STUDIA INFORMATICA 2009
Volume 30 Number 2A (83)

Małgorzata BACH, Stanisław KOZIELSKI, Michał ŚWIDERSKI
Politechnika Śląska, Instytut Informatyki

ZASTOSOWANIE ONTOLOGII DO OPISU SEMANTYKI RELACYJNEJ BAZY DANYCH NA POTRZEBY ANALIZY ZAPYTAŃ W JĘZYKU NATURALNYM

Streszczenie. W prezentowanej pracy przeprowadzono analizę możliwości wykorzystania ontologii dla usprawnienia analizy semantycznej zdań sformułowanych w języku polskim, traktowanych jako zapytania do bazy danych. Skoncentrowano się na zagadnieniach opisu semantyki zapytań oraz semantyki bazy danych na podstawie logiki opisowej (ang. *Description Logic*).

Słowa kluczowe: ontologie, logika opisowa, przetwarzanie języka naturalnego, lingwistyka komputerowa, bazy danych

APPLICATION OF ONTOLOGY TO RELATIONAL DATABASE SEMANTICS DESCRIPTION IN CONTEXT OF NATURAL LANGUAGE QUERIES ANALYSIS

Summary. The feasibility analysis of ontology incorporation into queries semantic analysis process, where queries to databases are formulated in natural polish language are contained in this paper. In presented work, the semantics of queries and the database was formalized with use of Description Logic.

Keywords: Ontology, Description Logic, Natural Language Processing, Computational Linguistic, Database

1. Wprowadzenie

W związku z tym, że niemal we wszystkich systemach informatycznych wspierających działalność człowieka wykorzystywane są systemy zarządzania bazą danych, problem ułatwienia komunikacji „człowiek – baza” jest wciąż bardzo aktualny. Niestety, dostęp do bazy

danych wymaga znajomości odpowiednich języków operowania danymi (zwanych też językami zapytań), które ze względu na sformalizowaną postać zapisu mogą być nieczytelne i trudne do zrozumienia, zwłaszcza dla osób niezajmujących się zawodowo informatyką. Stworzenie możliwości „porozumiewania się" z bazami danych za pomocą języka naturalnego, nawet leksykalnie ograniczonego i ukierunkowanego na określoną dziedzinę, będzie niewątpliwie krokiem w kierunku ułatwienia szerszemu gronu użytkowników korzystania z nich. Kluczem do realizacji systemów umożliwiających komunikację w wybranym języku naturalnym jest opracowanie skutecznych algorytmów analizy morfologicznej, składniowej i semantycznej wypowiedzi sformułowanej w tym języku. Dopiero po rozpoznaniu znaczenia (semantyki) wypowiedzi można zinterpretować to znaczenie i wywołać odpowiednią reakcję systemu informatycznego.

Celem prezentowanej pracy była analiza możliwości wykorzystania ontologii oraz języków stosowanych do ich opisu, do usprawnienia automatycznego tłumaczenia zadań wyszukiwania danych sformułowanych w języku polskim na język formalny SQL [187].

2. Ontologie i logiki opisowe

Pojęcie ontologii wywodzi się z filozofii, w której jest definiowane jako teoria bytu dotycząca badania charakteru i struktury rzeczywistości. Nieco inaczej ontologia jest postrzegana przez tych, którzy zajmują się sztuczną inteligencją, a w szczególności zagadnieniami związanymi z reprezentacją wiedzy. Konsorcjum W3C zaproponowało następującą definicję:

> *Ontologia definiuje pojęcia, używane do opisywania i reprezentacji gałęzi wiedzy. Ontologie są używane przez ludzi, bazy danych i aplikacje, które potrzebują informacji z danej dziedziny. Ontologie zawierają definicje pojęć z danej dziedziny i relacji zachodzących pomiędzy tymi pojęciami, które są czytelne dla komputera. Definicje te nie muszą być ścisłe w sensie rozumianym przez logikę, ale powinny być zrozumiałe dla aplikacji. Ontologie klasyfikują wiedzę w postaci gałęzi wiedzy, ale także wiedzę pochodzącą z różnych dziedzin. W ten sposób czynią tę wiedzę dostępną dla człowieka i aplikacji* [2].

Istnieją różne metody opisu ontologii: reprezentacja graficzna, reprezentacje wykorzystujące logikę (np. logikę opisową) oraz reprezentacje oparte na języku XML. W ramach prezentowanych badań, jako formalizm służący do zapisu ontologii, zastosowana została logika opisowa, zwana też logiką deskrypcyjną (ang. *Description Logic – DL*). Formalizm ten, rozwijany był od ponad 30 lat, jednak z uwagi na rozwój idei Sieci Semantycznej zyskał on ostatnio na popularności [3]. Logika opisowa jest rozstrzygalnym podzbiorem rachunku predykatów pierwszego rzędu. DL w istocie obejmuje całą rodzinę języków przeznaczonych do definiowania ontologii o różnym stopniu ekspresji i różnych możliwościach wnioskowania.

Baza wiedzy, zgodnie z założeniami logiki opisowej, składa się z dwóch części:

- TBox (ang. *terminology box, taxonomy box*) zawierającej opis terminologii i pojęć z danej dziedziny wiedzy wraz z ich cechami i wzajemnymi relacjami,
- ABox (ang. *assertional box*) zawierającej opis konkretnych faktów i obiektów z wykorzystaniem pojęć podanych w TBox.

Aby ontologie mogły być przedmiotem wnioskowania, baza wiedzy musi posiadać moduł (silnik) wnioskujący (ang. *inference engine*), a także musi istnieć określony język zapisu ontologii. Różne zbiory wykorzystywanych konstruktorów prowadzą do różnych dialektów logiki opisowej. Najmniejszy język, interesujący z praktycznego punktu widzenia, to $\mathcal{AL}$. W języku tym pojęcia budowane są za pomocą czterech konstruktorów: $\neg$,$\sqcap$, $\forall$, $\exists$.

W prezentowanej pracy do wnioskowania z ontologii zastosowane zostało darmowe narzędzie o nazwie RACER. Zbiór konstruktorów wspomaganych przez to narzędzie jest znacznie szerszy niż w przypadku wspomnianego wcześniej języka $\mathcal{AL}$. Istnieje między innymi możliwość korzystania z sumy (unii) pojęć ($C \sqcup D$), ograniczenia liczności roli ($(\geq nR$, $\leq nR$), ról odwrotnych i tranzytywnych [3].

3. Analiza języka naturalnego

Specyfika języka polskiego (fleksja, swobodny szyk zdania itd.) powoduje, iż dla prawidłowego tłumaczenia zdań konieczna jest dokładna analiza lingwistyczna. W kolejnych etapach interpretacji zdań należy dokonać:

- analizy morfologicznej – analizy (rozpoznania) słów w tekście,
- analizy składniowej – rozbioru zdania na części składowe oraz określenia gramatycznych relacji zachodzących między nimi,
- analizy semantycznej – analizy sensu części składowych każdego zdania na podstawie pewnej bazy wiedzy dla określonej dziedziny przedmiotowej,
- analizy pragmatycznej – analizy sensu zdania w realnym kontekście w oparciu o własną bazę wiedzy systemu informatycznego.

Opracowane w Instytucie Informatyki Politechniki Śląskiej eksperymentalne moduły analizy morfologicznej i syntaktycznej zostały przedstawione w publikacjach: [5, 6, 7]. W niniejszej pracy przedstawiono jedynie wybrane zagadnienia związane z analizą semantyczną i pragmatyczną.

3.1. Określanie struktury semantycznej na podstawie gramatyki przypadków

Zadaniem analizy semantycznej jest określenie znaczenia zdania, zadecydowanie, które poprawne syntaktycznie zdania, mają sens i przetłumaczenie ich na odpowiednią postać pośrednią. Analiza problemu [1] doprowadziła do wniosku, iż dla reprezentacji semantyki zapytań do baz danych wygodne będzie zastosowanie gramatyki przypadków (ang. *case grammar*) Fillmore'a [8]. W gramatyce tej zakłada się, iż głównym elementem zdania jest czasownik, a frazy rzeczownikowe, przyimkowe itd. są realizacją tzw. przypadków (kategorii) semantycznych, które określają, jaką rolę semantyczną pełnią poszczególne części zdania.

Analizując różnorodne propozycje, przytaczane w literaturze oraz biorąc pod uwagę specyfikę zadań wyszukiwania danych [9, 10], w procesie tworzenia systemu DIALOG (eksperymentalny system wyszukiwania danych z dostępem w języku naturalnym) uwzględniono następujący zestaw kategorii semantycznych:

- ACTION – czynność,
- AGENT – wykonawca czynności,
- OBJECT – obiekt, na którym wykonywana jest akcja,
- COUNTER_AGENT – coś lub ktoś, wbrew czemu (komu) wykonywana jest akcja,
- CO_AGENT – współwykonawca,
- EXPERIENCER – odbiorca,
- INSTRUMENT – instrument,
- LOCATION – miejsce akcji,
- TIME – czas akcji,
- SOURCE – miejsce, z którego coś jest przemieszczane (punkt początkowy akcji),
- DESTINATION – miejsce, do którego coś jest przemieszczane (punkt docelowy),
- RESULT – coś, co powstaje w wyniku akcji,
- MANNER – sposób wykonania czynności.

Przykładowo, zdanie: „*Piotr pisze książkę*." ma strukturę semantyczną: AGENT ACTION OBJECT, co oznacza, że rzeczownik „*Piotr*" należy kojarzyć z wykonawcą czynności, czasownik „*pisze*" określa czynność, natomiast „*książk*a" jest obiektem tej czynności.

W ogólnym przypadku nie jest niestety prostą sprawą określenie struktury semantycznej zdania, nie istnieje bowiem bezpośrednia zależność między przypadkami syntaktycznymi typu: podmiot, orzeczenie, dopełnienie a kategoriami semantycznymi.

W celu poprawnego określenia funkcji semantycznych pełnionych przez poszczególne grupy słów należy:

- zdefiniować ramę przypadków, czyli określić wzorce dopuszczalnych zdań dla głównego czasownika w zdaniu,

- dokonać podziału zdania na grupy słów z wyodrębnieniem rdzeni i określeniem cechy semantycznej rdzenia (*animate*, *inanimate*, *physical_object*...),
- określić kategorię syntaktyczną grupy (podmiot, dopełnienie, ...),
- wyodrębnić słowa o specjalnym znaczeniu (głównie przyimki),
- określić stronę zdania (czynna, bierna).

Dopiero po uwzględnieniu wszystkich wymienionych czynników można się pokusić o próbę definicji reguł pozwalających na określenie kategorii semantycznych pełnionych przez poszczególne części zdania. Kilka przykładowych reguł przedstawiono poniżej:

1. (**CS** = animate) ∧ (**KS** = podmiot) ∧ (**SZ** = czynna) => AGENT.
2. (**CS** = inanimate) ∧ (**KS** = dopełnienie) ∧ (**SZ** = czynna) => OBJECT.
3. (**CS** = inanimate) ∧ (**KS** = podmiot) ∧ (**SZ** = bierna) => OBJECT.
4. (**CS** = place) ∧ (**KS** = okolicznik) ∧ (**PS** = „z”) => SOURCE.
5. (**CS** = place) ∧ (**KS** = okolicznik) => LOCATION.
6. (**CS** = time) ∧ (**KS** = okolicznik) => TIME.
7. (**CS** = animate) ∧ (**KS** = dopełnienie) ∧ (**SZ** = bierna) ∧ (**PS** = „przez”) => AGENT.
8. (**CS** = animate) ∧ (**KS** = dopełnienie) ∧ (**SZ** = czynna) ∧ ((**PS** = „do”) ∨ (**PS** = „dla”)) => EXPERIENCER.

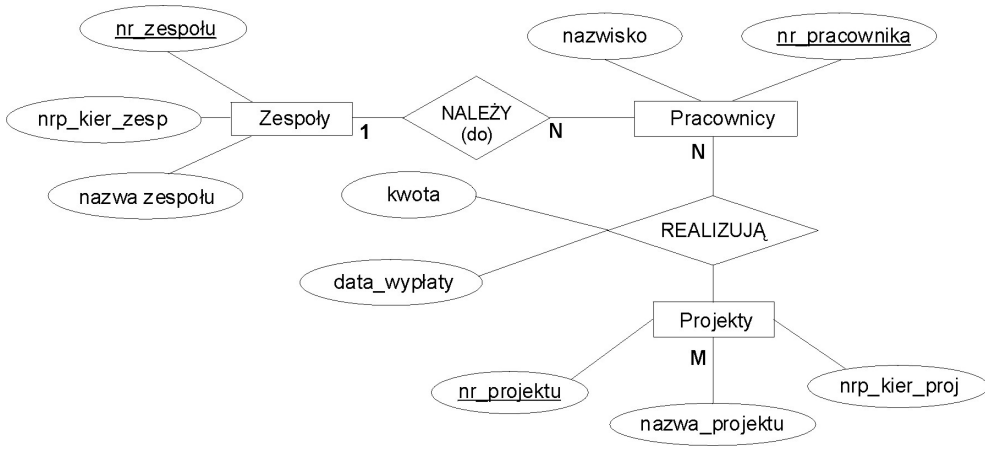
CS oznacza cechę semantyczną rdzeni, **KS** kategorię syntaktyczną grupy, **PS** przyimek specjalny, natomiast **SZ** stronę zdania.

3.2. Określanie struktury semantycznej bazy danych

Określanie znaczenia (semantyki) zapytania musi odbywać się w kontekście konkretnej bazy danych, do której to zapytanie jest kierowane. Dla realizacji tego zadania niezbędna jest jak najpełniejsza wiedza o strukturze bazy danych, w tym szczególnie o powiązaniach między danymi tworzącymi bazę. W systemie DIALOG wykorzystano, jako bogate źródło wiedzy w tym zakresie, diagramy związków encji (ang. ***E**ntity **R**elationship Diagram*), powstające w procesie modelowania konceptualnego baz danych. Określenie semantyki w tym przypadku odbywa się przez przypisanie poszczególnym elementom diagramu ER opisującego bazę, odpowiednich kategorii semantycznych. W pracy [1] zwrócono uwagę, że wyrazami wykorzystywanymi w odniesieniu do związków diagramu ER są czasowniki, natomiast w odniesieniu do encji – rzeczowniki. Ponieważ czasownik, zgodnie z zasadami gramatyki przypadków, wypełnia kategorię semantyczną ACTION, stąd w realiach schematu konceptualnego to związek odpowiada tej kategorii, czyli określa czynność. W kolejnym kroku należy określić role (kategorie) semantyczne, jakie pełnią encje połączone danym związkiem [1, 9]. Realizowane jest to przez analizę cech semantycznych rzeczowników wykorzystanych do nazwania poszczególnych encji oraz wymagań stawianych przez ramy

przypadków dla czasowników wykorzystywanych w odniesieniu do poszczególnych związków diagramu ER opisującego bazę.

Przykładowo, próba określenia struktury semantycznej bazy opisanej diagramem ER przedstawionym na rys. 1. prowadzi do wniosku, iż w bazie tej można wydzielić dwa fragmenty (rys. 2).



Rys. 1. Diagram ER przykładowej bazy PRZEDSIĘBIORSTWO
Fig. 1. Entity Relationship Diagram of Company database

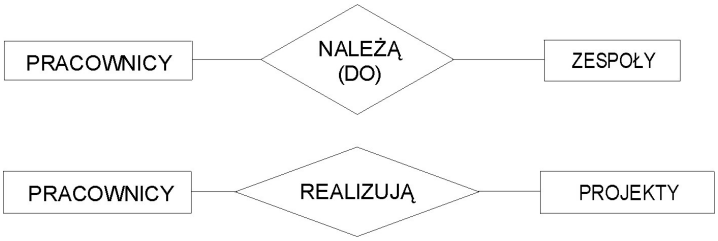
Obrazują one związki między danymi, które to związki można wyrazić zdaniami:

- Pracownicy należą do zespołów.
- Pracownicy realizują projekty.

Zdania te mają następujące struktury semantyczne:

- AGENT ACTION LOCATION
- AGENT ACTION OBJECT

Oznacza to, iż encja *„Pracownicy”* powinna być skojarzona z kategorią semantyczną AGENT, encja *”Zespoły”* z kategorią LOCATION, a *„Projekty”* z kategorią OBJECT.

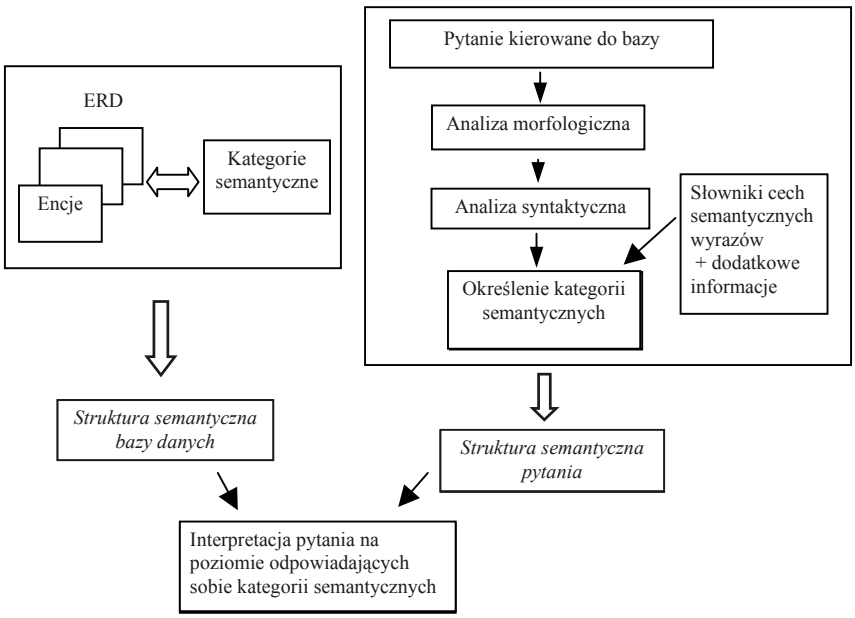


Rys. 2. Związki występujące w bazie PRZEDSIĘBIORSTWO
Fig. 2. Relationships in Company database

W ten sposób dla poszczególnych encji diagramu ER można określić funkcje semantyczne, jakie te encje pełnią w odniesieniu do danego związku. Można w ten sposób określić semantykę bazy.

Możliwość określenia struktury semantycznej pytań kierowanych do bazy danych (co pokazano w punkcie 3.1) oraz struktury semantycznej bazy danych nasunęła pomysł interpretacji zapytań przez porównanie obu struktur, tzn. interpretację na poziomie odpowiadających

sobie kategorii semantycznych. Ogólna koncepcja takiego odwzorowania została przedstawiona na rys. 3. Na rysunku tym część bloczków została wypełniona ciemniejszym kolorem, odpowiadają one tym fragmentom analizy, dla realizacji których stworzono odpowiednią ontologię i wykorzystano możliwości, jakie daje mechanizm wnioskowania.



Rys. 3. Ogólna koncepcja odwzorowania pytania na elementy diagramu ER
Fig. 3. Questions translation into Entity Relationship Diagram elements

4. Opis stworzonej ontologii

Pierwszym krokiem w tworzeniu ontologii było zdefiniowanie odpowiedniej hierarchii cech semantycznych. Wstępnie cechy semantyczne na najwyższym poziomie podzielono na: żywotne, nieżywotne, miejsce i czas. W logice zostało to zapisane w następujący sposób:

ZYWOTNE ⊔ NIEZYWOTNE ⊔ MIEJSCE ⊔ CZAS ≡ CS

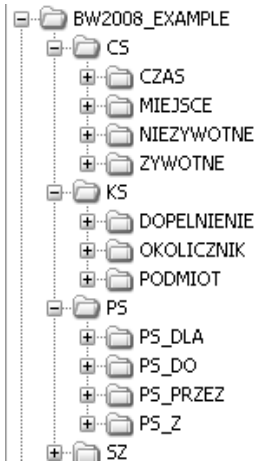
Następnie rozbudowano hierarchię pojęć (konceptów) dotyczących cech semantycznych. Poniżej przedstawiono przykłady zdefiniowanych aksjomatów:

OSOBOWE ⊔ NIEOSOBOWE ≡ ZYWOTNE
OSOBOWE ⊓ NIEOSOBOWE ≡ ⊥
MASZYNA ⊔ MATERIAL ⊔ NARZEDZIA ≡ NIEZYWOTNE
ZWIERZETA ⊔ ROSLINY ≡ NIEOSOBOWE

Zapisano również informacje o możliwych kategoriach syntaktycznych, stronach zdania oraz uwzględnianych przyimkach o „specjalnym” znaczeniu:

PODMIOT ⊔ DOPELNIENIE ⊔ OKOLICZNIK ⊔ PRZYDAWKA ≡ KS
CZYNNA ⊔ BIERNA ≡ SZ
PS_Z ⊔ PS_PRZEZ ⊔ PS_DO ⊔ PS_DLA ≡ PS

Przedstawione powyżej definicje zaowocowały powstaniem hierarchii konceptów zilustrowanej na rys. 4.



Rys. 4. Hierarchia konceptów
Fig. 4. Concepts hierarchy

W logice opisowej zapisane zostały również reguły stworzone dla określania kategorii semantycznych pełnionych przez poszczególne części zdania, czyli określania struktury semantycznej zdań (reguły te zostały omawiane w punkcie 3.1):

(ZYWOTNE ⊓ PODMIOT ⊓ CZYNNA) ⊔ (ZYWOTNE ⊓ DOPELNIENIE ⊓ BIERNA ⊓ PS_PRZEZ) ≡ **AGENT**
(NIEZYWOTNE ⊓ DOPELNIENIE ⊓ CZYNNA) ⊔ (NIEZYWOTNE ⊓ PODMIOT ⊓ BIERNA) ≡ **OBJECT**
MIEJSCE ⊓ OKOLICZNIK ⊓ PS_Z ≡ **SOURCE**
MIEJSCE ⊓ OKOLICZNIK ≡ **LOCATION**
CZAS ⊓ OKOLICZNIK ≡ **TIME**
(ZYWOTNE ⊓ DOPELNIENIE ⊓ CZYNNA ⊓ PS_DO) ⊔ (ZYWOTNE ⊓ DOPELNIENIE ⊓ CZYNNA ⊓ PS_DLA) ≡ **EXPERIENCER**

Przedstawione powyżej definicje były niezbędne dla usprawnienia procesu określania struktury semantycznej zdań wyrażonych w języku polskim.

Kolejnym etapem analizy (przedstawionym w punkcie 3.2) jest analiza pragmatyczna, czyli analiza sensu w konkretnym kontekście. W przypadku zadań wyszukiwania w bazach to właśnie zawartość bazy danych stanowi kontekst. Stąd też, w kolejnym kroku, w logice opisowej zapisane zostały informacje o semantyce bazy.

Dla bazy, której strukturę przedstawia rys. 1, opis tablic i ich atrybutów[1] wyglądał następująco:

ATR_NAZWISKO_PRACOWNICY ⊑ **TAB_PRACOWNICY**
ATR_NRPRAC_PRACOWNICY ⊑ **TAB_PRACOWNICY**
ATR_NRZESP_ZESPOLY ⊑ **TAB_ZESPOLY**
ATR_NAZWAZESPOLU_ZESPOLY ⊑ **TAB_ZESPOLY**
ATR_NUMKIER_ZESPOLY ⊑ **TAB_ZESPOLY**

1 Algorytm przekształcania elementów diagramu związków encji w schemat relacyjnej bazy danych został opisany dokładnie w pracy [1].

Następnie zdefiniowano role semantyczne, jakie mogą pełnić poszczególne tablice w odniesieniu do danego związku (rys. 2). Przykładowo, informacja, o tym, że tablica *Pracownicy* rozpatrywana w powiązaniu z tablicą *Zespoły,* w kontekście związku *Należy,* pełni rolę (kategorię) semantyczną AGENT, została zapisana następująco:

$$\text{TAB\_PRACOWNICY} \sqcap \forall \text{NALEZY.TAB\_ZESPOLY} \sqsubseteq \text{AGENT}$$

Natomiast tablica *Zespoły*, jeśli jest rozpatrywana w powiązaniu z tablicą *Pracownicy* przez odwrotną relację *Należy*, to ma kategorię semantyczną LOCATION:

$$\text{TAB\_ZESPOLY} \sqcap \forall \text{NALEZY}^{-}. \text{TAB\_PRACOWNICY} \sqsubseteq \text{LOCATION}$$

Zapisano również informacje o pewnych „niestandardowych" rolach, jakie mogą pełnić tablice lub ich atrybuty. Przykładowo, aby umożliwić realizację zapytania o numer lub nazwisko osoby pełniącej funkcje kierownika zespołu, utworzono definicję mówiącą, iż atrybut numer kierownika zespołu (należący do tablicy *Zespoły*) może pełnić rolę semantyczną AGENT w stosunku do tablicy *Zespoły* (która z kolei wypełnia kategorię semantyczną OBJECT w związku *Kieruje_zesp*):

$$\text{TAB\_ZESPOLY} \sqcap \forall \text{KIERUJE\_ZESP}^{-}. \text{ATR\_NUMKIER\_ZESPOLY} \sqsubseteq \text{OBJECT}$$

$$\text{ATR\_NUMKIER\_ZESPOLY} \sqcap \forall \text{KIERUJE\_ZESP.TAB\_ZESPOLY} \sqsubseteq \text{AGENT}$$

Ostatnim krokiem opisu semantyki bazy danych było zdefiniowanie dziedziny i przeciwdziedziny każdego związku. Dziedziną związku *Należy* jest tablica *Pracownicy*, a przeciwdziedziną tablica *Zespoły*:

$$\exists \text{ NALEZY.}\top \sqsubseteq \text{TAB\_PRACOWNICY}$$

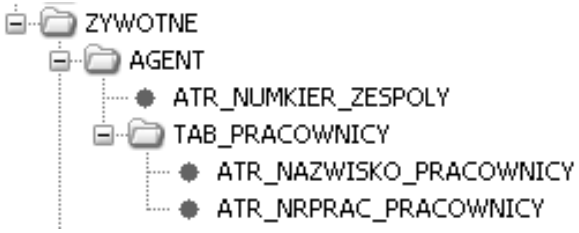
$$\top \sqsubseteq \forall \text{ NALEZY.TAB\_ZESPOLY}$$

Dziedziną relacji *Kieruje_Zesp* jest kolumna *NumKier*, a przeciwdziedziną tablica Zespoły:

$$\exists \text{ KIERUJE\_ZESP.}\top \sqsubseteq \text{ATR\_NUMKIER\_ZESPOLY}$$

$$\top \sqsubseteq \forall \text{ KIERUJE\_ZESP.TAB\_ZESPOLY}$$

Na podstawie powyższych opisów każda tablica i jej kolumny zostały zakwalifikowane do określonych kategorii semantycznych, w ten sposób określona została semantyka bazy danych (rys. 5).



Rys. 5. Fragment hierarchii opisującej semantykę bazy danych
Fig. 5. The hierarchy which describes the semantics of database

4.1. Przykład analizy zapytania na podstawie utworzonej ontologii

Dysponując wynikami analizy morfologicznej i składniowej oraz informacjami z odpowiednich słowników (opisujących m.in. cechy semantyczne), możemy przykładowo znaleźć odpowiedź na pytanie:

> Jaką rolę semantyczną może pełnić grupa słów, o której wiemy, że rdzeń grupy ma cechę semantyczną *żywotne*, na poziomie składni grupa ta pełni rolę *podmiotu*, a zdanie jest wyrażone w stronie *czynnej*?

Odpowiednie zapytanie w systemie RACER wygląda następująco:

Pytanie: (concept-synonyms (AND ZYWOTNE PODMIOT CZYNNA))

System zwrócił odpowiedź: (AGENT)

System pozwala również na zadawanie pytań o zawieranie się konceptów. Może to być wykorzystane w przypadku, gdy nie dysponujemy pełną informacją na temat analizowanej grupy słów. Przykładowo, jeśli nie znaleziono danego słowa w słowniku opisującym cechy semantyczne lub też z jakichś powodów nie znamy roli, jaką pełni analizowane słowo (lub grupa słów) na poziomie składni. Możemy wówczas zapytać:

> Czy można coś wnioskować jedynie na podstawie faktu, iż słowo ma cechę semantyczną *żywotne* i zdanie jest wyrażone w stronie *czynnej*? (nie jest znana kategoria syntaktyczna).

W systemie RACER zostało to zapisane w następującej postaci:

Pytanie: (concept-descendants (AND ZYWOTNE CZYNNA))

Odpowiedź systemu była następująca:

((ATR_NRPRAC_PRACOWNICY) (ATR_NAZWISKO_PRACOWNICY)
(ATR_NUMKIER_ZESPOLY) (TAB_PRACOWNICY) (EXPERIENCER) (AGENT))

Wynika z niej, iż słowo (grupa słów) spełniające powyższe warunki na poziomie semantyki może pełnić rolę wykonawcy (AGENT) lub odbiorcy (EXPERIENCER) czynności. Dysponując tą informacją oraz wiedzą na temat wymagań (tzw. ramy przypadków – o czym wspomniano w punkcie 3.1) głównego czasownika w zdaniu, zwykle możemy rozstrzygnąć, o którą z tych ról faktycznie chodzi.

Warto zaznaczyć, iż przypisanie roli semantycznej w opisanym powyżej przypadku (dysponowania niepełną informacją na temat danej grupy słów) nie było możliwe we wcześniejszej wersji systemu DIALOG. Możliwość ta wynika z zastosowania ontologii i mechanizmu wnioskowania, czyli odnajdywania niejawnych wniosków na podstawie jawnie przedstawionej wiedzy.

Po przypisaniu kategorii semantycznych wszystkim słowom (grupom słów) w zdaniu, w sposób opisany powyżej, można wykorzystać stworzoną ontologię do sprawdzenia, jakie elementy (tablice, atrybuty) bazy danych mogą być kojarzone z poszczególnymi kategoriami semantycznymi.

Przykładowo, można zapytać:

Jakie fragmenty bazy mogą dostarczyć danych dla rozpoznanej wcześniej kategorii AGENT?

Pytanie: (concept-descendants AGENT)

Odpowiedź: ((ATR_NRPRAC_PRACOWNICY) (ATR_NAZWISKO_PRACOWNICY)
(ATR_NUMKIER_ZESPOLY) (TAB_PRACOWNICY))

W kolejnym kroku należy ograniczyć wynik do tych elementów, które pełnią rolę AGENT w odniesieniu do konkretnego związku, np. związku *Należy*. Można to osiągnąć zadając pytanie o dziedzinę związku:

Pytanie: (role-domain NALEZY)

Odpowiedź: TAB_PRACOWNICY

Z odpowiedzi wynika, że to tablica *Pracownicy* powinna być kojarzona z rolą wykonawcy czynności (AGENT) w odniesieniu do związku *Należy*. Czasami konieczne jest jeszcze sprawdzenie hierarchii konceptu (jeśli przykładowo pytanie dotyczy konkretnego atrybutu). Można tę informację uzyskać zadając zapytanie:

Pytanie: (concept-descendants TAB_PRACOWNICY)

Odpowiedź: ((ATR_NRPRAC_PRACOWNICY) (ATR_NAZWISKO_PRACOWNICY))

Przedstawiony przykład pokazuje, jak korzystając ze zdefiniowanej wcześniej ontologii, można wnioskować, który fragment bazy danych zawiera dane, będące odpowiedzią na zadane pytanie.

5. Podsumowanie

W ramach prezentowanej pracy przeanalizowano możliwość wykorzystania ontologii dla usprawnienia analizy semantycznej zdań sformułowanych w języku polskim, traktowanych jako zapytania do bazy danych. Skoncentrowano się głównie na zagadnieniach opisu semantyki zapytań oraz semantyki bazy danych na podstawie logiki opisowej.

Zastosowanie ontologii i mechanizmu wnioskowania, czyli odnajdywania niejawnych wniosków na podstawie jawnie przedstawionej wiedzy, pozwoliło na usprawnienie działania eksperymentalnego systemu wyszukiwania danych z dostępem w języku naturalnym – DIALOG. Umożliwiło przypisanie odpowiedniej roli semantycznej grupie słów nawet w sytuacji, kiedy nie dysponujemy pełną informacją na temat wszystkich elementów tej grupy.

Przeprowadzone badania wniosły również nowe idee do szeroko rozumianego problemu interpretacji semantycznej zdań języka polskiego (w szerszym kontekście niż tylko analiza zapytań do bazy danych), zagadnienie to jest jednak bardzo szerokie i wymaga dalszych prac.

BIBLIOGRAFIA

1. Bach M.: Metody konstruowania zadań wyszukiwania w bazach danych w procesie translacji zapytań sformułowanych w języku naturalnym, Rozprawa doktorska, Gliwice 2004.
2. Strona internetowa konsorcjum W3C (stan na rok 2008): http:/www.w3c.org/.
3. Baader F. A., McGuiness D. L., Nardi D., Patel-Schneider P. F.: The Description Logic Handbook: Theory, implementation, and applications, Cambridge University Press, 2003.
4. OWL Web Ontology Language Reference. strona internetowa (stan na rok 2008): www.w3.org/TR/owl-ref.
5. Fabian P., Migas A., Suszczańska N.: Zastosowanie analizy morfologicznej i składniowej w procesie rozpoznawania mowy. Jassem W., Basztura C., Demenko G., Jassem K. (eds): Speech and Language Technology, Vol. 3, Poznań 1999.
6. Suszczańska N.: Charakteryzacia systemy syntaksycznych grup dla automatycznogo syntaksycznogo analizu tekstiw ukrajińskoju mowoju, Ukraina 1994.
7. Suszczańska N., Lubiński M.: POLMORPH, Polish Language Morphological Analysis Tool. 19th IASTED International Conference APPLIED INFORMATICS - AI'2001, Innsbruck (Austria) 2001.
8. Fillmore C.: The case for case, Universals in Linguistic Theory, New York 1968.
9. Bach M., Gruszka W.: Wykorzystanie gramatyki przypadków w analizie zadań wyszukiwania danych formułowanych w języku naturalnym. Materiały IV Krajowej Konferencji Naukowej „Inżynieria Wiedzy i Systemy Ekspertowe", Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2000, s. 195÷203.
10. Sabbagh S., An Application of Entity Relationship Models to Natural Language User Interface, Entity–Relationship Approach, Lausanna, Switzerland 1990.

Recenzent: Dr hab. inż. Krzysztof Goczyła, prof. Pol. Gdańskiej

Wpłynęło do Redakcji 10 lutego 2009 r.

Abstract

The research on computer system, which are able to process natural language, have been conducted for many years. In creating such system the key issue is to develop effective

algorithms of semantic analysis of an utterance in natural language. Only after the meaning of the utterance is recognized, it becomes possible to interpret it in a proper way and then cause the desired response of computer system.

Access to databases requires knowledge of specific query languages. Having access to morphological, syntactical and semantical analyzers we can think of building a translator that transforms queries stated in polish language into database query language.

The feasibility analysis of ontology incorporation into queries semantic analysis process, where queries to databases are formulated in natural polish language are contained in this article. In presented work, the semantics of queries and the database was formalized with use of Description Logic.

Adresy

Małgorzata BACH: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, malgorzata.bach@polsl.pl.
Stanisław KOZIELSKI: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, stanislaw.kozielski@polsl.pl.
Michał ŚWIDERSKI: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, michal.swiderski@polsl.pl.