

Sosnowiec, 30 sierpnia 2013 r.

Prof. dr hab. inż. Alicja Wakulicz – Deja
Instytut Informatyki
Wydział Informatyki i Nauki o Materiałach
Uniwersytet Śląski

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

Rozprawa doktorska mgr Piotra Grochowalskiego składa się z 7 rozdziałów, 4 załączników, bibliografii, spisu rysunków, spisu treści.

Rozdział I jest wstępem zawierającym wprowadzenie do rozprawy i omawiającym dotychczasowe prace o podobnej tematyce oraz główne wyniki rozprawy. Wstęp zawiera bardzo ogólne sformułowania dotyczące tematyki rozprawy, a sformułowania bardziej konkretne dotyczą zagadnień implementacji oraz projektu i realizacji systemu RSDS. Omówienie dotychczasowych prac dotyczących systemów informacyjnych i procesów wyszukiwania informacji zawiera wiele cytowań różnorodnych prac dotyczących tej tematyki powstałych w różnych latach, które zostały bardzo skrótowo omówione. Wydaje się, że bardziej interesujące byłoby szersze omówienie wybranych prac najbliższych tematyce rozprawy. W tym rozdziale zastrzeżenia budzą pewne sformułowania autora np.: „zostały uruchomione funkcjonalności”, używane są też pojęcia „słowa”, „termy”, „składniki zapytania” o których nie wiadomo, czym się różnią, czy też są synonimami.

Rozdział II zawiera definicje podstawowych pojęć takich jak: relacja binarna, pojęcia z zakresu teorii zbiorów przybliżonych, podstawowe definicje dotyczące grafów i struktur grafowych, definicje ciągów, metryk i miar odległości. Dodatkowo wprowadzono definicje pojęć z zakresu statystyki ograniczone do pojęć niezbędnych przy prowadzonych eksperymentach.

Rozdział III rozpoczyna właściwą merytorycznie część rozprawy doktorskiej. W rozdziale tym autor wprowadza pojęcie ontologii wykorzystywane w pracy, gdzie definiuje główne relacje takie jak: relacja powiązania, relacja zawierania, relacja generalizacji, relacja specyfikacji i relacja synonimi.

We wprowadzonej ontologii o typie relacji pomiędzy pojęciami decyduje ekspert. Wprowadzone zostało pojęcie grafu ontologii, który jest tworzony zgodnie z przyjętymi relacjami. Dodatkowo wprowadzone zostały znane z teorii zbiorów przybliżonych pojęcia dolnego i górnego przybliżenia zbioru pojęć, oraz pojęcie współczynnika aproksymacji. Określone zostało również pojęcie grafu współpracy dla danych bibliograficznych. Wszystkie badania skupiają się na konkretnym systemie bibliograficznym, dla którego tworzona jest ontologia (definiowane pojęcia i relacje) i budowana jest sieć semantyczna. Pojęcia i relacje wprowadzone są przez eksperta (autor). Weryfikacja powiązań przeprowadzana jest automatycznie (oprogramowanie autora).

W pierwszym kroku tworzona jest ontologia szczegółowa, znajduwane są relacje i powiązania między pojęciami. Następnie włącza się ontologię szczegółową do ogólnej tworząc nowe powiązania. Wyznaczany jest stopień dopasowania opisów bibliograficznych do pojęć (metoda zbiorów przybliżonych). Dopasowanie pojęć obliczane jest dla wszystkich typów relacji, a następnie dokonuje się agregacji dopasowań. Kolejnym krokiem jest tworzenie rankingu publikacji związanych z określonymi pojęciami. Rangowanie może w niektórych punktach

budzić zastrzeżenia, bo niekoniecznie np. późniejszy rok wydania świadczy o ważności publikacji.

Proces wyszukiwania w systemie „ontologicznym” polega na znalezieniu współczynnika dopasowania opisów bibliograficznych do pytań, a odpowiedź zawiera ranking publikacji stanowiących odpowiedź na pytanie. Dla ułatwienia procesów wyszukiwania tworzone są grupy publikacji. Ten sposób przeszukiwania jest bardzo zbliżony do metody Saltona, gdzie obliczany jest współczynnik korelacji pytania najpierw z reprezentantami grup, a potem z dokumentami w wybranych grupach. Odpowiedzią są dokumenty o współczynniku podobieństwa z pytaniem większym od założonego minimum, co odpowiada założonemu minimalnemu współczynnikowi dopasowania w przedstawionym podejściu.

W tym rozdziale brakuje odsyłaczy do literatury (publikacje autora).

Rozdział IV przedstawia opis systemu bibliograficznego RSDS (The Rough Set Database System), który w dużej mierze był realizowany przez autora rozprawy. System zawiera opisy bibliograficzne publikacji z zakresu teorii zbiorów przybliżonych i jej zastosowań, dane bibliograficzne autorów prac, dane kontaktowe. System zawiera ponad 4 000 opisów. W systemie można prowadzić wyszukiwanie alfabetyczne (tytuł, autor, wydawnictwo, itp.) lub wyszukiwanie zaawansowane (dla określonych kryteriów wyszukiwania). W wyszukiwaniu zaawansowanym następuje dopasowanie danych do wzorca metodą opartą na ontologii. Obliczane jest przybliżenie i współczynnik dopasowania publikacji do pytania kierowanego do systemu. Odpowiedzią są publikacje, dla których współczynnik dopasowania przekracza wartość progową. System dostarcza również szereg statystyk dotyczących wyników analiz danych zawartych w systemie.

Rozdział V omawia szczegółowo dodatkowe analizy, które są realizowane w systemie np.: tempo rozwoju systemu, statystyki dotyczące użytkowania systemu, autorów prac, liczby prac zbiorowych itp. System pozwala na przedstawienie grafu współpracy, co pozwala na prezentację badaczy i zespołów pracujących w zakresie teorii zbiorów przybliżonych, współpracy między nimi, znalezienie liderów grup, liczby publikacji w różnych okresach czasu itp.

Kolejny rozdział VI omawia przeprowadzone przez autora eksperymenty.

Eksperyment I dotyczy weryfikacji poprawności formalnej ontologii ogólnej podanej przez eksperta (zwrotność, symetryczność, przechodniość). Sprawdzane były relacje zdefiniowane w pracy. W przypadku niespełnienia wymogów formalnych relacja była modyfikowana. Eksperyment II dotyczy weryfikacji pracy systemu przez jego użytkowników. Eksperyment dotyczący oceny procesów wyszukiwania polegał na kierowaniu przez użytkowników określonego typu zapytań do systemu i oceny uzyskanych wyników.

Eksperyment III polega na porównaniu podejścia ontologicznego z modelem wektorowym przestrzeni dokumentów i zapytań (VSM). Trochę dziwi wybór metody wektorowej do porównań, bo ta metoda jest daleka od podejścia ontologicznego. Eksperyment polega na porównaniu wyników wyszukiwania z zastosowaniem modelu wektorowego i cosinusowej miary podobieństwa z metodą ontologiczną podaną przez autora. W tym eksperymencie dla metody ontologicznej wykorzystano różne parametry dotyczące rodzaju ontologii, ustalonych pojęć, wielkości i rodzaju współczynników aproksymacji, operatorów agregacji oraz generowanych wag (ranking).

Eksperyment I pozwolił na uzupełnienie brakujących powiązań, co prowadzi do znacznego zwiększenia liczby informacji w systemie.

Eksperyment II potwierdził, że wyniki wyszukiwania prowadzone w systemie są akceptowalne przez użytkowników.

Eksperyment III wykazał, że metoda oparta na ontologii daje znacznie szersze odpowiedzi (wykaz publikacji) niż metoda wektorowa. Szczegółowe wyniki eksperymentów autor przedstawił na wielu wykresach, gdzie podano również zakresy rozbieżności w ocenie

wyników w zależności od współczynnika korelacji oraz od przyjętej miary odległości. Porównanie metody wektorowej z ontologiczną wykazało wiele rozbieżności. Autor je omówił, nie stara się jednak analizować przyczyn uzyskanych rozbieżności (np. różnice w metodzie obliczania podobieństwa, różnice w rangowaniu itd.). Już prosta analiza podejścia ontologicznego pozwala zauważyć, że model wektorowy nie jest właściwy dla porównania z proponowanym podejściem. Brakuje tu jakiegokolwiek semantyki. Znacznie lepsze byłoby porównanie tego podejścia z podejściem opartym na grupowaniu dokumentów i obliczaniu współczynników podobieństwa z pytaniem. Eksperymenty są opisane przez autora niekiedy zbyt dokładnie i wiele elementów opisów się powtarza. Niektóre opisy są zbędne np.: analiza miary cosinusowej.

Rozdział VII zawiera podsumowanie i wnioski końcowe. Podsumowanie jest bardzo lakoniczne. Autor znacznie lepiej wypunktował wyniki swojej pracy we wstępie.

Rozprawa zawiera bogatą bibliografię dobrze powiązaną z tematyką pracy, chociaż część prac nie jest cytowana w rozprawie.

Reasumując, rozprawa doktorska mgr Piotra Grochowalskiego jest niewątpliwie wartościową rozprawą dotyczącą systemów informatycznych opartych na ontologii. Istotnym wkładem autora jest zaproponowanie metody tworzenia ontologii dla danych bibliograficznych, wprowadzenie własnych sensownych relacji i tworzenia grafu powiązań. Dodatkowym atutem jest wprowadzeniem pojęć z zakresu teorii zbiorów przybliżonych, dla przybliżenia zbioru pojęć nieostrych. Wszystkie badania autor przeprowadza na danych rzeczywistych dotyczących publikacji z zakresu teorii zbiorów przybliżonych. Autor jest jednym z głównych projektantów systemu RSDS, w którym opisane w pracy podejście do tworzenia systemu informacyjnego i procesu wyszukiwania w takim systemie zostało zaimplementowane. Ta część rozprawy wymagała ogromnego nakładu pracy i jest niewątpliwym atutem autora.

Szereg uwag krytycznych umieściłam przy omawianiu rozdziałów. Najważniejsze z nich dotyczą pewnych zawężeń w opisie wprowadzonej metodologii. Wprowadzone definicje są słuszne nie tylko dla konkretnych, podanych danych bibliograficznych, a nawet nie tylko dla systemów bibliograficznych. Można by tu szeroko nawiązać do systemów dokumentacyjnych. Autor zbyt często skupia się na konkretnym systemie. Zastrzeżenia budzi też „prześlizgnięcie” się po wprowadzonych pojęciach (rozdział II) i cytowanych pracach (rozdział I). Dodatkowo, uważam za niewłaściwy wybór metody wektorowej do porównania z podejściem ontologicznym, co zresztą potwierdziły uzyskane wyniki porównania. Podsumowanie i wnioski końcowe są lakoniczne i niedopracowane.

Przedstawione uwagi krytyczne nie zmieniają faktu, że rozprawa doktorska mgr Piotra Grochowalskiego prezentuje dobry poziom merytoryczny. Uzyskane wyniki stanowią interesujący przyczynek do badań związanych z systemami informacyjnymi (systemami wyszukiwania informacji). Rozprawa spełnia wymagania stawiane rozprawom doktorskim. Rekomenduję dopuszczenie mgr Piotra Grochowalskiego do dalszych etapów przewodu doktorskiego.

