



РЕПУБЛИКА СРПСКА  
УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ  
ПРИРОДНО-МАТЕМАТИЧКИ ФАКУЛТЕТ  
Број: 19/4.963-1/23  
Датум: 13.11.2023 198  
БАЊАЛУКА

## ИЗВЈЕШТАЈ

о оцјени урађене докторске дисертације

### 1. ПОДАЦИ О КОМИСИЈИ

Орган који је именовано комисију: Сенат Универзитета у Бањој Луци, на основу приједлога одлуке Наставно-Научног вијећа Природно-математичког факултета (Број 19/3.2121/23, дана 06.09.2023.)

Датум именовања комисије: 21.09.2023.

Број одлуке: 02/04-3.2036-91/23

Чланови комисије:

- |   |                  |                                 |
|---|------------------|---------------------------------|
| 1. Др Душко Богданић                                    | редовни професор | Алгебра и геометрија            |
| Презиме и име   | Звање            | Научно поље и ужа научна област |
| Природно-математички факултет Универзитета у Бањој Луци |                  | предсједник                     |
| Установа у којој је запослен-а                          |                  | Функција у комисији             |
| 2. Др Бојан Башић                                       | редовни професор | Дискретна математика            |
| Презиме и име   | Звање            | Научно поље и ужа научна област |
| Природно-математички факултет Универзитета у Новом Саду |                  | члан                            |
| Установа у којој је запослен-а                          |                  | Функција у комисији             |
| 3. Академик Зоран Митровић                              | редовни професор | Математичка анализа и примјене  |
| Презиме и име   | Звање            | Научно поље и ужа научна област |
| Електротехнички факултет, Универзитета у Бањој Луци     |                  | члан                            |
| Установа у којој је запослен-а                          |                  | Функција у комисији             |

4. Др Марко Ђукановић	доцент	Информационе науке и биоинформатика (развој софтвера)
Презиме и име	Звање	Научно поље и ужа научна област
Природно-математички факултет Универзитета у Бањој Луци		члан
Установа у којој је запослен-а		Функција у комисији

## 2. ПОДАЦИ О СТУДЕНТУ

Име, име једног родитеља, презиме: Бојан, Милисав, Николић

Датум рођења: 05.06.1980.

Мјесто и држава рођења: Тузла, БиХ

### 2.1. Студије првог циклуса или основне студије или интегрисане студије

Година уписа:	1999.	Година завршетка:	2004.	Просјечна оцјена током студија:	8,19
---------------	-------	-------------------	-------	---------------------------------	------

Универзитет: Универзитет у Бањој Луци

Факултет/и: Природно-математички факултет

Студијски програм: Математика и информатика

Стечено звање: Дипломирани математичар и информатичар

### 2.2. Студије другог циклуса или магистарске студије

Година уписа:	2005.	Година завршетка:	2010.	Просјечна оцјена током студија:	9,67
---------------	-------	-------------------	-------	---------------------------------	------

Универзитет: Универзитет у Новом Саду

Факултет/и: Природно-математички факултет

Студијски програм: Департман за математику и информатику

Назив завршног рада другог циклуса или магистарске тезе, датум одбране: Простори топологија, октобар 2010.

Ужа научна област завршног рада другог циклуса или магистарске тезе: Математика

Стечено звање: Магистар математичких наука

### 2.3. Студије трећег циклуса

Година уписа:	2017.	Број ECTS остварених до сада:	120	Просјечна оцјена током студија:	10,00
---------------	-------	-------------------------------	-----	---------------------------------	-------

Факултет/и: Природно-математички факултет

Студијски програм: Докторске студије математике

2.4. Приказ научних и стручних радова студента		
РБ		Категорија <sup>1</sup>
1.	Б. Николић, А. Картељ, М. Ђукановић, М. Грбић, С. Blum, G. Raidl, <i>Solving the Longest Common Subsequence Problem Concerning Non-Uniform Distributions of Letters in Input Strings</i> , Mathematics, Vol. 9, No. 13, 1515, Jun, 2021.	SCI листа IF=2,4
<p><i>Кратак опис садржине:</i></p> <p>Проблем најдужег заједничког подниза (LCS проблем) представља налажење најдужег подниза, који је заједнички за све улазне низове. Овај проблем има различите примјене у биоинформатици, молекуларној биологији и провјери плагијата, између осталог. Сви претходни приступи из литературе се заснивају на претпоставци да су улазне LCS инстанце изабране из униформних или скоро униформних дистрибуција вјероватноће слова. У овом раду представљамо приступ који је у стању да се ефикасно бави општијим случајевима, гдје појављивање слова у улазним стринговима прати неке друге случајеве полиномијалне расподјеле. Предложени приступ користи временски ограничену претрагу бима, вођену новом хеуристиком по имену GMPSUM. Ова хеуристика комбинује двије комплементарне статистике у облику конвексне комбинације. У раду је дата свеобухватна емпиријска процјена у два различита подешавања: (1) краткотрајно извршавање са фиксном величином бима како би се процијениле способности вођења упоређене хеуристике претраге и (2) дуготрајна извршења са фиксним циљним временима трајања како би се добила висококвалитетна рјешења. У оба подешавања, новопредложен поступак је у случају униформне расподјеле у рангу познатих метода из литературе, док их у случају разматраних неуниформних дистрибуција надмашује.</p>		
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		<b>ДА</b> <b>НЕ</b>
РБ		Категорија
2.	B. Nikolic, B. Sobot, "Measures of string similarities based on Hamming distance", Preprint, <a href="https://arxiv.org/abs/2211.14615">https://arxiv.org/abs/2211.14615</a> , 2022.	
<p><i>Кратак опис садржине:</i></p> <p>У раду се уводе нове мјере сличности фамилија стрингова уз помоћ техника и метода истрајне симплицијалне хомологије. Посматра се Чехова филтрација посматране фамилије стрингова и уз помоћ удаљености уског грла се дефинише мјера сличности фамилија стрингова која, поред упаривања бар-код линија, приоритет даје упаривању оних парова линија које одговарају квалитативно сличним хомолошким класама. Да би се ово постигло, у раду је развијена нова техника раздвајање радијуса симплекса.</p>		
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		<b>ДА</b> <b>НЕ</b>
РБ		Категорија

<sup>1</sup> Категорија се односи на оне часописе и научне скупове који су категорисани у складу са Правилником о публикавању научних публикација („Службени гласник РС”, бр. 77/10) и Правилником о мјерилима за остваривање и финансирање Програма одржавања научних скупова („Службени гласник РС”, бр. 102/14) односно припадност рада часописима индексираним у свјетским цитатним базама.

3.	Б. Николић, <i>Degenerated topological spaces</i> , NOVI SAD JOURNAL OF MATHEMATICS, Vol. 49, No. 2, pp. 95-107, 2019.	Scopus	
<i>Кратак опис садржине:</i>			
Коришћењем композиција оператора унутрашњости и затворења тополошког простора могуће је добити 7 различитих оператора. Тополошки простори у којима међу овим операторима има једнаких називају се дегенерисани тополошки простори. Овај рад даје класификацију и карактеризацију дегенерисаних тополошких простора.			
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		ДА	<u>НЕ</u>
РБ		Категорија	
4.	В. Nikolic, М. Djukanovic, D. Matic, <i>New mixed-integer linear programming model for solving the multidimensional multi-way number partitioning problem</i> , Computational and Applied Mathematics 41 (3), 119, 2022.	SCI листа IF=2,6	
<i>Кратак опис садржине:</i>			
Проблем мултидимензионалног вишеструког проблема партиционисања подразумева прављење партиције коначног скупа вектора, тако да суме одговарајућих координата у оквиру појединачних скупова из партиције не одступају значајно. Овај проблем има значајне практичне примјене у ситуацијама када треба извршити уједначење скупа елемената у погледу два или више посматраних одлика. У раду је предложен нови MILP модел за рјешавање овог проблема који у већој мјери надмашује постојеће моделе из литературе којима се рјешава овај проблем.			
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		ДА	<u>НЕ</u>
РБ		Категорија	
5.	Д. Богданић, Б.Николић, D.A.Romano, <i>Аксиоме теорије скупова</i> , МАТ-КОЛ, Vol. XV(1), pp. 17-25, 2009.		
<i>Кратак опис садржине:</i>			
Постоји више од једне могућности аксиоматизације теорије скупова. У овом раду размотрени су неки аспекти опште прихваћеног ZFC аксиоматског система теорије скупова.			
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		ДА	<u>НЕ</u>
РБ		Категорија	
6.	Б. Николић, Н. Елез, <i>О оператору границе</i> , МАТ-КОЛ, No. XIII(1), pp. 67-72, 2007.		
<i>Кратак опис садржине:</i>			
У раду су испитане особине оператора границе скупа у тополошком простору, посматрају се скупови добијени примјеном тог оператора и њихов међуоднос у смислу релације инклузије. Такође, доказане су неке нове формуле у вези границе скупа у тополошком			

простору.		
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		<b>ДА</b> <b><u>НЕ</u></b>
РБ		Категорија
7.	Б. Николић, Н. Елез, <i>Оператори пик, каро, треф, срце</i> , МАТ-КОЛ, No. XIII(1), pp. 73-80, 2007.	
<i>Кратак опис садржине:</i>		
У раду су, комбиновањем оператора унутрашњости и затворења, уведена 4 нова оператора, испитани су њихови међуодноси, као и топологије које ови оператори генеришу.		
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		<b>ДА</b> <b><u>НЕ</u></b>
РБ		Категорија
8.	Б. Николић, <i>О математичкој нотацији</i> , МАТ-КОЛ, Vol. XXVI, No. 1, pp. 1-15, 2020.	
<i>Кратак опис садржине:</i>		
Записивање математичких замисли подразумијева употребу неког облика математичке нотације. Избором симбола и начином њиховог записивања могу се јасније истаћи концепти који представљају суштину идеје која се жели записати. У овом раду је направљен осврт на коришћење математичке нотације, са акцентом на препознавање ситуација у којима је могуће изабрати нотацију којом се интуитивније изражава семантика текста који се жели записати.		
<b>Припадност рада ужој научној области којој припада предмет истраживања докторске дисертације</b>		<b>ДА</b> <b><u>НЕ</u></b>

### 3. УВОДНИ ДИО ОЦЈЕНЕ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Тема докторске дисертације под насловом “Мјере сличности на фамилијама стрингова” кандидата Бојана Николића, након давања сагласности на Извјештај о оцјени подобности теме и кандидата и испуњености услова за менторство за израду докторске дисертације на Природно-математичком факултету, прихваћена је одлуком Сената Универзитета у Бањој Луци број 02/04-3.2670-34/21 од 25.11.2021. године. Истом одлуком је проф. Др Борису Шоботу, ванредном професору на Одсјеку за математику и информатику Природно-математичког факултета Универзитета у Новом Саду, дата сагласност за менторисање кандидата на задату тему.

Садржај докторске дисертације изложен је у сљедећим главама:

- 1) Увод (странице: 3-6)
- 2) Тополошки методи (странице: 7-86)

- 3) Вјероватносни методи (странице: 87-144)
- 4) Примјена при рјешавању ЛЦС проблема (странице: 145-178)
- 5) Закључак (странице: 179-184)
- 6) Литература (странице: 185-192)

Дисертација је написана на српском језику, латиничним писаним фонтом Times New Roman на 192 странице А4 формата и садржи шест глава: Увод, Тополошки методи, Вјероватносни методи, Примјена при рјешавању ЛЦС проблема, Закључак, Литература (100 референци). Такође, дисертација садржи 22 слике и 12 табела.

**У уводној глави** (Увод, стр. 3-6) дат је кратак преглед студираних проблема, као и основне идеје кориштене у наставку. Такође, дат је и кратак опис садржаја осталих поглавља.

**У другој глави** (Тополошки методи, стр. 7-86) проучавано је неколико врста метрика, као што су Хамингова, ЛЦС и Хауздорфова метрика. Дате су и основе симплицијалне хомологије, а посебна пажња је посвећена Чеховом комплексу. Затим су уведени појмови филтрације, истрајног модула, те теорема о интервалној декомпозицији која представља основ за коришћење резултата о бар кодовима. На крају поглавља су уведени појмови удаљености уског грла, те је доказана теорема стабилности. Користећи фамилије стрингова, кандидат разматра аутоморфизме одговарајућих метричких простора и описује методу шеме регистрацијских комплекса. У овој глави су дати резултати у вези са методом раздвајања радијуса симплекса.

**У трећој глави** (Вјероватносни методи, стр. 87-144) су проучавани стрингови из угла вјероватносних мјера. Фамилије стрингова су искориштене да би се дефинисали одређени стохастички процеси и одговарајући простори вјероватноћа. Главни резултат овог поглавља се састоји из предложене мјере сличности ових фамилија стрингова. Ова мјера сличности је заснована на релативној ентропији. На крају поглавља су описана и два приступа за моделовање непознатих расподела. То су фреквенционистички приступ и Бејзовско закључивање.

**У четвртој глави** (Примјена при рјешавању ЛЦС проблема, стр. 145-178) су изведене рекурентне формуле за рачунање најдужега заједничког подниза два стринга. На почетку овог поглавља је дат преглед постојећих метода за општији случај фамилије стрингова и описан је начин на који се рјешења проблема моделирају као путеви у одговарајућем графу. У наставку поглавља се уводе нове статистике GM и PSUM, као и статистика GMPSUM која је њихова конвексна комбинација. На крају овог поглавља је дата процјена сложености добијених алгоритама, наведени су експериментални резултати који су упоређени са резултатима извршавања постојећих алгоритама.

**У закључку дисертације** (Закључак, стр. 179-184) је дат резиме добијених резултата и предложених поступака. Такође су наведени и евентуални будући правци разматраних тема.

#### 4. УВОД И ПРЕГЛЕД ЛИТЕРАТУРЕ

У истраживању се посматрају стрингови који се користе за математичко моделовање

објеката дискретног секвенцијалног типа. У већини случајева је потребна вишедимезионална анализа ових објеката. Стога је корисно да при анализи ових података имамо мјере поређења фамилија стрингова који моделују те податке. Проблем поређења фамилија стрингова је много компликованији од проблема поређења појединачних стрингова, тако да је неопходно укључити технике и хеуристике неколико математичких области. Ту се прије свега издвајају методе истрајне хомологије, хијерархијски Бејзовски модели и претрага бима. Ове методе су веома актуелне, нпр. методе истрајне хомологије су се почеле развијати тек крајем прошлог и почетком овог вијека ([5, 17, 36]). Садржај ове дисертације је усмјерен не само ка добијању теоријских резултата већ има и широк спектар потенцијалних примјена. Нпр, у анализи биолошких секвенци, претраживању текста, алгоритмима за препознавање говора итд.

Циљ дисертације је да се дефинишу мјере сличности скупова стрингова које би у што већој мјери одражавале заједничке релевантне особине тих скупова. Кориштене су технике и методе алгебарске топологије ([6, 18, 25]), вјероватносно-статистичке методе ([58, 75, 90]) и рачунске методе ([24, 33, 55]).

### Лигература:

- [1] J. Albert, J. Hu, "Probability and Bayesian modeling", CRC press, 2020.
- [2] S. Bacallado, S. Favaro, S. Power, L. Trippa, "Perfect sampling of the posterior in the hierarchical Pitman–Yor process", International Society for Bayesian Analysis, 2021.
- [3] R. A. Baeza-Yates, R. Gavaldà, G. Navarro, R. Scheihing, "Bounding the expected length of longest common subsequences and forests", Theory Comput. Syst., 32(4), pp. 435–452, 1999.
- [4] D. Bakkelund, "An LCS-based string metric", University of Oslo, 2009.
- [5] S.A. Barannikov, "The framed Morse complex and its invariants," Advances in Soviet Mathematics, 21, pp. 93-115, 1994.
- [6] U. Bauer i M. Lesnick, "Induced matchings and the algebraic stability of persistence barcodes," Journal of Computational Geometry, 6(2), pp. 162-191, 2015.
- [7] R. Beal, T. Afrin, A. Farheen, D. Adjeroh, "A new algorithm for the LCS problem with application in compressing genome resequencing data", BMC Genom., 17, 544, 2016.
- [8] R. Begleiter, R. El-Yaniv, G. Yona, "On prediction using variable order Markov models", Journal of Artificial Intelligence Research 22, pp. 385-421, 2004.
- [9] Y. Bengio, "Markovian models for sequential data", Neural computing surveys, 1999.
- [10] L. Bergroth, H. Hakonen, T. Raita, "A survey of longest common subsequence algorithms", Proceedings of the SPIRE 2000—The 7th International Symposium on String Processing and Information Retrieval, Coruna, Spain, pp. 39-48, 2000.
- [11] P. Billingsley, "Ergodic Theory and Information", New York: Wiley, 1965.
- [12] C. Blum, M. J. Blesa, "Probabilistic beam search for the longest common subsequence problem", In Proceedings of the International Workshop on Engineering Stochastic Local Search Algorithms, Brussels, Belgium, 2007., Springer: 200, pp. 150-161, 2007.
- [13] C. Blum, M. J. Blesa, M. López-Ibáñez, "Beam search for the longest common subsequence problem", Comput. Oper. Res. 2009, 36, pp. 3178–3186.
- [14] C. Blum, P. Festa, "Longest common subsequence problems", Metaheuristics for String Problems in Bioinformatics, Wiley:Hoboken,NJ, USA, 2016, Chapter 3, pp.45–60.
- [15] P. Bubenik i J. A. Scott, "Categorification of persistent homology", Discrete and Computational Geometry, Vol. 51, pp. 600–627, 2014.
- [16] B. Calvo, G. Santafé-Rodrigo, "Statistical comparison of multiple algorithms in multiple problems", The R Journal, Vol. 8/1, Aug. 2016.
- [17] G. Carlsson, "Topology and data", Bulletin (New Series) of the American Mathematical Society, 2009.

- [18] G. Carlsson, A. Zomorodian, A. Collins i L. Guibas, "Persistence barcodes for shapes", Eurographics Symposium on Geometry Processing, 2004.
- [19] H. T. Chan, C. B. Yang, Y. H. Peng, "The generalized definitions of the twodimensional largest common substructure problems", Proceedings of the 33rd Workshop on Combinatorial Mathematics and Computation Theory, Taipei, Taiwan, 13–14 May 2016, pp. 1–12.
- [20] F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas i S. Oudot, "Proximity of persistence modules and their diagrams," Research Report RR-6568, INRIA, 2008.
- [21] F. Chazal i B. Michel, "An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists", preprint 2021.
- [22] C. Chen, L. Du, W. Buntine, "Sampling table configurations for the hierarchical Poisson-Dirichlet process", ECML PKDD 2011: Machine Learning and Knowledge Discovery in Databases pp. 296–311, 2011.
- [23] S.F. Chen, J. Goodman, "An empirical study of smoothing technique for language modeling", Computer Science Group, Harvard University, Cambridge, Massachusetts, 1998.
- [24] V. Chvatal i D. Sankoff, "Longest common subsequences of two random sequences", Journal of Applied Probability, Vol. 12, No. 2, pp. 306-315, 1975.
- [25] D. Cohen-Steiner, H. Edelsbrunner i J. Harer, "Stability of persistence diagrams," Discrete and Computational Geometry, 37, pp. 103–120, 2007.
- [26] T. M. Cover i J. A. Thomas, "Elements of Information Theory", Second edition, John Wiley and Sons, Inc., 2006.
- [27] W. Crawley-Boevey, "Decomposition of pointwise finite-dimensional persistence modules", Journal of Algebra and Its Applications, Vol. 14, No. 5, 2015.
- [28] V. Dančik, "Expected length of longest common subsequences", PhD thesis, Department of Computer Science, University of Warwick, September 1994.
- [29] V. de Silva i V. Nanda, "Geometry in the space of persistence modules", Proceedings of the twenty-ninth annual symposium on Computational geometry, 2013.
- [30] T. K. Dey i Y. Wang, "Computational topology for Data Analysis", Cambridge University Press, 2022.
- [31] M. Djukanovic, A. Kartelj, D. Matic, M. Grbic, C. Blum, G. Raidl, "Solving the generalized constrained Longest Common Subsequence problem with many pattern strings", Technical Report AC-TR-21-008, AC, 2021.
- [32] M. Djukanovic, G. R. Raidl, C. Blum, "Anytime algorithms for the longest common palindromic subsequence problem", Comput. Oper. Res. 2020, 114, 104827.
- [33] M. Djukanovic, G. Raidl, C. Blum, "A Beam Search for the Longest Common Subsequence problem guided by a novel approximate expected length calculation", Proceedings of the LOD 2019—The 5th International Conference on Machine Learning, Optimization, and Data Science, Siena, Italy, 10–13 September 2019.
- [34] M. Djukanovic, G. Raidl, C. Blum, "Finding longest common subsequences: new anytime A\* search results", Appl. Soft. Comput. 2020, 95, 106499.
- [35] T. Easton, A. Singireddy, "A large neighborhood search heuristic for the longest common subsequence problem", J. Heuristics, 14, pp. 271-283, 2008.
- [36] H. Edelsbrunner i J. Harer, "Computational Topology: An Introduction", American Mathematical Society, 2010.
- [37] H. Edelsbrunner i J. Harer, "Persistent homology - a survey," Surveys on Discrete and Computational Geometry: Twenty Years Later, American Mathematical Society, 2008., pp. 257–282.
- [38] R. Forman, "A user's guide to discrete Morse theory", Seminaire Lotharingien de Combinatoire 48, Article B48c, 2002.
- [39] C.B. Fraser, "Subsequences and supersequences of strings", Ph.D. Thesis, University of



- Glasgow, Glasgow, UK, 1995.
- [40] A. Gabadinho i G. Ritschard, "Analyzing state sequences with probabilistic suffix trees: The PST R package", *Journal of Statistical Software*, 2016.
- [41] J.A. Gasthaus, "Hierarchical Bayesian nonparametric models for power-law sequences", PhD Thesis, 2020.
- [42] J.A. Gasthaus i Y. W. Teh, "Improvements to the sequence memoizer", *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems*, 2010.
- [43] R. M. Gray, "Entropy and Information Theory", Second edition, Springer New York, 2011.
- [44] D. Gusfield, "Algorithms on Strings, Trees, and Sequences", *Computer Science and Computational Biology*, Cambridge University Press: Cambridge, UK, 1997.
- [45] R. W. Hamming, "Error detecting and error correcting codes," *The Bell System Technical Journal*, vol. 29, no. 2, pp. 147-160, April 1950.
- [46] A. Hatcher, "Algebraic Topology", Cambridge University Press, 2001.
- [47] D. S. Hirschberg, "A linear space algorithm for computing maximal common subsequences", *Communications of the Association for Computing Machinery*, 18(6), pp. 341-343, 1975.
- [48] D. S. Hirschberg, "Algorithms for the Longest Common Subsequence problem", *Journal of the Association for Computing Machinery.*, 24(4), pp. 664-675, 1977.
- [49] L. C. Hsu, "A unified approach to generalized Stirling numbers", *Advances in Applied Mathematics* 20, pp. 366-384, 1998.
- [50] K. Huang, C. Yang, K. Tseng, "Fast algorithms for finding the common subsequences of multiple sequences", *Proceedings of the ICS 2004—The 9th International Computer Symposium*, Funchal, Portugal, 13–16 January 2004.
- [51] T. Jiang, M. Li, "On the approximation of shortest common supersequences and longest common subsequences", *SIAM J. Comput.*, 24, pp. 1122-1139, 1995.
- [52] C. Kermorvant, P. Dupont, "Improved smoothing for probabilistic suffix trees seen as variable order Markov chains", Springer-Verlag Berlin Heidelberg, 2002.
- [53] H. Kesten, N. Morse, "A property of the multinomial distribution", *Ann. Math. Stat.*, 30, pp. 120-127, 1959.
- [54] M. Kiwi, J. Soto, "On a speculated relation between Chvatal-Sankoff constants of several sequences", *Combin. Probab. Comput.*, 18(4), pp. 517–532, 2009.
- [55] M. Kiwi, M. Loeb, J. Matousek, "Expected length of the longest common subsequence for large alphabets", *Adv. Math.*, 197(2), pp. 480–498, 2005.
- [56] R. Kneser, H. Ney, "Improved backing-off for m-gram language modeling", *International Conference on Acoustics, Speech, and Signal Processing*, 1995.
- [57] J. B. Kruskal, "An overview of sequence comparison: Time warps, string edits, and macromolecules", *SIAM Rev.*, 25, pp. 201-237, 1983.
- [58] S. Kullback, R. A. Leibler, "On information and sufficiency", *The Annals of Mathematical Statistics*, 22, pp. 79-86, 1951.
- [59] D.A. Levin, Y. Peres, "Markov Chains and Mixing Times: Second Edition", American Mathematical Society, 2017.
- [60] Y. Li, Y. Wang, Z. Zhang, Y. Wang, D. Ma, J. Huang, "A novel fast and memory efficient parallel MLCS algorithm for long and large-scale sequences alignments", *Proceedings of the IEEE 32nd International Conference on Data Engineering*, Helsinki, Finland, 16–20 May 2016, pp. 1170–1181.
- [61] K. W. Lim, W. Buntine, C. Chen, L. Du, "Nonparametric Bayesian topic modelling with the hierarchical Pitman–Yor processes", *International Journal of Approximate Reasoning*, 2016.
- [62] G. S. Lueker, "Improved bounds on the average length of longest common subsequences", *Journal of the ACM*, 56(3), pp. 1–38, May 2009.

- [63] D. Maier, "The complexity of some problems on subsequences and supersequences", *J. ACM*, 25, pp. 322–336, 1978.
- [64] M. Maltenfort, "New definitions of the generalized Stirling numbers", *Aequationes mathematicae* volume 94, pp. 169–200, 2020.
- [65] G. Máté, A. Hofmann, N. Wenzel i D. W. Heermann, "A topological similarity measure for proteins", 2013.
- [66] P. Minkiewicz, M. Darewicz, A. Iwaniak, J. Sokołowska, P. Starowicz, J. Bucholska, M. Hryniewicz, "Common amino acid subsequences in a universal proteome—relevance for food science", *Int. J. Mol. Sci.*, 16, pp. 20748–20773, 2015.
- [67] S.R. Mousavi, F. Tabataba, "An improved algorithm for the longest common subsequence problem". *Comput. Oper. Res.* 2012, 39, pp. 512–520.
- [68] J. R. Munkres, "Elements of Algebraic Topology", CRC Press, 1993.
- [69] B. Nikolic, A. Kartelj, M. Djukanovic, M. Grbic, C. Blum, G. Raidl, "Solving the Longest Common Subsequence problem concerning non-uniform distributions of letters in input strings", *Mathematics* 9, 1515, 2021.
- [70] B. Nikolic, B. Sobot, "Measures of string similarities based on Hamming distance", Preprint, <https://arxiv.org/abs/2211.14615>, 2022.
- [71] A. Panconesi, "The stationary distribution of a Markov chain", Unpublished note, Sapienza University of Rome, 2005.
- [72] Z. Peng, Y. Wang, "A novel efficient graph model for the Multiple Longest Common Subsequences (MLCS) problem", *Front. Genet.*, 8, 104, 2017.
- [73] T. Petrie, "Probabilistic functions of finite state Markov chains," *Ann.Math.Statist.*, 40, No.1, 1969, pp. 97-115.
- [74] J. Pitman, "Coalescents with multiple collisions", *The Annals of Probability*, Vol. 27, No. 4, pp. 1870–1902, 1999.
- [75] J. Pitman, M. Yor, "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator", *The Annals of Probability*, vol 25., No.2, 1997.
- [76] T. Pohlert, "The pairwise multiple comparison of mean ranks package (PMCMR)", *R Package*, 27, 9, 2014.
- [77] L. Polterovich, D. Rosen, K. Samvelyan and J.Zhang, "Topological Persistence in Geometry and Analysis," American Mathematical Society, 2020.
- [78] Y. Reani i O. Bobrowski, "Cycle registration in persistent homology with applications in topological bootstrap," Preprint, 2021.
- [79] G. O. Roberts, "Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms", *Stochastic Processes and their Applications*, 49, pp. 207-216, 1994.
- [80] D. Sankoff, J. Kruskal, "Time warps, string edits, and macromolecules, *The Theory and Practice of Sequence Comparison*", CSLI Publication, 1999.
- [81] R. Serfozo, "Basics of Applied Stochastic Processes, Probability and its Applications", Springer-Verlag Berlin Heidelberg, 2009.
- [82] J.P. Serre, "Linear Representations of Finite Groups", Springer-Verlag New York, 1977.
- [83] C. E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal*, 1948.
- [84] D. J. Sheskin, "Handbook of parametric and nonparametric statistical procedures", Chapman and Hall/CRC, 2000.
- [85] S. J. Shyu, C.Y.Tsai, "Finding the longest common subsequence for multiple biological sequences by ant colony optimization", *Comput. Oper. Res.*, 36, pp. 73-91, 2009.
- [86] J. Storer, "Data Compression: Methods and Theory", Computer Science Press: MD, USA, 1988.
- [87] F. S. Tabataba, S. R. Mousavi, "A hyper-heuristic for the Longest Common Subsequence

- problem", *Comput. Biol. Chem.* 2012, 36, pp. 42–54.
- [88] Y. W. Teh, "A Bayesian interpretation of interpolated Kneser-Ney", NUS School of Computing Technical Report, 2006.
- [89] Y. W. Teh, "A hierarchical Bayesian language model based on Pitman-Yor processes", *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, 2006.
- [90] Y. W. Teh i M. I. Jordan, "Bayesian nonparametrics: hierarchical Bayesian nonparametric models with applications", *Cambridge Series in Statistical and Probabilistic Mathematics*, 2010.
- [91] S. G. Vadlamudi, S. Aine, P. P. Chakrabarti, "Anytime pack search", *Nat. Comput.*, 15, pp. 395–414. 2016.
- [92] S. G. Vadlamudi, P. Gaurav, S. Aine, P. P. Chakrabarti, "Anytime column search", *Proceedings of the AI'12—The 25th Australasian Joint Conference on Artificial Intelligence*, Sydney, Australia, 4–7 December 2012, pp. 254–265.
- [93] Q. Wang, D. Korkin, Y. Shang, "A fast multiple longest common subsequence (MLCS) algorithm", *IEEE Trans. Knowl. Data Eng.* 2011, 23, pp.321–334.
- [94] C. Wang, Y. Wang, Y. Cheung, "A branch and bound irredundant graph algorithm for large-scale MLCS problems", *Pattern Recognit.*, 119, 108059, 2021.
- [95] S. Wei, Y. Wang, Y. Yang, S. Liu, "A path recorder algorithm for Multiple Longest Common Subsequences (MLCS) problems", *Bioinformatics*, 36, pp 3035–3042, 2020.
- [96] F. Wood, C. Archambeau, J. Gasthaus, L. James, Y. W. Teh, "A stochastic memoizer for sequence data", *Proceedings of the 26 th International Conference on Machine Learning*, Montreal, Canada, 2009.
- [97] X. Xie, W. Liao, H. Aghajan, P. Veelaert, W. Philips, "Detecting road intersections from GPS traces using longest common subsequence Algorithm", *ISPRS Int. J. Geo-Inf.*, 1, 6, 2017.
- [98] J. Yang, Y. Xu, G. Sun, Y. Shang, "A new progressive algorithm for a Multiple Longest Common Subsequences Problem and its efficient parallelization", *IEEE Trans. Parallel Distrib. Syst.*, 24, pp. 862–870, 2013.
- [99] A. Zomorodian i G. Carlsson, "Computing persistent homology," *Discrete and Computational Geometry*, 33(2), pp. 249–274, 2004.
- [100] S. Zürcher, "Smallest enclosing ball for a point set with strictly convex level sets", MSc thesis, ETH Zurich, 2007.

## 5. МАТЕРИЈАЛ И МЕТОДОЛОГИЈА РАДА

У истраживању су кориштене методе алгебарске топологије и вјероватносно-статистичке методе. Осим тога, значајну улогу у анализи проблема заузимају и напредне рачунске методе које су имале улогу у упоређивања сложености алгоритама.

Кандидат је показао да адекватно користи поменути теоријски апарат за рјешавање посматраних проблема. Није дошло до промјене плана истраживања који је дат приликом пријаве докторске тезе.

## 6. РЕЗУЛТАТИ И НАУЧНИ ДОПРИНОС ИСТРАЖИВАЊА

### 6.1. Резултати истраживања

Оригинални и најзначајнији научни резултати овог истраживања се огледају у:

- 1) Развијене су нове технике и доказани одређени резултати који омогућавају не само увођење нових мјера сличности фамилија стрингова, већ и додатно осигуравају да ове мјере посједују својство стабилности. Описана је нова техника раздвајања радијуса симплекса, као кључно средство које омогућава модификацију методе уског грла путем хибридног упаривања.
- 2) Матрица транзиционих вјероватноћа одређује вјероватносну мјеру која одговара датом скупу стрингова. За двије такве мјере можемо рачунати релативну ентропију и помоћу ње дефинисати мјеру сличности за дате фамилије стрингова. Ове расподеле вјероватноћа у пракси нису познате, тако да се моделују разним методама. У тези су размотрене идеје које, коришћењем погодних апроксимација, знатно увећавају ефикасног овог поступка.
- 3) У тези је представљено рјешавање проблема проналажења најдужег заједничког подниза дате фамилије скупова над истим алфабетом (ЛЦС проблема) уз помоћ претраге бима (Beam Search или BS) који користи специјално дизајниране новоуведене хеуристике које усмјеравају претрагу ка перспективнијим рјешењима. Предности конструисаног метода у односу на постојеће су показане релативним статистичким тестовима.
- 4) Описани су нови правци истраживања дате тематике који су од изузетне важности у области проучавања структуре сложених скупова података.

### 6.2. Критичност и коректност тумачења резултата

Резултати истраживања су приказани веома компетентно, на јасан и прегледан начин.

### 6.3. Теоријски допринос и нови истраживачки резултати.

У дисертацији су развијене нове технике и доказани одређени резултати који омогућавају не само увођење нових мјера сличности фамилија стрингова, већ и додатно осигуравају да ове мјере посједују својство стабилности. Описана је нова техника раздвајања радијуса симплекса, као кључно средство које омогућава да поменуто хибридно упаривање има смисла. (B. Nikolic, B. Sobot, "Measures of string similarities based on Hamming distance", Preprint, <https://arxiv.org/abs/2211.14615>, 2022.). У тези је представљено и рјешење ЛЦС проблема уз помоћ претраге бима (Beam Search) уз коришћење специјално дизајниране новоуведене хеуристике која усмјерава претрагу ка перспективнијим рјешењима (B. Nikolic, A. Kartelj, M. Djukanovic, M. Grbic, C. Blum, G. Raidl, "Solving the Longest Common Subsequence problem concerning non-uniform distributions of letters in input strings", Mathematics 9, 1515, 2021.). На тај начин, ова теза представља оригиналан допринос посматраној теми. Презентовани резултати имају значајну примјену у већини природних наука, као и у одређеном броју друштвених наука гдје се проучавају дискретни објекти секвенцијалног типа.


## 7. ЗАКЉУЧАК И ПРИЈЕДЛОГ

На основу свега што је наведено у Извјештају, Комисија закључује да је докторска дисертација магистра Бојана Николића под насловом “Мјере сличности на фамилијама стрингова” израђена у складу са образложењем које је кандидат приложио приликом пријаве ове теме. Докторска дисертација је урађена према правилима и принципима научно-истраживачког рада и резултат је оригиналног научног рада кандидата. Развијене су нове технике и доказани одређени резултати који омогућавају не само увођење нових мјера сличности фамилија стрингова, већ и додатно осигуравају да ове мјере посједују својство стабилности. У тези је представљено и рјешење ЛЦС проблема уз помоћ претраге бима (Beam Search) која користи специјално дизајниране новоуведене хеуристичке функције које усмјеравају претрагу ка перспективнијим рјешењима. На тај начин, ова теза представља оригиналан допринос посматраној теми.

Будући да је кандидат показао темељно познавање предмета истраживања, те у потпуности одговорио на проблематику која се разматра у дисертацији, Комисија предлаже Научно-наставном вијећу Природно-математичког факултета Универзитета у Бањој Луци и Сенату Универзитета у Бањој Луци да прихвате овај извјештај и одобре јавну одбрану докторске дисертације.


Мјесто и датум:

Бања Лука,  
13. 11. 2023.

  
Др Душко Богданић, редовни професор,  
предсједник комисије

  
Др Бојан Башић, редовни професор, члан

  
Академик Зоран Митровић, редовни професор,  
члан

  
Др Марко Ђукановић, доцент, члан

ИЗДВОЈЕНО МИШЉЕЊЕ: Члан комисије који не жели да потпише извјештај јер се не слаже са мишљењем већине чланова комисије дужан је да у извјештај унесе образложење, односно разлоге због којих не жели да потпише извјештај.

