



Recenzja rozprawy doktorskiej
mgr. inż. Krzysztofa Kutta
zatytułowanej:

Collaborative Knowledge Engineering. Methods and Tools for System Design
(Metody projektowania systemów z użyciem kolaboratywnej inżynierii wiedzy)

1. Problem badawczy i jego znaczenie

Temat rozprawy dotyczy wspierania procesu tworzenia baz wiedzy. Autor rozprawy koncentruje się na procesach kolaboratywnych, to jest takich, w których w tworzeniu bazy wiedzy uczestniczą zarówno eksperci dziedzinowi, jak i inżynierowie wiedzy. Procesy kolaboratywne stoją niejako pośrodku pomiędzy procesami kooperatywnymi, w których uczestniczy heterogeniczna społeczność użytkowników Sieci Semantycznej pracująca nad wybranymi przez nich częściami bazy wiedzy, a procesami kolektywnymi, w których uczestnikami są członkowie grupy homogenicznej, złożonej z osób o podobnych kwalifikacjach i doświadczeniu. Aktualnie procesy kolaboratywne zyskują na znaczeniu z uwagi na fakt, że współczesne ogólnodostępne bazy wiedzy stają się coraz bardziej specjalistyczne, a jednocześnie wymagają logicznej spójności i konstrukcji zgodnej z przyjętymi powszechnie paradygmatami inżynierii wiedzy. Przykładami mogą być bazy medyczne dotyczące określonych rodzajów chorób. Bazy takie muszą być tworzone przy udziale specjalistów z zakresu medycyny, a jednocześnie przez inżynierów wiedzy, obeznanych z zasadami i dobrymi praktykami tworzenia ontologii.

Autor postawił przed sobą problem sformalizowania i częściowo zautomatyzowanego wspomagania procesu kolaboratywnej inżynierii wiedzy. Jest to zadanie o tyle niełatwe, że nie poddaje się prostej proceduralnej formalizacji z uwagi na fakt, że uczestniczą w nim ludzie, którzy ze swej natury subiektywnie i krytycznie podchodzą do prób formalizacji procesów twórczych. W zamierzeniu Autora zaproponowana metodyka ma w sposób istotny ułatwić tworzenie i rozwój systemów opartych na wiedzy przeznaczonych do konkretnych zastosowań praktycznych. W tym kontekście tematyka rozprawy ma potencjalnie duże znaczenie praktyczne, szczególnie zważywszy na fakt, że w aktualnym stanie Sieci Semantycznej tempo tworzenia zasobów danych znacznie wyprzedza tempo ich inteligentnego wykorzystywania.

Reasumując tę część recenzji, stwierdzam, że tematyka podjęta przez Autora jest ważna i aktualna dla współczesnej informatyki, szczególnie dla jej gałęzi zwanej inżynierią wiedzy (lub zarządzaniem wiedzą), a prace nad sformalizowaniem i wspomaganiem procesów kolaboratywnej inżynierii wiedzy – za celowe i potencjalnie twórcze.

2. Cele i tezy rozprawy

W podrozdziale 1.2 Autor określa główne cele swojej rozprawy jako zaprezentowanie opisu procesu kolaboratywnej inżynierii wiedzy, w którym określone są role jego uczestników oraz etapy składające się na ten proces, a także zaproponowanie metod i narzędzi wspierających ten proces w utworzeniu bazy wiedzy o wysokiej jakości, w rozsądnym czasie oraz wygodnej w użytkowaniu i rozwoju (moje wolne tłumaczenie z tekstu angielskiego). Do tak sformułowanego celu nie mam zastrzeżeń, choć w pierwszej części ww. celu zamiast sformułowania „to describe” użyłbym „to propose”, a w drugiej części zamiast „to propose” – „to develop”. Istotnie, celem rozprawy doktorskiej nie może być opisanie czegoś, chyba że podkreśli się, że to „coś” ma charakter innowacyjny. Zarazem wspomniane metody i narzędzia zostały przez Autora zaimplementowane i sprawdzone w praktyce, a nie tylko zaproponowane.

Autor precyzuje też cztery cele podrzędne, na które składają się:

- zdefiniowanie pojęcia kolaboratywnej inżynierii wiedzy w porównaniu z innymi procesami tworzenia baz wiedzy w sposób rozproszony (kooperatywnym i kolektywnym),
- zaprezentowanie autorskiego procesu kolaboratywnej inżynierii wiedzy (CKE – *Collaborative Knowledge Engineering*) – co stanowi zasadniczą część pracy i oryginalny wkład Autora,
- określenie wymagań względem narzędzi wspierających proponowany proces CKE,
- zaimplementowanie tych narzędzi.

Dodałbym tu jeszcze piąty cel podrzędny, jawnie nie wymieniony przez Autora – jakim jest przeprowadzenie eksperymentalnej walidacji zaproponowanego rozwiązania.

Autor nie definiuje tezy (lub tez) rozprawy, co wymagane jest przez rady naukowe niektórych wydziałów. Jednak z tak sformułowanych celów rozprawy można łatwo taką tezę wywieść. Mogłaby ona na przykład (w skrócie) głosić, że zaproponowany w rozprawie proces kolaboratywnej inżynierii wiedzy umożliwia tworzenie baz wiedzy, które spełniają określone wymagania jakościowe i użytkowe.

3. Wkład autora

Oryginalny dorobek i wkład w dziedzinę Autor zaprezentował w rozdziałach od 4. do 8. pracy. W rozdziale 4. Autor prezentuje zasadniczy wynik swojej rozprawy jakim jest zwinny (*agile*) kolaboratywny proces inżynierii wiedzy. Proces ten wykorzystuje znane metodyki inżynierii oprogramowania i inżynierii ontologii, dostosowując je do procesu CKE. Sam proces został zilustrowany na tle projektu pod nazwą *European Wiki*. Już w tym miejscu warto podkreślić, że w całej rozprawie Autor stara się ilustrować swoje wywody praktycznymi przykładami, co bardzo dobrze świadczy o podejściu inżynierskim Autora do podejmowanych problemów.

W rozdziałach od 5. do 7. Autor rozpatruje trzy aspekty baz wiedzy powstałych w wyniku proponowanego procesu CKE: aspekt jakościowy, aspekt dokonywania zmian i aspekt użytkowników. Na płaszczyźnie kontrolowania jakości powstającej bazy wiedzy Autor zaproponował podejście polegające na generowaniu zapytań SPARQL do tworzonej bazy wiedzy, które sprawdzają jej spójność lub inne aspekty jakościowe. Autor nazywa te zapytania jednostkowymi testami wnioskowania (*reasoning unit tests*). Na płaszczyźnie kontrolowania zmian (a zmienność jest cechą immanentną baz wiedzy tworzonych jakimikolwiek metodami opartymi na współpracy) Autor zaproponował specjalną ontologię zmian. W efekcie konstruowany jest semantyczny dziennik (*log*) zmian, który pozwala na śledzenie zmian i ewentualne wycofywanie stanu bazy wiedzy. Na płaszczyźnie oceny użytkowników uczestniczących w proponowanym procesie CKE Autor

zaproponował podejście oparte na konkurencyjnych grach, w których każdy z użytkowników otrzymuje nagrodę za czynności dokonywane na tworzonej w procesie bazie wiedzy.

W rozdziale 8. Autor opisuje eksperymenty walidacyjne przeprowadzone na kilku grupach użytkowników. Eksperymenty zostały przeprowadzone w zaimplementowanym przez Autora środowisku o nazwie BiFróST, opartym na środowisku Loki – tzw. semantycznym *wiki*, opracowanym przez grupę badawczą, której członkiem jest Autor. Eksperymenty te potwierdzają, że proponowany przez Autora zwinny proces CKE wraz ze wspierającymi narzędziami wspierającymi pozytywnie wpływa na – rozumianą wieloaspektowo - jakość baz wiedzy tworzonych na zasadach współpracy kolaboratywnej i może być stosowany w praktyce.

Za najważniejsze elementy oryginalnego wkładu Autora w dziedzinę uważam;

- zdefiniowanie i zaimplementowanie w istniejącym środowisku typu *wiki* zwinnego, iteracyjnego procesu kolaboratywnej inżynierii wiedzy, obejmującego określenie ról uczestników procesu, poszczególnych jego etapów oraz metod i narzędzi wspierających ten proces;
- zdefiniowanie ontologii zmian i skonstruowanie wynikającego z tej ontologii semantycznego dziennika zmian dokonywanych na tworzonej bazie wiedzy;
- opracowanie koncepcji i realizacja nadzoru nad jakością tworzonej bazy wiedzy za pomocą jednostkowych testów wnioskowania;
- opracowanie koncepcji oceny użytkowników-twórców bazy wiedzy metodą grywalizacji;
- przeprowadzenie testów i eksperymentów walidacyjnych zaproponowanego procesu.

Warto podkreślić, że Autor wykazał się przy tym bardzo dobrą znajomością istniejących metodyk i narzędzi z zakresu objętego rozprawą i dla osiągnięcia wyznaczonych celów nie usiłował wyważać drzwi już otwartych przez innych badaczy i realizatorów systemów.

Podsumowując tę część recenzji, stwierdzam, że Autor rozprawy w pełni zrealizował postawione w rozprawie cele i wykazał prawdziwość wynikającej z nich – domyslniej – tezy. Uczynił to w sposób wyczerpujący i dojrzały z naukowego punktu widzenia.

4. Ocena rozprawy i uwagi krytyczne

Rozprawa napisana jest w dobrym języku angielskim. W całym tekście natrafiłem na jedynie kilka błędów, głównie literowych. Zamieszczona bibliografia jest bardzo obszerna – zawiera 184 pozycje, w większości z ostatnich kilku lat. Wśród tych pozycji nie znalazłem jednak żadnej, w której autorem lub współautorem byłby Autor rozprawy. Czy wyniki Autora nie zostały do tej pory rozpropagowane? Jeśli tak, to na pewno na to zasługują. Mam też niejaką obawę, że mogłem nie zauważyć w tym spisie poszukiwanych publikacji, gdyż studiowanie obszernego spisu pod kątem poszukiwania konkretnego nazwiska okazało się bardzo kłopotliwe.

Poniższe uwagi dotyczące rozprawy należy traktować jako uwagi dyskusyjne, które mogą służyć ulepszeniu dalszych prac w obszarze objętym rozprawą, wskazanych przez Autora w rozdziale 9.

1. W analizie istniejących metodyk tworzenia baz wiedzy (w tym baz nieustrukturalizowanych) Autor nie wspomina techniki znanej (w różnych odmianach) pod nazwą Delphi, polegającej na kilku-, zazwyczaj dwuetapowym uzgadnianiu przez ekspertów zdania na pewien temat, z możliwością odrzucania opinii ostatecznie sprzecznych. Wydaje się, że ta technika

znakomicie nadaje się jako element wzbogacający zwinny proces CKE (np. w module „Opinie i dyskusje”), szczególnie dla baz wiedzy o silnym charakterze eksperckim.

2. Autor używa terminu „semantic knowledge base” (np. str. 57 u dołu). Słowo „semantic” jest tu zbyt techniczne (bazy wiedzy są z natury „semantyczne”). Prawdopodobnie chodzi Autorowi o takie bazy wiedzy, które są zgodne z pewnym paradygmatem, np. z RDF.
3. Opisując implementację w systemie Loki jednostkowych testów wnioskowania (str. 60) w kroku 2. procedury użytkownik ma podać oczekiwany wynik zapytania. Czy taki wynik zawsze można przewidzieć, nawet dla bazy poprawnie skonstruowanej? Czy konieczność podania wyniku zapytania nie ogranicza zakresu testów?
4. Prezentując metryki bazy wiedzy w podrozdziale 5.2 (str. 70), warto było sprecyzować, co Autor rozumie przez atrybuty. Czy są to atrybuty w sensie języka OWL (czyli właściwości o zakresach będących dziedzinami konkretnymi), czy są to właściwości (*properties* w sensie OWL) wiążące klasy, czy też może i to, i to? Jest to istotne np. dla znaczenia metryki 5.3. Ponadto opis wspomnianej metryki na stronie 72 jest błędny.
5. Na stronach 72 i 75 używane są dwa różne terminy: *coherence* i *consistency* odnoszące się do tej samej miary WA_2 .
6. Śledzenie tekstu miejscami utrudnia odnoszenie się do kolorów na rysunkach (np. rys. 6.3), które nie są kolorowe.

5. Podsumowanie

Stwierdzam, że mgr inż. Krzysztof Kutt w swojej rozprawie doktorskiej wykazał się bardzo dobrą znajomością inżynierii wiedzy, a także nowoczesnej inżynierii oprogramowania. Przedstawił w sposób wyczerpujący aktualny stan wiedzy w tym zakresie, a także istniejące narzędzia, które potrafił wykorzystać dla osiągnięcia celów swojej pracy. Swoją głęboką znajomość obszaru objętego tematyką rozprawy potwierdził szerokim doбором literatury przedmiotu.

Oceniam, że Kandydat jest badaczem zdolnym do prowadzenia samodzielnych badań w obszarach związanych z zarządzaniem wiedzą. W swojej rozprawie wykazał się wiedzą i umiejętnościami wymaganymi do uzyskania stopnia doktora nauk technicznych w dyscyplinie **informatyka** zgodnie z obowiązującą *Ustawą z dn. 14 marca 2003 r. o stopniach naukowych i tytule naukowym* (z późn. zmianami). W konsekwencji wnoszę o przekazanie recenzowanej rozprawy do dalszych etapów przewodu doktorskiego.

