

Andrzej Bytniewski, Marcin Hernes

Uniwersytet Ekonomiczny we Wrocławiu

ALGORYTM WYZNACZANIA CONSENSUSU W SYTUACJI KONFLIKTU WIEDZY NIEUSTRUKTURALIZOWANEJ W ROZPROSZONYCH INFORMATYCZNYCH SYSTEMACH WSPOMAGAJĄCYCH ZARZĄDZANIE

Streszczenie: W artykule przedstawiono problematykę rozwiązywania konfliktów wiedzy nieustrukturalizowanej w systemach wspomagających zarządzanie. W pierwszej części artykułu zwrócono uwagę, że wiedza nieustrukturalizowana stanowi istotny element funkcjonowania organizacji, zwłaszcza w aspekcie jej konkurencyjności. Następnie scharakteryzowano źródła i istotę konfliktów wiedzy nieustrukturalizowanej gromadzonej w przedsiębiorstwach w postaci szeroko pojętych dokumentów tekstowych. W końcowej części artykułu, przyjmując reprezentację dokumentów tekstowych w postaci binarnej macierzy częstości słów, opracowano algorytm consensusu umożliwiający rozwiązywanie konfliktów wiedzy nieustrukturalizowanej, co stanowiło podstawowy cel artykułu.

Słowa kluczowe: wiedza nieustrukturalizowana, konflikty wiedzy, macierz częstości słów, algorytmy consensusu.

1. Wstęp

Informatyczne systemy wspomagające zarządzanie, funkcjonujące w środowisku rozproszonym, odgrywają obecnie bardzo dużą rolę w szeroko pojętej działalności człowieka i różnego rodzaju organizacji. Rozwój sieci komputerowych spowodował, że aktualnie systemy służące wspomaganiu zarządzania są systemami rozproszonymi [Coulouris i in. 1998], funkcjonującymi właśnie w sieci komputerowej, głównie w Internecie. Coraz częściej rozpatrywane systemy realizowane są w ramach usług chmury obliczeniowej, a wobec tego widoczny jest proces zanikania granicy pomiędzy lokalnymi systemami funkcjonującymi w przedsiębiorstwie a systemami funkcjonującymi w sieci globalnej. Rozpatrywane systemy pozwalają na zbieranie i przetwarzanie bardzo dużej ilości informacji, ale również znacznie ułatwiają zarządzanie wiedzą w organizacji [Bytniewski 2005; Kisielnicki 2008; Adamczewski 2009], ponieważ na przykład umożliwiają automatyczne wyciąganie wniosków

na podstawie posiadanych informacji. Z tego też względu w niniejszym artykule przyjęto zasadę, że informacja pozyskana między innymi z zasobów internetowych traktowana jest jako wiedza, ponieważ celem funkcjonowania rozproszonych informatycznych systemów zarządzania jest przede wszystkim generowanie nowej wiedzy na podstawie posiadanej informacji oraz zarządzanie tą wiedzą. W systemie informatycznym funkcjonującym w środowisku rozproszonym, oprócz rozproszenia sprzętu, występuje również rozproszenie zasobów informacyjnych, zawierających często wiedzę nieustrukturalizowaną.

Podstawowym problemem okazuje się jednak obecnie przetwarzanie tej wiedzy w omawianych systemach. Zauważmy bowiem, że wiedza zawarta w informatycznym systemie zarządzania jest z reguły ustrukturalizowana i w systemie tym używane są różnego rodzaju metody przetwarzania wiedzy ustrukturalizowanej oraz rozwiązywania konfliktów takiej wiedzy. Coraz częściej jednak przedsiębiorstwa wykorzystują wiedzę nieustrukturalizowaną, przede wszystkim właśnie zawartą w cyberprzestrzeni, która daje możliwość uzyskania większej elastyczności i konkurencyjności organizacji. Wiedza nieustrukturalizowana wspiera więc w znacznym stopniu wiedzę ustrukturalizowaną organizacji. Głównie jest to wiedza zapisana w języku naturalnym, a więc wiedza przetwarzana symbolicznie (a nie numerycznie), na przykład opinie użytkowników forum na dany temat. Można przyjąć, że wiedza ta zawarta jest w różnego rodzaju dokumentach tekstowych.

Należy jednocześnie zauważyć, że częstym zjawiskiem jest występowanie konfliktów w obszarze wiedzy nieustrukturalizowanej. Przykładem może być opinia użytkowników na temat danego produktu oferowanego w sklepach internetowych. Niektórzy użytkownicy mogą wypowiadać się pozytywnie, inni negatywnie, a jeszcze inni mogą nie mieć zdania na temat danego produktu. Prawidłowe rozwiązywanie tego typu konfliktów jest niezwykle trudne, a zarazem istotne, ponieważ może prowadzić do poprawy funkcjonowania systemu wspomagającego zarządzanie, a w konsekwencji spowodować, że organizacja wykorzystująca system może być bardziej elastyczna i konkurencyjna.

W literaturze przedmiotu (np. [Dyk, Lenar 2006; Barthlemy 1992; Michalczyk, Waścicki 2009]) spotyka się różne metody rozwiązywania konfliktów wiedzy, zwłaszcza ustrukturalizowanej. Mają one jednak pewne niedociągnięcia. Na przykład metody negocjacji gwarantują osiągnięcie pożądanego kompromisu pomiędzy stronami konfliktu, jednak realizowane jest to kosztem zwiększonej komunikacji pomiędzy elementami systemu, co oczywiście wpływa na szybkość jego funkcjonowania. Natomiast metody z grupy dedukcyjno-obliczeniowych (na przykład metody oparte na teorii gier, mechanice klasycznej, wywodzące się z dziedziny badań operacyjnych, wywodzące się z nauk behawioralnych i socjologii, metody wielokryterialne, metody wyboru czy też metody consensusu) nie wpływają w dużym stopniu na szybkość działania systemu, jednak metody te, oprócz metod consensusu, nie gwarantują osiągnięcia dobrego kompromisu. Użytkownik natomiast wymaga od systemu dobrej wydajności (często działania w czasie zbliżonym do rzeczywistego) i sprawnego rozwiązywania konfliktów wiedzy.

Dlatego jako metody rozwiązywania konfliktów wiedzy lepiej zastosować metody consensusu, mające najczęściej charakter heurystyczny, które należą do grupy metod dedukcyjno-obliczeniowych. W odróżnieniu od innych metod z tej grupy gwarantują one osiągnięcie dobrego kompromisu, głównie ze względu na fakt, że w consensusie pod uwagę brane są opinie wszystkich stron konfliktu (consensus jest reprezentacją wszystkich stron konfliktu), a jednocześnie wyznaczanie consensusu nie wymaga ingerencji w wewnętrzne stany wiedzy węzłów systemu (na przykład wiedzy agentów programowych funkcjonujących w systemie), co ma miejsce na przykład w przypadku metod negocjacji. Jednocześnie zastosowanie metod consensusu nie wymaga dużej mocy obliczeniowej procesora oraz zwiększonej komunikacji pomiędzy elementami systemu, dzięki czemu pozwala na funkcjonowanie systemu w czasie zbliżonym do rzeczywistego.

Dotychczas jednak teoria consensusu nie znalazła zastosowania w rozwiązywaniu konfliktów wiedzy nieustrukturalizowanej w rozproszonych informatycznych systemach wspomagających zarządzanie, między innymi ze względu na trudności związane z przetwarzaniem tego typu wiedzy.

Zatem celem artykułu jest opracowanie algorytmu wyznaczania consensusu w celu rozwiązywania konfliktów wiedzy nieustrukturalizowanej w informatycznych systemach zarządzania posiadających charakter rozproszony, reprezentowanych w postaci binarnych macierzy częstości słów¹.

2. Konflikty wiedzy nieustrukturalizowanej

Wiedza nieustrukturalizowana w rozproszonych systemach wspomagających zarządzanie występuje w postaci niesformatowanych tekstowych baz danych, przechodzących różnorodne typy dokumentów tekstowych, takie jak artykuły gazetowe, książki (e-book), wiadomości e-mail, strony WWW i szeroko rozumiane dokumenty tekstowe. Dokumenty te opisują pewne zjawiska występujące w świecie rzeczywistym, w środowisku funkcjonowania danej organizacji. Dokumenty tekstowe nie posiadają żadnej wewnętrznej struktury, inaczej mówiąc – wiedza w nich zawarta jest nieustrukturalizowana lub słabo ustrukturalizowana. Na przykład struktura e-maila zawiera część informacji sformatowanych, dotyczących na przykład nadawcy, odbiorcy, rozmiaru, tematu wiadomości. Natomiast już pozostała jego część (właściwa treść) jest łańcuchem znaków pozbawionym wewnętrznej struktury. Trzeba wyraźnie podkreślić, że dokumenty tekstowe często są źródłem ważnej i użytecznej wiedzy.

Często dokumenty tekstowe reprezentowane są w bazach danych na podstawie słów kluczowych zawartych w dokumencie i ontologii (reprezentacji symbolicznej wiedzy). Przy takiej reprezentacji jednak bardzo utrudnione jest porównywanie dokumentów, a szczególnie mierzenie odległości pomiędzy dokumentami, przy czym odległość rozumiana jest tu jako stopień podobieństwa dokumentów. Alternatywnym

¹ Macierz ta reprezentuje częstość występowania słów kluczowych w danym dokumencie.

więc podejściem do reprezentacji dokumentu tekstowego jest podejście oparte na reprezentacji wektorowej dokumentu (reprezentacji numerycznej wiedzy). Podstawowa idea reprezentacji wektorowej sprowadza się do tego, że dowolny dokument jest reprezentowany w postaci wektora częstości występowania słów kluczowych, nazywanych również terminami indeksującymi². Stąd zbiór M przechowywanych dokumentów tekstowych można przedstawić w postaci macierzy częstości słów, określanej często w literaturze jako *Term Frequency Matrix* (TFM), której element $\text{TFM}[di, ti]$ reprezentuje liczbę wystąpień słowa kluczowego ti (gdzie: $1 < i < N$) w dokumencie di (gdzie: $1 < i < M$). Dowolny dokument di jest reprezentowany w postaci wektora częstości występowania słów kluczowych. Element $\text{TFM}[di, ti]$ nazywamy wagą słowa ti w dokumencie di . W najprostszej reprezentacji boolowskiej wagi słów w wektorze dokumentu mogą przyjmować tylko dwie wartości: 0 lub 1 (tab. 1). Waga słowa ti w dokumencie di równa 1 oznacza, że słowo ti występuje w danym dokumencie di , jeżeli natomiast słowo ti nie występuje w dokumencie di , waga słowa ti w dokumencie di jest równa 0. Należy zauważyć, że reprezentacja boolowska dokumentów określa tylko, czy słowo kluczowe występuje, czy też nie występuje. Zatem na przykład dokument, w którym słowo kluczowe występuje jeden raz, jest równy dokumentowi, w którym to samo słowo kluczowe występuje wiele razy. Wynika z tego, że reprezentacja boolowska dokumentów odpowiada ich reprezentacji na podstawie słów kluczowych.

Tabela 1. Przykładowa reprezentacja boolowska macierzy TFM

	t_1	t_2	t_3	t_4	t_5
d_1	0	1	1	1	0
d_2	1	1	0	1	0
d_3	1	0	1	1	1
d_4	1	0	1	1	0

Źródło: opracowanie własne.

Zbiór słów kluczowych, wykorzystywanych w podejściu reprezentacji dokumentów w postaci macierzy TFM, może być bardzo duży. Macierz ta może być wykorzystana również do reprezentacji wiedzy nieustrukturalizowanej w systemach zarządzania, które mają charakter rozproszony.

Informatyczne systemy zarządzania mające charakter rozproszony wychodzą na przeciw oczekiwaniom współczesnych użytkowników dlatego, że mają takie własności, jak dzielenie zasobów, otwartość, współbieżność, skalowalność, przezroczystość i tolerowanie uszkodzeń [Coulouris i in. 1998]. Jednakże systemy rozproszone, w tym również systemy zarządzania, mają także cechy heterogeniczne, które generują problemy wynikające ze stosowania tego rodzaju systemów. Należą do nich

² Terminy indeksujące to uporządkowany zbiór słów kluczowych.

między innymi trudności w zachowaniu bezpieczeństwa zasobów, generowanie różnych wyników dla tego samego zadania (na przykład w wyniku stosowania różnych algorytmów rozwiązania tego zadania) w sytuacji, kiedy tak naprawdę potrzebne jest tylko jedno rozwiązanie, lub też występowanie niejednorodności struktur danych.

Należy jednocześnie zauważyć, że najczęstszym zjawiskiem jest występowanie konfliktów w obszarze wiedzy nieustrukturalizowanej. Przykładem może być opinia użytkowników na temat danego produktu oferowanego w sklepach internetowych. Niektórzy użytkownicy mogą wypowiadać się pozytywnie, inni negatywnie, a jeszcze inni mogą nie mieć zdania na temat danego produktu. Poza tym w innym sklepie internetowym opinii na temat tego samego produktu mogą być zupełnie odwrotne. Kolejnym przykładem mogą być dokumenty tworzone przez pracowników opisujące czynności lub zjawiska zachodzące w organizacji. Te same czynności lub zjawiska mogą być przez każdego z pracowników opisane w różny sposób, a jeszcze w inny przez osobę trzecią (klienta, dostawcę).

Prawidłowe rozwiązywanie tego typu konfliktów jest niezwykle trudne, a zarazem istotne, ponieważ może prowadzić do poprawy funkcjonowania systemu wspomagającego zarządzanie, a w konsekwencji spowodować, że organizacja wykorzystująca system może być bardziej elastyczna i konkurencyjna.

Konflikty wiedzy występują wtedy, gdy do tych samych obiektów świata przyporządkowane są różne atrybuty lub do tych samych atrybutów (cech) przyporządkowane są różne wartości [Katarzyniak, Nguyen 2000].

Z analizy literatury przedmiotu [Manning, Shutze 1999; De Long, Seeman 2000] wynika również, że dotychczas nie istnieją uznane metody rozwiązywania konfliktów wiedzy nieustrukturalizowanej, reprezentowanej symbolicznie, co ma związek z problemem przetwarzania tego rodzaju wiedzy. Obecnie stosuje się hybrydowe metody przetwarzania wiedzy nieustrukturalizowanej, polegające na strukturalizacji wiedzy i następnie przetworzeniu symbolicznym (na przykład z wykorzystaniem systemów ekspertowych czy też algorytmów genetycznych) lub też na przekształceniu wiedzy w reprezentację numeryczną i następnie przetworzeniu numerycznym (na przykład z wykorzystaniem sieci neuronowych czy też systemów logiki rozmytej). W obu przypadkach w celu ekstrakcji wiedzy dokumenty mogą podlegać eksploracji danych [Lula 2005; Ziemia 2005]. Często w procesie analizy dokumentów tekstowych wykorzystuje się takie metody, jak uczenie maszynowe [Frank, Bouckaert 2006], lub reguły, na podstawie których dokonuje się identyfikacji (adnotacji) fragmentów tekstu dotyczących konkretnego tematu [Abramowicz i in. 2013].

W obliczu przedstawionych problemów w praktyce gospodarczej dokonuje się również transformacji „w drugą stronę” – dane ustrukturalizowane transformuje się w dokumenty zapisane w języku naturalnym, a następnie przeprowadza się „ręczną” analizę tych dokumentów. Jednakże podejście to jest mało efektywne ze względu na jego pracochłonność i czasochłonność, a przecież turbulentne otoczenie gospodarcze wymusza na decydentach podejmowanie decyzji w czasie zbliżonym do rzeczywistego.

Wykorzystanie metod consensusu w rozwiązywaniu konfliktów wiedzy nieustrukturalizowanej umożliwi przedstawienie przez system użytkownikowi jednej, wiarygodnej wersji reprezentacji zbioru dokumentów tekstowych opisujących ten sam obiekt lub zjawisko, przez co zostaną usunięte rozterki decydenta w procesie szybkiego podejmowania decyzji.

3. Algorytm wyznaczania consensusu

Consensus w ogólnym pojęciu oznacza porozumienie. Consensus pewnego zbioru (profilu) dokumentów tekstowych może być nowym dokumentem (hipotetycznym), utworzonym na podstawie dokumentów znajdujących się w tym profilu.

Wyznaczanie consensusu składa się z trzech podstawowych etapów. W pierwszym etapie należy określić sposób reprezentacji dokumentów tekstowych. W niniejszym artykule założono, że dokumenty te reprezentowane są w postaci binarnych wektorów częstości z wykorzystaniem macierzy TFM. W drugim etapie niezbędne jest zdefiniowanie funkcji obliczania odległości pomiędzy poszczególnymi wariantami. Trzeci etap to opracowanie algorytmów wyznaczania consensusu, czyli wyznaczania takiej reprezentacji zbioru dokumentów (profilu), że odległość pomiędzy tą reprezentacją (consensusem) a poszczególnymi dokumentami profilu (przechowywanymi w rozproszonym informatycznym systemie wspomagającym zarządzanie) jest minimalna (według różnych kryteriów). Należy zauważyć, że profilu nie stanowią wszystkie dokumenty tekstowe przechowywane w systemie, lecz dokumenty powiązane ze sobą tematycznie, na przykład jeden profil mogą stanowić dokumenty zawierające opinie użytkowników na temat produktu p_1 , zaś inny profil mogą stanowić dokumenty zawierające opinie użytkowników na temat produktu p_2 .

Formalna definicja profilu dokumentów tekstowych przedstawia się następująco:

Definicja 1

Dany jest zbiór N terminów indeksujących (słów kluczowych) $T = \{t_1, t_2, \dots, t_N\}$, gdzie $t_i \in \{0,1\}$.

Profilem $D = \{d_1, d_2, \dots, d_M\}$ nazywamy zbiór M dokumentów tekstowych opisanych za pomocą wektorów częstości terminów indeksujących skończonego zbioru T , taki że:

$$\begin{aligned} d_1 &= \langle t_{1(d_1)}, t_{2(d_1)}, \dots, t_{N(d_1)} \rangle \\ d_2 &= \langle t_{1(d_2)}, t_{2(d_2)}, \dots, t_{N(d_2)} \rangle \\ &\dots\dots\dots \\ d_M &= \langle t_{1(d_M)}, t_{2(d_M)}, \dots, t_{N(d_M)} \rangle, \end{aligned} \tag{1}$$

gdzie $t_{i(d_x)}$ oznacza wartość terminu indeksującego t_i w dokumencie d_x .

Główną zaletą reprezentacji wektorowej dokumentów, w stosunku do reprezentacji opartej na zbiorze słów kluczowych, jest możliwość zdefiniowania miary odległości pomiędzy dokumentami lub zapytaniem użytkownika, co jest istotą drugiego etapu wyznaczania consensusu. Jeżeli określona jest wektorowa reprezentacja dokumentu, to dokumenty o podobnej tematyce powinny charakteryzować się podobną częstością występowania tych samych słów kluczowych. Mając do dyspozycji wektorową reprezentację dokumentów, można interpretować każdy dokument jako punkt w T -wymiarowej przestrzeni, której wymiary odpowiadają poszczególnym słowom kluczowym.

Zatem do oceny odległości pomiędzy dokumentami oraz pomiędzy dokumentami a zapytaniem można wykorzystać dowolne miary stosowane do mierzenia odległości w przestrzeni euklidesowej. Miary te spełniają zawsze warunki metryczne (funkcja odległości jest metryką). Przyjęcie jednak odległości euklidesowej związane jest z niedogodnościami wynikającymi z dużego wpływu zmian skali współrzędnych na wyniki grupowania elementów przestrzeni cech, na przykład słów kluczowych (aby uniknąć tego problemu, należy dokonać normalizacji przestrzeni wartości cech, co powoduje w konsekwencji większą złożoność algorytmu obliczania odległości). Poza znanymi miarami odległości, stosowanymi w wielowymiarowych przestrzeniach euklidesowych, dla potrzeb systemów wyszukiwania informacji opracowano wiele specyficznych miar, takich jak: odległość kosinusowa czy odległość Hamminga. W niniejszym artykule w celu obliczania odległości zostanie wykorzystana metryczna miara odległości Hamminga. W literaturze przedmiotu (np. [Cholewa 1983]) stwierdza się bowiem, że zaletą tego typu odległości jest brak wpływu niezależnego powiększania współrzędnych na jej wartość.

Odległość Hamminga pomiędzy dwoma ciągami tekstowymi o równej długości jest to liczba miejsc, w których odpowiadające im symbole są różne. Innymi słowy, mierzy się minimalną liczbę podstawień wymaganych, by zmienić jeden ciąg w drugi, lub liczbę błędów, które zmieniają jeden ciąg w drugi [Hamming 1950]. Odległość Hamminga spełnia wszystkie warunki metryczne i formalne, a w odniesieniu do wektorów binarnych definiowana jest następująco:

Definicja 2

Niech d_1, d_2 będą binarnymi wektorami częstości charakteryzującymi dokumenty tekstowe. Wówczas:

$$\omega(d_1, d_2) = \sum_{i=1}^N [d_1[i] \circ d_2[i]], \quad (2)$$

gdzie $d_x[i] = \{0, 1\}$ ($i = \{1, \dots, M\}$) oznacza wartość i -tego terminu indeksującego w wektorze d_x , natomiast symbol \circ oznacza alternatywę wykluczającą, czyli operację, w wyniku której otrzymujemy następujące wartości:

$$[d_1[i] \circ d_2[i]] = 0 \quad \Leftrightarrow \quad d_1[i] = d_2[i],$$

$$[d_1[i] \circ d_2[i]] = 1 \Leftrightarrow d_1[i] \neq d_2[i]. \quad (3)$$

Przykład 1

Niech dane będą binarne wektory częstości charakteryzujące dokumenty tekstowe:

$$d_1 = \langle 1, 0, 0, 1, 0, 1, 1 \rangle,$$

$$d_2 = \langle 0, 1, 1, 1, 0, 0, 1 \rangle.$$

Odległość Hamminga liczona jest w następujący sposób:

$$[d_1[1] \circ d_2[1]] = 1,$$

$$[d_1[2] \circ d_2[2]] = 1,$$

$$[d_1[3] \circ d_2[3]] = 1,$$

$$[d_1[4] \circ d_2[4]] = 0,$$

$$[d_1[5] \circ d_2[5]] = 0,$$

$$[d_1[6] \circ d_2[6]] = 1,$$

$$[d_1[7] \circ d_2[7]] = 0.$$

Odległość wynosi zatem $\omega(d_1, d_2) = 1 + 1 + 1 + 0 + 0 + 1 + 0 = 4$.

Mając zdefiniowaną funkcję odległości, można przejść do trzeciego etapu, czyli opracowania algorytmu wyznaczania consensusu profilu dokumentów tekstowych.

Początkowo teoria consensusu dotyczyła prostych struktur, takich jak porządek liniowy lub częściowy porządek. Następnie zaczęto zajmować się bardziej złożonymi strukturami, takimi jak podziały, hierarchie, n -drzewa [Barthlemy 1992]. Obecnie prace nad teorią consensusu dotyczą struktur wieloatrybutowych i wielowartościowych, jak na przykład struktura wiedzy agenta czy też struktura decyzji ekonomicznej. Tak więc dotychczasowe prace dotyczące teorii consensusu odnoszą się tylko do rozwiązywania konfliktów wiedzy dobrze ustrukturalizowanej.

W niniejszym artykule postanowiono rozszerzyć teorię consensusu o rozwiązywanie konfliktów wiedzy nieustrukturalizowanej. Funkcja consensusu na ogół jest funkcją minimalizującą sumę odległości tego consensusu do wszystkich elementów profilu. W pracach [Nguyen 2002; Sobieska-Karpińska, Hernes 2012] stwierdzono, że lepszą funkcją odległości, ze względu na fakt większej równomierności consensusu (czyli consensus jest równomiernie zbliżony do wszystkich elementów profilu), jest funkcja minimalizująca sumę kwadratów odległości consensusu do wszystkich elementów profilu. Wyznaczanie consensusu przy zastosowaniu tego typu funkcji jest problemem NP-zupełnym. W odniesieniu do reprezentacji dokumentów tekstowych za pomocą binarnych wektorów częstości funkcję tę można zdefiniować następująco:

Definicja 3

Niech dany będzie profil D , a W oznacza zbiór wszystkich wartościowań zbioru terminów indeksujących T .

$$Con(D) = \left\{ c \in W : \omega^2(c, D) = \min \sum_{i=1}^M \omega(c, d_i)^2 \right\}. \quad (4)$$

Algorytm wyznaczający consensus zgodnie z funkcją określoną w definicji 3 jest algorytmem heurystycznym (ponieważ mamy do czynienia z problemem NP-zupełnym) i przedstawia się następująco:

Algorytm 1

Dane: Profil $D = \{d_1, d_2, \dots, d_M\}$ składający się z M wektorów częstości charakteryzujących dokumenty.

Wynik: Consensus $c = Con(D) = \langle t_1, t_2, \dots, t_n \rangle$ względem D .

START

Krok 1: Przyjmujemy $j:=1$.

Krok 2: $s = \sum_{i=1}^M d_i[j]$.

Krok 3: Jeżeli $s = M / 2$, to $c[j] = random$.

Jeżeli $s < M / 2$, to $c[j] = 0$.

Jeżeli $s > M / 2$, to $c[j] = 1$.

Krok 4: Jeżeli $j < N$, to $j = j + 1$. Przechodzimy do: kroku 2.

Jeżeli $j = N$, to przechodzimy do: kroku 5.

Krok 5: $o = \omega^2(c, D)$.

Krok 5: Przyjmujemy $j:=1$.

Krok 7: $c[j] = -c[j]$.

Krok 8: Jeżeli $\omega^2(c, D) < o$, to $o = \omega^2(c, D)$.

Jeżeli $\omega^2(c, D) > o$, to $c[j] = -c[j]$.

Krok 9: Jeżeli $j < N$, to $j = j + 1$. Przechodzimy do: kroku 7.

Jeżeli $j = N$, to STOP.

STOP.

Biorąc pod uwagę na przykład macierz przedstawioną w tab. 1, w odniesieniu do dokumentów tekstowych opisanych za pomocą wektorów charakterystycznych:

$$d_1 = \langle 0, 1, 1, 1, 0 \rangle, \quad d_2 = \langle 1, 1, 0, 1, 0 \rangle, \quad d_3 = \langle 1, 0, 1, 1, 1 \rangle, \quad d_4 = \langle 1, 0, 1, 0, 0 \rangle,$$

consensus, obliczony z wykorzystaniem algorytmu 1, reprezentuje na przykład następujący wektor charakterystyczny:

$$c = \langle 1, 0, 1, 1, 0 \rangle.$$

Zauważmy, że wygenerowany został nowy wektor, zawierający taki układ słów kluczowych, który nie wystąpił w żadnym z dokumentów, na podstawie których wyznaczany był consensus.

4. Podsumowanie

Dotychczas, co należy wyraźnie podkreślić, teoria consensu nie znalazła zastosowania w rozwiązywaniu konfliktów wiedzy nieustrukturalizowanej w rozproszonych informatycznych systemach wspomagających zarządzanie. Trzeba jednak zauważyć, że teoria consensu może okazać się przydatna w rozwiązywaniu tego typu konfliktów, ze względu na fakt, że każda ze stron konfliktu jest brana pod uwagę w consensu i „traci” najmniej, jak tylko to jest możliwe, każda ze stron wnosi swój wkład w consensus, ze względu na to, że consensus jest reprezentacją wszystkich stron konfliktu.

Jeżeli na przykład w systemie występują różne opisy danego zjawiska, to stosując metody consensu, na podstawie tych opisów, można wyznaczyć jeden wariant prezentowany następnie użytkownikowi. Wariant ten nie musi być jednym z opisów znajdujących się w systemie. Może to być całkiem nowy wariant utworzony na podstawie opisów istniejących w systemie. Dzięki temu wszystkie opisy danego zjawiska mogą być wzięte pod uwagę. Takie postępowanie pozwala między innymi skrócić czas wyznaczenia opisu docelowego (użytkownik nie musi analizować poszczególnych opisów i zastanawiać się nad ich wyborem – system wykona te czynności automatycznie za niego) oraz zmniejszyć ryzyko wyboru opisu najgorszego (ponieważ wszystkie opisy brane są pod uwagę w consensu). W konsekwencji proces zarządzania organizacją może być realizowany szybciej i skuteczniej.

Należy równocześnie zauważyć, że zdarzają się sytuacje konfliktowe, w których nie można zastosować metod consensu. Na przykład jeżeli rozpatrywane są dokumenty opisujące czynności wykonywane przez pracowników, związane z realizacją procesów biznesowych przedsiębiorstwa, to jeżeli część z pracowników wykonuje te czynności nieprawidłowo, to ich opis nie powinien być uwzględniany w consensu. Jest to problem związany z podatnością profilu na consensus, scharakteryzowany na przykład w pracy [Hernes, Sobieska-Karpińska 2009].

Istotnym problemem jest również reprezentacja dokumentów tekstowych za pomocą macierzy TFM, która stanowi próbę częściowej formalizacji wiedzy nieustrukturalizowanej, dlatego autorzy podjęli badania dotyczące rozwiązywania konfliktów wiedzy w sytuacji reprezentacji dokumentów tekstowych za pomocą ontologii, ze szczególnym uwzględnieniem mapy pojęć.

Dodatkowo w niniejszym artykule główny nacisk położony został na aspekty ekonomiczno-informatyczne przetwarzania wiedzy, natomiast w dalszych pracach badawczych uwzględnione zostaną również aspekty społeczne i psychologiczne związane z tym problemem.

Autorzy podjęli również prace dotyczące opracowania koncepcji i prototypu kognitywnego zintegrowanego systemu informatycznego zarządzania [Bytniewski, Hernes 2013], umożliwiającego między innymi przetwarzanie dokumentów tekstowych reprezentowanych za pomocą macierzy TFM oraz mapy pojęć, w ramach którego zostanie przeprowadzona weryfikacja (badanie skuteczności) algorytmu

wyznaczania konsensusu opracowanego w niniejszym artykule, a wyniki tej weryfikacji zostaną opublikowane w następnych pracach.

Literatura

- Abramowicz W., Bukowska E., Filipowska A., *Zapewnienie bezpieczeństwa przez semantyczne monitorowanie cyberprzestrzeni*, „E-mentor” 2013, nr 3(50).
- Adamczewski P., *Evolution in ERP-expanding Functionality by Bi-Modules in Knowledge-Based Management Systems*, [w:] B.F. Kubiak, A. Korowicki (red.), *Information Management*, Gdansk University Press, Gdańsk 2009.
- Barthlemy J.P., *Dictatorial consensus function on n-trees*, “Mathematical Social Science” 1992, nr 25.
- Bytniewski A. (red.), *Architektura zintegrowanego systemu informatycznego zarządzania*, Wydawnictwo AE we Wrocławiu, Wrocław 2005.
- Bytniewski A., Hernes M., *Wykorzystanie agentów kognitywnych w budowie zintegrowanego systemu informatycznego zarządzania*, [w:] T. Porębska-Miąc, H. Sroka, *Systemy wspomagania organizacji*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2013, s. 44-61.
- Cholewa W., *Metoda diagnozowania maszyn z zastosowaniem zbiorów rozmytych*, Zeszyty Naukowe – Mechanika, z. 79, Politechnika Śląska, Gliwice 1983.
- Coulouris G., Dollimore J., Kindberg T., *Systemy rozproszone. Podstawy i projektowanie*, WNT, 1998.
- De Long D., Seemann P., *Confronting conceptual confusion and conflict in knowledge management*, “Organizational Dynamics” 2000, nr 29(1).
- Dyk P., Lenar M., *Applying Negotiation Methods to Resolve Conflicts in Multi-Agent Environments*, [w:] *Multimedia and Network Information Systems, MISSI 2006*, A. Zgrzywa (red.), Oficyna Wydawnicza PWr, Wrocław 2006.
- Frank E., Bouckaert R., *Naive bayes for text classification with unbalanced classes*, Knowledge Discovery in Databases: PKDD, 2006.
- Hamming R.W., *Error detecting and error correcting codes*, “Bell System Technical Journal” 1950, 29(2).
- Hernes M., Sobieska-Karpińska J., *Susceptibility to Consensus of Conflict Situation in Intelligent Multi-Agent Decision Support System*, [w:] B.F. Kubiak, A. Korowicki (red.), *Information Management*, Gdansk University Press, Gdańsk 2009.
- Katarzyna R., Nguyen N.T., *Model systemu wieloagentowego z procedurami grupowej aktualizacji wiedzy opartymi na metodach teorii konsensusu*, Raport z serii SPR nr 3, ISiTS PWr, Wrocław 2000.
- Kisielnicki J., *MIS systemy informatyczne zarządzania*, Wydawnictwo Placet, Warszawa 2008.
- Lula P., *Text mining jako narzędzie pozyskiwania informacji z dokumentów tekstowych*, StatSoft, 2005.
- Manning C., Schütze H., *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA, 1999.
- Michalczyk A., Waściński T., *Metody i techniki sztucznej inteligencji w zarządzaniu wiedzą i doświadczeniem*, [w:] W. Chmielarz, J. Turyna (red.), *Komputerowe systemy zarządzania*, Wydawnictwo Naukowe Wydziału Zarządzania UW, Warszawa 2009.
- Nguyen N.T., *Metody wyboru konsensusu i ich zastosowanie w rozwiązywaniu konfliktów w systemach rozproszonych*, Oficyna Wydawnicza PWr, Wrocław 2002.
- Sobieska-Karpińska J., Hernes M., *Determining Consensus in Distributed Computer Decision Support System*, [w:] *Ekonometria nr 31, Zastosowania metod ilościowych*, J. Dziechciarz (red.), Wydawnictwo UE we Wrocławiu, Wrocław 2012.
- Ziemia E., *Zastosowanie eksploracji danych do analizy i oceny serwisów internetowych*, Konferencja SWO, Prace Naukowe Akademii Ekonomicznej w Katowicach, Katowice 2005.

CONSENSUS DETERMINING ALGORITHM IN A SITUATION OF UNSTRUCTURED KNOWLEDGE CONFLICTS IN DISTRIBUTED INFORMATION MANAGEMENT SUPPORT SYSTEMS

Summary: The problem of unstructured knowledge conflicts solving in the management support systems is presented in the article. The first part of the article pays attention to the fact that unstructured knowledge is an important element of organizations functioning, especially in terms of competition. Next, the sources and essence of unstructured conflicts of unstructured knowledge are characterized. This knowledge is stored at wide mining text documents. In the final part of the article, assuming the representation of text document in the binary form of frequency matrix, the consensus determining algorithm is elaborated allowing to solve unstructured knowledge conflicts, which is the basic purpose of the article.

Keywords: unstructured knowledge, knowledge conflicts, term frequency matrix, consensus algorithms.