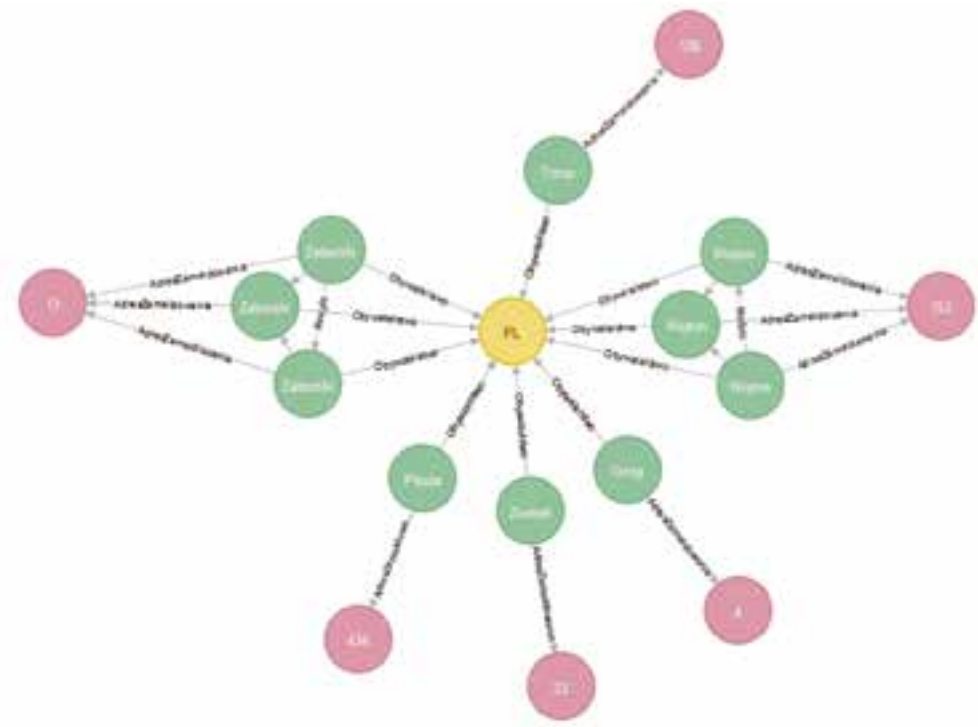
Rysunek nr 3. Konceptualny model danych IAFEC

Źródło: opracowanie własne

5. Sieciowa baza danych systemu IAFEC

Analizy dokonywane w systemie IAFEC w dużej mierze opierają się na mechanizmach poszukiwania dróg i powiązań pomiędzy obiektami w bazie danych. Stosowanie tego typu analiz dla relacyjnych modeli danych jest jednak nieefektywne. W celu zapewnienia wysokiej efektywności przetwarzania danych o dużej złożoności powiązań zaleca się wykorzystanie baz danych o modelu sieciowym. Implementację sieciowej bazy danych w systemie IAFEC wykonano za pomocą systemu zarządzania bazą danych Neo4j.

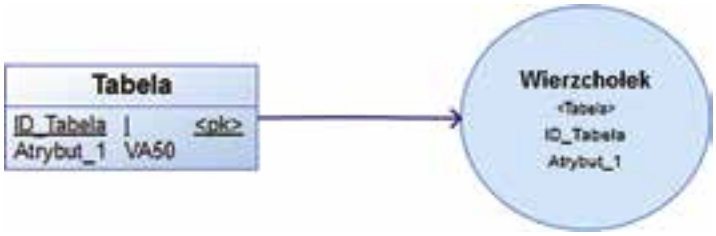
Sieciową bazę danych tworzą dwa rodzaje elementów – encje (węzły) oraz powiązania między nimi (łuki), przy czym każda encja odwzorowuje pojedynczy byt, natomiast łuki przedstawiają zależności pomiędzy poszczególnymi bytami. Rozróżnienie bytów jest możliwe dzięki zastosowaniu etykiet oraz typów. W systemie IAFEC są to np.: osoba fizyczna, adres, pojazd, rodzicielstwo, posiadanie pojazdu.



Rysunek nr 4. Graficzna reprezentacja sieciowej bazy danych systemu IAFEC

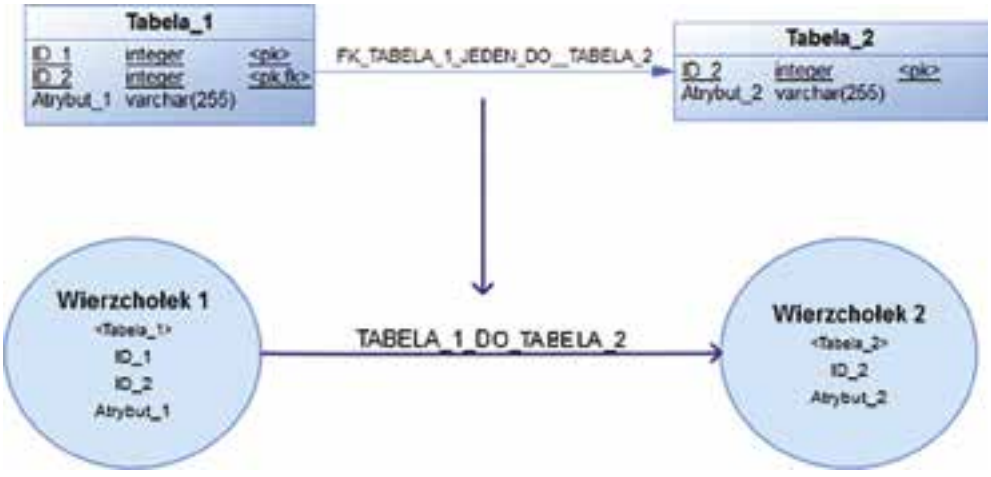
Źródło: opracowanie własne

Implementacja sieciowego modelu danych dla systemu IAFEC wykonana została w oparciu o opracowane wzorce transformacji z modelu relacyjnego zbiorczej bazy danych. Wzorce te opisują sposób budowania modelu sieciowego za pomocą zestawienia liczby związków wychodzących z tabeli oraz referencji z innych tabel. Opracowane na potrzeby systemu wzorce transformacji przedstawiono na rysunkach nr 5–11.



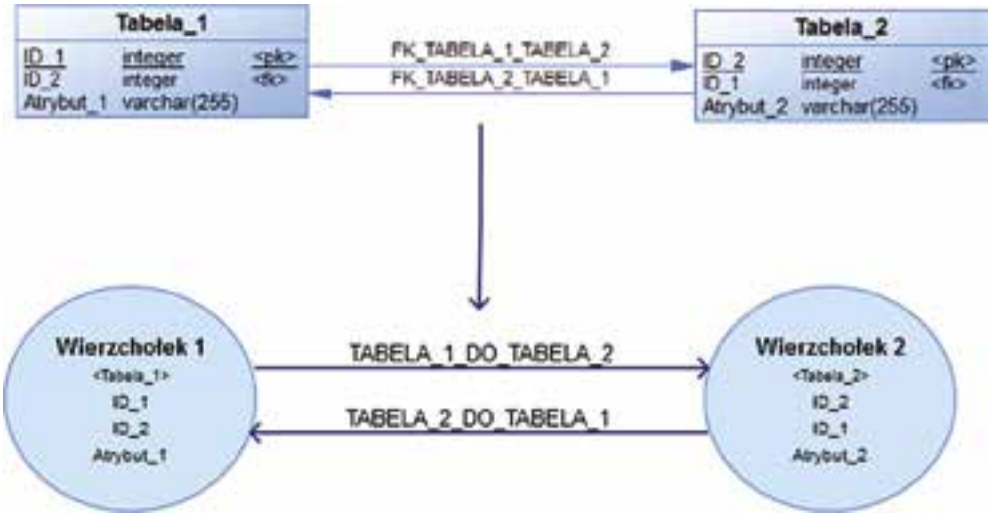
Rysunek nr 5. Przykład wzorca transformacji modelu relacyjnego na sieciowy dla przypadku, gdy tabela nie posiada żadnych powiązań

Źródło: opracowanie własne



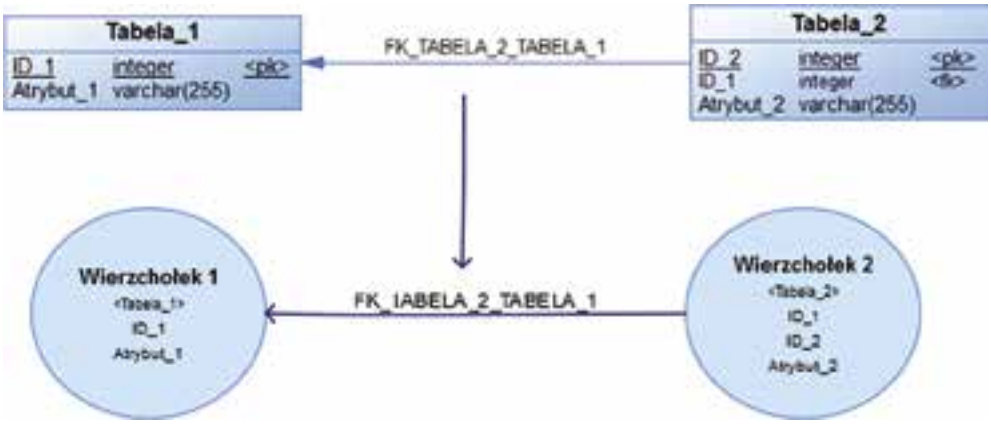
Rysunek nr 6. Przykład wzorca transformacji modelu relacyjnego na sieciowy dla przypadku, gdy dla Tabela_1 istnieje dokładnie jedna referencja wskazująca z innej tabeli (Tabela_2)

Źródło: opracowanie własne



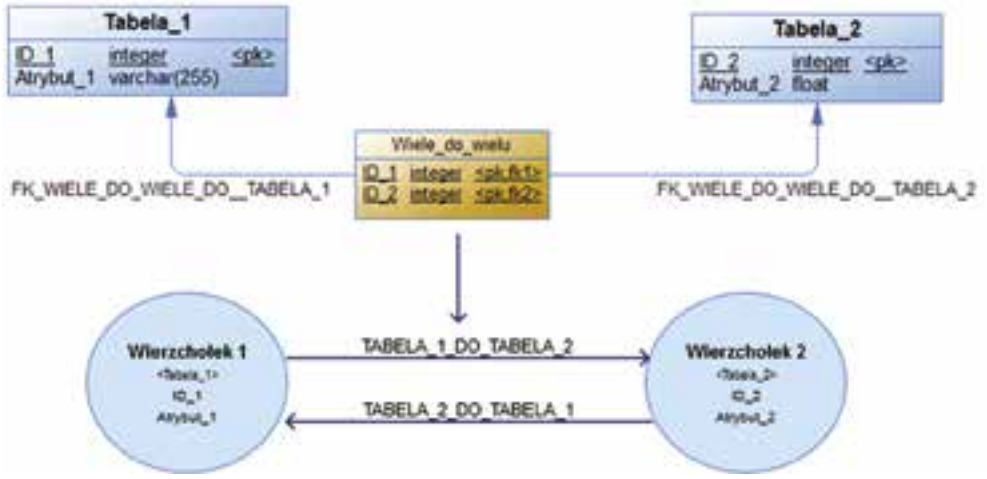
Rysunek nr 7. Przykład wzorca transformacji modelu relacyjnego na sieciowy dla przypadku, gdy Tabela_1 posiada dokładnie jedną referencję wychodzącą do innej tabeli (Tabela_2)

Źródło: opracowanie własne



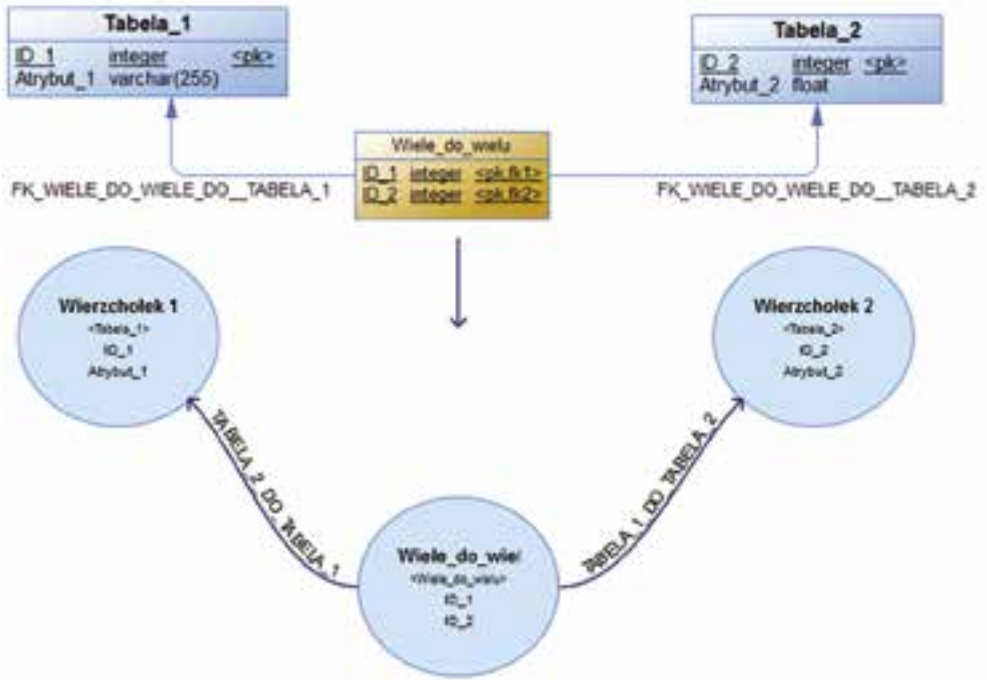
Rysunek nr 8. Przykład wzorca transformacji modelu relacyjnego na sieciowy dla przypadku, gdy istnieje dokładnie jeden związek wychodzący z tabeli, a także dokładnie jedno odwołanie z innej tabeli

Źródło: opracowanie własne



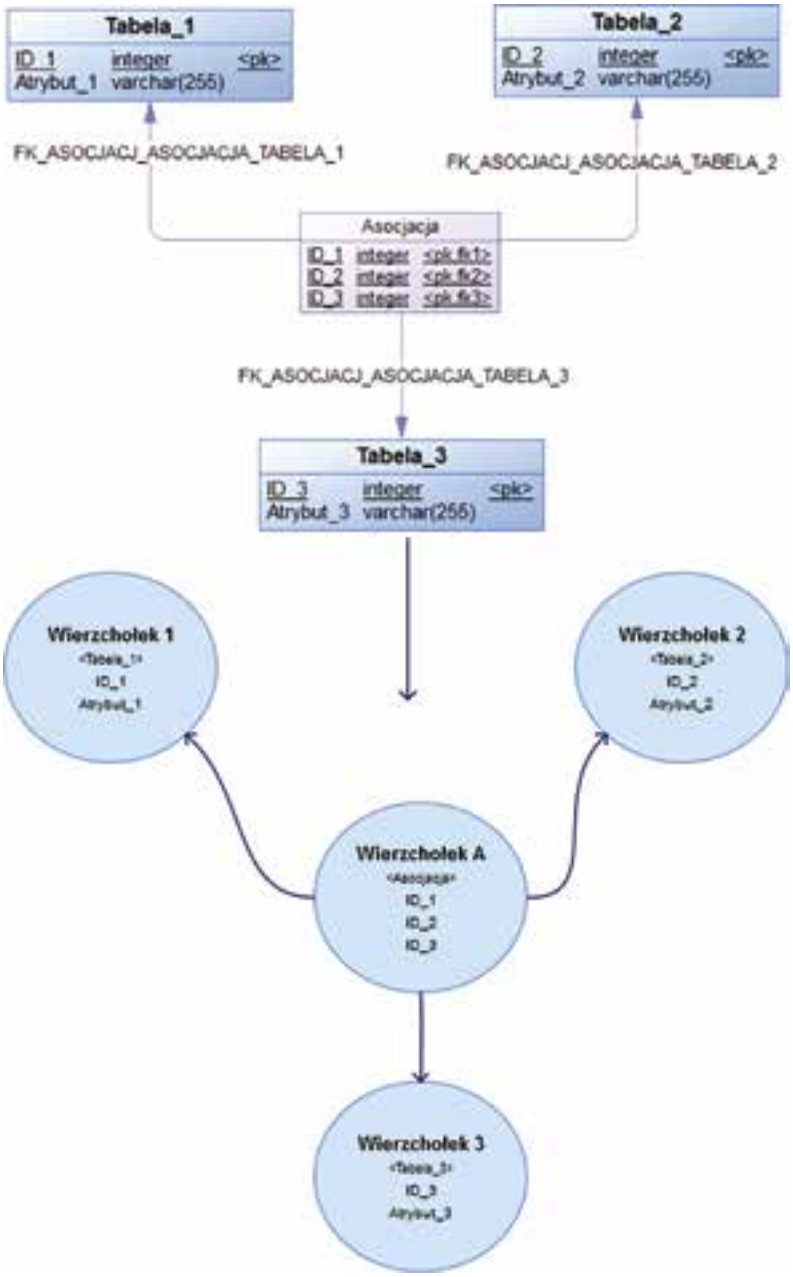
Rysunek nr 9. Przykład wzorca transformacji modelu relacyjnego na sieciowy dla przypadku, gdy tabela intersekcji posiada dokładnie dwa związki wychodzące do innych tabel i zostanie przekształcona w dwa łuki

Źródło: opracowanie własne



Rysunek nr 10. Przykład wzorca transformacji modelu relacyjnego na sieciowy dla przypadku, gdy tabela intersekcji posiada dokładnie dwa związki wychodzące do innych tabel i zostanie przekształcona w węzeł

Źródło: opracowanie własne



Rysunek nr 11. Przykład wzorca transformacji modelu relacyjnego na sieciowy dla przypadku, gdy tabela intersekcji posiada więcej niż dwa odwołania z innych tabel

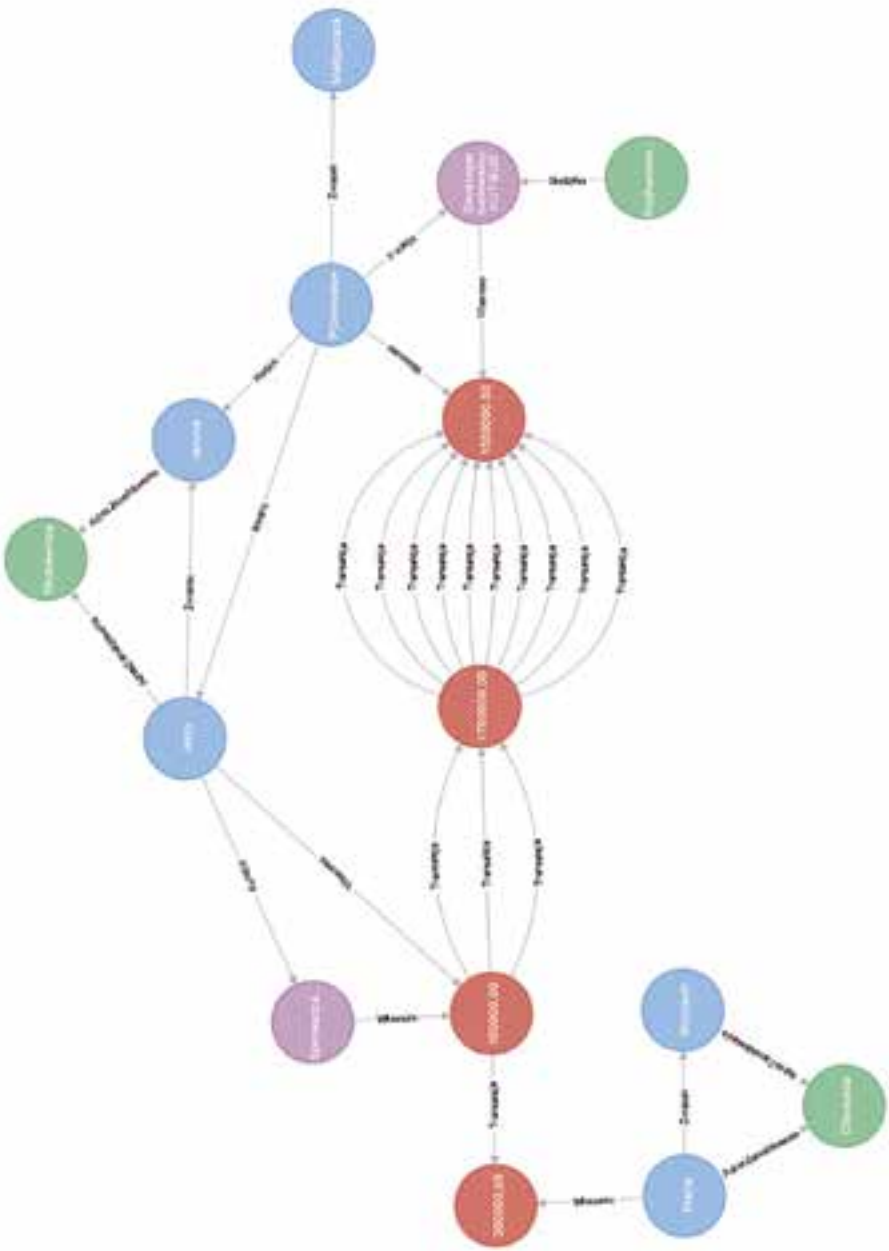
Źródło: opracowanie własne

Istotny dla procesu transformacji danych pozostaje fakt, iż sieciowa baza danych – w przeciwieństwie do bazy relacyjnej – należy do grupy baz bezschematowych. Oznacza to, że nie wymaga się utworzenia modelu danych (schematu) przed implementacją bazy danych, lecz jest on tworzony dynamicznie w trakcie gromadzenia danych w bazie. Rozwiązanie to zapewnia wysoką elastyczność baz danych o modelu sieciowym, utrudniając jednocześnie utrzymanie integralności oraz spójności danych.

Dodatkowym atutem wykorzystania środowiska Neo4j jest stosowany w nim deklaratywny język zapytań – Cypher Query Language (CQL). Umożliwia on efektywne przetwarzanie danych opartych na sieciowych modelach danych. Przykładowe zapytania używane w systemie przedstawiono poniżej:

- Dla dwóch numerów PESEL odszukać „odległość rodzinną" pomiędzy nimi:
  MATCH (n1: OsobaFizyczna {PESEL: *'pesel1'*}), (n2:OsobaFizyczna {PESEL: *'pesel2'*}), p=shortestPath((n1)-[r:Rodzic| Zwiazek *]-(n2)) RETURN p;
- Dla podanego numeru PESEL znaleźć osoby spokrewnione „odległe rodzinnie" nie dalej niż *maxOdleglosc*
  MATCH (n1: OsobaFizyczna {PESEL: *'pesel'*}) – [r:Rodzic| Zwiazek *..*maxOdleglosc*]-(n2:OsobaFizyczna) RETURN r.

Wykorzystanie języka CQL w systemie pozwala na efektywne wyszukiwanie dróg i zależności wśród zgromadzonych danych. Ma to szczególne znaczenie w wykrywaniu przestępstw finansowych. Narzędzia analizy przeszukują zgromadzone dane w celu rozpoznania wzorców mogących świadczyć o praniu pieniędzy. Na rysunku nr 12 przedstawiono przykładowy przypadek testowy wykrycia podejrzanych transakcji.



Rysunek nr 12. Wizualizacja przykładowych danych świadczących o „praniu pieniędzy”
Źródło: opracowanie własne

6. Podsumowanie

Przedstawione mechanizmy są tylko jednym ze sposobów zastosowania nowoczesnych technologii do wykrywania przestępczości finansowej. System IAFEC może stanowić skuteczne narzędzie analiz dla organów ścigania. Zastosowanie sieciowych baz danych pozwoliło na uzyskanie wysokiej efektywności przetwarzania sieci i grafów obrazujących przetwarzane dane. Jednak ze względu na ciągły rozwój technologii, a także zmieniające się metody działania sprawców zaleca się dalszy rozwój systemu. Jednym z dalszych kierunków rozwoju, ze względu na charakter przechowywanych danych, może być implementacja w systemie mechanizmów temporalności. Rozwiązanie to pozwoliłoby zarówno na odwzorowanie zmian danych w czasie, jak i rozważenie w analizie zdarzeń domniemanych (alternatywnych).

Bibliografia

Beynon-Davies P., *Systemy baz danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 1998.

Chałon M., *Systemy Baz Danych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2001.

Date C.J., *Wprowadzenie do systemów baz danych*, Wydawnictwa Naukowo-Techniczne, Warszawa 2000.

Delobel C., Adiba M., *Relacyjne bazy danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 1989.

Garcia-Molina H., Ullman J.D., Widom J., *Implementacja systemów baz danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 2003.

Robinson I., Webber J., Eifrem E., *Graph Databases, New Opportunities for Connected Data*, wyd. II, Wydawnictwo O'Reilly, Sebastopol 2015.

Vukotic A., Watt N., *Neo4j in Action*, Manning, Shelter Island 2014.

Źródła sieciowe

Portal webowy Neo4j, https://neo4j.com (data odczytu: 12.11.2017).

* * *

Identification and Prevention of Financial Fraud with the Use of Graph Databases: The IAFEC System

Abstract

The article presents the developed and implemented elements of the identification and prevention of the financial fraud system – IAFEC. The IAFEC System based on the featured concept uses also a graph database for analysis. The article shows the general architecture of the IAFEC system and its information assets together with the applied method of data acquisition for the proposal data model.

Keywords: financial fraud identification, money laundering, data analysis, database, graph database