

29 listopada 2015 r.

dr hab. inż. Mikołaj Morzy

Instytut Informatyki Politechniki Poznańskiej
ul. Piotrowo 2, 60-965 Poznań

KWESTIONARIUSZ – RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: **Reputation Systems on The Web**

Autorka rozprawy: **Paulina Adamska**

- 1. Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?**

Recenzowana rozprawa dotyczy konstrukcji systemów reputacyjnych dla szerokiej gamy usług oferowanych w internecie. W szczególności Autorka zaprezentowała nowy model reputacji dla oceny jakości i wiarygodności zasobów internetowych. Zaprezentowany model ma charakter ogólny i może być z powodzeniem zaaplikowany do wielu różnych klas zasobów internetowych. Wprowadzony model został przestudiowany teoretycznie oraz eksperymentalnie na drodze symulacji.

Konstrukcja systemów reputacyjnych ma bardzo duże znaczenie dla jakości wielu usług oferowanych w internecie. Ze względu na zagrożenia pojawiające się w sieci internet (w szczególności związane z licznymi bodźcami finansowymi zachęcającymi do nieuczciwych zachowań) trudno jest przecenić praktyczne znaczenie systemów reputacyjnych dla ochrony interesów użytkowników. Systemy reputacyjne od lat znajdują się w centrum zainteresowania osób zajmujących się naukowo usługami internetowymi i są postrzegane jako niezbędny składnik nowoczesnych platform usługowych. Badania nad systemami reputacyjnymi są aktualnie prowadzone aktywnie w licznych ośrodkach badawczych na całym świecie i recenzowana rozprawa doktorska bardzo dobrze wpisuje się w ten trend. W mojej ocenie zagadnienia badawcze, które są rozważane w rozprawie, stanowią istotne wyzwanie naukowe, a aktualnie dostępne rozwiązania są dalekie od satysfakcjonujących.

Rozprawa doktorska Pauliny Adamskiej liczy 102 strony i jest napisana w języku angielskim. Na rozprawę składa się osiem rozdziałów, spis literatury liczący 113 pozycji, oraz jeden dwustronicowy dodatek. Praca nie została zaopatrzona w streszczenie w języku polskim. W rozdziale 1 przedstawiono krótkie wprowadzenie do problemu. Rozdział 2 zawiera podstawowe definicje związane z obszarem zarządzania zaufaniem. W rozdziale 3 zamieszczono przegląd i klasyfikację systemów reputacyjnych wykorzystywanych współcześnie. Rozdział 4 poświęcony jest przeglądowi możliwych sposobów ataku na system reputacyjny, przedstawiono w nim charakterystykę popularnych metod obchodzenia systemów reputacyjnych i związanych z nimi niebezpieczeństw. W rozdziale 5 zaprezentowano sposób wzmocnienia systemu reputacyjnego dla konkretnej klasy usług (w tym przypadku serwisów aukcyjnych) za pomocą tzw. dowodów. Główny oryginalny wkład autorki przedstawiono w rozdziałach 6 i 7, w których wprowadzono nowy model systemu reputacyjnego służącego do oceny jakości dowolnych zasobów internetowych (w przeciwieństwie do wcześniejszych propozycji, które skupiały się na obliczaniu reputacji dla uczestników usług internetowych, ze szczególnym naciskiem na obszar handlu internetowego). Rozdział 6 zawiera propozycję nowego modelu reputacyjnego dla oceny jakości zasobów internetowych, a rozdział 7 przedstawia wykorzystanie tego modelu do oceny wiarygodności i jakości zasobów internetowych. Rozdział 8 stanowi podsumowanie rozprawy. W dodatku A zamieszczono także tabele zawierające syntetyczne podsumowanie przeprowadzonych badań symulacyjnych.

Autorka nie sformułowała jednoznacznie tezy rozprawy, podobnie jak nie zostały w sposób jawny podane cele rozprawy. Biorąc pod uwagę ogólną strukturę rozprawy można się domyślić, że celem Autorki było zaprojektowanie uniwersalnego systemu reputacyjnego, którego efektem byłoby promowanie zasobów internetowych charakteryzujących się wysoką jakością i wiarygodnością, przy jednoczesnej ochronie systemu przed próbami promowania zasobów niewiarygodnych i o niskiej jakości (ale np. charakteryzujących się atrakcyjną formą prezentacji). Szkoda, że cele te nie zostały jasno sformułowane w rozdziale 1. Recenzowana rozprawa ma zarówno charakter teoretyczny – poprzez konstrukcję modelu systemu reputacyjnego, jak i konstrukcyjny – poprzez przeprowadzenie badań symulacyjnych sugerujących możliwy wpływ implementacji proponowanego systemu reputacyjnego w rzeczywistym środowisku. W kontekście aktualnego stanu badań nad systemami reputacyjnymi i oceną jakości zasobów internetowych należy stwierdzić, że badania przeprowadzone przez Autorkę i zaprezentowane w rozprawie dobrze wpisują się w bieżące kierunki badań w dziedzinie systemów reputacyjnych oraz stanowią ważny cel naukowy spełniający wymagania stawiane rozprawom doktorskim.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle / świadczącej o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Autorka nie zdecydowała się na zamieszczenie analizy wcześniejszych prac w dyscyplinie w formie jednego spójnego rozdziału, lecz umieściła odnośniki do wcześniejszych badań we wszystkich rozdziałach rozprawy. Taki wybór wydaje się być podyktowany dużą różnorodnością odnośników literaturowych zamieszczonych w bibliografii. Większość odnośników literaturowych pojawia się w rozdziałach 2, 3 i 4, w których znajdują się opisy wcześniej zaproponowanych rozwiązań. Opisy te uznają za wystarczające do właściwego umieszczenia recenzowanej pracy w kontekście aktualnego stanu wiedzy.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Wyniki przedstawione w rozprawie (mam na myśli zawartość rozdziałów 6 i 7) zostały eksperymentalnie zweryfikowane za pomocą symulacji. Proponowany system reputacji dla oceny jakości zasobów internetowych został poddany symulacji w postaci iterowanej gry z tabelą wypłat. System reputacji dla oceny wiarygodności zasobów internetowych został zasymulowany przy pomocy systemu wieloagentowego. W obu przypadkach nie przeprowadzono próby ewaluacji systemu na rzeczywistych zbiorach danych. Najprawdopodobniej związane jest to z trudnością uzyskania dostępu do rzeczywistego systemu, w którym można by zaimplementować proponowane rozwiązania. Niemniej jednak przy ocenie poprawności przeprowadzonych eksperymentów należy mieć na uwadze fakt, że nie udało się oszacować, jak sensowne są przyjęte w symulacji założenia (choćby tabela wypłat). W ogólności metoda rozwiązania problemu eksperymentalnej ewaluacji zaproponowanych modeli jest poprawna i nie budzi większych zastrzeżeń (pewne drobne uwagi krytyczne dotyczące eksperymentu sformułowałem w dalszej części recenzji). Niezależnie jednak od drobnych wątpliwości należy stwierdzić, że w rozprawie przedstawiono dwa oryginalne modele systemów reputacyjnych dla oceny jakości i wiarygodności zasobów internetowych. Przedstawione modele mają duże znaczenie praktyczne.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Oryginalny dorobek Autorki został zaprezentowany w rozdziałach 5, 6 i 7. Polega on na opracowaniu metody ulepszenia klasycznych systemów reputacyjnych wykorzystywanych w środowiskach aukcyjnych o wielokryterialną ocenę reputacji, ocenie wpływu systemu reputacyjnego na jakość tworzonych zasobów internetowych, oraz na zaprojektowaniu systemu reputacyjnego dla oceny wiarygodności zasobów internetowych. W rozdziale 5 Autorka definiuje pojęcie *dowodu* (ang. proof) oraz przedstawia operatory do agregacji i selekcji dowodów, a następnie proponuje funkcję do skalowania wagi poszczególnych kryteriów dowodów. Użyteczność proponowanej metody jest przedstawiona za pomocą prostego wywodu analitycznego i jednego przykładu. W rozdziale 6 znaleźć można nowy autorski model wykorzystania reputacji do oceny jakości zasobów internetowych, i tę właśnie część rozprawy uważam za zdecydowanie najbardziej wartościową pod kątem naukowym. Autorka konstruuje teoriogrowy schemat iterowanej interakcji między producentami i konsumentami zasobów internetowych, wprowadzając szerokie spektrum możliwych strategii: począwszy od produkcji zasobów wysokiej jakości i prezentacji, poprzez zasoby wysokiej jakości jednak o uboższej prezentacji, aż po produkcję zasobów o niskiej jakości i niskiej prezentacji. Następnie analizuje grę polegającą na konsumpcji zasobów przez konsumentów przy założeniu, że produkcja wysokiej jakości zasobów jest droższa niż prezentacja zasobów, zatem producenci w sposób naturalny mają tendencję do produkowania zasobów o niskiej jakości, za to atrakcyjnie prezentowanych. Seria eksperymentów dość jednoznacznie pokazuje, w jaki sposób włączanie mechanizmów reputacyjnych chroni konsumentów przed konsumowaniem zasobów niskiej jakości i wymusza na producentach inwestowanie w jakość zasobów. Eksperyment jest starannie przeprowadzony i opisany, wnioski wynikające z eksperymentu są opatrzone komentarzem i odniesione do możliwych rzeczywistych scenariuszy. Rozdział 7 zawiera model systemu reputacyjnego dla oceny wiarygodności zasobów internetowych. Model ten został przez Autorkę poddany prostej ewaluacji eksperymentalnej za pomocą jednej symulacji. Zaproponowane w rozprawie rozwiązania mają charakter oryginalny i nowatorski, zgodnie z moją wiedzą nie były wcześniej prezentowane w literaturze naukowej ani implementowane w innych systemach reputacyjnych.

5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników /zwięzłość, jasność, poprawność redakcyjna rozprawy/ ?

Recenzowana rozprawa jest skonstruowana poprawnie. Można mieć wręcz wrażenie, że niektóre elementy zostały potraktowane przez Autorkę zbyt skrótowo i po lekturze pozostaje pewien niedosyt. Przykładowo, symulacja przedstawiona w rozdziale 7 jest relatywnie prosta i odnosi wrażenie, że wprowadzony model systemu reputacyjnego dla oceny wiarygodności zasobów internetowych mógłby być przetestowany w sposób zdecydowanie bardziej szczegółowy. Tekst pracy jest napisany staranną angielszczyzną, znalezione (nieliczne) błędy składniowe lub gramatyczne zamieściłem w postaci osobnego dokumentu. Jakość użytych w rozprawie rycin jest bardzo różna, część z nich jest zdecydowanie zbyt mała (np. Ryciny 7.4 lub 7.6). Użyte w rozprawie oznaczenia matematyczne nie są, w mojej opinii, dobrane zbyt szczęśliwie (więcej na ten temat w uwagach krytycznych). Poważnym niedociągnięciem rozprawy jest brak streszczenia w języku polskim (wymaganie sformułowane w Artykule 13 pkt. 6 Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki¹).

¹ Dz.U.2003 Nr 65 poz. 595 (tekst ujednolicony)

6. Jakie są słabe strony rozprawy i jej główne wady?

Poniżej zamieszczam uwagi dotyczące merytorycznej strony pracy. Dotyczą one fragmentów, które są niejasne, a także fragmentów, które, w mojej opinii, zawierają błędy.

- s. 41: Przyjęte oznaczenie funkcji (μ) jest nieszczęśliwe, oznaczenie to jest powszechnie utożsamiane z wartością średnią zmiennej losowej. We wzorze brak indeksu dolnego i przy odchyleniu od średniej oceny (δ). Zupełnie nie jest jasne, dlaczego δ_i^{\max} jest połową różnicy między oceną maksymalną i minimalną (w praktyce nie będzie to miało znaczenia, bo przy dużej liczbie ocen jest w zasadzie pewne, że s_i^{\max} i s_i^{\min} przyjmą wartości odpowiednio największe i najmniejsze na skali. Formuła zaproponowana dla funkcji μ jest całkowicie arbitralna i podana bez żadnego wyjaśnienia.
- s. 41: Oznaczenie przyjęte dla funkcji skalującej (σ) jest równie nietrafne, ponieważ jednoznacznie kojarzy się z odchyleniem standardowym. Podana na stronie 42 definicja tej funkcji jest błędna, ponieważ posiada zaszytą w sobie konkretną skalę ocen.
- s. 42: W definicji następstwa etykiet jeden z indeksów przy funkcji σ powinien być j
- s. 43: Rycina 5.3 została użyta bez wskazania na źródło pochodzenia (doktorat T. Kaszuby)
- s. 46: Wynik opisany w ostatnich paragrafach wydaje się być oczywisty, równanie 5.1 jest w istocie schematem ważenia kryteriów. Wstawiając proponowane wartości parametrów α i β z Tablicy 5.1 do równania 5.1 otrzymujemy kolejne wagi różniące się o 0.5 między kolejnymi etykietami.
- s. 63: Nie rozumiem ustalenia prawdopodobieństwa porzucenia skompromitowanej tożsamości i przyjęcia nowej tożsamości na 80% po każdej iteracji symulacji. Czy takie działanie nie powinno być przeprowadzone dopiero wówczas, gdy reputacja związana z jakąś tożsamością spadnie poniżej określonego progu?
- s. 76: Zapis funkcji celu L (7.2) zupełnie niepotrzebnie miesza dwie notacje. Z jednej strony mamy operatory sumowania sumujące po zbiorze użytkowników i obiektów, z drugiej strony wewnątrz sumowania mamy iloczyny macierzowe, przy czym mnożeniu podlegają wektory (fragmenty macierzy indeksowane aktualnym indeksem pętli związanej z operatorem sumowania). Taki zapis nie tylko wprowadza bardzo duże zamieszanie, ale również nie pozwala na określenie, czy pochodne cząstkowe w równaniach 7.3 i 7.4 na stronie 77 są poprawne. Generalnie nie da się tego sprawdzić, ponieważ Autorka nie precyzuje, co kryje się za symbolem e_{uc} kluczowym dla sprawdzenia poprawności obu pochodnych (można przynajmniej sprawdzić, że zapis nie jest poprawny, bo w równaniu 7.2 nie występują zmienne U_u^p i V_c^p po których następuje różniczkowanie). Dokładnie te same uwagi dotyczą funkcji celu i jej pochodnych cząstkowych przedstawionych na stronie 79.
- s. 77: Według Autorki iloczyn macierzy zaprezentowany w równaniu 7.5 reprezentuje prawdopodobieństwo interakcji między użytkownikiem a zasobem. Nie rozumiem, w jaki sposób z oryginalnej macierzy reprezentującej oceny wiarygodności zasobów przez użytkowników przeszliśmy nagle do prawdopodobieństwa oceny zasobu. Zupełnie nie jest też dla mnie jasne zdanie w ostatnim paragrafie, rozpoczynające się od słów „For the target user u , a list of candidate [...]”. Dlaczego system rekomendacyjny miałby być wykorzystywany do tego celu? Gdyby chodziło o, jak pisze Autorka, znalezienie listy tych zasobów, które nie posiadają wielu ocen wiarygodności i nie były ocenione przez użytkownika, to rozwiązanie jest trywialne: posortuj zasoby według rosnącej liczby ocen i pokaż użytkownikowi pierwsze n zasobów których jeszcze nie ocenił.
- s. 85: Gdyby prawdopodobieństwo kopiowania fałszywej oceny faktycznie było losowane z $N(0,1,0,1)$ to w wielu przypadkach to prawdopodobieństwo byłoby negatywne...

- s. 85: Prezentowana na Rycinie 7.4 precyzja jest bardzo niska i nie przekracza 15%. W tekście rozprawy zabrakło mi odniesienia się do tego wyniku i wyjaśnienia, czym jest spowodowany.
- s. 87: Na Rycinie 7.6 z kolei precyzja gwałtownie rośnie i operuje w zakresie 60-99%. Czym podyktowana jest ta nagła zmiana?

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Recenzowana praca dotyczy problemu konstrukcji systemów reputacyjnych odpornych na ataki. W przeciwieństwie do większości prac prowadzonych w obszarze systemów reputacyjnych, rozprawa mgr Pauliny Adamskiej nie koncentruje się jedynie na systemach dedykowanych dla środowisk handlu elektronicznego (systemy aukcyjne, sklepy internetowe, platformy typu *crowd-funding*), ale proponuje oryginalne rozwiązania dla uniwersalnych zasobów i usług internetowych. W zamierzeniu proponowane rozwiązania mają służyć podnoszeniu jakości i wiarygodności zasobów, a także być odporne na typowe ataki przeprowadzane przez internetowych oszustów. Prezentowane w pracy modele zostały wstępnie przetestowane eksperymentalnie i wydaje się, że mogłyby w przyszłości znaleźć zastosowanie w przemyśle i handlu. Wnioski płynące z przeprowadzonych badań są obiecujące i mogą zaowocować dalszymi, wartościowymi badaniami naukowymi. Na pochwałę zasługuje także praktyczny aspekt badań prezentowanych w rozprawie. Zainteresowanie wykazywane przez przemysł w odniesieniu do systemów reputacyjnych pozwala mieć nadzieję, że przedstawione w rozprawie pomysły zostaną w przyszłości zaimplementowane w rzeczywistych środowiskach.

W podsumowaniu recenzji stwierdzam, że rozprawa doktorska mgr Pauliny Adamskiej spełnia wymagania stawiane rozprawom doktorskim przez obowiązującą Ustawę o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 r. (Dz.U. Nr 65 poz. 595 z późniejszymi zmianami) i wnoszę o jej dopuszczenie do publicznej obrony.

8. Do której z następujących kategorii Recenzent zalicza rozprawę:

- a) nie spełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy
- b) wymagająca wprowadzenia poprawek i ponownego recenzowania
- c) spełniająca wymagania
- d) spełniająca wymagania z wyraźnym nadmiarem
- e) wybitnie dobra, zasługująca na wyróżnienie

