

Krzysztof GOCZYŁA, Aleksander WALOSZEK, Wojciech WALOSZEK
Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki

O SEMANTYCE KONTEKSTOWYCH BAZ WIEDZY¹

Streszczenie. W artykule omawiany jest problem poprawnego projektowania kontekstowych baz wiedzy. Rozszerzamy w nim definicje wybranych pojęć wprowadzonych przez metodę OntoClean – jedną z najbardziej znanych metod oceny jakości ontologii niekontekstowych. Pozwala nam to na sformułowanie wymagań jakościowych dla projektów uwzględniających semantykę kontekstów. Następnie analizujemy pod kątem tych wymagań metodę projektowania kontekstowych baz wiedzy SIM.

Słowa kluczowe: ontologie modularne, bazy wiedzy, kontekst, konceptualizacja

ABOUT SEMANTICS OF CONTEXTUAL KNOWLEDGE BASES

Summary. The paper discusses the issue of designing well-founded contextual knowledge bases. Basing on OntoClean, the well-known method of assessment of taxonomies, a set of desirable qualities for *context-semantic knowledge bases* is introduced. Then it is shown that SIM method of organizing contextual knowledge bases is in accordance with the introduced desiderata.

Keywords: modular ontologies, knowledge bases, context, conceptualization

1. Wstęp

Najnowsze trendy obowiązujące w dziedzinie baz wiedzy mających zastosowanie w semantycznym opisywaniu danych znajdujących się w sieci globalnej wskazują na konieczność narzucenia im architektury modularnej. Główną tego przyczyną jest konieczność powtórnego wykorzystania już istniejących ontologii i włączania ich do nowo tworzonych systemów. W sytuacji gdy dołączane ontologie również korzystają z innych, powstaje sieć wzajemnie na siebie oddziałujących modułów. W takiej sieci pojawia się wiele nowych problemów, na

¹ Praca częściowo finansowana z funduszy NCBiR, grant nr SP/I/1/77065/10, projekt SYNAT.

przykład problemy: zarządzania przepływem wniosków między modułami, zarządzania nowymi wersjami, przydzielania uprawnień i in. Z czysto technicznych powodów sytuacja ta wymusza wprowadzenie pewnych standardów oraz metodyk projektowania ontologii modularnych. Próbą uporządkowania sytuacji w tej dziedzinie jest projekt NeOn (<http://www.neon-project.org/>), w ramach którego opracowano pewne standardy dla sieci ontologii oraz oparte na nich techniki i narzędzia.

Choć oczywiście tego typu rozwiązania są bardzo ważne, nie podejmują one w naszej opinii najpoważniejszego problemu związanego z tego rodzaju sieciami. Nie istnieje do dziś żaden sposób opisu kontekstu semantycznego, w którym umieszczony jest dany moduł.

Twórcy każdego z węzłów sieci ontologii kierują się w trakcie prac pewnym punktem widzenia. W specyficzny sposób, zgodny z narzuconymi celami, postrzegają dziedzinę, którą opisują. Rzutuje to z kolei na sposób wyodrębniania obiektów i relacji między nimi oraz na sposób ich sklasyfikowania; innymi słowy, na sposób przeprowadzenia konceptualizacji. Tak więc punkt widzenia jest istotnym czynnikiem wpływającym nie tylko na kształt powstałej ontologii, lecz także na przyszłe kierunki jej rozwoju.

Oczywiste jest więc, że chcąc daną ontologię powtórnie wykorzystać, należy dokładnie poznać punkt widzenia jej autorów. Użytkownik, który podejmie decyzję o powtórnym wykorzystaniu ontologii bez dokładnej znajomości tego punktu widzenia, podejmuje duże ryzyko. Jest bardzo prawdopodobne, że projekt się nie powiedzie lub też któraś z kolejnych wersji powtórnie użytej ontologii nie będzie nadawała się do zastosowania.

W chwili obecnej nie ma innego sposobu, by poznać punkt widzenia autorów danej ontologii, niż dokładne poznanie i zrozumienie jej zawartości. W niektórych przypadkach może to być trudne lub nawet niewykonalne.

Aby pokonać tę trudność, należy opracować formalizm umożliwiający wystarczająco precyzyjne opisanie czegoś, co nazywamy *semantycznym kontekstem modułu*, by zastąpić szerokie pojęcie punktu widzenia. Co rozumiemy przez semantyczny kontekst modułu, w jaki sposób powinien zostać zdefiniowany i jak opisany? Odpowiedź na to pytanie nie jest łatwa. Dotychczasowe definicje kontekstu nie są wystarczające. W każdym razie zastosowanie ich do opisu węzłów istniejących obecnie sieci ontologicznych albo jest niewykonalne (gdy dokonywane są próby stosowania metod formalnych), albo niewiele daje (gdy wykorzystywany jest język naturalny). Ich semantyczny kontekst był ustalany w sposób czysto intuicyjny, często nieujawniony, a więc trudny do opisania.

Uważamy, że aby poprawnie odpowiedzieć na postawione wyżej pytanie, należy wrócić do badań nad pojęciem kontekstu. Wiemy z doświadczenia, że człowiek w procesie myślenia i komunikacji korzysta z kontekstowego ujęcia wiedzy. Wiemy, że im dyskurs jest bardziej fachowy, tym ważniejsze jest precyzyjne zdefiniowanie kontekstu oraz że w wąskich dziedzinach specjalistycznych istnieją pewne pojęcia pozwalające ten kontekst bardzo uściślić.

Wiemy też, że potrafimy dość szybko zmienić kontekst i doskonale orientujemy się w tym, jak ta zmiana wpływa na sposób ujęcia tematu.

W naszych pracach nad bazami wiedzy przyjmujemy, że procesowi konceptualizacji zawsze towarzyszy proces kontekstualizacji. Dlatego preferujemy takie podejście do procesu tworzenia baz wiedzy, które zakłada, że są one od początku bazami kontekstowymi. Podczas projektowania baz kontekstowych projektant, podejmując każdą decyzję projektową, umieszcza ją w odpowiednim kontekście, a definiując nowy kontekst, umieszcza go w strukturze innych kontekstów, mówiąc tym samym, w jaki sposób jego istnienie wpływa na wiedzę zawartą w innych kontekstach. Tylko taki sposób projektowania pozwoli nam zaobserwować, jak podział na konteksty wpływa na proces konceptualizacji. Taki rodzaj projektowania baz wiedzy możemy nazwać *semantycznym projektowaniem kontekstowym*.

Niniejszy artykuł jest rozwinięciem tematyki poruszanej już w [1]. Chcemy w nim przedstawić próbę opisanie wzajemnego wpływu zależności między kontekstami oraz kształtu taksonomii klas. Podejmując tę próbę, wyszliśmy od metody opisu zależności między kontekstami, zwanej metaforą kontekstu jako pudełka (ang. *metaphor of context as a box*), oraz metody OntoClean, jednej z najbardziej znanych metod semantycznej oceny jakości ontologii niekontekstowych. Pojęcia definiowane przez tę ostatnią rozszerzamy w ten sposób, aby objąć nimi zjawiska charakterystyczne dla środowiska kontekstowego. W ten sposób stworzymy propozycję *metody OntoClean-C*, która być może przybliży nas do formalizmu będącego w stanie opisać semantyczny kontekst modułu, widziany jako miejsce w pewnej strukturze kontekstowej.

W dalszej części artykułu z metodą OntoClean-C konfrontujemy znane techniki tworzenia kontekstowych baz wiedzy, m.in. metodę SIM, i pokazujemy, dlaczego właśnie metoda SIM wyznacza właściwy kierunek organizowania modularnych baz wiedzy.

2. Preliminaria

W punkcie tym opiszemy niektóre wyniki badań nad kontekstowym sposobem przedstawiania wiedzy oraz naszkicujemy metodę OntoClean na tle wyników prac dotyczących procesu konceptualizacji.

2.1. Kontekstualizacja wiedzy

Logicy zajmują się pojęciem kontekstu już od dawna. Wynika to stąd, że poszczególne słowa, zwroty czy nawet fragmenty wypowiedzi formułowanych w języku naturalnym mają zmienne znaczenie, zależne od kontekstu. Samo słowo *kontekst* nie ma ścisłej definicji. Naj-

częściej używana jest jego definicja lingwistyczna: „tekst, w którym użyto danego (zwłaszcza cytowanego gdzie indziej) wyrazu albo zwrotu” [2]. Szersza definicja podaje cztery znaczenia, z których jedno wychodzi poza terminy lingwistyczne: „zespół czynników współistniejących, powiązanych z czymś” [3]. To „coś” może być sytuacją, otoczeniem, dziedziną, układem, tłem lub środowiskiem i może zawierać encje, obiekty lub cokolwiek innego, co jest przedmiotem zainteresowania.

Początkowo logicy byli zainteresowani wąskim rozumieniem słowa kontekst. Rozwój logiki i jej zastosowań spowodował jednak, że konieczne stało się sformalizowanie takich zjawisk, jak występowanie różnych punktów widzenia czy istnienie różnych przekonań (postaw propozycyjalnych). W szczególności dotyczyło to prac nad systemami agentowymi oraz związanego z tymi systemami procesu dowodzenia twierdzeń. Ogólnie mówiąc, logicy dostrzegli, że istnieje potrzeba uwzględnienia różnego sposobu wartościowania danego zdania nie tylko w ramach różnych interpretacji, lecz także w ramach jednej interpretacji. Formalizmem, który jako pierwszy uwzględnił tę potrzebę, jest logika modalna [4], opracowana na początku XX wieku przez Clarena Irvinga Lewisa i rozwinięta przez Saula Kripkego. Podobną rolę odgrywają prace Jaakko Hintikki, który zaproponował system zwany dziś systemem modelowym Hintikki (ang. *Hintikka's model system*) [5].

Wprowadzenie powyższych typów logiki pozwoliło ująć wiele zjawisk dotychczas nieuchwytnych i w rezultacie umożliwiło reprezentowanie różnych punktów widzenia i różnych postaw propozycyjnych. Jednak wielu badaczy problematyki sztucznej inteligencji uznało to rozwiązanie za niewystarczające i obarczone wadami. Konstrukcje, jakimi są światy możliwe i zbiory modelowe, pozwalają wyrazić różne modele świata, ale modele te tyczą się światów jako całości. To daje możliwość opisanie np. zbioru systemów agentowych, z których każdy dysponuje inną bazą wiedzy o danej dziedzinie.

Związek z postawą intencjonalną nie jest jednak jedynym czynnikiem warunkującym punkt widzenia. W nowszych pracach zauważano, że wartość logiczna zdań zależy również od innych uwarunkowań, przede wszystkim od czasu i miejsca wypowiedzenia zdania. Dokładnie problem ten ujmuje logika zaimków wskazujących (ang. *logic of demonstratives*), zaproponowana przez Davida Kaplana [6]. Inni autorzy wymieniają jeszcze inne czynniki, takie jak zakres dziedziny dyskursu lub przyjęty poziom szczegółowości [7].

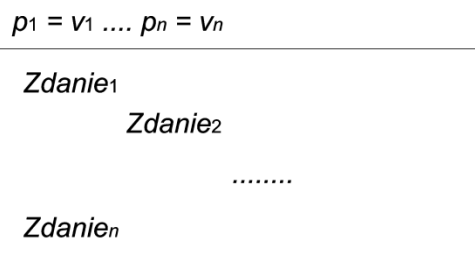
W naszych rozważaniach celem jest znalezienie sposobu formalizacji semantycznego kontekstu modułu. Czynniki determinujące kontekst są więc bardzo ważne. Dlatego wrócimy do nich w dalszej części tekstu. Będziemy je nazywać *parametrami kontekstowymi*.

Formalizm, na którym opiera się najbardziej obecnie znana technika tworzenia modularnych baz wiedzy, zwana DDL (ang. *Distributed Description Logic*) [8], abstrahuje od parametrów kontekstowych, skupia się natomiast na mechanizmach wnioskowania kontekstowego. Formalizm ten powstał na podstawie prac Fausto Giunchiglii i współpracowników i nosi

nazwę DFOL (*Distributed First Order Knowledge*) [9]. Bazuje on na dwóch *zasadach wnioskowania kontekstowego*, zdefiniowanych w [7]. Zasada pierwsza, nazywana *pierwszą zasadą wnioskowania kontekstowego* (ZWK1) lub *zasadą lokalności*, mówi, że wnioskowanie musi być zawsze dokonywane z punktu widzenia pojedynczego modułu i wszelkie wnioski wyciągane podczas wnioskowania muszą być spójne z każdym z aksjomatów zawartych w tym module. Zgodnie z zasadą drugą, zwaną *drugą zasadą wnioskowania kontekstowego* (ZWK2) lub *zasadą kompatybilności*, wszystkie wyciągane z danego modułu wnioski muszą uwzględniać *ograniczenia strukturalne*, które wynikają z wpływu innych modułów współtworzących wraz z danym modulem kontekstową bazę wiedzy.

DDL jest adaptacją tego kontekstowego formalizmu do logiki opisowej. W jego ramach, podobnie jak w DFOL, ograniczenia strukturalne są realizowane za pomocą *reguł pomostowych*. Reguły tworzone za pomocą odrębnego języka rozszerzają zbiór podstawowych reguł wnioskowania przez szczegółowy opis semantyki pojęć zdefiniowanych w jednym module względem semantyki pojęć zdefiniowanych w drugim module. Inaczej mówiąc, są niskopoziomowym opisem zależności między pojęciami zdefiniowanymi wewnątrz łączonych modułów. Reguł tych mogą być setki, nawet dla stosunkowo prostych modułów. Nie są więc one opisem zależności między modułami, lecz opisem zależności między zawartościami modułów i jako takie nie nadają się do realizacji celu, jaki został postawiony we wstępie tego artykułu.

Wróćmy zatem do koncepcji parametrów kontekstowych. Bez względu na to, ile czynników uwzględnimy, można powiedzieć, że tworzą one pewną przestrzeń parametrów wyznaczających kontekst, a co za tym idzie – wartość logiczną zdań. Konstatacja ta jest punktem wyjścia dla metafory przedstawiającej kontekst w postaci pudełka [10]. Tutaj opisujemy ją w postaci, w jakiej ujęto ją w [7].



Rys. 1. Kontekst jako pudełko (na podst. [7])

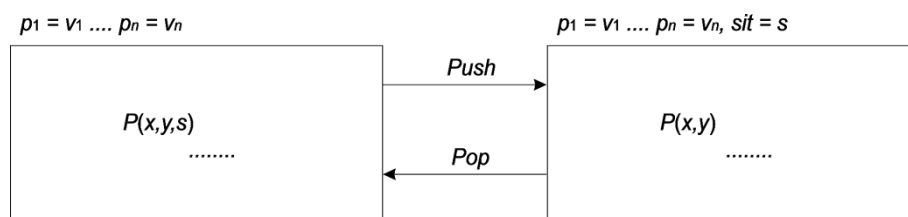
Fig. 1. Context as a box (based on [7])

Kontekst jest przedstawiany jako pudełko (rys. 1), wewnątrz którego znajdują się zdania. Poza pudełkiem znajdują się parametry p_1, \dots, p_n . Parametrom tym przypisane są wartości v_1, \dots, v_n . Symbolizują one informację, która niejako została „odjęta” od informacji zawartej w zdaniach znajdujących się wewnątrz kontekstu. Przykładowo, wartość logiczna zdania: „W pobliżu znajduje się leśniczówka” jest uzależniona od miejsca, gdzie to zdanie zostało

wypowiedziane. Podobnie jest ze zdaniem: „Najnowszym odkryciem w dziedzinie transplantologii jest przeszczep twarzy”, tylko że tym razem jego wartość logiczna jest uzależniona od chwili, w której je sformułowano.

Sposób interpretowania zdań wewnątrz kontekstu jest więc funkcją zależną od tego, jakie parametry ten kontekst opisują oraz jakie są ich wartości. Dzięki takiemu opisowi można abstrahować od całej złożoności sytuacji na zewnątrz kontekstu.

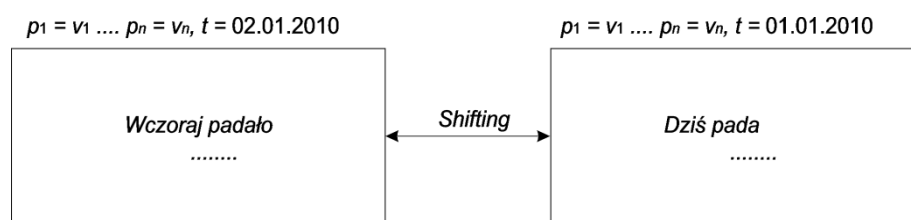
Podstawowym celem stosowania kontekstów jest umożliwienie ograniczenia wiedzy branej pod uwagę podczas rozwiązywania konkretnego problemu. Chodzi więc głównie o wnioskowanie lokalne. Jednakże nie ulega kwestii, że to lokalne wnioskowanie musi być uzależnione od informacji zewnętrznej reprezentowanej przez param p_1, \dots, p_n . Autorzy [7] wyodrębniają dwie metody uwzględniania informacji zewnętrznej: *wstawianie i usuwanie* (ang. *push and pop*) oraz *przesuwanie* (ang. *shifting*). Pierwsza z nich zakłada, że istnieje jakiś stan bezkontekstowy, w którym lista parametrów jest pusta. Jest to stan idealny, który w praktyce nigdy nie istnieje. Proces *kontekstualizacji* polega na stopniowym usuwaniu ze zbioru zdań kolejnych informacji i umieszczaniu ich w zbiorze parametrów. Wynik tego procesu po n -tym kroku jest zilustrowany na rys. 2 jako lewy kontekst. Jednak ten proces może trwać dalej. W wyniku kolejnego kroku ze zdań wewnątrz kontekstu została usunięta informacja s (jest to jedynie symboliczne przedstawienie i nie należy odnosić tego np. do zmiennych), a na liście parametrów pojawił się dodatkowy parametr o nazwie sit i wartości s . Opisana czynność jest odwracalna, więc możliwa jest *dekontekstualizacja*, czyli usunięcie parametru sit i wstawienie s do wnętrza kontekstu (jest to możliwe jedynie dla parametru jawnego).



Rys. 2. Metoda wstawiania i pobierania (na podst. [7])

Fig. 2. Push and pop method (based on [7])

Natomiast metoda przesuwania polega na tym, że wybrany parametr nie znika, lecz zmienia się jego wartość. Na rys. 3 widoczna jest zmiana wartości parametru t . Jeśli rzeczywiście w dniu 1 stycznia 2010 roku padał śnieg lub deszcz, to w kontekście określonym tą datą prawdziwe jest zdanie „Dziś pada”, a w kontekście określonym datą o jeden dzień późniejszą zdanie „Wczoraj padało”.



Rys. 3. Metoda przesuwania (na podst. [7])

Fig. 3. Shifting method (based on [7])

Opierając się na metaforze kontekstu jako pudełka, możemy próbować opisać semantyczne cechy kontekstów. Jeśli na przykład zbiór parametrów kontekstu k_1 zawiera się w zbiorze parametrów kontekstu k_2 i wartości parametrów wspólnych są równe, to możemy powiedzieć, że kontekst k_2 jest bardziej wyspecjalizowany od kontekstu k_1 . Podobnie jest wówczas, gdy zbiory parametrów są równe, ale wartości niektórych z parametrów dla kontekstu k_2 są bardziej precyzyjne. Jednakże aby mówić o wzajemnej relacji wartości parametrów, musimy jakoś uporządkować ich dziedziny.

Narzucając pewne ograniczenia na zbiory parametrów oraz na ich dziedziny, możemy łatwo przekształcić relację między kontekstami w częściowy porządek. Takie uporządkowanie bardzo pomogłoby opisać semantyczne cechy kontekstów.

2.2. Konceptualizacja w świetle metody OntoClean

Proces konceptualizacji był tematem wielu publikacji Nicoli Guarina i Chrisa Welty'ego. Skupiały się one na problemie poprawnego tworzenia taksonomii pojęć. W rezultacie autorzy ci opracowali metodę semantycznej oceny jakości ontologii, zwaną OntoClean [11]. Podstawowym celem tej metody jest dostarczenie twórcom ontologii zbioru precyzyjnie zdefiniowanych reguł, mówiących o tym, w jaki sposób nakładać na koncepty (klasy) relację dziedziczenia (subsumcji).

Proces zaczyna się od przypisania każdemu z konceptów (oryg. ang. *property*) czterech metawłaściwości. Metawłaściwości te mówią, czy dany koncept jest *permanentny* (ang. *rigid*), czy *dostarcza* lub *przenosi kryterium tożsamości* (ang. *supplying or carrying identity*), czy opisuje encje będące całością złożoną z części (ang. *carrying unity*) oraz czy jego istnienie jest zależne od istnienia innych konceptów (ang. *dependent*).

Permanentność zdefiniowano w następujący sposób:

Definicja 2.1 (właściwość permanentna, niepermanentna i przeciwpermanentna)

Koncept C jest *permanentny* wtedy i tylko wtedy, gdy $\forall x (C(x) \rightarrow \Box C(x))$.

Koncept C jest *niepermanentny* wtedy i tylko wtedy, gdy $\exists x (C(x) \wedge \neg \Box C(x))$.

Koncept C jest *przeciwpermanentny* wtedy i tylko wtedy, gdy $\forall x (C(x) \rightarrow \neg \Box C(x))$. ■

W powyższej definicji zastosowano notację z kwantyfikowanej logiki modalnej. Należy je interpretować w następujący sposób: Koncept jest permanentny, jeśli dla każdego osobnika, który jest jego wystąpieniem, jest on *koniecznie* (czyli we wszystkich światach możliwych) jego wystąpieniem. Przy odniesieniu tego do danej ontologii oznacza to, że koncept jest permanentny, jeśli twórcy tej ontologii wykluczają jakąkolwiek możliwość takiego jej rozwoju, w wyniku którego pewne osobniki, zanim minąłby czas ich istnienia, przestałyby być jego wystąpieniami. Takim konceptem jest na przykład *Człowiek*. Jeśli dany osobnik jest jego wystąpieniem (jeśli dany obiekt jest człowiekiem), to będzie nim zawsze.

Koncept jest niepermanentny, jeżeli istnieje taki osobnik będący jego wystąpieniem, dla którego nie jest on koniecznie (czyli nie we wszystkich światach możliwych) jego wystąpieniem. Po prostu gdy istnieje jakakolwiek wątpliwość, że koncept jest permanentny, to jest on niepermanentny. Przykładem może być tutaj koncept *Arcydzieło literackie*. Najczęściej arcydzieło literackie pozostaje arcydziełem na zawsze, a upływający czas jedynie potwierdza jego znaczenie. Może się jednak zdarzyć, że jakieś dzieło literackie, mimo że początkowo wysoko cenione, przestaje być takim po upływie jakiegoś czasu.

Koncept jest przeciwpermanentny, jeśli dla każdego osobnika będącego jego wystąpieniem nie jest on koniecznie jego wystąpieniem. W przypadku takiego konceptu istnieje pewność, że każdy osobnik będący jego wystąpieniem musi po jakimś czasie wyjść poza zakres jego odniesienia. Przykładem może być koncept *Student*. Wiadomo, że każdy student jest nim jedynie w pewnym przedziale czasu.

Po nadaniu konceptom odpowiednich metawłaściwości podczas tworzenia hierarchii dziedziczenia należy zastosować odpowiednie reguły. I tak w przypadku permanentności odpowiednia reguła mówi, że koncept permanentny nie może dziedziczyć od przeciwpermanentnego.

Ze względu na ograniczoną ilość miejsca nie będziemy opisywać innych metawłaściwości. W tym artykule interesuje nas jedynie, w jaki sposób można zapożyczyć pewne pojęcia z metody OntoClean i dostosować je do środowiska kontekstowego. Wystarczy, jeśli prześledzimy to na przykładzie permanentności. Dla innych metawłaściwości wywód byłby w ogólnych zarysach podobny.

3. Propozycja metody OntoClean-C

W tym podrozdziale w taki sposób przedstawimy propozycję modyfikacji niektórych pojęć zaczerpniętych z metody OntoClean tak, by ujmowały one zjawiska występujące podczas projektowania kontekstowych baz wiedzy.

Metoda *OntoClean* jest dla nas wartościowa nie dlatego, że systematyzuje proces oceny ontologii, lecz dlatego, że nakłada pewien formalny system pojęciowy na proces konceptualizacji. Dla naszych rozważań istotne jest, jak ten system się zmieni, gdy odniesiemy go do sytuacji, w której mamy do czynienia z ontologią kontekstową. Najpierw więc spróbujemy opisać taką sytuację.

Aby scharakteryzować strukturę kontekstową, należy, stosując metaforę kontekstu jako pudełka, nałożyć relacje porządku na dziedziny, z których będą pobierane wartości parametrów kontekstowych. Niech $P = \{p_1, \dots, p_n\}$ będzie skończonym zbiorem parametrów kontekstowych branych pod uwagę podczas projektowania kontekstowej bazy wiedzy K . Zdefiniujmy funkcję $f_p: K \rightarrow \mathcal{P}(P)$ (gdzie $\mathcal{P}(P)$ jest zbiorem potęgowym zbioru P), która każdemu kontekstowi $k \in K$ przyporządkuje podzbiór P . Przyporządkowanie parametrom z P wartości polega na stworzeniu *krotki wartościującej* $\sigma = \{(p_1, v_1), \dots, (p_n, v_n)\}$; każda wartość v_i ($i \in [1..n]$) pochodzi z dziedziny Δ_{p_i} dopuszczalnych wartości dla parametru p_i . Symbolem $\sigma(p_i)$ będziemy oznaczać wartość odpowiadającą parametrowi p_i w krotce σ . Wszystkie dopuszczalne krotki wartościujące tworzą relację $\Sigma \subset \prod_{p \in P} \Delta_p$.

Jeśli teraz naszym celem jest stworzenie struktury kontekstów, musimy na każdą dziedzinę Δ_{p_i} nałożyć porządek częściowy. Załóżmy, że taki porządek jest realizowany przez relację kompatybilności \Rightarrow_{p_i} z elementem maksymalnym $\max(\Delta_{p_i})$ takim, że każda wartość $v \in \Delta_{p_i}$ jest kompatybilna z $\max(\Delta_{p_i})$ (co zapisujemy $v \Rightarrow_{p_i} \max(\Delta_{p_i})$).

Mówimy, że krotka wartościująca σ_2 jest kompatybilna z krotką σ_1 ($\sigma_1 \Rightarrow \sigma_2$) wtedy i tylko wtedy, gdy $\forall p \in P (\sigma_1(p) \Rightarrow_p \sigma_2(p))$.

Nałożenie takiego porządku pozwala nam utworzyć przestrzeń, w której każdy punkt reprezentuje jeden kontekst możliwy do utworzenia w bazie K . Nazwijmy ją *przestrzenią kontekstową*. Element maksymalny wyznacza w niej najszerzy kontekst interpretacji zdań, czyli minimalną liczbę dodatkowych założeń, które musimy przyjąć, aby w ogóle dobrze zrozumieć jakiegokolwiek intencje projektanta bazy wiedzy (można także powiedzieć, że taki kontekst wyznacza minimalną liczbę dodatkowych zdań „odjętej informacji” z pudełka). Jednak niektóre informacje zawarte w bazie (w wybranych modułach) do poprawnego (zgodnego z intencją projektantów) zinterpretowania będą wymagały więcej dodatkowej informacji, czyli będą wyrażone w mniejszych kontekstach, gdzie „mniejszość” należy rozumieć jako dopuszczanie mniejszej liczby możliwych interpretacji.

Przestrzeń kontekstów bardzo dobrze współgra z koncepcją możliwych światów wyrażoną w metodzie *OntoClean*. Rozważmy ponownie przykład permanentności: Guarino i Welty definiują ją w terminach logiki modalnej za pomocą warunku: $\forall x (C(x) \rightarrow \Box C(x))$, który musi być spełniony, jeśli koncept C jest permanentny. Powstaje jednak pytanie, co rozumieć

przez konieczność (operator modalny „ \Box ”), czyli, innymi słowy, jaki rozważać zbiór możliwych światów.

W ontologiach niemodularnych możliwe światy można rozumieć jako skutek aktualizacji ontologii dostosowującej ją do stanu świata rzeczywistego. Zatem jeśli jakiś koncept C zostaje uznany przez projektanta za permanentny, należy założyć, że przyjmuje on, że osobnik nie przestanie do niego przynależać w toku kolejnych aktualizacji ontologii. Jeśli potraktujemy zmieniające się warunki w świecie zewnętrznym (np. wpływający czas) jako zmieniające się wartości wektora parametrów kontekstowych $P = \{p_1, \dots, p_n\}$, uzyskamy przełożenie idei *OntoClean* na ideę pudełek kontekstowych i przestrzeni kontekstów.

Każda zmiana ontologii odzwierciedla zatem jeden z możliwych światów i może być reprezentowana przez relację Σ . Różne wersje naszej ontologii O możemy więc powiązać z różnymi wektorami $\sigma \in \Sigma$ i oznaczyć jako $O(\sigma)$. Takie posunięcie umożliwia nam wyrażenie zasady permanentności bez operatorów logiki modalnej (w poniższej definicji $\mathcal{S}(O)$ oznacza sygnaturę O):

Definicja 3.1 (właściwość permanentna, użycie wektora parametrów kontekstowych)

Koncept C jest permanentny w ontologii O , jeśli dla każdej pary wersji tej ontologii $O(\sigma)$, $O(\sigma')$ dla każdego x należącego do sygnatury obu wersji $C(x)$ jest albo wnioskiem płynącym z obu wersji, albo z żadnej:

$$\forall x (\forall \sigma, \sigma' \in \Sigma (x \in \mathcal{S}(O(\sigma)) \cap \mathcal{S}(O(\sigma')) \rightarrow (O(\sigma) \models C(x) \leftrightarrow O(\sigma') \models C(x)))) \quad \blacksquare$$

Takie przekształcenie definicji permanentności pozwala nam na znacznie łatwiejsze powiązanie tego pojęcia z przestrzenią kontekstów. Można łatwo zauważyć, że permanentność w oryginalnym rozumieniu autorów koncepcji (Guarina i Welty’ego) obowiązuje w całej przestrzeni dopuszczalnych kontekstów (wartości parametrów kontekstowych). Wprowadzając model *OntoClean-C*, postulujemy uelastycznienie tego pojęcia, tak aby mogło obowiązywać w pewnej wyznaczonej podprzestrzeni przestrzeni kontekstów. Podprzestrzeń tę najłatwiej wyznaczyć przez wskazania największego wektora parametrów kontekstowych σ , dla którego będzie obowiązywać właściwość permanentności (zawężamy tym samym zbiór możliwych światów). Tym samym, zamiast o permanentności w ogóle (globalnej), zaczynamy mówić o permanentności lokalnej (obowiązującej w pewnej podprzestrzeni kontekstowej).

Definicja 3.2 (właściwość lokalnie permanentna, wektor parametrów kontekstowych)

Koncept C jest *lokalnie permanentny* w ontologii O dla podprzestrzeni kontekstów wyznaczonej przez wektor parametrów kontekstowych σ , jeśli dla każdej pary wersji tej ontologii $O(\sigma)$, $O(\sigma')$, takiej że $\sigma' \Rightarrow \sigma$, dla każdego x należącego do sygnatury obu wersji, $C(x)$ jest albo wnioskiem płynącym z obu wersji, albo z żadnej:

$$\forall x (\forall \sigma, \sigma' \in \Sigma ((x \in \mathcal{S}(O(\sigma)) \cap \mathcal{S}(O(\sigma'))) \wedge \sigma' \Rightarrow \sigma \rightarrow (O(\sigma) \models C(x) \leftrightarrow O(\sigma') \models C(x)))) \quad \blacksquare$$

Ponieważ wszystkie definicje metawłaściwości wyrażone są za pomocą logiki modalnej, można zastosować do nich dokładnie to samo rozumowanie. Podprzestrzeń kontekstów wyznaczona przez σ zawęża wówczas liczbę rozważanych możliwych światów i ogranicza zakres obowiązywania definicji metawłaściwości.

Wprowadzone powyżej pojęcia można także zastosować do ontologii niemodularnych. Wówczas stosowane mogą być do procesu aktualizacji ontologii do nowej wersji, „lokalność” metawłaściwości zaś oznacza jej zachowywanie dla pewnych rodzajów aktualizacji (oczywiście brak „globalności” oznacza jej utratę dla innych rodzajów aktualizacji). Jednak w przypadku ontologii niemodularnych rozróżnienie między różnymi ontologiami a różnymi wersjami tej samej ontologii jest trudne do uchwycenia i sformalizowania.

Z kolei dla ontologii modularnych zastosowanie pojęć lokalnych metawłaściwości wydaje się bardzo naturalne i bardzo potrzebne. Przestrzeń kontekstów narzuca również porządek na moduły ontologii, a deklaracja umieszczenia modułu w konkretnej podprzestrzeni jest jednocześnie deklaracją zachowania niezmienności pewnych cech modułu we wszystkich jego nowych wersjach. Jest to informacja krytycznie ważna dla procesu ponownego użycia.

Biorąc pod uwagę powyższe rozważania, możemy zastosować przedstawione rozumowanie również do uściślenia pojęcia kontekstu. Przez *semantykę kontekstową* rozumiemy funkcję przypisującą każdej parze (k, ϕ) , gdzie k jest kontekstem, a ϕ właściwością (czyli konceptem lub rolą), wartości metawłaściwości. Zatem proces konceptualizacji uwzględniający kontekstualizację polega na ustalaniu zakresu obowiązywania (lokalnych) metawłaściwości. Natomiast *semantyczna kontekstowa baza wiedzy* (SKBW) utrzymuje obowiązywanie lokalnych metawłaściwości przez cały okres życia (czyli poprzez wszystkie aktualizacje jej zawartości).

4. W stronę SKBW: model SIM

W trakcie naszych wcześniejszych prac opracowany został model bazy wiedzy SIM (tutaj zaprezentujemy jedynie skrótowy opis koncepcji; zainteresowanego Czytelnika odsyłamy do [12, 13]. W modelu tym baza podzielona jest na konteksty (moduły), które zorganizowane są hierarchicznie (według częściowego porządku \leq). Założenie przy tym jest takie, że każdy kontekst znajdujący się wyżej w hierarchii opisuje szerszą dziedzinę, ale w mniejszych szczegółach. Z tym związane są inne specyficzne założenia dotyczące przepływu nazw i wniosków. Każdy kontekst „niższy” musi zawierać wszystkie terminy (nazwy konceptów i ról), które zawiera kontekst „wyższy”, kontekst „wyższy” zaś musi zawierać wszystkie nazwy osobników użyte w którymkolwiek kontekście „niższym”. Wnioski między modułami przepływają według zasady, że wszystkie wnioski możliwe do wyprowadzenia w kontekście

„wyższym” (przy ograniczonej terminologicznie sygnaturze) muszą obowiązywać też w kontekstach „niższych”.

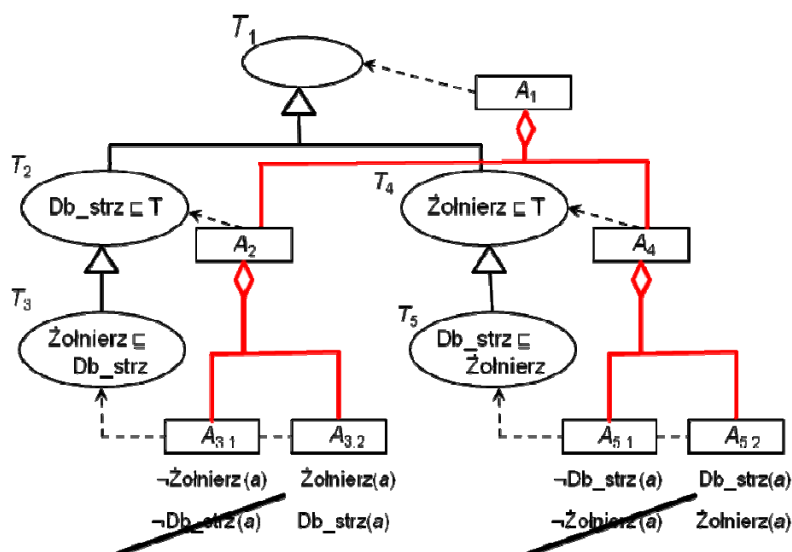
Można pokazać, że przez zastosowanie prostej metody jesteśmy w stanie sprawić, że baza wiedzy SIM zacznie spełniać warunki nałożone na SKBW. Zarys metody można przedstawić w poniższych krokach:

1. Określamy zbiór relewantnych parametrów kontekstowych P .
2. Każdemu kontekstowi K struktury bazy SIM przypisujemy wektor parametrów kontekstowych $\sigma(K)$.
3. Weryfikujemy prawidłowość: $K_1 \leq K_2 \rightarrow (\sigma' \Rightarrow \sigma)$.
4. Ustalamy metawłaściwości dla poszczególnych par (kontekst, koncept), zachowując zasadę lokalności (tj. metawłaściwość ustalona dla konkretnego konceptu w „wyższym” kontekście musi obowiązywać też w kontekstach „niższych”).

Po przeprowadzeniu tego procesu można zauważyć bardzo ciekawą prawidłowość: struktura bazy wraz z regułami wnioskowania narzuca, że zasada lokalnej permanentności będzie spełniona pod rygorem niespójności bazy. Innymi słowy, użytkownik bazy wiedzy, aktualizując ją do kolejnej wersji, nie może naruszyć metawłaściwości bez naruszenia spójności bazy.

Efekt ten możemy zilustrować na prostym przykładzie. Na rys. 4 pokazano fragment struktury bazy SIM. Jej cechą charakterystyczną jest to, że tworzą ją dwie nakładające się na siebie hierarchie: modułów terminologicznych, zwanych *typami kontekstów*, oraz modułów opisujących fakty, zwanych *wystąpieniami kontekstów*. Każda z hierarchii jest tworzona za pomocą innego typu relacji: hierarchię typów kontekstów tworzy *relacja dziedziczenia* (strzałki narysowane linią ciągłą, zakończone grotami trójkątnymi), podczas gdy hierarchię wystąpień kontekstów wyznacza *relacja agregacji* (strzałki narysowane linią ciągłą, zakończone rombami). Obie relacje wyznaczają kierunki i sposób przepływu wniosków: relacja dziedziczenia – w dół, relacja agregacja – w górę. Strzałkami przerywanymi zaznaczono przyporządkowania wystąpień kontekstów właściwym typom (jest to *relacja konkretyzacji*).

W obu gałęziach hierarchii umieszczone są koncepty *Żołnierz* i *Db_strz* (skrót od *Dobry Strzelec*). W lewej gałęzi koncept *Db_strz* został zdefiniowany na wyższym piętrze hierarchii terminologii, w ramach typu kontekstu T_2 . Oznacza to, że we wszystkich wystąpieniach kontekstów, które są agregowane w ramach wystąpienia A_2 (są to wystąpienia $A_{3.1}$ i $A_{3.2}$), koncept ten musi być permanentny. Zilustrowano to, skreślając zdanie $\neg Db_strz(a)$, które wprowadziłoby do ontologii niespójność. Natomiast koncept *Żołnierz*, zdefiniowany w module T_3 , nie jest w wystąpieniach $A_{3.1}$ i $A_{3.2}$ permanentny. Definicja relacji agregacji powoduje, że zdania $Żołnierz(a)$ i $\neg Żołnierz(a)$ w ogóle do A_2 nie docierają (koncept *Żołnierz* nie jest tam widoczny).



Rys. 4. Metoda przesuwania (na podst. [7])

Fig. 4. Shifting method (based on [7])

Lokalność permanentności jest na rysunku ukazana w ten sposób, że w prawej gałęzi hierarchii panuje sytuacja odwrotna: tu koncept *Żołnierz* jest lokalnie (od modułu A_4 w dół) permanentny.

Jeśli w pokazanej ontologii mielibyśmy decydować o kierunku dziedziczenia obu konceptów, to w lewej gałęzi poprawną decyzją byłoby, aby koncept *Żołnierz* dziedziczył od konceptu *Db_strz*, natomiast w prawej gałęzi decyzja powinna być odwrotna.

Można postawić tezę, że oznacza to, iż bazy SIM są bliskie idei SKBW i dobrze nadają się do praktycznej realizacji postawionych w tej pracy postulatów: uwzględniania obecności kontekstowości w procesie konceptualizacji i wykorzystania tej kontekstowości do lepszego opisu modularnych baz wiedzy.

Warto też zauważyć, że sposób postępowania w przypadku baz modularnych opartych np. na idei DDL nie mógłby być tak bezpośredni, jak w prezentowanej tu metodzie. W DDL zależności między modułami wyrażone są w skomplikowany sposób (za pomocą, zazwyczaj bardzo licznych, zbiorów reguł pomostowych). Tym samym możliwa jest ocena, czy dana baza DDL spełnia postulaty SKBW, ale ocena ta musi być przeprowadzona indywidualnie dla każdej pojedynczej bazy, a wręcz dla każdej jej wersji. Nie ma tu możliwości takiej jak w modelu SIM, czyli przyporządkowania pewnych rodzajów relacji między modułami do pewnych aspektów SKBW. Tym samym nie ma możliwości, aby mechanizm badania spójności bazy DDL dbał jednocześnie o zachowanie lokalności metawłaściwości.

5. Podsumowanie

W niniejszej artykule podjęliśmy zagadnienia rozszerzenia metody *OntoClean* w kierunku uwzględnienia problemów kontekstualności wiedzy. Zagadnienie to podjęliśmy, motywowani głównie chęcią zbadania wpływu uwzględniania kontekstów na proces konceptualizacji.

Prace te doprowadziły do zdefiniowania pojęć lokalnych metawłaściwości (w tym permanentności). Lokalność metawłaściwości jest silnym narzędziem pozwalającym na ujednoczenie opisu wiedzy modelowanej z różnych punktów widzenia w ramach pojedynczej przestrzeni kontekstowej. Pojęcie lokalności stało się też dobrym punktem wyjścia do zdefiniowania cech semantycznej kontekstowej bazy wiedzy, czyli bazy wiedzy jawnie uwzględniającej istnienie kontekstów już na etapie jej projektowania.

Wszystkie podjęte kroki wydają się dobrym wstępem do dalszych prac, których zadaniem będzie utworzenie środków opisu modularnych baz wiedzy (czy też sieci ontologii) pozwalających na przyspieszenie procesu ich rozbudowy, zmniejszenie związanego z nim ryzyka utraty spójności pomiędzy kolejnymi wersjami modułów oraz znaczne zwiększenie możliwości ponownego użycia istniejących ontologii.

BIBLIOGRAFIA

1. Goczyła K., Waloszek A., Waloszek W.: Towards context-semantic knowledge bases. Proc. of 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), 2012, s. 475÷482.
2. Kopaliński W.: Słownik wyrazów obcych. Świat Książki, Warszawa 2000.
3. Szymczak M. (red.): Słownik języka polskiego. PWN, Warszawa 1978.
4. Garson J. W.: Modal Logic for Philosophers. Cambridge University Press, 2006.
5. Hintikka J.: Knowledge and Belief. An Introduction to the Logic of the Two Notions. Cornell University Press, 1962.
6. Kaplan D.: On the Logic of Demonstratives. Journal of Philosophical Logic, No. 8, 1978, s. 81÷98.
7. Benerecetti M., Bouquet P., Ghidini C.: Contextual reasoning distilled. Journal of Experimental & Theoretical Artificial Intelligence, Vol. 12, 2000, s. 279÷305.
8. Borgida A., Serafini L.: Distributed Description Logics: Directed Domain Correspondence in Federated Information Sources. On the Move to Meaningful Internet Systems 2002, LNCS, Vol. 2519, Springer Verlag, 2002, s. 36÷53.

9. Ghidini C., Serafini L.: Distributed First Order Logics, [in:] Gabbay D., de Rijke M. (red.): *Frontiers of Combining Systems 2. Studies in Logic and Computation*. Research Studies Press, 1998, s. 121÷140.
10. Bar-Hillel Y.: Indexical Expressions. *Mind*, No. 63, 1954, s. 359÷379.
11. Guarino N., Welty C. A.: Identity, Unity, and Individuality: Towards a formal toolkit for ontological analysis. *Proceedings of ECAI-2000: The European Conference on Artificial Intelligence*, IOS Press, 2000, s. 219÷223.
12. Goczyła K., Waloszek A., Waloszek W.: Hierarchiczny podział przestrzeni ontologii na konteksty, [w:] Kozielski S., Małysiak B., Kasprowski P., Mrozek D. (red.): *Bazy danych. Nowe technologie – architektura, metody formalne i zaawansowany analiza danych*. Wydawnictwa Komunikacji i Łączności, Warszawa 2007, s. 247÷260.
13. Goczyła K., Waloszek A., Waloszek W.: Ontologia PIPSDrugs: praktyczne zastosowanie metody SIM. *Studia Informatica*, Vol. 30, No. 2A (83), Gliwice 2009, s. 157÷168.

Wpłynęło do Redakcji 16 stycznia 2013 r.

Abstract

In the paper we discuss the issue of integrating and building modular knowledge bases. While the number of ontologies is still growing, the process of their reuse is more difficult than expected. One of the factors inhibiting this process is lack of a standard detailed description of assumptions behind design of particular modules. As a result, every ontology engineer face substantial risk when deciding about reuse, as lack of such knowledge may lead to losing compatibility between modules, especially after their updates.

We propose to look for a way of handling this problem by using the notion of context. We perceive conceptualization as a process conducted always within some contextual assumptions, and therefore we formulate the postulate of considering the contextualization as an integral part of conceptualization.

In order to investigate this relation more deeply, we analyzed *OntoClean*, the well-known method of evaluating conceptual assumptions for ontologies. We extended the theory behind *OntoClean* towards contextual modular knowledge bases by introducing definitions of local metaproperties (like local rigidity) and we define a structure for contexts to describe such locality. This leads us to formulation of an introductory definition of a context-semantic knowledge base, i.e. a base designed with the awareness of contextual issues.

We show that SIM model, proposed in earlier papers, may be perceived as a step towards context-semantic knowledge base, as it is possible to enforce maintainability of locality of metaproperties by means of verifying consistency of SIM knowledge base. In other words it is impossible to violate the locality rule without making the base inconsistent.

The presented results seem to be a good introduction towards further work which would lead to formulation of a consistent standard of describing modular ontologies, facilitating processes of their integration and greatly increasing their potential to be reused.

Adresy

Krzysztof GOCZYŁA: Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, ul. Narutowicza 11/12, 80-233 Gdańsk, Polska, kris@eti.pg.gda.pl.

Aleksander WALOSZEK: Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, ul. Narutowicza 11/12, 80-233 Gdańsk, Polska, alwal@eti.pg.gda.pl.

Wojciech WALOSZEK: Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, ul. Narutowicza 11/12, 80-233 Gdańsk, Polska, wowal@eti.pg.gda.pl.