

prof. dr hab. inż. Andrzej Skowron
Instytut Matematyki UW
Banacha 2, 02-097 Warszawa

Warszawa, dnia 17 września 2013 roku

Recenzja rozprawy doktorskiej
pt. *Zastosowanie wiedzy ontologicznej w wyszukiwaniu danych bibliograficznych*
mgr. inż. Piotra Grochowalskiego

Niniejsza opinia została przygotowana w odpowiedzi na pismo Prof. dr hab. Danuty Stróż, Dziekana Wydziału Informatyki i Nauki o Materiałach Uniwersytetu Śląskiego, w związku z powołaniem mnie przez Radę Wydziału Informatyki i Nauki o Materiałach Uniwersytetu Śląskiego na recenzenta w przewodzie doktorskim mgr. inż. Piotra Grochowalskiego.

Tematyka rozprawy

Jeden z kierunków intensywnie rozwijanych na świecie badań wiąże się z poszukiwaniem metod semantycznego wyszukiwania. W szczególności metody takie winny umożliwiać wyszukiwanie, w bazach danych tekstowych czy też w Internecie, dokumentów tekstowych relewantnych dla zadawanych zapytań czy też uzyskiwanie odpowiedzi na bardziej zaawansowane ('inteligentne') pytania. Mimo, że do pełnego rozwiązania takich zadań jest jeszcze daleka droga to w ostatnich latach, dzięki ogromnym wysiłkom w wielu ośrodkach badawczych na świecie i w firmach komputerowych, dokonano w tym zakresie istotnego postępu. Recenzowana rozprawa doktorska wpisuje się w tę tematykę. Autor rozprawy wykorzystuje jako źródło danych tekstowych bibliograficzną bazę danych tekstowych. Celem rozprawy jest opracowanie metod i algorytmów wydobywania ze zgromadzonych rzeczywistych danych tekstowych w tej bazie dodatkowej wiedzy, która okazałaby się użyteczna w procesie semantycznego wyszukiwania pożądaných informacji. Ta dodatkowa wiedza winna wspomagać konstruowanie odpowiedzi na pytania kierowane do baz danych tekstowych poprzez umożliwienie wydobywania z niej struktur w zadawalającym stopniu dopasowanych semantycznie do zapytań. Autor rozprawy proponuje wydobywanie tej wiedzy dodatkowej w postaci różnego typu ontologii. Cel rozprawy został określony jasno. Nie ulega dla mnie wątpliwości, że podjęta w rozprawie tematyka może stanowić przedmiot badań rozprawy doktorskiej.

Omówienie treści rozprawy

Rozprawa doktorska mgr. inż. Piotra Grochowalskiego liczy 129 stron i składa się siedmiu rozdziałów, czterech dodatków, spisu literatury oraz spisów rysunków oraz tabel.

Wstęp do rozprawy przedstawia wprowadzenie do rozprawy wraz ze szczegółowym omówieniem problematyki rozprawy na tle innych badań. Zawiera on też opis zawartości rozprawy oraz głównych wyników w niej zawartych. W rozdziale 2 podane zostały podstawowe pojęcia dotyczące relacji binarnych, zbiorów przybliżonych, grafów, ciągów, przestrzeni metrycznych, w szczególności stosowanych metryk, wektorów oraz współczynnika korelacji rang Spearmana. Rozdział trzeci przedstawia proces wydobywania dodatkowej wiedzy z danych oraz to w jaki sposób będzie ona wykorzystywana. Istotnym aspektem dyskutowanym w tym rozdziale jest reprezentacja danych tekstowych oraz przygotowanie systemu bibliograficznego do wykorzystania tej dodatkowej wiedzy. Reprezentacja wiedzy w rozprawie bazuje na ontologiach. W rozdziale tym omawiane jest zagadnienie agregacji pozyskanej wiedzy dodatkowej oraz wykorzystanie tej wiedzy w wyszukiwaniu informacji. Pozyskana dodatkowa wiedza pozwala dzięki metodom zbiorów przybliżonych na określenie stopnia dopasowania danych tekstowych do poszukiwanej informacji wyrażonej w zapytaniach. Ważnym zagadnieniem tu omawianym jest też przyjazna dla użytkownika wizualizacja informacji. Rozdział czwarty poświęcony jest omówieniu systemu komputerowego RSDS (ang. *Rough Set Database System*) wykorzystywanego w rozprawie jako system bibliograficzny tekstów, do którego kierowane są zapytania użytkownika. Omówiono architekturę systemu, strukturę logiczną, funkcjonalności, dane wejściowe i wyjściowe oraz wymagania sprzętowe. System ten jest wynikiem wieloletniej pracy, w której doktorant miał wiodący udział. Użycie tego systemu do wyszukiwania informacji semantycznej wymagało dodatkowych rozszerzeń, w szczególności tych omówionych w rozdziale trzecim. Rozdział piąty zawiera metody analizy współpracy między autorami z RSDS, w szczególności bazujące na grafie współpracy. Uzyskane wyniki charakteryzują zarówno poszczególnych autorów jak i trendy rozwoju zbiorów przybliżonych. Rozdział szósty zawiera omówienie przeprowadzonych eksperymentów wraz z wynikającymi z ich wnioskami. Ostatni rozdział jest podsumowaniem rozprawy. Zawiera on też informację o kierunkach dalszych badań.

Ocena rozprawy

Doktorant w pełni zrealizował cel rozprawy. Wykazał się bardzo dobrą znajomością literatury oraz samodzielnością w formułowaniu i rozwiązywaniu nietrywialnych problemów z bardzo aktualnej dziedziny badań. Redakcja rozprawy jest bardzo staranna.

Do głównych wyników rozprawy zaliczam:

1. Opracowanie ontologii dziedzinowej dla zbiorów przybliżonych. Wymagało to poważnej pracy w związku z tym, że dla około 160 pojęć występuje tysiąc kilkaset powiązań między nimi, dla rozważanych relacji semantycznych między pojęciami.

2. Zaprojektowanie i zaimplementowanie systemu RSDS, profesjonalnego bibliograficznego systemu bazodanowego dla publikacji dotyczących zbiorów przybliżonych oraz rozszerzenie funkcjonalności tego systemu zapewniające istotne wsparcie przy rozwiązywaniu problemów rozważanych w rozprawie. Prace nad tym systemem były systematycznie realizowane przez wiele lat.

3. Opracowanie metod ekstrakcji wiedzy z opisów bibliograficznych zawartych w RSDS oraz jej reprezentacji na użytek semantycznego wyszukiwania. Pozwoliło to, w szczególności usprawnić współpracę z użytkownikiem. Ekstrahowane z opisów bibliograficznych lokalne ontologie z jednej strony ułatwiają ocenę dopasowania opisu bibliograficznego do zadawanego pytania, a z drugiej strony pozwalają wzbogacić ogólną ontologie.

4. Opracowanie metody aproksymacji zbioru pojęć relatywnie do wyróżnionej relacji między pojęciami i wykorzystanie tej metody w ocenie dopasowania rozważanej informacji do informacji poszukiwanej.

5. Opracowanie metod analizy sieci społecznościowej utworzonej przez autorów prac z RSDS, w szczególności prowadzących do wykrywania różnych wzorców zachowań obejmujących zależności między autorami, wykrywanie liderów grup, czy też trendów zmian w tej sieci. Opracowane metody bazują na analizie grafów współpracy i jego zmian w czasie.

6. Wnikliwe eksperymenty, które doprowadziły do potwierdzenia jakości opracowanych przez doktoranta metod.

Uwagi krytyczne

1. Zabrakło mi szerszej dyskusji o skalowalności opracowanych w rozprawie metod na większe zbiory danych, co obecnie stanowi wyzwania dla metod analizy danych.

2. Mimo przeprowadzonych wielu doświadczeń uzasadniających wartość opracowanych metod pozostaje pytanie odnośnie jakości całego systemu jako systemu otwartego. Czy system ten będzie w zbliżonym stopniu równie sprawny, gdy zostanie poddany aktualizacji wymagającej dołączenie wielu nowych tekstów? Jeśli nie, to jakie strategie adaptacji można byłoby zastosować.

3. Zabrakło mi przynajmniej wstępnej dyskusji dotyczącej tego, czy opracowane metody mogłyby się okazać skuteczne dla innych dziedzin, w szczególności dla dziedzin mających nieporównywalnie więcej dokumentów tekstowych jak w rozważanym przypadku.

Te uwagi w niczym nie zmieniają mojej pozytywnej ocen rozprawy i mogą być traktowane jako propozycje kierunków dalszych badań.

Mgr inż. Piotr Grochowalski ma poważny dorobek publikacyjny. Jest on współautorem piętnastu publikacji o zasięgu międzynarodowym, w tym trzech publikacji w czasopiśmie z listy filadelfijskiej (Fundamenta Informaticae), dwóch prac w czasopiśmie Transactions on Rough Sets (w serii LNCS), dwóch rozdziałów w książkach o charakterze monograficznym wydanych przez wydawnictwo Springer oraz ośmiu prac opublikowanych w sprawozdaniach z konferencji międzynarodowych.

Konkluzja końcowa

Uważam, że rozprawa doktorska mgr inż. Piotra Grochowalskiego pt. *Zastosowanie wiedzy ontologicznej w wyszukiwaniu danych bibliograficznych* spełnia wszelkie wymogi stawiane przy ubieganiu się o nadanie stopnia doktora (Ustawa o Stopniach i Tytule Naukowym oraz o Stopniach i Tytule w Zakresie Sztuki z dnia 14 marca 2003 roku wraz z późniejszymi zmianami) w dziedzinie nauk technicznych. W związku z tym zwracam się do Rady Wydziału Informatyki i Nauki o Materiałach Uniwersytetu Śląskiego z wnioskiem o dopuszczenie mgr. inż. Piotra Grochowalskiego do dalszych etapów przewodu doktorskiego.

Jednocześnie z uwagi na wartość merytoryczną wyników udokumentowaną licznymi publikacjami, w tym kilkoma w czasopiśmie z listy filadelfijskiej wnioskuję o wyróżnienie tej rozprawy.

