

SYSTEM DO PROGNOZOWANIA ZANIECZYSZCZEŃ ŚRODOWISKIEM WERYFIKACJI ONTOLOGII

Adam CZARNECKI, Cezary ORŁOWSKI

Streszczenie: W rozdziale zaprezentowano stan badań prowadzonych nad zastosowaniem ontologii w modelu wieloagentowego systemu wspomagania decyzji. Przedstawiono model systemu, dobór środowiska weryfikacji ontologii oraz procesy weryfikacji z wykorzystaniem danych opisujących środowisko zanieczyszczeń powietrza w Gdańsku. Dobór środowiska weryfikacji wynikał z potrzeby wykorzystania powtarzalnych danych oraz ilościowej ich reprezentacji. Procesy weryfikacji ontologii odniesiono do oceny zapytań systemu oraz oceny poprawności ilościowej i jakościowej gromadzonych zasobów. Na zakończenie wskazano na potrzeby zmian w opracowanym systemie zarówno odniesionych do ontologii jak też głównych jego podmiotów — agentów systemowych.

Słowa kluczowe: ontologie, technologie informatyczne, zarządzanie wiedzą.

1. Wprowadzenie do idei systemu

W Zakładzie Zarządzania Technologiami Informatycznymi na Wydziale Zarządzania i Ekonomii Politechniki Gdańskiej realizowane jest przedsięwzięcie skoncentrowane na wytworzeniu systemu do oceny technologii informatycznych. Elementami tego systemu mają być narzędzia sztucznej inteligencji, takie jak systemy ekspertowe oraz sztuczne sieci neuronowe, a także ontologie. Elementy te spinać ma architektura agentowa.

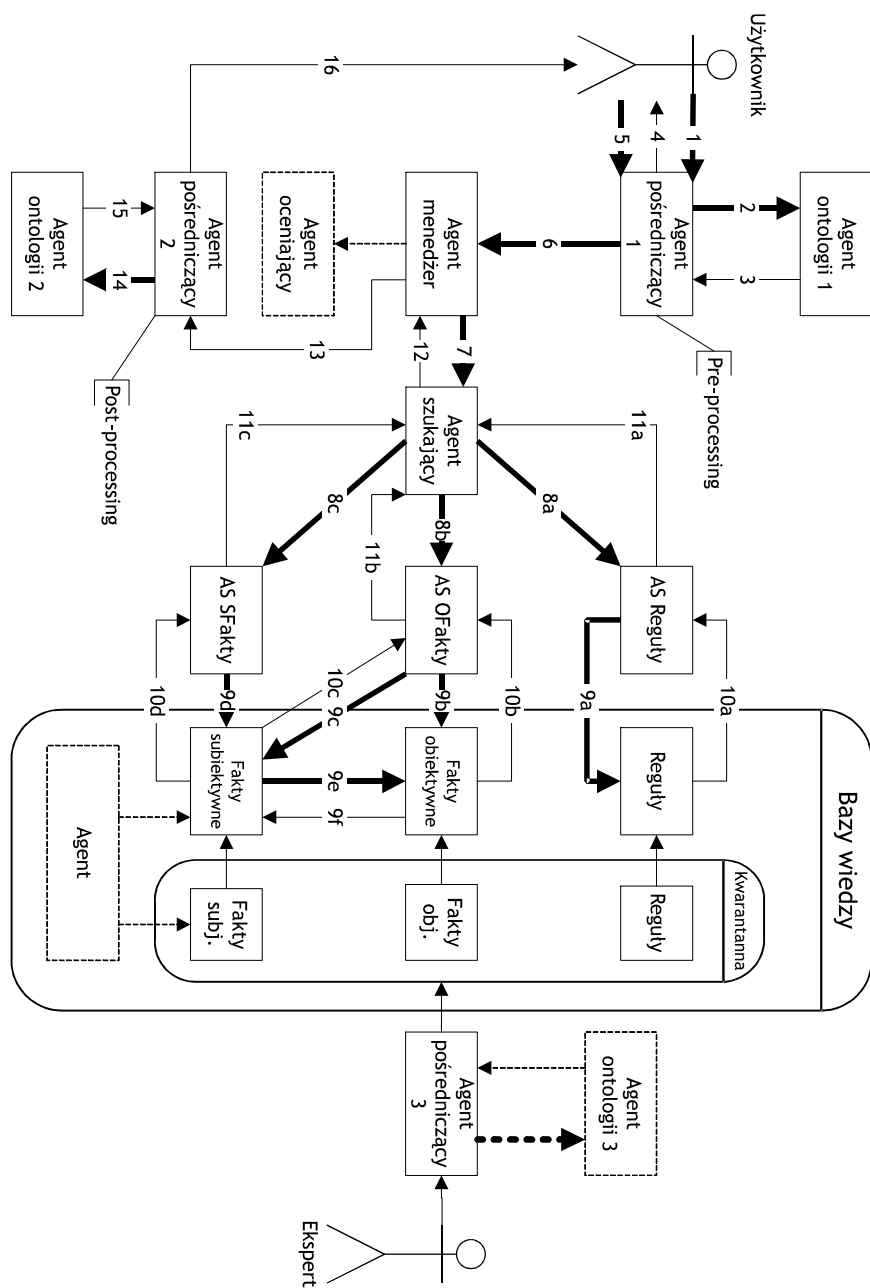
Celem działania proponowanego systemu ma być z jednej strony gromadzenie wiedzy o technologiach informatycznych, zdobywanej głównie od ekspertów, a z drugiej — udostępnianie tej wiedzy osobom zainteresowanym danymi technologiami, głównie na potrzeby doboru właściwych narzędzi informatycznych do kontekstu ich użycia. Obecnie system ma postać modelu zaimplementowanego w środowisku VBA (ang. *Visual Basic for Application*) i wymaga weryfikacji poprawności funkcjonowania jego składników.

2. Schemat funkcjonowania systemu

W toku prac nad systemem opracowano schemat jego funkcjonowania z perspektywy przepływu danych między aktorami i modułami (rys. 1) [1].

Ogólnie, sesja użytkownika zainteresowanego uzyskaniem wiedzy z systemu (klienta) polega na wprowadzaniu zapytania, przy czym agent pośredniczący odpowiada za interfejs użytkownika, a zadaniem ontologii jest zapewnienie poprawności zapytania (kroki 1–5 na rys. 1). Zapytanie trafia do agenta menedżera (6), który decyduje o miejscu wysłania zapytania. W przypadku przeszukania baz wiedzy komunikat trafia do agenta szukającego (7). Agent ten wybiera wyspecjalizowany program, który poszukiwać będzie odpowiedzi czy to w bazie reguł, faktów obiektywnych, czy bazie faktów subiektywnych (8–9). Wynik poszukiwań przekazywany jest poprzez agenty do użytkownika (10–16). W trakcie tych działań następuje sprawdzenie formalnej poprawności uzyskanej odpowiedzi przez ontologię.

Powyższy opis celowo pomija użycie bardziej złożonego modelu ocenowego oraz sesję eksperta z systemem. Są to osobne zagadnienia, na których rozwinięcie nie ma w tej publikacji miejsca.



Rys. 1. Proponowany schemat przepływu danych w systemie wspomagania decyzji

3. Domena ontologii

Kolejnym krokiem po opracowaniu struktury modułów — w szczególności agentów — powinno być stworzenie modelu prototypu systemu, by móc zweryfikować założenia jego działania oraz proponowaną architekturę.

Założoną domeną dla docelowego systemu jest obszar technologii informatycznych, głównie wspierających proces zarządzania przedsięwzięciami wytwarzania oprogramowania. Niestety, ilość zgromadzonej wiedzy na ten temat okazała się niewystarczająca, by móc w stopniu satysfakcjonującym zweryfikować działanie prototypu. Z tego powodu zdecydowano się na wykorzystanie danych na temat zanieczyszczenia środowiska opracowanych przez uczestników studium doktoranckiego przy Wydziale Chemicznym Politechniki Gdańskiej. Dane te pochodzą z sieci Agencji Regionalnego Monitoringu Atmosfery Aglomeracji Gdańskiej (ARMAAG, <http://armaag.gda.pl/>).

Udostępnione dane (ze stacji monitoringu rozlokowanych w 10 punktach aglomeracji trójmiejskiej) dotyczą miesięcznych pomiarów stężeń związków grupy BTEX: benzenu, toluenu, etylobenzenu, o-ksylenu oraz p,m-ksylenu. Tym danym towarzyszą pomiary meteorologiczne informujące o średnich w miesiącach: temperaturze, wilgotności, sile wiatru oraz ciśnieniu.

O wyborze domeny, jaką są pomiary związków BTEX, zdecydował — poza ilością i dostępnością danych — charakter gromadzonej wiedzy. Informacje pochodzące z systemu czysto technicznego, a z takim mamy tu do czynienia, są prostsze w pozyskaniu i interpretacji niż wiedza ekspercka z dziedziny technologii informatycznych. Stąd uznano, że testy prototypu w takim przypadku będą mniej narażone na problemy wynikające z doboru samej domeny. Oczywiście, gdy rozwiązanie zostanie przetestowane na mniej złożonym przykładzie, podjęte zostaną prace nad wykorzystaniem go w docelowej domenie technologii informatycznych.

Podsumowując, elementy prototypu systemu — w tym ontologię — zdecydowano się przetestować w oparciu o wiedzę z zakresu zanieczyszczeń środowiska.

4. Struktura ontologii

Jednym z najczęściej podnoszonych problemów związanych z wytworzeniem oraz oceną ontologii, jest możliwość ich ponownego wykorzystania [2]. Ma to szczególne znaczenie w środowisku sieci semantycznych budowanych na potrzeby dużych projektów.

Konstruując prototyp opisywanego tu modelu systemu, a zatem i ontologii, zdecydowano, że priorytetem jest czas wytworzenia oraz dopasowanie struktury ontologii do minimalnych wymagań przewidywanych funkcjonalności. Dodatkowym ograniczeniem było przyjęte środowisko implementacji — aplikacja MS Excel 2003 wraz z językiem VBA, dość dobrze rozpoznane przez zespół, niemniej nie wspierające ani środowiska agentowego, ani mechanizmów wnioskowania systemu ekspertowego, ani — co tu najważniejsze — żadnej metodyki tworzenia ontologii. Powstała w takich warunkach struktura nie ma elastyczności potrzebnej do efektywnej rozbudowy, czy — ogólniej — zarządzania jej eksploatacją.

Dla opisanych w punkcie 5. przypadków użycia wykorzystanie ontologii powinno zapewniać semantykę pojęć, narzucać określoną strukturę wartości opisujących i ograniczenia na pojęcia (instancje klas) oraz służyć jako słownik.

4.1. Opis semantyczny zmiennych

Na potrzeby systemu analizy zanieczyszczeń ważne było rozróżnienie, które dane mają charakter wejściowy w modelu prognostycznym, a które wyjściowy. Stąd też utworzono dwie klasy:

- Parametr wejściowy,
- Parametr wyjściowy.

Każdej z nich przyporządkowano odpowiednie instancje, czyli konkretne dane zasilające system (tab. 1).

Tab. 1. Klasy i instancje ontologii

Klasa	Instancja
Parametr wejściowy	nr stacji
	źródło
	rok
	miesiąc
	średnia temperatura
	średnia wilgotność
	średnia siła wiatru
Parametr wyjściowy	średnie ciśnienie
	benzen
	toluen
	etylobenzen
	o-ksylen
p,m-ksylen	

Rozróżnienie na parametry wejściowe i wyjściowe zostało wykorzystane przy przekazywaniu przez agenty danych do systemu ekspertowego i sztucznej sieci neuronowej.

4.2. Własności parametrów i ograniczenia

Dla parametrów przyjmujących wartości liczbowe przyjęto jednolitą strukturę własności, którymi można je opisać (tab. 2.):

- Jednostka — miara stosowana przy określaniu wartości parametru; w przypadku związków BTEX chodzi stężenie, dla miar czasu jednostka określa format zapisu,
- Limit dolny — najmniejsza zmierzona wartość parametru,
- Limit górny — największa zmierzona wartość parametru,
- Alert dolny — ustalona przez eksperta dolna granica wartości parametru, którą można uznać za błędną.
- Alert górny — ustalona przez eksperta górna granica wartości parametru, którą można uznać za błędną,
- Typ — potrzebna na etapie kodowania informacja o typie zmiennej, jakiego należy użyć.

Tab. 2. Parametry instancji i ich wartości

Instancja	Jednostka	Limit dolny	Limit górny	Alert dolny	Alert górny	Typ
benzen	□g/m ³	0,39722	10,88890	0	100	Double
toluen	□g/m ³	0,44987	24,91114	0	100	Double
etylobenzen	□g/m ³	0,02569	11,59907	0	100	Double
o-ksylen	□g/m ³	0,07549	8,76802	0	100	Double
p,m-ksylen	□g/m ³	0,21020	29,14535	0	100	Double
średnia temperatura	°C	-1,71308	20,22831	-50	50	Double
średnia wilgotność	%	59,82765	93,01933	0	100	Double
średnia siła wiatru	m/s	0,75620	5,32151	0	50	Double
średnie ciśnienie	hPa	991,17583	1020,13493	950	1050	Double
rok	RRRR	2006	2008	0	9999	Integer
miesiąc	MM	1	12	1	12	Byte

Jak widać w tab. 2., prawie wszystkie pokazane w niej parametry za wyjątkiem instancji rok i miesiąc mają charakter zjawisk ciągłych. Instancje rok i miesiąc to parametry dyskretne. Poza pokazanymi tu instancjami używany jest także nr stacji, którego zbiorem wartości jest lista dziesięciu stacji pomiarowych od AM1 do AM10 oraz źródło określające urządzenie pomiarowe.

4.3. Słownik

Na potrzeby wprowadzania przez użytkownika systemu zapytań o dane historyczne udostępniono interfejs pozwalający wybierać różne operatory porównań. Użytkownik widzi opis w formie tekstowej, zaś zastosowanie ontologii zapewni jego przełożenie na odpowiedni operator zrozumiały dla mechanizmu odpowiedzialnego za przeszukiwanie bazy wiedzy. Prostą strukturę słownika zaprezentowano w tab. 3.

Tab. 3. Słownik operatorów

Indeks	Opis słowny	Operator
1	równa się	
2	nie równa się	<>
3	większe niż	>
4	większe niż lub równe	>=
5	mniejsze niż	<
6	mniejsze niż lub równe	<=
7	brak	

Dla opisu „równa się” pole operatora jest puste, gdyż użyty moduł przeszukiwania brak operatora traktuje jako żądanie wyszukania dokładnej wartości.

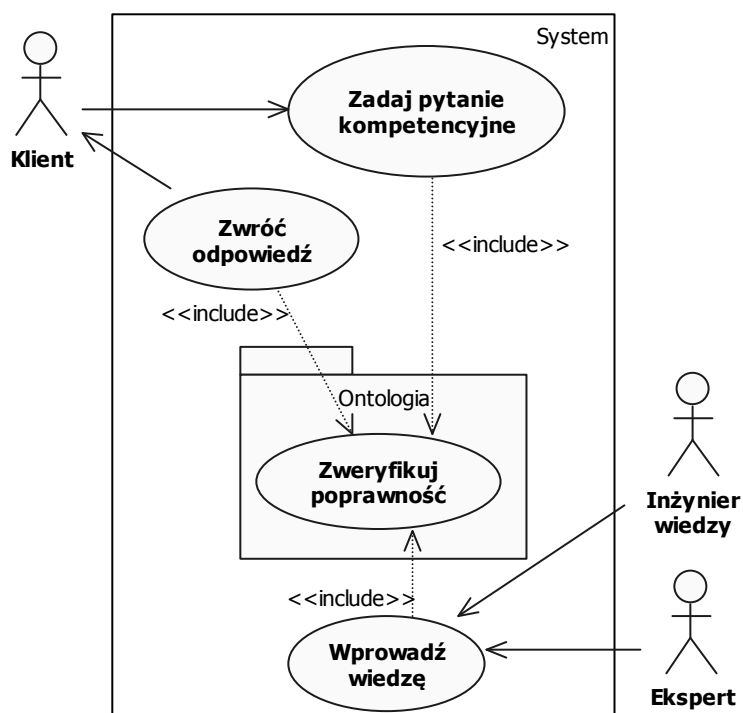
5. Przypadki użycia ontologii

Niezależnie od wybranej domeny, zakładane najważniejsze przypadki użycia ontologii w opisanym na początku systemie pozostają takie same.

Jak widać na diagramie przypadków użycia ontologii (rys. 2), jej funkcjonalność nie jest bezpośrednio udostępniana użytkownikowi. Nie jest to konieczne, albowiem ontologia stanowi rodzaj wewnętrznego mechanizmu kontroli poprawności danych, informacji czy wiedzy, jakie do systemu dopływają i jakie z niego są przekazywane na zewnątrz. Ontologia jest używana przez agenty na etapie:

8. wprowadzania zapytania przez użytkownika,
9. zwracania użytkownikowi odpowiedzi na zadane pytanie,
10. zasilania systemu wiedzą przez eksperta lub inżyniera wiedzy.

Dla potrzeb systemu wykorzystującego dane na temat zanieczyszczeń BTEX pojawił się jeszcze jeden wymóg związany z semantyką — wskazanie, które parametry mają charakter wejściowy, a które wyjściowy na potrzeby sztucznej sieci neuronowej.



Rys. 2. Przypadki użycia ontologii w systemie

5.1. Sprawdzenie zapytania

System udostępnia dwa rodzaje zapytań:

1. o prognozowane stężenie substancji BTEX w oparciu o dane meteorologiczne,
2. o zarejestrowane dane historyczne.

W pierwszym przypadku rolą ontologii jest sprawdzenie, czy podane nazwy parametrów wejściowych (dane meteorologiczne: temperatura, wilgotność, siła wiatru, ciśnienie) są zgodne z nazwami instancji klasy parametr wejściowy zawartych w samej ontologii. Innym możliwym sposobem realizacji tej funkcjonalności jest przedstawianie użytkownikowi wyłącznie pól klasy parametr wejściowy dającymi możliwość wpisania odpowiednich wartości dla prognozy. Poza porównaniem nazw następuje też sprawdzenie, czy podane wartości parametrów wejściowych nie znajdują się poza dozwolonymi zakresami alarmowymi zdefiniowanymi dla nich w ontologii (por. 4.2.). Poza wartościami alarmowymi są też wartości ostrzegawcze informujące, że zapytanie nie mieści się w granicach do tej pory zmierzonych.

Drugi przypadek sprawdzenia zapytania dotyczy sytuacji, w której użytkownik chce uzyskać z systemu dane (fakty) o pomiarach stężeń substancji BTEX. Wówczas system udostępnia interfejs pozwalający dla każdego z parametrów wejściowych zbudować warunek wyszukiwania na zasadzie: parametr wejściowy | operator | wartość. Operatory porównania wyświetlane są w formie opisu tekstowego, zaś wykorzystanie ontologii przekłada je na potrzeby wyszukiwania na operatory języka zapytań (por. 4.3.).

5.2. Sprawdzenie odpowiedzi

System w trybie prognostycznym zwraca użytkownikowi listę parametrów wyjściowych wraz z oszacowanymi wartościami. Podobnie jak przy sprawdzaniu zapytania (por. 5.1.) zadaniem ontologii jest zweryfikowanie nazw parametrów oraz obliczone wartości w oparciu o podane limity i alerty.

5.3. Sprawdzenie nowej wiedzy

Na obecnym stanie prac nad systemem agentowym „parasol” ontologii rozłożony jest nad przepływem danych w sesjach odpytywania systemu. Dalszym krokiem jest weryfikacja danych, które zasilają jego bazy wiedzy. Nie będzie to wymagać rozbudowy samej ontologii, gdyż wszystkie dane, które trafiać mają do systemu, są już zdefiniowane na poziomie instancji. Wystarczy opracować i wdrożyć algorytm sprawdzający kolejne napływające rekordy i przenoszący je do odpowiedniej sekcji bazy wiedzy — jej części „produkcyjnej”, gdy dane nie budzą zastrzeżeń, do „bufora”, gdy na przykład przekraczają dotychczas zmierzone limity, oraz do „kosza”, gdy nie spełniają warunków zapisanych w polach wartości alarmowych.

6. Weryfikacja funkcjonowania ontologii

Dla przedstawionego w tekście systemu (będącego wciąż na etapie prototypu) przeprowadzono zestaw sesji-eksperymentów sprawdzających działanie ontologii, jak opisano je w punktach 5.1. i 5.2.

6.1. Tryb prognostyczny

W pierwszym scenariuszu wybrano tryb prognostyczny. Użytkownikowi system zaprezentował interfejs do wprowadzenia wartości parametrów wejściowych (rys. 3). Po wypełnieniu pól i zatwierdzeniu danych mechanizm odwołuje się do ontologii, by sprawdzić, czy podane parametry wejściowe są tożsame ze zdefiniowanymi instancjami oraz czy wartości dla nich zadane nie wykraczają poza stany alarmowe, co mogłoby skutkować nieprawidłowym działaniem systemu ekspertowego lub sztucznej sieci neuronowej.

Podaj parametry do prognozy			
parametr	operator	wartość	jednostka
średnia temperatura	=	16	°C
średnia wilgotność	=	75	%
średnia siła wiatru	=	1,2	m/s
średnie ciśnienie	=	1012	hPa

Prześlij zapytanie

Rys. 3. Interfejs użytkownika dla wprowadzania parametrów wejściowych w module prognostycznym

Pozytywna weryfikacja parametrów uruchamia jeden z modułów inteligentnych, który zwraca prognozowane wartości (stężenia) substancji BTEX. Wyniki te ma otrzymać użytkownik, jednak przed ich wyświetleniem jeden z agentów uruchamia weryfikację wartości parametrów wyjściowych z poziomami zapisanymi w ontologii. Ewentualne ich przekroczenie może świadczyć o błędnych danych użytych do prognozowania lub wadach algorytmu wyliczającego spodziewane wartości. Wtedy agent decyduje o zakończeniu procesu i poinformowaniu użytkownika o błędzie. Pozytywna weryfikacja powoduje przekazanie prognozy aktorowi inicjującemu proces.

6.2. Tryb wyszukiwania danych historycznych

Scenariusz drugi odpytywania systemu zakładał sięgnięcie do danych historycznych znajdujących się w bazie wiedzy w oparciu o formularz filtrowania (rys. 4).

Użytkownik wprowadza wartości oraz określa operatory, które mają być użyte dla każdego z parametrów. Zatwierdzenie ich uaktywnia agenta weryfikującego parametry i ich wartości z treścią ontologii. Ponieważ operatory, z jakich korzysta użytkownik budując zapytanie, są bezpośrednio sprzęgnięte z ontologią, nie ma dodatkowej potrzeby sprawdzania ich poprawności. Ewentualne błędy są zgłaszane za pomocą okna dialogowego. Prawidłowo przeprowadzona sesja powinna dostarczyć użytkownikowi tabelę z listą wszystkich danych spełniających warunki zapytania, przy czym nagłówki

tabeli są identyczna jak w formularzu filtrowania. Zwracane dane nie podlegają żadnej dodatkowej obróbce przez narzędzia sztucznej inteligencji, dlatego nie ma konieczności weryfikacji uzyskanego wyniku przez ontologię.

Podaj parametry wejściowe

parametr	operator	wartość	jednostka
nr stacji	równa się	am 1	-
miesiąc	brak		MM
średnia temperatura	większe niż	8	°C
średnia wilgotność	mniejsze niż	60	%
średnia siła wia	równa się nie równa się		m/s
średnie ciśnieni	większe niż		hPa
rok	większe niż lub równe mniejsze niż mniejsze niż lub równe brak	2007	RRRR
Prześlij zapytanie			

Rys. 4. Interfejs użytkownika dla wprowadzania parametrów wejściowych w module przeszukiwania danych historycznych

6.3. Otrzymane rezultaty i planowane zmiany

W efekcie system jest w stanie rozpoznać parametry oraz porównywać podawane na wejściu lub wyjściu wartości z dozwolonymi. Ewentualne błędy lub ostrzeżenia są komunikowane użytkownikowi.

W najbliższym czasie dodana zostanie możliwość sprawdzania danych napływających do systemu, jako że zespół ma otrzymać kompletne wyniki pomiarów dla 2008 roku. Będą one zawierać nowy parametr wejściowy, do tej pory nie uwzględniany, to jest kierunek wiatru. Najpewniej nowym zadaniem dla ontologii będzie przekształcenie danych wyrażonych w stopniach na 16-elementowy układ róży wiatrów.

7. Podsumowanie

Zaprezentowany system do analizowania danych o stężeniach zanieczyszczeń stanowi, z punktu widzenia pracującego nad nim zespołu, rodzaj „poligonu” dla przetestowania koncepcji systemu wieloagentowego wykorzystującego ontologie oraz moduły inteligentne takie jak systemy ekspertowe oraz sztuczne sieci neuronowe. Ma także wymiar praktyczny, gdyż — po koniecznej rozbudowie i strojeniu — zostanie wykorzystany w pracach prowadzonych na Wydziale Chemicznym Politechniki Gdańskiej.

Dla powstałego prototypu udało się zbudować ontologię i wykorzystać ją w zdefiniowanych przypadkach użycia systemu. Powstała ontologia nie została zapisana w żadnym powszechnie używanym języku, takim jak OWL, chociaż układ tabelaryczny też można interpretować jako formalny sposób reprezentacji domeny [3]. Niemniej, nie wydaje się, by znalazło się dla niej zastosowanie poza rozpatrywanym systemem.

Zależnie od zainteresowania zlecającego wykonanie systemu, możliwa jest przyszła implementacja z użyciem któregoś z popularnych języków programowania (Java, C#). Wówczas być może nastąpi też zmiana sposobu zapisu ontologii na częściej spotykany. Może to dostarczyć wniosków na temat sposobów przenoszenia wiedzy pomiędzy różnymi formatami. Powinno też pozwolić na uzyskanie rozwiązania cechującego się efektywniejszym zarządzaniem ontologią podczas tworzenia, rozbudowy i utrzymania, co sprawia problemy w obecnym prototypowym systemie.

Dalsze prace, poza dokończeniem systemu na potrzeby analizy danych o substancjach BTEX, będą dotyczyły powrotu do domeny technologii informatycznych i próby przeniesienia tam modelu oraz doświadczeń płynących z weryfikacji w opisywanym tu środowisku.

Tekst powstał w ramach projektu *Zastosowanie metod inteligentnych do oceny technologii informatycznych* finansowanego przez Ministerstwo Nauki i Szkolnictwa Wyższego.

Literatura

1. Czarnecki A., Orłowski C.: *Ontology Management in the Multi-Agent System for the IT Evaluation*. [W:] *Information Systems Architecture and Technology. Information Systems and Computer Communication Networks*. Red. nauk. A. Grzech, L. Borzowski, J. Świątek, Z. Wilmowska. Wrocław University of Technology, Wrocław 2008, s. 37–48.
2. Ganemi A., Catenacci C., Ciaramita M., Lehmann J.: *A Theoretical Framework for Ontology Evaluation and Validation*. *Semantic Web Applications and Perspectives*, Springer, 2005. <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-166/9.pdf>
3. Jung S.W., Kang M.Y., Kwon H.C.: *Constructing Domain Ontology Using Structural and Semantic Characteristics of Web-table Head*. W: *New Trends in Applied Artificial Intelligence*. Vol. 4570/2007, Berlin/Heidelberg, 2007.

Mgr inż. Adam CZARNECKI
Dr hab. inż. Cezary ORŁOWSKI, prof. PG
Zakład Zarządzania Technologiami Informatycznymi
Politechnika Gdańska
80-952 Gdańsk, ul. Narutowicza 11/12
tel.: (0-58) 347 24 55
fax: (0-58) 348 60 24
e-mail: adam.czarnecki@zie.pg.gda.pl
cezary.orlowski@zie.pg.gda.pl