

Integracja ontologii z wykorzystaniem teorii consensusu

Michał Rusin¹, Ngoc Thanh Nguyen¹

Streszczenie: Ważnym elementem współczesnych systemów informatycznych stają się ontologie. W wielu przypadkach tworząc nowy system niezbędne jest przeprowadzanie procesu integracji już istniejących ontologii. Proces ten często wymaga rozwiązywania konfliktów pomiędzy tymi ontologiami. W niniejszej pracy zostaje przedstawione podejście proponujące wykorzystanie teorii wyboru consensusu w celu rozwiązywania konfliktów wielowartościowych, pojawiających się w procesie integracji ontologii na poziomie instancji klas.

Słowa kluczowe: ontologie, integracja ontologii, teoria wyboru consensusu, rekomendacje

1. Wprowadzenie

Wraz z szybkim rozwojem technologii informatycznych sieć WWW zyskała na znaczeniu jako bardzo dobre medium pozyskiwania wiedzy. W dobie społeczeństwa informatycznego, sfery życia, gdzie wymiana informacji staje się bardzo ważnym aspektem codziennej egzystencji, sukcesywnie powiększają się. W zależności od przedmiotu posługującego się informacją może ona mieć odmienny charakter oraz znaczenie. Najważniejsze jest jednak, aby raz uzyskana i zgromadzona wiedza mogła zostać zapisana i udostępniona w takiej formie, która pozwalałaby na skorzystanie z niej w przyszłości.

Jednym z podstawowych problemów ponownego wykorzystania wiedzy wśród systemów ze sztuczną inteligencją jest brak jednego wspólnego słownictwa, które reprezentowałoby zgromadzoną wiedzę w sposób spójny. Wśród naukowców zajmujących się inżynierią wiedzy, przetwarzaniem języka naturalnego czy zarządzaniem wiedzą, pojęcie ontologii w ostatnich czasach zyskało na znaczeniu i popularności. To właśnie ontologia jako definicja klas, relacji, funkcji i innych obiektów (patrz Gruber, 1993) ma posłużyć jako narzędzie umożliwiające opisanie rzeczywistości, wymianę wiedzy w globalnej sieci Internet. Z tego właśnie powodu w ostatnich latach naukowcy stworzyli wiele różnych ontologii, niestety niekompatybilnych. Ta różnorodność formatów, specyfikacji oraz metodologii tworzenia ontologii wymusiła na naukowcach wyznaczenie nowych kierunków badań, których celem jest dostarczenie wystarczającej wiedzy umożliwiającej opracowanie algorytmów, narzędzi czy też metodologii, które mogłyby wspomóc poprzez zautomatyzowanie proces integracji różnych ontologii.

W dzisiejszych czasach komputery stają się punktami wejściowymi – bramami do cyfrowego świata wymiany informacji i transakcji biznesowych (patrz Fensel,

¹ Wydział Informatyki i Zarządzania, Politechnika Wrocławska, Wyb. Wyspiańskiego 27, 50-370 Wrocław
e-mail: {michal.rusin, thanh}@pwr.wroc.pl

2001). Sama wymiana danych, informacji oraz wiedzy staje się ważnym obszarem badań naukowców nad nowymi technologiami komputerowymi. Ontologie, dostarczając wspólnych struktur wiedzy dla opisywanych dziedzin, stają się powoli nieodzownym kluczowym elementem związanym z teorią wymiany informacji.

Kolejnym ważnym aspektem związanym z ontologią jest jej integracja. Do chwili obecnej istnieją spory naukowców dotyczące pojęcia integracji oraz sposobu podejścia do tego problemu. Sam termin "integracja" pojawia się w różnym znaczeniu w obrębie dziedziny ontologii zbyt często, co świadczy o nadużywaniu tego pojęcia. W pracy Pinto (1999) zostały wyróżnione trzy główne obszary, w których znaczenie terminu "integracja" jest różne:

- pierwszy obszar dotyczy budowania nowej ontologii, wykorzystując w tym celu już istniejące i dostępne inne ontologie, poprzez np. selektywne łączenie, rozszerzanie, specjalizację czy też adopcję
- drugi obszar dotyczy budowania ontologii poprzez łączenie kilku ontologii o podobnej tematyce w jedną, która unifikuje je wszystkie
- trzeci obszar związany jest z budową aplikacji wykorzystującej jedną bądź wiele ontologii

Należy zwrócić uwagę na fakt, iż integracja nie tylko tworzy i specyfikuje specjalne operacje umożliwiające budowanie ontologii z już istniejących w konkretnym środowisku programistycznym, ale także wyznacza zbiór działań, które razem tworzą swoistego rodzaju metodologię precyzującą w jaki sposób budować ontologię używając innych publicznie dostępnych ontologii. Celem niniejszej pracy jest zaprezentowanie możliwości użycia algorytmów wyboru consensusu w celu zintegrowania dwóch lub większej liczby ontologii na poziomie instancji pojedynczych klas. Na potrzeby prac i eksperymentów została zaprojektowana i zaimplementowana bazowa ontologia, dla której zostały następnie opracowane oraz zdefiniowane konkretne algorytmy wyboru consensusu. Dziedziną analizowanej ontologii jest sieć Internet i mogłaby być ona wykorzystana przez rozproszony system agentowy, którego głównymi zadaniami byłoby:

- tworzenie profilu użytkownika
- badanie preferencji użytkownika
- rekomendowanie dokumentów zgodnie z jego zainteresowaniami

Plan niniejszej pracy wygląda następująco. W sekcji drugiej zostaną przedstawione główne aspekty dotyczące problemu integracji ontologii. W sekcji trzeciej zostanie opisany consensus jako narzędzie do rozwiązywania konfliktów. W sekcji czwartej zostanie zaprezentowany przykładowy algorytm integracji macierzy trójnarnych.

2. Integracja ontologii

W ostatnich latach wiele wysiłku kosztowało naukowców rozwijanie technik, które miały na celu stworzenie sieci semantycznej, będącej następczynią sieci Internet,

pozwalającej komputerom na częściowe rozumienie informacji w niej zawartej. Niektóre z nowych technik wymagały od naukowców zaprojektowania różnych ontologii, co spowodowało, iż zagadnienia związane z ich integracją stały się kluczowe w obszarze współdzielenia i wymiany wiedzy. Dopasowywanie i łączenie istniejących ontologii, w większości wypadków dokonywane ręcznie, jest procesem żmudnym. Pomimo, iż dużo czasu poświęca się na zdefiniowanie metodologii ułatwiających ten proces wiele pytań związanych z integracją pozostaje bez odpowiedzi. Wg Noy (1999) mamy dwie możliwości podejścia do tego zagadnienia:

- łączenie ontologii w celu stworzenia pojedynczej spójnej ontologii
- dopasowanie poprzez ustalanie odnośników pomiędzy nimi tak aby można było wykorzystywać wiedzę z obu ontologii

2.1. Różne podejścia

Przytłaczająca ilość informacji zawarta w różnych źródłach danych - szczególnie w sieci Internet - gdzie wiedza nie jest usystematyzowana wywiera wpływ na opracowanie metodologii oraz technik integracji wiedzy tak aby ułatwić współdzielenie ogólnie dostępnych danych. O ile pojedyncza aplikacja zawierająca określony zbiór terminów może wykonywać zadania bez większych problemów o tyle szybki rozwój komputerów oraz sieci teleinformatycznych spowodował, iż pojedyncze aplikacje są coraz częściej zintegrowane z innymi zewnętrznymi systemami. Taka sytuacja wiąże się często z konfliktami znaczeniowymi i przeszkodami, które utrudniają swobodną wymianę wiedzy (patrz Gangemi, 1999). Rozwiązaniem tych problemów może być stworzenie jednej ustandaryzowanej konceptualizacji dziedziny w której aplikacje te są używane bądź też stworzenie dziedziny niezależnych, stałych odwołań, które pozwolą reprezentować każdą aplikację w jednoznaczny sposób. Należy zwrócić uwagę, iż ze względu na ilość zmian w strukturze jaka musi zostać dokonana w ontologiach podczas procesu ich integracji wg Gangemi (1998) możemy wyróżnić trzy poziomy:

- dopasowania – to najslabsza forma integracji wymagająca przeprowadzenia minimalnych zmian struktury ale wspierająca jedynie ograniczone rodzaje zadań
- częściowej zgodności – wymaga więcej zmian aby wspierać bardziej skomplikowane zadania
- unifikacji albo całkowitej zgodności – może ona wymagać poważnych zmian a nawet konieczność przeorganizowania całej ontologii; w rezultacie otrzymujemy najbardziej kompletną interoperacyjność

Jedną z metodologii wykorzystywana zarówno do analizy i integracji ontologii jest metodologia ONIONS (patrz Gangemi, 1998, 1999). Była ona tworzona od wczesnych lat dziewięćdziesiątych w związku z pojawiającymi się problemami różnorodności konceptualnej różnych źródeł danych. ONIONS został wdrożony i zastosowany w projekcie związanym z wielkimi zbiorami terminologii medycznej. Główne cele metodologii zawierają, wg Gangemi (1999):

- tworzenie dobrze dopasowanego zbioru ogólnych ontologii wspierającego integrację ontologii związanych dziedzinowo z medycyną
- zintegrowanie zbioru ontologii dziedzinowych z zadawalającym poziomem formalizmu oraz konceptualizacji biblioteki ontologii w celu wspierania różnych zadań zawierających m.in. dostęp do danych, integrację zawartości czy też generowanie przewodników
- dostarczenie wyraźnie określonych procedur budowania ontologii

W większości prac poświęconych integracji ontologii autorzy skupiają się na problemie wyboru właściwego języka bądź metody pozwalającej zbudować globalną ontologię z ontologii lokalnych. W pracy Calvanese (2001) autorzy skupiają się na kluczowym problemie a mianowicie na wyspecyfikowaniu mapowania pomiędzy globalną ontologią a ontologiami lokalnymi. Aspekt ten jest szczególnie ważny w przypadku użycia globalnej ontologii do odpowiedzi na pytania w kontekście sieci semantycznej. Autorzy zwracają szczególną uwagę na fakt, iż nie chodzi o podejście w którym lokalne ontologie są traktowane jako krok pośredni w kierunku ontologii globalnej, ale o użycie globalnej ontologii w celu dostania się do informacji zgromadzonych w ontologiach lokalnych.

Wiele ontologii jest budowanych przez organizacje w różnych celach a co się z tym wiąże mogą one reprezentować te same informacje zapisane w różnej formie i na różnym poziomie abstrakcji. Tworzenie mapowania pojęć pomiędzy różnymi ontologiami, wg Calvanese (2001) oznacza, iż pojęcie w jednej ontologii odpowiada "widokowi" w innej ontologii. Warto zwrócić uwagę, iż aby wyrazić mapowanie pomiędzy pojęciami w różnych ontologiach należy zdefiniować odpowiedni język zapytań, który powinien zostać dodany do języka specyfikacji ontologii i rozważony byłby w różnych zadaniach wnioskowania. W pracy Calvanese (2001) autorzy zwracają dodatkowo uwagę na różnicę pomiędzy integracją danych a integracją ontologii. O ile w przypadku danych można założyć, iż źródłem jest baza danych - pojedynczy model wiedzy o tyle takie założenie nie może zostać poczynione w przypadku ontologii, gdzie dowolna teoria logiczna może być przedstawiona za pomocą różnych modeli. Metodologie wykorzystujące słownik WordNet jako narzędzie wspomagające proces łączenia ontologii są nowym podejściem mającym na celu zautomatyzowanie technik tworzenia powiązań pomiędzy różnymi ontologiami oraz zredukowanie ingerencji ludzkiej w ten proces. WordNet opisuje m.in. synonimy, antonimy, uogólnienia oraz uszczegółowienia, które mogą zostać wykorzystane do obliczenia podobieństwa pomiędzy pojęciami w różnych ontologiach.

W pracy Kong (2005) dla łączenia heterogenicznych ontologii dziedzinowych, słownik WordNet został użyty do uszczegółowienia relacji pomiędzy pojęciami. Warto zwrócić uwagę, iż autorzy mówiąc łączenie mają na myśli połączenie różnych ontologii odnośnie tej samej dziedziny w jedną, która unifikuje je wszystkie, co różni się od procesu integracji, w którym buduje się nową ontologię używając do tego celu istniejących ontologii. Wykorzystanie słownika WordNet pozwala na rozwiązanie problemu łączenia ontologii związanego z różnym punktem widzenia inżynierów ontologii na opisywany wycinek świata. Użycie WordNet do głębszej analizy taksonomicznej terminów występujących w ontologiach źródłowych pozwala

na przeprowadzenie bardziej kompletnego procesu łączenia ontologii i zbudowania właściwej hierarchii relacji pojęć.

2.2. Konflikty

Złożoność procesu integracji ontologii zależy w dużej mierze od problemów, które są wynikiem niedopasowania pomiędzy istniejącymi ontologiami. W pracy Klein (2001) została przedstawiona w sposób bardzo klarowny klasyfikacja problemów utrudniających użycie niezależnie stworzonych ontologii. Autor podzielił mogące występować niezgodności na dwie niezależne grupy, tj.:

- różnice pojawiające się na poziomie języka użytego do wyspecyfikowania ontologii albo metamodelu; niezgodności na tym poziomie związane są ze sposobem definiowania elementów składowych ontologii takich jak klasy albo relacje, można wyznaczyć cztery następujące grupy konfliktów:
 - składniowe – są to najbardziej oczywiste niezgodności wynikające z różnego sposobu definiowania elementów
 - w reprezentacji logicznej – są one bardziej skomplikowane i zależą od sposobu definiowania pojęć logicznych
 - znaczenia składników języka – ta sama nazwa w konstrukcji języka może mieć zupełnie inne znaczenie oraz interpretację, może się tak zdarzyć, iż dwie ontologie używają tej samej składni która różni się znaczeniowo
 - ekspresywności języka – niezgodności na poziomie meta modelu związane są z różnymi możliwościami ekspresyjności języków, w których one zostały wyrażone
- drugą grupę stanowi płaszczyna ontologii albo modelu, którą ontologia aktualnie opisuje, w tym przypadku różnice polegają na sposobie w jaki jest zamodelowana dziedzina; konflikty na tym poziomie można podzielić na:
 - konflikty związane ze specyfikacją ale także konceptualizacją dziedziny (np problemy pokrycia, ziarnistości)
 - konflikty związane z stylem modelowania, niezgodności terminologicznych czy też kodowania

Wraz ze wzrostem zainteresowania ontologiami i zadaniami z nimi związanymi (ich przetwarzanie, łączenie oraz integracja) poszukuje się coraz to nowych mechanizmów, które pozwoliłyby na rozwiązywanie pojawiających się konfliktów.

Jednym z takich proponowanych rozwiązań w niniejszej pracy jest właśnie użycie metod wyboru consensusu. Sam proces integracji ontologii na potrzeby consensusu możemy sformułować następująco wg Nguyen (2006): dla danych ontologii O_1, O_2, \dots, O_n proces integracji ontologii powinien zdeterminować jedną ontologię, która najlepiej je reprezentuje. Warto jednak zauważyć, iż integracja ontologii nie zawsze wiąże się z koniecznością rozwiązywania konfliktów. Konflikty pomiędzy integrowanymi ontologiami pojawiają się tylko wtedy, gdy opisują one tą samą rzeczywistość.

Rozwiązywanie konfliktów w ontologiach jest w pewnym sensie związane z procesem przywracania spójności danych. Metody consensusu mogą być nie tylko użyteczne w przypadku występowania różnych wersji wiedzy albo danych o tych samych obiektach w świecie rzeczywistym ale także w przypadku kiedy wspomniane wersje odwołują się do różnych części opisywanego świata. Sam consensus jest natomiast realizowany przez postulaty zdefiniowane dla funkcji wyboru consensusu (patrz Nguyen, 2006).

Konflikty pomiędzy ontologiami mogą być rozważane na płaszczyznach:

- konflikty występujące na poziomie instancji – zakłada się, iż dwie ontologie różnią się wartościami atrybutów dla tego samego pojęcia, w niniejszej pracy zostaną zaproponowane metody wyboru consensusu właśnie dla konfliktów pojawiających się na poziomie instancji
- konflikty na poziomie konceptu – mamy z nimi do czynienia w sytuacji, gdy w różnych ontologiach to samo pojęcie ma różną strukturę, tzn. jest opisane za pomocą innych atrybutów
- konflikty na poziomie relacji – relacje dla tych samych pojęć są różne w różnych ontologiach

2.3. Proponowany model

Tak, jak zostało wcześniej wspomniane dziedziną proponowanej ontologii jest sieć Internet a dokładniej obszar związany z wyszukiwaniem oraz rekomendacją dokumentów webowych. Ontologia hipotetycznie mogłaby być wykorzystana przez rozproszony system agentowy, którego pojedynczy autonomiczny agent zajmowałby się trzema głównymi zadaniami.

- tworzeniem profilu użytkownika
- badaniem preferencji użytkownika – na podstawie analizy historii wyszukiwań oraz analizy dokumentów, które zostały ocenione zarówno przez system jak i przez użytkownika jako relewantne
- rekomendowaniem dokumentów zgodnie z zainteresowaniami użytkownika – na podstawie zgromadzonej wiedzy w ontologii oraz grafu zainteresowań, system mógłby lepiej dopasować się do indywidualnych potrzeb użytkownika poprzez rekomendowanie dokumentów, które byłyby w kręgu jego zainteresowań

3. Consensus jako metoda rozwiązywania konfliktów

Teoria consensusu powstała w obrębie nauk socjologicznych i swoje korzenie posiada w ogólnej teorii wyboru, którą można zdefiniować w następujący sposób wg (Nguyen, 2002): jeśli rozpatrujemy pewien konkretny zbiór obiektów będący podzbiorem uniwersum, oraz dokonujemy wyboru pewnych elementów z tego zbioru na podstawie ustalonych wcześniej kryteriów oraz nasz wybór jest dokonywany w

sposób jednoznaczny, to powtarzając go dla tego samego zbioru źródłowego powinniśmy móc uzyskać ten sam wynik. Dokonując, więc selekcji w sposób deterministyczny możemy mówić o zdefiniowaniu pewnej funkcji wyboru. Natomiast w przypadku teorii consensusu nie wymaga się aby wynik dokonywanego wyboru był podzbiorem elementów należących do zbioru źródłowego a nawet aby miał on tą samą strukturę co elementy analizowanego zbioru (patrz Nguyen, 2006).

Tak samo jak w przypadku eksploracji danych teoria consensusu dotyczy problemów pojawiających się podczas analizowania zbiorów danych w celu wydobycia potrzebnych informacji. Różnica między tymi technikami polega jednak na wynikach jakich nam dostarczają. W przypadku użycia metod eksploracji danych, będących jednym z etapów procesu odkrywania wiedzy z baz danych, znajdujemy ukryte dla człowieka (z uwagi na ograniczone możliwości czasowe) prawidłowości w danych tzw. związki “przyczynowo - skutkowe” występujące w analizowanych danych. Natomiast techniki wyboru consensusu pozwalają na cyt. (Nguyen, 2002): “wyznaczenie takiej wersji dla pewnego zestawu wersji danych, która najlepiej reprezentuje dane wersje lub jest dobrym kompromisem, który powinien być akceptowany przez strony w konflikcie, będące autorami tych wersji danych”.

3.1. Własności funkcji consensusu

W przypadku rozwiązywania konfliktów z wykorzystaniem algorytmów wyboru consensusu, możemy wg Nguyen (2002) spotkać się z dwoma ich rodzajami a mianowicie:

- konfliktami w których szukane rozwiązanie jest niezależne od opinii uczestników konfliktu
- szukane rozwiązanie zależy od stron biorących udział w konflikcie

Konflikty należące do pierwszej grupy związane są z sytuacją, w której uczestnicy konfliktu nie znają prawidłowego rozwiązania z powodu własnej niewiedzy albo innych czynników uniemożliwiających im poznanie, a co za tym idzie ich opinie określają rozwiązanie ale cyt. Nguyen (2002): “nie wiadomo, czy w sposób poprawny i kompletny”. Za przykład takiego konfliktu może wg Nguyen (2002) posłużyć podanie przez różne stacje meteorologiczne różniących się informacji dotyczących prognozy zachmurzenia, opadów, temperatury oraz siły wiatru na konkretnym obszarze geograficznym w określonym czasie. Rozwiązaniem tego problemu będzie jednoznaczny obraz pogody, który dopiero nadejdzie w tym czasie, niezależnie od opinii stacji meteorologicznych. Druga grupa konfliktów, to sytuacje w których strony konfliktu mają bezpośredni wpływ na jego rozwiązanie. Za przykład można podać popularną jazdę figurową na lodzie, w której oceny sędziów decydują, kto zdobędzie wymarzone pierwsze miejsce.

Aby consensus stanowił akceptowalny kompromis wykorzystuje się funkcje odległości pomiędzy wyznaczonym consensusem a opiniami uczestników w celu wyznaczenia struktury obiektu złożonego, czyli wspomnianej wcześniej makrostruktury. Zakłada się więc, iż makrostrukturą uniwersum U jest następująca funkcja opisana w Nguyen (2002): $d : U \times U \rightarrow [0, 1]$, która spełnia następujące warunki:

- $(\forall x, y \in U) (d(x, y) \geq 0)$
- $(\forall x, y \in U) [d(x, y) = 0 \Leftrightarrow x = y]$
- $(\forall x, y \in U) [d(x, y) = d(y, x)]$

Jak widać, funkcja d spełnia wszystkie warunki funkcji odległości, nie jest jednak konieczne aby spełniała ona warunek nierówności trójkąta co oznacza, iż funkcja d nie musi być metryką. Powinna ona jednak spełniać pewne aksjomaty dla dowolnego profilu X , aby była ona zgodna z teorią wyboru consensusu opisaną dokładnie w Nguyen, (2002):

- Niezawodności(Re) – oznaczający, iż dla każdego profilu istnieje rozwiązanie i można wyznaczyć consensus
- Spójności(Co) – postulat wymaga, iż jeśli x jest consensusem dla pewnego profilu, to po jego rozszerzeniu o ten element, x powinien być nadal consensusem dla nowego profilu
- Quasi - jednomyślności(Qu) – jeśli element x nie jest consensusem pewnego profilu to powinien on być consensusem dla profilu, który zawiera n elementów x
- Proporcjonalności(Pr) – jeśli mamy do czynienia z dwoma profilami i jeden z nich jest większy od drugiego, to także suma odległości wybranego dla niego consensusu będzie większa
- 1 - optymalności(O_1) – wymaga on aby consensus był jak najbardziej zbliżony do elementów tworzących profil

3.2. Potrzeba integracji oraz identyfikacja konfliktów

Proces integracji ontologii został potraktowany jako sposób wymiany informacji pomiędzy agentami reprezentującymi użytkowników o podobnych profilach, tzn. takich, które np. różnią się jedynie wektorami preferencji. Analizując obszar zastosowania systemu możemy wyodrębnić dwie sytuacje, w których zaistniałaby potrzeba przeprowadzenia procesu integracji, a mianowicie:

- w momencie tworzenia nowego profilu dla grupy osób, która wcześniej nie była utworzona w systemie, np. nowy użytkownik chcący korzystać z systemu jest zainteresowany zarówno informatyką jak i chemią a co za tym idzie można wykorzystać już istniejące ontologie stworzone tylko dla profilu zainteresowania chemią bądź informatyką w celu utworzenia nowej ontologii stanowiącej początkowy stan wiedzy (podglądu na wycinek świata) dla nowego agenta
- w sytuacji, gdy wielu agentów zajmuje się podobnym profilem (czyli np. tylko informatykami), wtedy istnieje potrzeba aby agenci ci mogli wymieniać się wiedzą między sobą - każdy z agentów przed zintegrowaniem posiadałby inny

zbiór dokumentów, terminów oraz powiązań między nimi. W wyniku przeprowadzenia integracji powstałaby ontologia najbardziej zbliżona do ontologii wejściowych, która stałaby się nowym stanem wiedzy agentów biorących udział w procesie integracji ontologii

3.3. Integracja grafów zainteresowań

Zaproponowany algorytm integracji będzie wyznaczał graf g^* będącym consensusem dla zbioru grafów $G = \{g_1, g_2, \dots\}$ z integrowanych ontologii wejściowych, przy zachowaniu kryterium optymalności $\partial(g^*, G) \rightarrow \min$ tzn. suma odległości między grafem g^* powstałym w wyniku wyboru consensusu a elementami zbioru G jest minimalna. Zdefiniujemy zatem funkcję $d' : \vec{X} \times \vec{X} \rightarrow [0, 1]$ będącą funkcją odległości pomiędzy dwoma wektorami \vec{X} o pewnej długości $N > 1$, zakładając, iż zbiór wartości binarnych zostanie dodatkowo powiększony o wartość ? oznaczającą niepewność oraz zakładając, iż istnieje co najmniej jedna para wartości pozycji wektorów taka, że żadna z tych wartości nie jest równa ? tzn. gdzie x_i i y_i są i -tymi wartościami wektorów \vec{X} i \vec{Y} . Możemy zatem zapisać $\forall_x x \in \{0, 1, ?\}$ gdzie $\vec{X} = \langle x_1, \dots, x_n \rangle$ dla $n > 0$. Funkcję możemy zatem zdefiniować w następujący sposób:

$$d'(\vec{X}, \vec{Y}) = \frac{\sum_{i=1}^N t(x_i, y_i)}{\sum_{i=1}^N s(x_i, y_i)} \quad (1)$$

gdzie

$$t(x_i, y_i) = \begin{cases} 1, & \text{gdy } x_i \neq y_i \wedge x_i, y_i \in \{0, 1\}, \\ 0, & \text{w przeciwnym wypadku.} \end{cases} \quad (2)$$

$$s(x_i, y_i) = \begin{cases} 1, & \text{gdy } x_i, y_i \in \{0, 1\}, \\ 0, & \text{w przeciwnym wypadku.} \end{cases} \quad (3)$$

Oznacza to, że odległością dwóch wektorów \vec{X} i \vec{Y} będziemy nazywać iloraz liczby oznaczającej liczbę pozycji na których wektory te różnią się - pomijając pary, w których jedna z wartości wynosi ? - do liczby pozycji, dla których wektory te mają wartości określone.

3.3.1. Wyznaczanie consensusu dla wektorów trójnarych

Wejście: zbiór wektorów trójnarych X , gdzie $\forall_{x \in X} \text{length}(x) = N$ oraz $\text{card}(X) = M$
Wyjście: wektor x^* dla którego spełnione jest założenie $d'(x^*, X) \rightarrow \min$
Poszczególne pozycje wektora x^* obliczane są w następujący sposób:

1. Niech $i = 1$
2. $x_i^* = \left[\frac{\sum_{x \in X} s(x_i)}{M - \sum_{x \in X} n(x_i)} \right]$ (4)

gdzie

$$s(x_i) = \begin{cases} 1, & \text{gdy } x_i = 1, \\ 0, & \text{w przeciwnym wypadku.} \end{cases} \quad (5)$$

$$n(x_i) = \begin{cases} 1, & \text{gdy } x_i = ?, \\ 0, & \text{w przeciwnym wypadku.} \end{cases} \quad (6)$$

	Termin 1	Termin 2	Termin 3	Termin 4
Termin 1	0	1	?	0
Termin 2	0	0	?	1
Termin 3	?	?	?	?
Termin 4	1	0	?	0

Rysunek 1. Schemat przykładowej ontologii

- Niech $i = i + 1$, jeśli $i \leq N$ idź do punktu 2

Otrzymany wektor x^* spełnia warunek $d'(x^*, X) = \sum_{x \in X} d'(x^*, x) \rightarrow \min$

3.3.2. Procedura wyznaczania consensusu dla macierzy trójnarnych

Skierowany graf g będący reprezentacją zainteresowań użytkownika korzystającego z Internetu można przedstawić jako macierz trójnarną o wartościach $?, 0, 1$, czyli innymi słowy jako uporządkowany ciąg wektorów \vec{X} . Każdemu wierszowi oraz kolumnie odpowiada jeden termin, zatem dla macierzy $N \times N$ rozpatrujemy N różnych terminów. Jeśli na pozycji x_{ij} mamy zapisaną wartość 1 oznacza ona, że istnieje powiązanie pomiędzy terminem t_i a t_j . Wartość $?$ oznaczająca niepewność zostaje wpisana w kolumnie oraz wierszu odpowiadającemu terminowi, który nie występuje w konkretnej ontologii - oznacza to, iż każdy graf zainteresowań jest przedstawiony za pomocą macierzy kwadratowej, której kolumny oraz wiersze reprezentują terminy ze wszystkich integrowanych ontologii. Rozpatrując graf g jako macierz $N \times N$ mamy n wektorów o długości n . Niech funkcja będzie funkcją odległości pomiędzy dwoma macierzami trójnarnymi $N \times N$ zdefiniowana w następujący sposób

$$\partial(X, Y) = \frac{\sum_{x \in X} d'(x_i, y_i)}{N} \quad (7)$$

gdzie x_i jest i -tym wierszem (wektorem) w macierzy X o długości N . Ostatecznie procedura wyznaczenia grafu g^* mogłaby wyglądać następująco: Wejście: zbiór macierzy trójnarnych G , gdzie $\forall_{g \in G} \text{length}(G) = N$ oznacza liczbę wierszy w macierzy g . Wyjście: macierz g^* .

- Niech $i = 1$
- Wyznacz wektor g_i^* korzystając z procedury wyznaczania consensusu dla wektorów trójnarnych, gdzie $X = \bigcup_{g \in G} g_i$
- Niech $i = i + 1$ jeśli $i \leq N$ idź do punktu 2

W celu weryfikacji proponowanego podejścia została opracowana funkcja rankingująca pozwalająca ocenić poprawność wyników. Została ona wraz ze szczegółami

proponowanego podejścia (formalnymi dowodami) zaprezentowana w pracy Rusin, (2006) oraz Rusin, Nguyen (2006).

4. Podsumowanie

Zagadnienia związane z ontologiami stają się w dzisiejszych czasach coraz bardziej popularne i wzbudzają coraz większe zainteresowanie wśród naukowców zajmujących się zarówno sposobami gromadzenia oraz wymiany wiedzy pomiędzy systemami komputerowymi jak również tych mających na uwadze budowanie i rozwijanie sieci semantycznej. Ontologie już teraz znajdują zastosowanie w wielu sferach życia m.in. począwszy od zagadnień związanych z wyszukiwaniem informacji poprzez handel elektroniczny, medycynę etc. Wiele grup badaczy tworzy ontologie do własnych celów. Zapisane są one z wykorzystaniem różnych formalizmów uwzględniających konkretne cechy obiektów i opisują różne fragmenty otaczającego nas świata. Aby umożliwić swobodną wymianę informacji, istnieje zazwyczaj potrzeba dokonania integracji częściowo pokrywających się ontologii, aby uzyskać spójny i kompletny obraz przedstawianego w nich świata. Niestety, proces integracji ontologii został rozpoznany jako skomplikowany do przeprowadzenia i trudny do zautomatyzowania ze względu na niezgodności zarówno na poziomie języka jak i całego modelu ontologii.

W pracy zostało w skrócie scharakteryzowane nowe podejście do problemu integracji ontologii wykorzystujące algorytmy wyboru consensusu do rozwiązywania pojawiających się konfliktów wielowartościowych. Proponowana procedura obejmuje integrację grafów zainteresowań i może zostać dostosowana i użyta do integracji dowolnych grafów, które można przedstawić jako macierz trójkątną. Wydaje się, po przeprowadzeniu wstępnych eksperymentów (patrz Rusin, 2006), iż proponowane rozwiązanie daje zadawalające rezultaty i nakreśla kolejne kierunki badań, np wykorzystania teorii wyboru consensusu do procesu integracji struktur hierarchicznych.

Literatura

- CALVANESE, D. and DE GIACOMO, G. and LENZERINI, M. (2001) A framework for ontology integration *In Proceeding of the First Semantic Web Working Symposium*, 303–316
- FENSEL, D. (2001) *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer.
- GANGEMI, A. and PISANELLI, D. M. and STEVE, G. (1998) Ontology Integration: Experiences with Medical Terminologies *Proceedings of FOIS'98, Formal Ontology in Information Systems*, 163–178
- GANGEMI, A. and PISANELLI, D. M. and STEVE, G. (1999) An Overview of the ONIONS project: Applying Ontologies to the Integration of Medical Terminologies *Data and Knowledge Engineering* **31**, 183–220

- GRUBER, T. R. (1993) A translation approach to portable ontology specifications. *Knowledge Acquisition* **5**, 2, 199–220.
- KLEIN, M. (2001) Combining and relating ontologies: an analysis of problems and solutions *Workshop on Ontologies and Information Sharing, IJCAI'01*, 53–62.
- KONG, H. and HWANG, M. and KIM, P. (2005) A New Methodology for Merging the Heterogeneous Domain Ontologies based on the WordNet *In Proceeding of International Conference on Next Generation Web Services Practices*, 235–240.
- NGUYEN, N. T. (2002) *Metody wyboru consensusu i ich zastosowanie w rozwiązywaniu konfliktów w systemach rozproszonych*. Oficyna Wydawnicza Politechniki Wrocławskiej.
- NGUYEN, N. T. (2006) Conflicts of Ontologies - Classification and Consensus - based Methods for Resolving *Proceedings of KES2006 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems* **2**, 267–274.
- NGUYEN, N. T. and RUSIN, M. (2006) A Consensus-Based Approach for Ontology Integration *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06)*, 514–517.
- NOY, N. F. and MUSEN, M. A. (1999) SMART: Automated Support for Ontology Merging and Alignment *In Proceedings of the 12th Workshop on Knowledge Acquisition, Modelling and Management (KAW'99)*, 1–20.
- PINTO, H. S. and GÓMEZ-PÉREZ, A. and MARTIS, J. P. (1999) Some issues on ontology integration. *IJCAI-99 workshop on Ontologies and Problem-Solving: Lesson learned and Future Trends* **18**, 7.1–7.11.
- RUSIN, M. (2006) Praca magisterska *Metody integracji ontologii z wykorzystaniem algorytmów wyboru consensusu*. Politechnika Wroclawska.

Methods of the ontology integration based on consensus choice theory

Ontologies become a very important element of modern information systems. In many cases creating new system involves performing integration of existing ontologies. In this process it is often necessary to resolve the conflicts between ontologies. The paper proposes an approach which uses the theory of consensus choice to solve the multivalent conflicts which appear in integration process on an instance level.