

RECENZJA
rozprawy doktorskiej mgra Łukasza Rogowskiego
„Implementacja systemu wnioskującego za pomocą DNA”
Promotor: prof. dr hab. inż. Tadeusz Krasieński

1. Podstawowe informacje o rozprawie

Przedstawiona do recenzji rozprawa mgra Ł. Rogowskiego „Implementacja systemu wnioskującego za pomocą DNA” napisana pod kierunkiem prof. dr. hab. inż. Tadeusza Krasieńskiego jest poprawioną wersją dysertacji z roku 2016. Wraz z rozprawą Recenzent otrzymał 22 marca br. opracowane przez Kandydata do stopnia doktora „Odpowiedzi na zarzuty recenzentów”, sformułowane przez prof. Jana Mulawkę i Recenzenta do pierwszej wersji rozprawy. Dodatkowo, na prośbę Recenzenta, Kandydat przesłał odpowiedzi na szczegółowe uwagi sformułowane w pierwszej recenzji przez Recenzenta.

2. Tematyka rozprawy, przydatność rozprawy dla nauk matematycznych

Praca Leonarda M. Adlemana z 1994 roku pt. „Molecular computation of solutions to combinatorial problems”, w której autor na niespełna 4 stronach (owszem, 3-szpaltowych) opisał, w jaki sposób przy użyciu powszechnie znanych metod inżynierii genetycznej rozwiązał zadanie poszukiwania cyklu Hamiltona, ma dzisiaj prawie 6 tys. cytowań wg Google Scholar. Artykuł ten stał się w istocie początkiem nowej dziedziny nauki, leżącej na pograniczu kilku dziedzin, między innymi informatyki, matematyki i biologii, nazywanej obliczeniami DNA (ang. DNA computing). Gwoli ścisłości L.M. Adleman nie był pierwszym, który pokazał możliwości obliczeniowe tkwiące w takich strukturach biologicznych, jak ciągi DNA czy RNA, że wspomnieć tutaj pracę T. Heada z 1987 roku, ale dopiero jego udany (aczkolwiek żmudny i niepraktyczny w istocie) eksperyment przyciągnął zainteresowanie innych. Co prawda Kandydat podkreśla w rozprawie niezwykle możliwości obliczeń masowych i równoległych wykonywanych przez sekwencje biomolekularne, ale wydaje się, że współczesne zainteresowanie badaczy zajmujących się tą tematyką skupia się raczej na molekularnym poziomie obliczeń (patrz bioautomaty czy bioczuJNIKI).

Pisząc o obliczeniach molekularnych nie sposób nie wspomnieć o „komputerach” chemicznych, biologicznych, optycznych, czy dzisiaj szczególnie popularnych komputerach kwantowych. Wszystkie one są próbą złamania ograniczeń związanych z rozwojem komputerów „krzemowych” (pisanych w cudzysłowie, bo przecież dzisiaj stosuje się całą gamę pierwiastków w budowie procesorów) i ciężących na nich przekleństwie Moorea.

Recenzowana i poprawiona rozprawa podejmuje zagadnienie konstrukcji w pełni autonomicznego, molekularnego systemu wnioskującego za pomocą ciągów DNA, o nowych możliwościach, o których będzie więcej w dalszej części recenzji. Tematyka rozprawy wpisuje się zatem bardzo dobrze w aktualny nurt światowych badań w dziedzinie biomolekularnych obliczeń, a poprzez swoje wybitnie teoretyczne i sformalizowane rozważania może mieć dużą przydatność dla nauk matematycznych. Co więcej, ma szansę powiększyć wkład polskiej nauki, poza pracami wymienionymi w rozprawie, w rozwój tej nowoczesnej i dynamicznie rozwijającej się dziedziny.

W świetle powyższych uwag można stwierdzić, że **problematyka rozprawy jest ważna i w pełni uzasadniona, o dużej przydatności dla nauk matematycznych.**

3. Kompozycja, zawartość i redakcja rozprawy

Recenzowana rozprawa składa się z sześciu rozdziałów poprzedzonych nienumerowanym Wstępem oraz Celem i tezą pracy, a zakończona Podsumowaniem, Dodatkiem A, Bibliografią, wreszcie Spisem rycin i tabel. Rozprawa obejmuje 154 numerowanych stron. W porównaniu do pierwszej wersji z 2016 roku, liczba stron powiększyła się o 54 oraz pojawił się nowe rozdziały: Cel i teza pracy, którego ewidentnie brakowało w wersji pierwszej, rozdział 6 zatytułowany „Symulacja eksperymentów laboratoryjnych” oraz Spisy rycin i tabel. Bibliografia powiększyła się z 40 pozycji do 60, Kandydat uzupełnił – zgodnie z uwagami recenzentów - rozdział 3 o opis prac zespołu prof. J. Mulawki oraz wspomniął krótko o artykułach wykorzystujących technikę DSD (and. DNA strand displacement). W Dodatku A, zawierającym spis tzw. „lepkich końców” o długości 4, zaznaczono także pary komplementarnych sekwencji, które w proponowanym systemie wnioskowania nie można wykorzystywać ze względu na własności restryktazy.

W kilkustronicowym Wstępie Kandydat skrótowo porusza zagadnienia klasycznego rachunku zdań, obliczeń na sekwencjach DNA, wspomina eksperyment Adlemana, propozycje molekularnych obliczeń zespołów Shapiry, Recenzenta i prof. J. Mulawki, słusznie zauważając brakującej w tych rozwiązaniach implementacji negacji. Wspomniana jest też technika DSD

(tym razem Kandydat zauważa, że to podejście nie używa enzymów restrykcyjnych). Wstęp kończy wymienienie autorskich publikacji, schemat dysertacji i podziękowania.

W rozdziałach 1-3 Autor wprowadza podstawowe informacje na temat, odpowiednio wnioskowania formalnego, obliczeń DNA i znanych (Kandydatowi) implementacji systemów wnioskowania. Z uwag do tych rozdziałów, mających charakter informacji wprowadzających do właściwego rozwiązania, należałoby wymienić niepoprawiony w stosunku do pierwszej wersji dysertacji opis eksperymentu Adlemana z sekcji 2.3 (opis jest co najmniej niepełny), może drobiazg, ale wart zauważenia, że Yaakov Benenson był członkiem zespołu prof. Shapiry (a dokładnie jego doktorantem), a nie odwrotnie (patrz str. 27), dodany akapit z referencjami do technik DSD w sekcji 2.5 oraz – najważniejsze – uzupełnienie znanych implementacji systemów wnioskowania o prace zespołu prof. Mulawki. In plus zamieszczonych opisów modeli jest ich autorska i krytyczna analiza oraz porównanie możliwości systemów wnioskujących opracowanych przez zespoły Unolda i Shapiry. Niestety, Autor opisując dzisiaj bardzo szerokie zagadnienie obliczeń DNA, nie uwzględnił uwagi Recenzenta sformułowanej w recenzji do pierwszej wersji dysertacji, że podany opis nie wyczerpuje zagadnienia. Rozdział 2 mógłby, a nawet powinien być uzupełniony o odwołania do najnowszej literatury, w której poza wspomnianymi w sekcji 2.5 modelami opartymi na samoskładaniu, membranach czy obliczeniach bez użycia restryktaz, można spotkać modele stosujące (mi)RNA, białka, nanopory, nie mówiąc już o coraz szerszym wachlarzu zastosowań (diagnostyka, biosensory czy bioczuJNIKI).

Istotne ze względu na ocenę dysertacji są trzy rozdziały 4-6, w których Autor przedstawia koncepcję własnego systemu wnioskującego, jego analizę i wreszcie (dodaną w stosunku do pierwszej wersji) symulację eksperymentów biologicznych. Rozdziały 4 i 5 nie różnią się praktycznie od ich wersji z 2016 roku, rozdział 6 jest próbą wyjścia naprzeciw oczekiwaniom recenzentów, a dotyczących laboratoryjnej implementacji zaproponowanego modelu biomolekularnego.

W rozdziale 4 zaprezentowany jest autorski model systemu wnioskującego na ciągach DNA. Proponowany system indukcji oparty jest o splatanie, enzymy ligazy i restryktazy BbvI, ściśle określony molekularny zapis aksjomatów początkowych (literałów, klauzul), zawsze w koniunkcyjnej postaci normalnej, wreszcie molekułę terminalną i mechanizm wnioskowania *modus ponens* (regułę odrywania). Kandydat nigdzie nie podkreśla faktu, że w stosunku do pierwszej wersji rozprawy zastąpił restryktazę BseXI jej izoschizomerą BbvI, czyli enzymą restrykcyjną rozpoznającą tę samą sekwencję DNA, ale wymagającą innych warunków reakcji chemicznej. Warto by się odnieść do tej zmiany i ją uzasadnić, prawdopodobnie

przewidywanymi wymogami doświadczalnymi. Zapytania mogą dotyczyć klauzuli faktu zawartej w zbiorze aksjomatów początkowych lub możliwej do wywiedzenia, koniunkcji bądź alternatywy kilku faktów, zdania zapisanego w dysjunkcyjnej postaci normalnej, a nawet kilku zdań jednocześnie. Prawdopodobnie największą zaletą modelu jest możliwość zastosowania negacji, której brakowało w omawianych wcześniej systemach wnioskowania. Dzięki tej innowacji system wnioskowania zdolny jest do wykrycia sprzeczności, tak w zbiorze aksjomatów początkowych, jak i jako wynik pośrednich operacji odrywania. W krótkim podrozdziale opisana jest ogólna zasada działania systemu indukcji i szkoda, że Kandydat nie poprawił tego rozdziału wprowadzając – za sugestią Recenzenta - ogólnego schematu reprezentacji molekuł wprowadzonego dopiero w sekcji 5.3.2. Dalej przedstawione zostały podstawowe kroki działania autonomicznego działania modelu i rozpoznawania odpowiedzi, a w kolejnej sekcji podano dwa przykłady „złożonego” wnioskowania (cudzysłów sugeruje, że ta złożoność jest względna). Rozdział 4 kończy opis, w jaki sposób model realizuje wybrane prawa klasycznego rachunku zdań, czyli prawo podwójnego przeczenia, regułę odrywania, kontrapozycji, przechodniości implikacji i prawa de Morgana.

W rozdziale 5 Autor dokonuje szczegółowej analizy własnego systemu wnioskowania. W pierwszej kolejności opisuje ograniczenia systemu, które wynikają z przyjętej reprezentacji molekuł oraz zastosowanej restryktazy. W tym miejscu Recenzent powtórzy swoje wątpliwości co do podanej liczby 66 zmiennych (w pierwszej wersji była to liczba 65), gdyż najprawdopodobniej możliwości laboratoryjne istotnie tę liczbę by ograniczyły. Dalej, Autor zauważa, że można równocześnie zakodować 16 różnych zapytań (w jednej próbówce), rozważyć oczywiste w każdym systemie wnioskującym ograniczenia dla przypadku, gdy pytamy jednocześnie o fakt i jego negację (również w sytuacji pytań bardziej złożonych). Analizowane są również ograniczenia dotyczące zapytań o klauzule złożone oraz wnioskowania faktu ze zbioru alternatyw. W rozdziale znajdujemy również wskazanie możliwości wartościowania funkcji logicznych, w tym również, gdy tych funkcji jest więcej. Naturalną konsekwencją możliwości obliczania wartości logicznych dla zadanych argumentów jest wykorzystanie jej w tzw. częściowym rozwiązaniu problemu SAT. Rozdział 5 kończy opis możliwych rozszerzeń modelu wnioskowania, w tym przeprowadzania dowodu nie wprost. Zaslужującą na uwagę jest analiza możliwości zastosowania innych enzymów restrykcyjnych we wnioskowaniu i wreszcie propozycja zastąpienia elektroforezy na żelu fluorescencją, wykorzystywaną w innych systemach wnioskujących.

Rozdział 6 mógłby być opisem rzeczywistej implementacji opisanego wcześniej i przeanalizowanego molekularnego modelu wnioskowania, czego zresztą oczekiwali recenzenci

od pierwszej wersji dysertacji, jest natomiast jedynie ręczną, aczkolwiek bardzo staranną symulacją możliwej laboratoryjnej implementacji. Zgodnie z treścią rozdziału, zawarte w rozdziale techniczne aspekty obliczeń zostały skonsultowane ze specjalistami z Katedry Genetyki Molekularnej Wydziału Biologii i Ochrony Środowiska Uniwersytetu Łódzkiego. Według Autora brak eksperymentów laboratoryjnych spowodowany został niemożnością pozyskania odpowiednich środków finansowych. Recenzentowi trudno się odnieść do zaprezentowanych symulacji, choć zdają się być jak najbardziej poprawne. Trzeba mieć jednak świadomość, że nawet najbardziej dokładne symulacje nie zastąpią rzeczywistego eksperymentu laboratoryjnego, zdolnego potwierdzić wszystkie założenia nowego modelu wnioskowania. Rozdział podsumowuje zestawienie kroków niezbędnych do przeprowadzenia doświadczenia.

Rozprawę zamyka Podsumowanie podkreślające m.in. przewagę modelu Kandydata nad systemami zaproponowanymi przez zespół Recenzenta i Shapiry, a także podkreślające zastosowane innowacje dotyczące użytej reprezentacji molekularnej.

Kompozycja i struktura rozprawy nie budzą większych zastrzeżeń, podobnie zresztą jak jej pierwsza wersja z 2016 roku. Pozostają wątpliwości co do tytułu dysertacji, zapowiadającego implementację systemu wnioskującego, a po poprawkach opisującego jedynie jego precyzyjną symulację.

Język rozprawy jest poprawny, w pracy Recenzent znalazł nieliczne błędy literowe i składniowe, razi natomiast duża liczba błędów interpunkcyjnych, zaznaczonych w otrzymanym do recenzji egzemplarzu.

Bibliografia w stosunku do poprzedniej wersji rozprawy została poszerzona o 20 pozycji, ale ciągle daleko jej do chociażby szkicowego przeglądu najważniejszych (i najnowszych!) prac z dziedziny obliczeń biomolekularnych.

Kandydat w przeciągu prawie 5 lat, jakie minęły pomiędzy przedstawieniem obydwu wersji dysertacji, nie poszerzył swojego dorobku publikacyjnego, na który składa się 1 preprint, 1 praca konferencyjna, 2 artykuły zamieszczone w polskich czasopismach oraz 1 opublikowany w czasopiśmie zagranicznym (z zagranicznym współautorem).

4. Oryginalne osiągnięcia Autora

Zgodnie ze sformułowaną na str. 12 tezą rozprawy, możliwy jest projekt rozszerzonego, w pełni autonomicznego systemu wnioskującego za pomocą ciągów DNA, działającego w oparciu o mechanizm splatania i spójnik negacji. Udowodnienie tej tezy wymagało od jej Autora wykonania nietrywialnych badań teoretycznych oraz przeprowadzenia symulacji.

Do najważniejszych i oryginalnych wyników przedstawionych w recenzowanej rozprawie należy zaliczyć:

- Opracowanie koncepcji w pełni autonomicznego molekularnego systemu wnioskującego, istotnie rozszerzającego możliwości analogicznych systemów opisywanych w literaturze. Rozszerzenie wprowadzone jest zasadniczo przez spójnik negacji i związany z nim cały zestaw praw rachunku zdań. Dodatkowo, proponowany system m.in. umożliwia zadawanie pytań o fakt, alternatywę lub koniunkcję faktów, dowolne zdanie logiczne zapisane w formie dysjunkcyjnej oraz zbiór zdań. Zdolność systemu do wykrywania sprzeczności prowadzi do kolejnego, poza regułą odrywania, mechanizmu wnioskowania, czyli dowodu nie wprost.
- Zastosowanie opracowanego systemu wnioskowania do zadania wartościowania funkcji logicznych i częściowego rozwiązywania problemu k -SAT.
- Opracowanie reguł konstrukcji molekuł dla potencjalnych, nowych restryktaz.
- Wykazanie możliwości zastąpienia elektroforezy na żelu zjawiskiem fluorescencji.
- Przeprowadzenie symulacji eksperymentów laboratoryjnych.

Dodatkowo, w rozprawie uzyskano następujące oryginalne rezultaty badawcze, jednak o mniejszej randze i funkcji pomocniczej w stosunku do dowodzonej tezy:

- Krytyczna analiza literaturowych systemów wnioskujących, stanowiących podstawę wprowadzonego systemu wnioskowania.
- Porównanie krytyczne modeli wnioskujących zespołów Recenzenta i Shapiry.
- Zastosowanie cząstek symetrycznych, umożliwiających różne przekształcenie tej samej klauzuli.

Zgodnie z powyższym można sformułować wniosek, że uzyskane wyniki teoretyczne wraz z symulacją potwierdzają tezę rozprawy.

5. Uwagi krytyczne i dyskusyjne

Jak już wspomniano wcześniej, Kandydat do stopnia doktora odniósł się do większości uwag sformułowanych przez Recenzenta do pierwszej wersji rozprawy. Uwagi natury szczegółowej zostały w większości uwzględnione w poprawionej wersji, natomiast uwagi natury ogólnej już niestety nie wszystkie (wymienić tu można uwagi co do tytułu rozprawy, badań

eksperymentalnych, zawartości Bibliografii czy symulatora obliczeń). Podsumowując, lektura rozprawy skłania do sformułowania poniższych uwag natury ogólniejszej:

- Zasadniczą uwagą krytyczną, sformułowaną już do pierwszej wersji rozprawy, jest brak eksperymentalnego potwierdzenia opisanej koncepcji. Owszem, należy zauważyć staranną i precyzyjnie przedstawioną symulację wybranych doświadczeń laboratoryjnych, jednak nie jest ona w stanie zastąpić zdjęcia z wyniku elektroforezy na żelu. Trzeba jednak obiektywnie dodać, że wzmiankowany w rozprawie bardzo dobry, poparty solidnymi publikacjami doktorat Sainz de Muriety z 2013 roku „Inference models in DNA computing”, w którym wprowadzono różne modele indukcji biomolekularnych, również nie zawierał wyników badań eksperymentalnych (!).
- Kolejna uwaga, również już formułowana w poprzedniej recenzji, dotyczy zbyt szerokiego tytułu rozprawy, obiecującego więcej niż tylko koncepcję czy projekt.
- Przegląd literatury z dziedziny jest zdecydowanie za wąski.
- Recenzent podtrzymuje zasadność budowy symulatora obliczeń biomolekularnych.
- Warto by było rozważyć podczas kodowania zapytania w dysjunkcyjnej postaci normalnej stosowanie różnej długości łańcucha dowolnych nukleotydów dla składników alternatywy. Dałoby to szanse na precyzyjną identyfikację odpowiedzi systemu (patrz rys. 95 i inne).
- Schemat kodowania z rys. 145 (str. 99) powinien zostać wykorzystany już w rozdziale 4.

Poniżej wymieniono uwagi natury szczegółowej:

- Błąd edycyjny w odwołaniu do Wstępu (a nie wstępu) na str. 23.
- Brak koniecznej precyzji w opisie eksperymentu Adlemana na str. 24, gdyż nie wystarczy sprawdzenie długości ścieżki, w celu wykazania, że wszystkie kodowane wierzchołki na ścieżce się nie powtarzają.
- Błąd kodowania faktu p na rys. 32b. Jest GTGT, powinno być CGGC. Podobnie na rys. 32c.
- Błąd kodowania zapytania q? na rys. 34. Jest GGGC, powinno być CGGC.
- Literówka na str. 37. Jest „dysjunkcyjnej”, powinno być „dysjunkcyjnej”.
- Błąd edycyjny w legendzie do rys. 51 (a i b). Jest zapytanie, powinien być fakt.

- Nieprecyzyjny zapis czwórki komplementarnych nukleotydów na rys. 90. Jest $\frac{NNNN}{NNNN}$, powinno być $\frac{NNNN}{NNNN}$.
- Niezrozumiała i zaciemniająca obraz zmiana długości dowolnych nukleotydów w kodowaniu zapytań na rys. 94 i 95.
- Błąd edycyjny w legendzie do rys. 103a. Jest u, powinno być c.
- Błąd składni w drugim zdaniu sekcji 5.1.5 na str. 83.
- Nie pokazano alternatywnego rozwiązania w częściowym rozwiązaniu problemu SAT na str. 89.

Wszystkie wymienione wyżej krytyczne uwagi i spostrzeżenia nie podważają pozytywnej oceny pracy.

5. Wniosek końcowy

Zawarte w recenzji uwagi mają na ogół charakter dyskusyjny, a te krytyczne nie wpływają na pozytywną ocenę recenzowanej dysertacji. Uważam, że poprawiona rozprawa mgra Łukasza Rogowskiego spełnia wymogi formalne i te zwyczajowe stawiane rozprawom doktorskim. Recenzowana dysertacja niewątpliwie stanowi oryginalne rozwiązanie postawionego problemu naukowego, wprowadzając bardzo obiecującą koncepcję molekularnego systemu wnioskowania. Kandydat wykazał się właściwą wiedzą w dziedzinie nauk matematycznych, w dyscyplinie naukowej Informatyka, dowiódł też opanowania umiejętności samodzielnego prowadzenia pracy naukowej.

Wnioskuje o dopuszczenie mgra Łukasza Rogowskiego do dalszych etapów postępowania o nadanie stopnia doktora.

Olgierd Umda