

GRZEGORZ BLIŹNIUK, MARIUSZ CHMIELEWSKI, TOMASZ GZIK,
RAFAŁ KASPRZYK, JAROSŁAW KOSZELA, ANDRZEJ NAJGEBAUER

Wydział Cybernetyki
Wojskowa Akademia Techniczna w Warszawie

Spójność informacji o historii leczenia pacjenta w heterogenicznej infrastrukturze elektronicznych rekordów medycznych

1. Wstęp

Niniejsze opracowanie jest powiązane przyczynowo z zamieszczoną w niniejszej publikacji pracą, w której przedstawiono główne uwarunkowania dotyczące interoperacyjności zapisów w rekordach medycznych¹. Kolejnym krokiem na drodze do uzyskania efektywnego wspomaganie procesu medycznego na bazie komputerowych ścieżek klinicznych jest odpowiednie wyszukiwanie i interpretacja danych w celu ich korelacji z poszczególnymi krokami w ścieżce. Na tej podstawie możliwe jest zbudowanie właściwych mechanizmów wspomaganie decyzji medycznych wykorzystywanych w procesie diagnozy, terapii i rehabilitacji. Jak wspomniano w pracy *Interoperacyjność zapisów historii leczenia pacjenta w heterogenicznej infrastrukturze elektronicznych rekordów medycznych*², zapewnienie wskazanych powyżej możliwości jest uwarunkowane sprawnym działaniem systemu udostępniającego repozytorium komputerowych ścieżek klinicznych (tzw. RSK³) oraz historii ich wykonania (tzw. hurtowni procesów⁴),

¹ G. Bliźniuk et al., *Interoperacyjność zapisów historii leczenia pacjenta w heterogenicznej infrastrukturze elektronicznych rekordów medycznych*.

² Ibidem.

³ *Raport końcowy projektu POIG.01.03.01-00-145/08*, red. G. Bliźniuk, Wydawnictwo WAT, Warszawa 2010, s. 93–99, 311–348, 495–506, 733–739.

⁴ T. Gzik, P. Kędzierski, J. Koszela, *Hurtownie procesów i procesy dynamiczne jako narzędzia wspierające wspomaganie podejmowania decyzji medycznych*, w: *Modelowanie i zastosowanie komputerowych systemów medycznych*, red. M. Cieciora, W. Olchownik, Vizja Press&IT,

wraz z efektywnymi mechanizmami zapewnienia ich interoperacyjności z systemami EHR⁵. Konieczne jest ponadto szybkie przeprowadzanie odpowiedniej analizy semantycznej pozyskiwanych danych i właściwe umieszczanie ich w przestrzeni decyzyjnej odnoszącej się do procesu leczenia⁶. Dotyczy to sytuacji, kiedy te same informacje medyczne są zapisywane wielokrotnie w różnych miejscach. Oznacza to, że dla zapewnienia spójności prezentacji informacji konieczne jest odpowiednie ich odfiltrowanie i ustalenie istotności poszczególnych węzłów informacyjnych w sieci systemów EHR, z którym współpracuje system ścieżek klinicznych.

2. Semantyczna ekstrakcja danych

Zastosowanie metod związanych z modelowaniem semantyki i algorytmów przetwarzania modeli semantycznych zapewnia narzędzia wykorzystywane do unifikacji przetwarzanych zbiorów danych. Dziedziną zastosowania tego typu metod jest integracja zbiorów danych opisanych różnymi modelami danych (schematami). Modele semantyczne mogą również być wykorzystywane do automatycznej unifikacji danych przekazywanych w komunikatach wymieniających pomiędzy usługami systemów informatycznych⁷. Cechami pozwalającymi na realizację tych celów jest w głównej mierze korzystanie z języków opisu bazujących na formalizmach języków logiki zarówno rachunku predykatów pierwszego rzędu, jak i logiki opisowej (ang. *Description Logic*)⁸. Integracja

Warszawa 2009, s. 50–57; T. Gzik, *Analiza rozwiązań informatycznych wykorzystywanych do wspierania modelowania procesów wspomagania podejmowania decyzji medycznych*, w: *Raport z realizacji zadania 1 projektu POIG.01.03.01-00-145/08*, Wydawnictwo WAT, Warszawa 2009, s. 207–219.

⁵ G. Bliźniuk, *Koncepcja implementacji warunków interoperacyjności systemu ścieżek klinicznych i elektronicznego rekordu pacjenta*, „Biuletyn” Instytutu Systemów Informatycznych, nr 6, Wydawnictwo WAT, Warszawa 2010, s. 1–10.

⁶ M. Chmielewski, A. Gałka, *Semantic battlespace data mapping using tactical symbology*, „Advances in Intelligent Information and Database Systems” 2010, Springer, s. 157–168, s. 157–168.

⁷ M. Chmielewski, J. Koszela, *The concept of C4I systems data integration for planning joint military operations, based on JC3 standar*, Proceedings of the Military Communications and Information Systems Conference MCC2008, Kraków, 22–24.09.2008; M. Chmielewski, J. Koszela, *Semantyczna integracja danych*, w: *Platformy integracyjne. Zagadnienia wybrane*, red. T. Górski, Wydawnictwo Naukowe PWN, Warszawa 2012, s. 54–91.

⁸ F. Baader et al., *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press, Cambridge 2003.

danych może się odbywać na trzech poziomach: technicznym, syntaktycznym (składniowym – formatów danych) oraz semantycznym. Oczywiście jest to, że ostatni poziom jest najtrudniejszy do realizacji, gdyż wymaga opisu znaczenia danych w języku automatycznie przetwarzalnym przez automaty – stąd wybierane są języki logiki i mechanizmy wnioskujące.

Integracja zbiorów danych jest istotnym zagadnieniem poruszonym w niniejszym artykule, gdyż pozwala na unifikację i automatyczną transformację danych wyrażonych w różnych semantykach (terminologiach). Zbiory danych systemów medycznych możemy uważać za źródła heterogeniczne posługujące się generalnie zbliżoną terminologią, jednak w szczegółach – zwłaszcza opisów pacjentów, jednostek chorobowych – często odbiegające od siebie. Badania w zakresie unifikacji oznaczeń jednostek chorobowych są prowadzone od wielu lat i ich skutkiem jest opracowanie elementarzy, taksonomii, klasyfikacji dziedzinowej wiedzy medycznej. Analizując dostępne publikacje⁹, można odnaleźć bogaty zbiór takich zasobów formułujący terminologię i standardy opisu:

- ICD-11 (ang. *International Statistical Classification of Diseases and Related Health Problems*) – międzynarodowy system informacji nozologicznej (podział, klasyfikacja chorób i ich opis); standard obejmuje ewolucję opisów, począwszy od ICD-6 do ICD-11;
- SNOMED (11) (ang. *Systematized Nomenclature of Medicine*) – rozbudowany systematyczny zbiór terminów medycznych;
- ATC (ang. *Anatomical Therapeutic Chemical Classification System*) – klasyfikacja anatomiczno-terapeutyczno-chemiczna;
- DSM-IV (ang. *Diagnostic and Statistical Manual of Mental Disorders*) – klasyfikacja zaburzeń psychicznych wg Amerykańskiego Towarzystwa Psychiatrycznego;
- DSM-5 (DSM-V) – nowa wersja klasyfikacji zaburzeń psychicznych wg Amerykańskiego Towarzystwa Psychiatrycznego;
- DC:0-3R – (ang. *Diagnostic Classification of Mental Health And Development Disorders Of Infancy and Early Childhood – Revised Edition*) – klasyfikacja diagnostyczna zaburzeń psychicznych i rozwojowych w okresie niemowlęctwa i wczesnego dzieciństwa.

Przywołane zbiory to nie jedyne dostępne standardy reprezentacji usystematyzowanych, często encyklopedycznych danych. Tak dobrze opisane bazy

⁹ (CDC) Centers for Disease Control and Prevention, *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)*, National Center for Health Statistics, 2010.

ustrukturalizowanej wiedzy mogą posłużyć do zdefiniowania ontologii medycznych i wykorzystania ich do unifikacji danych przechowywanych w systemach medycznych. Wiele spośród dostępnych systemów wykorzystuje takie standardy, jednak w głównej mierze posługując się kodami standardów, np. ICD-10, i uzupełniając aktualnie przechowywane w systemie dane, co może prowadzić do pewnych niespójności. Proponowane rozwiązanie bazujące na opisach języków ontologicznych pozwala takie sytuacje wykrywać i wskazywać miejsca niespójności w zbiorze danych. Przedstawiane w tym miejscu rozważań zagadnienia są również powiązane z problematyką modelowania szczególnej klasy procesów biznesowych¹⁰, projektowania systemów z wykorzystaniem szczególnego rodzaju podejścia architektonicznego¹¹ oraz badania efektywności zaprojektowanych i uruchomionych procesów biznesowych¹².

Wzorując się na procesach technologicznych wykorzystywanych w dziedzinie hurtowni danych, można zidentyfikować szereg bardzo użytecznych narzędzi i technik wykorzystywanych przy naturalnej dla tego typu narzędzi migracji danych z wielu różnorodnych źródeł. W przypadku hurtowni integracja danych jest wykorzystywana w procesie przygotowywania danych znanym pod nazwą ETL (ang. *Extract, Transform and Load*). Używane są również inne powszechnie stosowane techniki integracji danych, wykorzystujące: bezpośrednią migrację i replikację danych pomiędzy bazami danych (wspomagana procedurami wbudowanymi) oraz dedykowane integracji usługi serwisowe pochodzące z podejścia SOA (ang. *Service Oriented Architecture*).

Mimo dość zaawansowanych dostępnych mechanizmów integracji danych w procesie ETL znaczną część wysiłku realizuje zwykle człowiek, który dzięki posiadanej wiedzy, umiejętnościom i doświadczeniu analizuje zawartość informacyjną źródeł danych, nie tylko pod kątem struktury danych, modelu, ale

¹⁰ R. Waszkowski, A. Chodowska, R. Popławski, *Automatyzacja procesów biznesowych jako element systemu wspomagania decyzji odpowiedzialny za sterowanie działań zgodnie z przyjętymi procedurami*, w: *Modelowanie i symulacja procesów oraz określenie komputerowo wspomaganych procedur w zakresie zarządzania ryzykiem bezpieczeństwa żywności i żywienia*, red. J. Bertrandt, K. Lasocki, BELStudio, Warszawa 2012, s. 1208–1221.

¹¹ T. Górski, *Architectural view model for an integration platform*, „Journal of Theoretical and Applied Computer Science” 2012, vol. 6, no. 1, s. 25–34; T. Górski, *Projektowanie platform integracyjnych w architekturze zorientowanej na usługi*, „Wiadomości Górnicze” 2012, nr 7–8, s. 407–417.

¹² M. Lignowska, T. Nowicki, *Symulacyjna metoda badania procedur medycznych*, w: *Technologie informatyczne i ich zastosowania*, red. A. Jastrebów, Politechnika Radomska, Radom 2010, s. 243–252; T. Nowicki, *Efficiency estimation of organization described by workflow model*, w: *Contemporary corporate management*, red. K. Grzybowska, A. Stachowiak, Publishing House of Poznan University of Technology, Poznań 2009, s. 87–100.

również ich aspekt semantyczny i pragmatyczny (dane pochodzące z systemów utrzymywanych przez ludzi o odmiennych światopoglądach, kulturach czy zwyczajach). Poprawność tak szerokiej analizy źródeł danych oraz określenia i zdefiniowania celu działań w sposób znaczący zminimalizuje wystąpienie błędów w procesie integracji danych.

Proponowana metoda¹³ zakłada przeprowadzenie dwóch odrębnych procesów na potrzeby integracji danych medycznych: ekstrakcji semantyki (ang. *semantics extraction*) oraz mapowania modeli semantycznych (ang. *semantic model alignment*).

Ekstrakcja semantyki na potrzeby tej pracy będzie definiowana jako proces analizy zbiorów danych (formatów danych) w celu pozyskania i wytworzenia opisów terminologii zawierających definicje wykorzystywanych pojęć i związków między nimi. Definicje te muszą być spójne logicznie i muszą zawierać precyzyjną aksjomatykę formułującą definicje poszczególnych pojęć. W tym celu proponuje się używać języków z rodziny języków konceptualnych (ang. *conceptual languages*) ALC zawierających podstawowe konstrukcje semantyczne logiki opisowej.

Mapowanie modeli semantycznych może być postrzegane jako proces dopasowania elementów modeli tak, aby wskazać odpowiadające sobie, najbardziej podobne elementy źródłowych terminologii. Proces mapowania zakłada wykorzystanie miar podobieństwa etykiet elementów danych terminologii, ale również strukturalnych cech ilościowych analizowanych modeli. Wielokryterialna analiza podobieństwa elementów ontologii jest realizowana w postaci wielokryterialnego zadania przydziału. Zadanie to zakłada wyznaczenie przydziału pojęć ontologii A do pojęć ontologii B przy maksymalizacji stopnia podobieństwa wierzchołków wyznaczanego jako metakryterium wytypowanych miar podobieństwa pojęć.

Opisany proces integracji danych medycznych w szczególności zakłada:

1. Ekstrakcję semantyki danych, polegającą na zbudowaniu ontologii opisującej rozpatrywany zbiór danych.
2. Migrację rekordów danych z zasobów danych i ich wprowadzenie do bazy wiedzy w postaci instancji zgodnie z opisem ontologicznym opracowanym w punkcie 1.

¹³ M. Chmielewski, *Ontology-based indirect association assessment method using graph and logic reasoning techniques. (Metoda oceny skojarzeń pośrednich w modelach semantycznych wykorzystująca techniki grafowe i wnioskowania)*, praca doktorska, Wydział Cybernetyki WAT, Warszawa 2012; M. Chmielewski, A. Gałka, op.cit., s. 157–168; G. Bliźniuk et al., *Hurtownie procesów*, „*Studia Informatica*” 2012, t. 33, nr 2A (105), Wydawnictwo Politechniki Śląskiej, s. 112–127.

3. Opracowanie lub wyznaczenie mapowania ontologii zakładające przeprowadzenie procesu dopasowania dwóch terminologii poprzez sformułowanie mapowań semantycznych (ang. *semantic bridges*) – faza mapowania modeli semantycznych.

4. Uruchomienie mechanizmów wnioskowania w celu zweryfikowania spójności migrowanych danych i weryfikacji poprawności opracowanych mapowań.

Poszczególne kroki opisywanej metody wspierane są narzędziowo przez opracowane środowisko przetwarzania semantyki SaaE¹⁴ (ang. *Semantic Association Analysis Environment*), zbudowane na bazie narzędzia Protégé 3.4. X oraz bibliotek analizy struktur grafowych JUNG, Prefuse oraz silników wnioskujących JESS, Pallet, Racer.

Identyfikowane w metodzie pomosty semantyczne są konstrukcjami definiującymi podobne lub tożsame elementy modeli. W tym celu używa się oferowanych w językach ontologicznych konstrukcji równoważności pojęć (ang. *Concept Equivalence Axiom*) oraz równoważności relacji (ról) (ang. *Concept Equivalence Role*). Konsekwencją użycia tych zapisów w modelu jest pozwolenie mechanizmowi wnioskującym na postrzeganie instancji danych pojęć jako równoważnych sobie, a więc dane wyrażone w jednej semantyce automatycznie będą interpretowane jako dane o znaczeniu definiowanym przez drugi model semantyczny. Ważnym aspektem budowy pomostów semantycznych jest dobór odpowiednich aksjomatów wyrażających związki pomiędzy poszczególnymi elementami modelu. Oznacza to, że inżynier wiedzy, biorąc pod uwagę rekomendacje metody dotyczące przydziału pojęć, może skonstruować pomost semantyczny (pojęcie) zawierający rozbudowane i złożone warunki logiczne odnoszące się do pojęć w ontologiach źródłowych.

Pomosty semantyczne¹⁵, będąc konstrukcjami dwuargumentowymi, mogą być również rozszerzane, gdyż są przechodnie. Opisywane podejście wykorzystuje następujące dyrektywy języka OWL:

- *owl:equivalentClass* – pozwalające na definicje tożsamych klas,
- *owl:equivalentProperty* – pozwalające na definicje tożsamych relacji,
- *owl:sameAs* – pozwalające na definicje tożsamych instancji.

Tożsamość przedstawionych elementów może być wykorzystana do bezpośredniego wskazania odpowiadających sobie elementów. W mapowanych modelach może jednak nie istnieć wymagany zbiór tożsamych pojęć, w takim

¹⁴ M. Chmielewski, op.cit.

¹⁵ Ibidem; G. Bliźniuk et al., *Hurtownie procesów...*, op.cit., s. 112–127.

przypadku należy dostarczyć dodatkowe pojęcia, które będą dołączone do jednej z mapowanych ontologii.

Formalnie zdefiniujemy ontologię jako trójkę $O = \langle C, R, A \rangle$, zawierającą zbiór pojęć C , relacji R , łączących pojęcia, oraz aksjomatów A , budowanych na podstawie zdefiniowanych relacji z wykorzystaniem języka formalnego, np. logiki opisowej. Rekomendacja dopasowania pojęć w rozpatrywanej parze ontologii zakłada wytworzenie dla pary ontologii O', O'' , zadanych jako $O' = \langle C', R', A' \rangle$, $O'' = \langle C'', R'', A'' \rangle$, macierzy podobieństwa pojęć:

$$Sim_{O', O''}^C = \begin{bmatrix} m_{c'_1, c''_1} & \dots & m_{c'_1, c''_j} \\ \dots & \dots & \dots \\ m_{c'_i, c''_1} & \dots & m_{c'_i, c''_j} \end{bmatrix}_{\bar{C}' \times \bar{C}''} \quad \begin{array}{l} c'_1, \dots, c'_i \in C' \\ c''_1, \dots, c''_j \in C'' \end{array} \quad m_{c'_i, c''_j} \in [0, 1] \\ M_{c'_i, c''_j} = [m_{c'_i, c''_j}^1, \dots, m_{c'_i, c''_j}^k]$$

Przy czym zakładamy, że macierz zawiera miary podobieństwa, które mogą być dobierane zgodnie z preferencjami decydenta, oddając potrzeby odwzorowania zarówno podobieństwa strukturalnego, jak i podobieństwa ilościowego modeli (np. podobieństwa etykiet pojęć, miar istotności wierzchołków-pojęć). Agregacja miary podobieństwa dwóch pojęć jest wyznaczana jako suma ważona składowych podobieństw wynikających z przyjętych w analizie miar oceny ilościowej i podobieństwa strukturalnego definiowanych w postaci wektora $M_{c'_i, c''_j}$.

$$m_{c'_i, c''_j} = \sum_{l=0}^k \alpha_l \cdot m_{c'_i, c''_j}^l, \quad \sum_{l=0}^k \alpha_l = 1, \quad \forall_{l \in 1..k} \alpha_l \in [0, 1].$$

Ekstrakcja semantyki jest procesem złożonym i tylko częściowo może być wspierana programowo. Zbudowane mechanizmy przygotowują rekomendację optymalnego przydziału elementów mapowanych ontologii, dostarczając decydentowi możliwość korekcji przydziału, a przede wszystkim wpływania na istotność wybieranych kryteriów dla zdefiniowanego zadania wielokryterialnego. Rekomenduje się rozwiązanie zadania z wykorzystaniem skalaryzacji, wprowadzając metakryterium lub hierarchizację rozpatrywanych kryteriów.

Prezentowane podejście ujednolicania semantyki zakłada definicję ontologii spinającej mapowane modele. Podejście to może być uważane za rodzaj strategii *bottom-up*, gdyż bazując na rozłącznych modelach dziedzinowych, staramy się wskazać odpowiadające sobie elementy tych modeli, często tworząc dedykowany model zawierający odpowiednie pomosty semantyczne. Odmiennym pojęciem jest stosowanie podejścia *top-down*, polegającego na wytwarzaniu modeli semantycznych korzystających z jednego modelu wyższego poziomu, np. modelu SUMO, CYC, DOLCE, lub w przypadku dziedziny medycyny – OGMS, OBO,

DiseaseOntology. W tym przypadku pojęcia definiowane w wytworzonych odrębnych ontologiach są bezpośrednio związane z pojęciami w ontologii bazowej, a więc istnieje wspólny podzbiór pojęć odwzorowany w integrowanych modelach semantycznych. W przypadku jednak, gdy wspólne pojęcia nie istnieją, należy stosować strategię *bottom-up*.

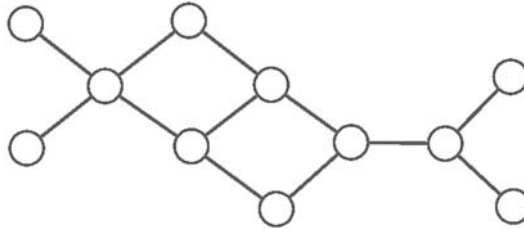
3. Dobór kluczowych węzłów w sieci systemów EHR

Już na wstępie niniejszego artykułu wskazano, że dla powodzenia przebiegu procesu leczenia kluczowe jest zdefiniowanie dla każdej jednostki chorobowej właściwej procedury postępowania (tzw. ścieżki klinicznej). Dodatkowo owe procedury postępowania w wielu przypadkach muszą ulec modyfikacji lub przynajmniej swego rodzaju profilowaniu z punktu widzenia konkretnego pacjenta. Konieczność profilowania może wynikać choćby z powodu współistnienia u pacjenta kilku jednostek chorobowych. Liczność możliwych ścieżek medycznych jest bardzo duża już ze względu na liczbę jednostek chorobowych. Liczba sklasyfikowanych jednostek chorobowych i innych dolegliwości w używanych w Polsce systemach kodowania jest określana na kilkadziesiąt tysięcy. Jeśli uwzględnimy możliwość ich modyfikacji praktycznie dla każdego pacjenta, wówczas przestrzeń możliwych stanów (etapów procedury medycznej) pacjenta „przyprawia o zawrót głowy” niejednego lekarza. Każdy ze stanów pacjenta opisywany może być z wykorzystaniem innego systemu EHR i składowany/zapisany na innym serwerze.

Powstaje zasadnicze pytanie, jak zdefiniować, a następnie wyszukać kluczowe systemy EHR z uwzględnieniem poszczególnych jednostek chorobowych i/lub pacjentów. Jednym ze sposobów może być opracowanie modelu zależności pomiędzy systemami EHR w postaci grafu/sieci i wykorzystanie do oceny istotności poszczególnych węzłów sieci tzw. miar centralności. Można zatem przyjąć, że pomiędzy dwoma systemami EHR występuje zależność (krawędź w grafie) w przypadku analizowanej jednostki chorobowej, jeśli dla dowolnego pacjenta leczonego na daną jednostkę chorobową istnieje taka ścieżka kliniczna, w której systemy te są wykorzystywane na sąsiadujących (następujących bezpośrednio po sobie) etapach procedury medycznej.

Miary centralności ułatwiają udzielenie odpowiedzi na pytanie o to, „kto(co) jest najważniejszy(e) w analizowanej sieci?”¹⁶. Okazuje się, że nie ma jednoznacznej odpowiedzi. Zależy ona bowiem od przyjętej semantyki słowa „istotny”, co obrazują wprowadzone dalej miary centralności (istotności) węzłów.

Na potrzeby zobrazowania interpretacji poszczególnych miar centralności wykorzystano graf przedstawiony na rysunku 1. Węzły grafu przyjmują na kolejnych rysunkach różne kolory, o intensywności proporcjonalnej do wartości rozpatrywanej miary centralności. Najistotniejsze wierzchołki zostały dodatkowo oznaczone za pomocą otaczającego ich okręgu (w celu zwiększenia czytelności rysunków).



Rysunek 1. Przykładowa sieć (graf) do zobrazowania interpretacji miar centralności

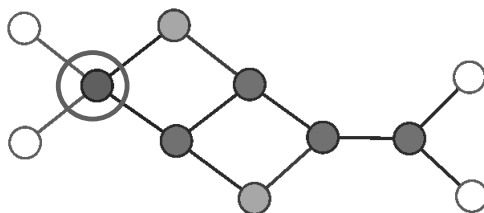
Źródło: opracowanie własne.

Miary centralności występują najczęściej w postaci znormalizowanej w przedziale $[0, 1]$. Dzięki normalizacji możliwa staje się interpretacja i porównywanie wartości miar centralności wierzchołków dla grafów o różnych rozmiarach (liczbie wierzchołków). Najprostszą miarą centralności wierzchołka, którą podpowiada intuicja, jest **znormalizowany stopień wierzchołka** (ang. *degree centrality*) (rysunek 2). Według tej miary, wierzchołek jest tym istotniejszy w chwili t , im wyższy stopień ma w danej chwili – więcej bezpośrednich połączeń z pozostałymi wierzchołkami w sieci:

$$dc_i(t) = \frac{k_i(t)}{|V_G(t)| - 1}.$$

W grafach skierowanych (digrafach) rozróżnia się znormalizowany stopień wejściowy $dc_i^{IN}(t)$ i znormalizowany stopień wyjściowy $dc_i^{OUT}(t)$.

¹⁶ R. Kasprzyk, *Modele ewolucji systemów złożonych i metody badania ich charakterystyk dla potrzeb komputerowej identyfikacji potencjalnych sytuacji kryzysowych*, praca doktorska, Wydział Cybernetyki WAT, Warszawa 2012; S. Wasserman, K. Faust, *Social Network Analysis: Methodology and Application*, Cambridge University Press, Cambridge–New York 1994.

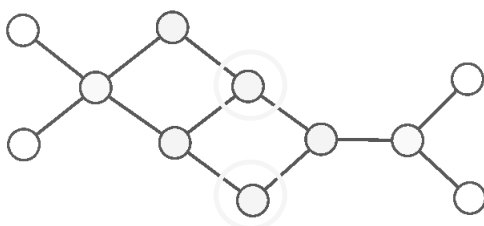


Rysunek 2. Zobrazowanie wartości *degree centrality* węzłów w przykładowej sieci

Źródło: opracowanie własne.

Kolejną miarą centralności jest tzw. **promień wierzchołka** (ang. *radius centrality*)¹⁷, według której wierzchołek jest tym ważniejszy, im mniejsza jest jego odległość do najdalszego wierzchołka (rysunek 3):

$$rc_i(t) = \frac{1}{\max_{v_j \in V_G(t)} d_{ij}(t)}.$$



Rysunek 3. Zobrazowanie wartości *radius centrality* węzłów w przykładowej sieci

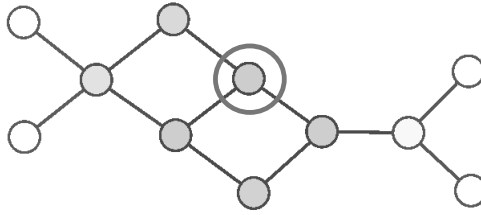
Źródło: opracowanie własne.

Często stosowaną miarą jest tzw. **bliskość wierzchołka** (ang. *closeness centrality*)¹⁸, która pozwala na wyliczenie średniej odległości rozpatrywanego węzła do wszystkich pozostałych węzłów w grafie (rysunek 4). Wierzchołek, który średnio jest najbliższy wszystkim wierzchołkom, jest wówczas najistotniejszy:

$$cc_i(t) = \frac{|V_G(t)| - 1}{\sum_{v_j \in V_G(t)} d_{ij}(t)}.$$

¹⁷ S. Wuchty, P.F. Stadler, *Centers of complex networks*, „Journal of Theoretical Biology” 2003, vol. 222, s. 45–53.

¹⁸ L.C. Freeman, *Centrality in Social Networks: Conceptual clarification*, „Social Networks” 1979, vol. 1(3), s. 215–239.

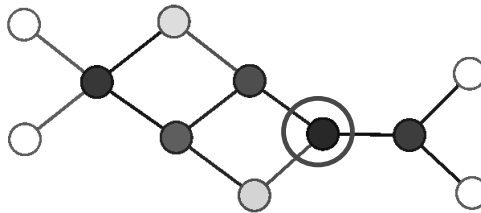


Rysunek 4. Zobrazowanie wartości *closeness centrality* węzłów w przykładowej sieci

Źródło: opracowanie własne.

Ciekawym sposobem oceny istotności węzła jest wyznaczenie pośrednictwa/obciążenia wierzchołka (ang. *betweenness/load centrality*)¹⁹, które można zdefiniować jako procent najkrótszych dróg pomiędzy dowolną parą wierzchołków przechodzących przez rozpatrywany węzeł (rysunek 5). Jeśli przez $p_{l,i,k}(t)$ oznaczymy liczbę najkrótszych dróg pomiędzy v_l i v_k przechodzących przez v_i w chwili t oraz przez $p_{l,k}(t)$ liczbę wszystkich dróg pomiędzy v_l i v_k , w chwili t , wówczas:

$$bc_i(t) = \frac{\sum_{v_l \in V_G(t)} \sum_{v_k \neq v_l \in V_G(t)} \frac{p_{l,i,k}(t)}{p_{l,k}(t)}}{(|V_G(t)| - 2)(|V_G(t)| - 1)}.$$



Rysunek 5. Zobrazowanie wartości *betweenness centrality* węzłów w przykładowej sieci

Źródło: opracowanie własne.

Usunięcie wierzchołka o największej wartości obciążenia powoduje znaczące zwiększenie średniej odległości pomiędzy węzłami, a tym samym największe utrudnienia komunikacji w grafie. Obciążenie wierzchołka jest niezwykle istotne dla odpowiedzi na pytanie: jak trudne i czasochłonne może być zadanie polegające na maksymalizacji rozspojenia grafu?

¹⁹ L.C. Freeman, *A set of measures of centrality based on betweenness*, „Sociometry” 1977, vol. 40, s. 35–41.

O ile stopień wierzchołka za wartość istotności przyjmuje liczbę wierzchołków przyległych, o tyle oczywiste jest to, że nie wszyscy sąsiedzi wierzchołka powinni zwiększać jego istotność o tę samą wartość. Jeśli wierzchołek ma znaczną liczbę połączeń, ale z wierzchołkami, które nie są istotne, to logiczne jest, że nie powinien on być uznany za tak ważny jak wierzchołek, który posiada choćby jedno połączenie, ale za to z najważniejszym wierzchołkiem w sieci.

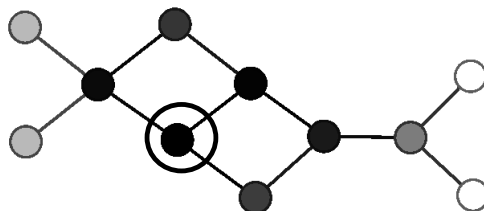
Ten rodzaj centralności, tzw. **prestież wierzchołka**, można wyliczyć na podstawie **wektora własnego macierzy sąsiedztwa** (ang. *eigenvector centrality*)²⁰ (rysunek 6). Jeśli istotność wierzchołka v_i w chwili t oznaczmy jako $ec_i(t)$, wówczas mamy:

$$ec_i(t) = \frac{1}{\lambda} \sum_{j=1}^{|V_G(t)|} a_{ij}(t) \cdot ec_j(t).$$

Używając notacji macierzowej, możemy zapisać:

$$\vec{ec}(t) = \frac{1}{\lambda} A(t) \cdot \vec{ec}(t).$$

Wektor $\vec{ec}(t)$ jest wektorem własnym macierzy $A(t)$ dla największej wartości własnej λ , czego dowodzi się z wykorzystaniem teorii Perrona–Frobeniusa²¹. Składowe wektora $\vec{ec}(t)$ po jego normalizacji odpowiadają wartością miar centralności wierzchołków grafu.



Rysunek 6. Zobrazowanie wartości *eigenvector centrality* węzłów w przykładowej sieci

Źródło: opracowanie własne.

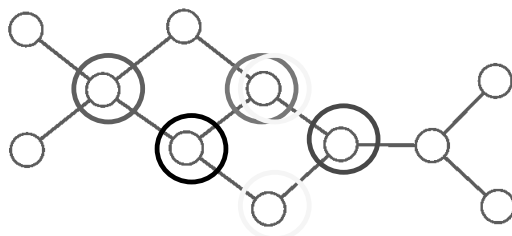
Do wyznaczenia wektora $\vec{ec}(t)$ wykorzystuje się najczęściej algorytm oparty na iteracyjnej metodzie potęgowej (ang. *power iteration method*)²².

²⁰ M.E.J. Newman, *The mathematics of networks*, Center of the Study of Complex Systems, University of Michigan, Ann Arbor 2006.

²¹ Ibidem.

²² G.H. Golub, C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore 1996.

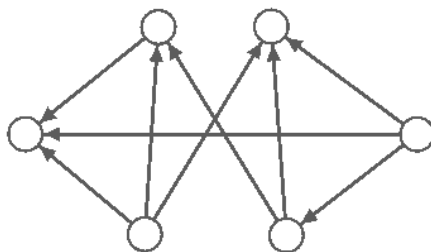
Wpływ miary centralności na wybór węzła najbardziej centralnego przedstawiony został na rysunku 7. Przykładowa sieć została tak dobrana, aby ułatwić zobrazowanie wpływu miary centralności na wybór wierzchołka centralnego.



Rysunek 7. Wpływ miary centralności na wybór węzła najbardziej centralnego w przykładowej sieci

Źródło: opracowanie własne.

W przypadku grafów skierowanych ze względu na możliwość wyróżnienia stopnia wejściowego i wyjściowego wierzchołka używa się często rozróżnienia między wartością **autorytetu** (ang. *authority centrality*) i **koncentratora** (ang. *hub centrality*) dla każdego wierzchołka²³ (rysunki 8 i 9). Wierzchołek jest tym lepszym autorytetem, im przez lepsze koncentratory jest wskazywany. Jednocześnie wierzchołek jest tym lepszym koncentratorem, im lepsze autorytety wskazuje.



Rysunek 8. Przykład grafu skierowanego do zobrazowania interpretacji wartości miary *hub centrality* i *authority centrality*

Źródło: opracowanie własne.

Jeśli przez $ac_i(t)$ i $hc_i(t)$ oznaczymy odpowiednio wartość autorytetu i koncentratora wierzchołka v_i w chwili t , wówczas mamy:

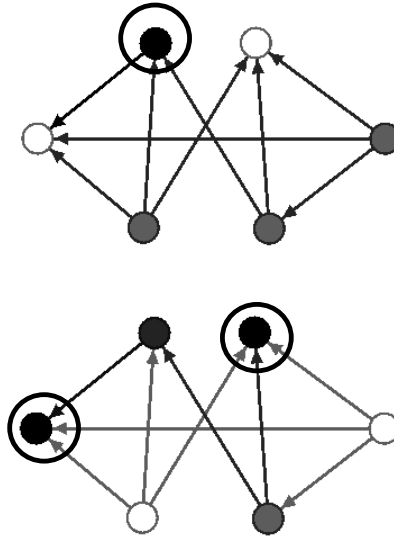
²³ J.M. Kleinberg, *Authoritative Sources in a Hyperlinked Environment*, „Journal of the ACM” 1999, vol. 46(5), s. 604–632.

$$\begin{cases} ac_i(t) = \frac{1}{\lambda_a} \sum_{j=1}^{|V_G(t)|} p_{ji}(t) \cdot hc_j(t) \\ hc_i(t) = \frac{1}{\lambda_h} \sum_{j=1}^{|V_G(t)|} p_{ij}(t) \cdot ac_j(t) \end{cases}$$

Używając notacji macierzowej, możemy zapisać:

$$\begin{cases} \vec{ac}(t) = \frac{1}{\lambda_a} P(t)^T \cdot \vec{hc}(t) \\ \vec{hc}(t) = \frac{1}{\lambda_h} P(t) \cdot \vec{ac}(t) \end{cases} \Rightarrow \begin{cases} \vec{ac}(t) = \frac{1}{\lambda_a} P(t)^T P(t) \cdot \vec{ac}(t) \\ \vec{hc}(t) = \frac{1}{\lambda_h} P(t) P(t)^T \cdot \vec{hc}(t) \end{cases}$$

Wektor $\vec{ac}(t)$ i $\vec{hc}(t)$ jest odpowiednio wektorem własnym macierzy $P(t)^T P(t)$ dla największej wartości λ_a i $P(t) P(t)^T$ dla największej wartości λ_h .



Rysunek 9. Zobrazowanie wartości *hub centrality* (sieć z lewej strony) i *authority centrality* (sieć z prawej strony) węzłów w przykładowej sieci

Źródło: opracowanie własne.

4. Podsumowanie

Przedstawiony koncept semantycznej ekstrakcji danych z systemów EHR oraz poszukiwania trafności doboru EHR źródłowego jest kluczowy dla zapewnienia spójności informacji o historii leczenia pacjenta. Rozważane zagadnienia należy rozpatrywać łącznie w powiązaniu z problematyką interoperacyjności przedstawioną w opracowaniu *Interoperacyjność zapisów historii leczenia pacjenta w heterogenicznej infrastrukturze elektronicznych rekordów medycznych*²⁴.

Oryginalność niniejszego podejścia do omawianych zagadnień wynika przede wszystkim z tego, że nie zaproponowano nowej koncepcji EHR-o centrycznej (CEN 13606) czy work-flowcentrycznej (HL7) dla zapewnienia interoperacyjności systemów i spójności informacji. Takie nowe podejście wynika z doświadczeń naukowców z WAT w realizacji jednego z projektów w dziedzinie informatyki medycznej²⁵, w którym wykazano kluczowość zapisu definicji komputerowo interpretowalnej ścieżki klinicznej dla zapewnienia właściwego ułożenia pozostałego otoczenia systemowego wspomaganie decyzji medycznych. Dokonania światowe są w tym zakresie stosunkowo niewielkie, a dziedzina wciąż nowa.

Nieco inna, aczkolwiek zmierzająca w kierunku wspomaganie decyzji medycznych, jest koncepcja profilu IHE XDW²⁶. Bazuje ona na innym profilu – IHE XDS, będącym koncepcją reguł dla obiegu elektronicznych dokumentów medycznych. Można tutaj zastosować standardy HL7 CDA, które są przydatne również do trwałego składowania informacji w rekordach medycznych oraz jako źródła informacji dla systemów ścieżek klinicznych. Ten zakres rozważań wykracza jednak poza niniejsze opracowanie.

Bibliografia

1. Baader F. et al., *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press, Cambridge 2003.

²⁴ G. Bliźniuk et al., *Interoperacyjność zapisów...*, op.cit.

²⁵ *Raport końcowy projektu POIG.01.03.01-00-145/08...*, op.cit., s. 93–99, 311–348, 495–506, 733–739.

²⁶ http://www.ihe.net/Technical_Framework/upload/IHE_ITI_Suppl_XDW_Rev2-1_TI_2011-10-03.pdf.

2. Bliźniuk G., *Koncepcja implementacji warunków interoperacyjności systemu ścieżek klinicznych i elektronicznego rekordu pacjenta*, „Biuletyn” Instytutu Systemów Informatycznych, nr 6, Wydawnictwo WAT, Warszawa 2010, s. 1–10.
3. Bliźniuk G. et al., *Hurtownie procesów*, „Studia Informatica” 2012, t. 33, nr 2A (105), Wydawnictwo Politechniki Śląskiej, s. 112–127.
4. (CDC) Centers for Disease Control and Prevention, *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)*, National Center for Health Statistics, 2010.
5. Chmielewski M., *Ontology-based indirect association assessment method using graph and logic reasoning techniques. (Metoda oceny skojarzeń pośrednich w modelach semantycznych wykorzystująca techniki grafowe i wnioskowania)*, praca doktorska, Wydział Cybernetyki WAT, Warszawa 2012.
6. Chmielewski M., Gałka A., *Semantic battlespace data mapping using tactical symbology*, „Advances in Intelligent Information and Database Systems” 2010, Springer, s. 157–168.
7. Chmielewski M., Koszela J., *Semantyczna integracja danych*, w: *Platformy integracyjne. Zagadnienia wybrane*, red. T. Górski, Wydawnictwo Naukowe PWN, Warszawa 2012, s. 54–91.
8. Chmielewski M., Koszela J., *The concept of C4I systems data integration for planning joint military operations, based on JC3 standar*, Proceedings of the Military Communications and Information Systems Conference MCC2008, Kraków, 22–24.09.2008.
9. Freeman L.C., *A set of measures of centrality based on betweenness*, „Sociometry” 1977, vol. 40, s. 35–41.
10. Freeman L.C., *Centrality in Social Networks: Conceptual clarification*, „Social Networks” 1979, vol. 1(3), s. 215–239.
11. Golub G.H., Van Loan C.F., *Matrix Computations*, The Johns Hopkins University Press, Baltimore 1996.
12. Górski T., *Architectural view model for an integration platform*, „Journal of Theoretical and Applied Computer Science” 2012, vol. 6, no. 1, s. 25–34.
13. Górski T., *Projektowanie platform integracyjnych w architekturze zorientowanej na usługi*, „Wiadomości Górnicze” 2012, nr 7–8, s. 407–417.
14. Gzik T., *Analiza rozwiązań informatycznych wykorzystywanych do wspierania modelowania procesów wspomaganie podejmowania decyzji medycznych*, w: *Raport z realizacji zadania 1 projektu POIG.01.03.01-00-145/08*, Wydawnictwo WAT, Warszawa 2009, s. 207–219.
15. Gzik T., Kędzierski P., Koszela J., *Hurtownie procesów i procesy dynamiczne jako narzędzia wspierające wspomaganie podejmowania decyzji medycznych*, w: *Modelowanie i zastosowanie komputerowych systemów medycznych*, red. M. Cieciora, W. Olchowik, Vizja Press&IT, Warszawa 2009, s. 50–57.

16. Kasprzyk R., *Modele ewolucji systemów złożonych i metody badania ich charakterystyk dla potrzeb komputerowej identyfikacji potencjalnych sytuacji kryzysowych*, praca doktorska, Wydział Cybernetyki WAT, Warszawa 2012.
17. Kleinberg J.M., *Authoritative Sources in a Hyperlinked Environment*, „Journal of the ACM” 1999, vol. 46(5), s. 604–632.
18. Lignowska M., Nowicki T., *Symulacyjna metoda badania procedur medycznych, w: Technologie informatyczne i ich zastosowania*, red. A. Jastrebow, Politechnika Radomska, Radom 2010, s. 243–252.
19. Newman M.E.J., *The mathematics of networks*, Center of the Study of Complex Systems, University of Michigan, Ann Arbor 2006.
20. Nowicki T., *Efficiency estimation of organization described by workflow model*, w: *Contemporary corporate management*, red. K. Grzybowska, A. Stachowiak, Publishing House of Poznan University of Technology, Poznań 2009, s. 87–100.
21. *Raport końcowy projektu POIG.01.03.01-00-145/08*, red. G. Bliźniuk, Wydawnictwo WAT, Warszawa 2010.
22. Wasserman S., Faust K., *Social Network Analysis: Methodology and Application*, Cambridge University Press, Cambridge–New York 1994.
23. Waszkowski R., Chodowska A., Popławski R., *Automatyzacja procesów biznesowych jako element systemu wspomagania decyzji odpowiedzialny za sterowanie działaniami zgodnie z przyjętymi procedurami*, w: *Modelowanie i symulacja procesów oraz określenie komputerowo wspomaganych procedur w zakresie zarządzania ryzykiem bezpieczeństwa żywności i żywienia*, red. J. Bertrandt, K. Lasocki, BELStudio, Warszawa 2012, s. 1208–1221.
24. Wuchty S., Stadler P.F., *Centers of complex networks*, „Journal of Theoretical Biology” 2003, vol. 222, s. 45–53.

Źródła sieciowe:

1. http://www.ihe.net/Technical_Framework/upload/IHE_ITI_Suppl_XDW_Rev2-1_TI_2011-10-03.pdf.
2. <http://www.ihtsdo.org/snomed-ct/>.

* * *

Consistency of information about the history of the patient's treatments in heterogeneous infrastructure of electronic health records

Summary

Consistent delivery of medical records for patients, doctors, and others, it is possible i.a. by providing opportunities to carry out "on the fly" adequate semantic analysis of data obtained from EHR systems and their proper place in the decision-

-making process regarding treatment. Often, the same medical information is stored in multiple places. This means that in order to ensure consistent presentation of information, it is necessary to filter out the relevant information, and determine the relevance of each information node in the EHR system network, which cooperates with the clinical pathways system. Analysis of connections between EHR systems on the network, conducted from the perspective of patient information, may be based on the achievements of research in social networks. On this basis it is possible to develop optimal methods for delivering relevant medical data for a specific description of the clinical path, based on knowledge of the data collected in different nodes of the EHR system network.

Keywords: information consistency, interoperability, electronic health records, semantic data extraction, process warehouse, clinical pathway