

Резюмета на статии, представени в изпълнение на
минималните национални изисквия за научна област
4.5 Математика

16 декември 2020 г.

1 Група В

1. Статия: Stefan Gerdjikov, A General Class of Monoids Supporting Canonisation and Minimisation of (Sub)sequential Transducers, Proceedings of the 12th International Conference on Language and Automata Theory and Applications, LATA 2018, pp. 143 – 155, 2018, ISBN:978-3-319-77312-4, doi:10.1007978-3-319-77313-1_11.

Резюме: В настоящата статия разглеждаме проблемите за канонизация и минимизация на последователни преобразуватели с изходи в произволен моноид. Показваме, че тези проблеми могат да бъдат ефективно решени в широк клас от моноиди, който включва: свободните моноиди, тропическите моноиди и групите и който е затворен относно декартово произведение. Разглежданият клас от моноиди е описан посредством пет прости аксиоми. Първите четири от тях изглеждат естествени. Що се отнася до последната, ние показваме, че тя всъщност е *необходима* за ефективното решаване на гореспоменатите проблеми.

Abstract:In this paper we consider the problems of canonisation and minimisation of subsequential transducer with output in an arbitrary monoid. We show that these problems can be efficiently solved for a large class of monoids that includes the free monoids, tropical monoid, and groups, and is closed under Cartesian Product. We describe this class of monoids in terms of five simple axioms. The first four of them seem to be natural. For the last one, we show that it is also necessary.

2. Статия: Stefan Gerdjikov, Stoyan Mihov, Klaus U. Schulz, A Simple Method for Building Bimachines from Functional Finite-State Transducers, Proceedings of the 22nd International Conference Implementation and Application of Automata, CIAA 2017, pp. 113 – 125, 2017, ISBN:978-3-319-60133-5, doi:10.1007978-3-319-60134-2_10

Резюме: Стандартната конструкция за построяване на бимашина от функционален преобразувател изисква трансформирането на преобразувателя в еднозначен (тоест такъв, в който за всяка входна дума има най-много един успешен път в преобразувателя). Тази трансформация се извършва посредством специализирана детерминизация. В тази статия, предлагаме нов принцип за конструкция на бимашини, при който от изходния преобразувател директно се строи еквивалентна бимашина. Идеята на този принцип е, че за всяка входна дума, която се приема от преобразувателя, бимашината имитира един от успешните пътища в преобразувателя. По този начин, за някои класове от (функционални) преобразуватели новата конструкция дава бимашина с

експоненциално по-малко състояния от стандартната. Първо представяме проста и обща схема на нашата конструкция. Доуточняване на тази схема води до по-добри оценки за пространствената сложност на резултатната бимашина.

Abstract:The standard construction of a bimachine from a functional transducer involves a preparation step for converting the transducer into an unambiguous transducer (A transducer is unambiguous if there exists at most one successful path for each label.). The conversion involves a specialized determinization. We introduce a new construction principle where the transducer is directly translated into a bimachine. For any input word accepted by the transducer the bimachine exactly imitates one successful path of the transducer. For some classes of transducers the new construction can build a bimachine with an exponentially lower number of states compared to the standard construction. We first present a simple and generic variant of the construction. A second specialized version leads to better complexity bounds in terms of the size of the bimachine.

3. **Статия:**Stefan Gerdjikov, Stoyan Mihov, Klaus U. Schulz, Space-Efficient Bimachine Construction Based on the Equalizer Accumulation Principle, Theoretical Computer Science, vol. 790, pp. 80 – 95, 2019, doi:10.1016/j.tcs.2019.04.027

Резюме: Съществуващите в литературата алгоритми за строене на бимашини от функционални преобразуватели се основават на следния принцип: всяко изпълнение на бимашината симулира конкретен успешен път на изходния преобразувател. Всеки отделен (частичен) изход на бимашината съответства на конкретен преход на преобразувателя. В тази статия въвеждаме алтернативен принцип за строене на бимашини, който наричаме *принцип за акумулиране на изравнители*. Според него, отделните стъпки на бимашината отчитат различните възможни (успешни) пътища на преобразувателя, като максимизират възможния изход по всички тях. Това води до конструкция, при която и левият и десният автомат на бимашината имат размер, ограничен от $2^{|Q|}$, където $|Q|$ е броят на състоянията в първоначалния преобразувател. За сравнение, предишните конструкции на бимашини водят до асимптотично по-големи автомати. В настоящата статия, показваме клас от реалновременни функционални преобразуватели с $n + 2$ състояния, за който стандартната конструкция генерира бимашина с поне $n!$ състояния докато конструкцията, която следва принципа за акумулиране на изравнители, води до $2^n + n + 3$ състояния. От друга страна, представяме клас от реалновременни функционални преобразуватели с $4(n + 1)$ състояния, който не допуска бимашина с по-малко от 2^n състояния. Това показва, че пространствената сложност на нашата конструкция се приближава до оптималната. Новопредложената конструкция може да се прилага за рационални функции с дефиниционна област в свободен моноид и област от стойности в “mge”¹ моноиди – широк клас от моноиди, който включва свободните моноиди, групите и други, и който е затворен относно декартово произведение.

Abstract: Algorithms for building bimachines from functional transducers found in the literature are based on the following principle: each run of the bimachine simulates a particular successful path of the input transducer. Every single bimachine output exactly corresponds to the output of a single transducer transition. Here we introduce an alternative construction principle called the equalizer accumulation principle. It suggests that the bimachine steps take into account alternative parallel transducer paths, maximizing the possible output at each step using a joint view. This results in a construction where the

¹mge е съкращение от most general equaliser, тоест, най-общ изравнител, и, следвайки аналогията за НОД и НОК, абривиатурата на български би била НОИ. Тъй като тази абривиатура вече се е установила с друго значение в българския език, предпочитаме да я избягваме употребата ѝ в математически контекст.

deterministic left and right automaton of the bimachine both have size bounded by $2^{|Q|}$ where $|Q|$ is the number of transducer states. In contrast, previous bimachine constructions lead to larger automata. We present a class of real-time functional transducers with $n + 2$ states for which the standard bimachine construction generates a bimachine with at least $n!$ states whereas the construction based on the equalizer accumulation principle leads to $2^n + n + 3$ states. On the other end we present a real-time functional transducers with $4(n+1)$ states that cannot be represented as a bimachine with less than 2^n states. Therefore the space complexity of our construction is close to optimal in terms of the number of states. The new construction can be applied to rational functions from free monoids to “mge monoids”, a large class of monoids including free monoids, groups, and others that is closed under Cartesian products.

4. Статия: Stefan Gerdjikov, Generalised Twinning Property, 23rd International Conference on Implementation and Applications of Automata, pp. 173 – 185, 2008, ISBN: 978-3-319-94811-9, doi:10.1007/978-3-319-94812-6_15

Резюме: В настоящата стаия разглеждаме проблема за секвенциализация на рационални функции $f : \Sigma^* \rightarrow M$. Въвеждаме клас от моноиди, който включва инфинитарните групи², свободните моноиди, тропическите моноиди и е затворен относно декартово произведение. За този клас от моноиди даваме конструкция за секвенциализация на преобразуватели и подходящо обобщаваме *близнашкото свойство*. Също така показваме конструкция, която проверява дали даден преобразувател над разглеждания клас от моноиди удовлетворява това свойство и доказваме, че то (близнашкото свойството) е необходимо и достатъчно условие за това конструкцията за секвенциализация да завърши.

Abstract: In this paper we consider the problem of sequentialisation of rational functions $f : \Sigma^* \rightarrow M$. We introduce a class of monoids that includes infinitary groups, free monoids, tropical monoids and is closed under Cartesian Product. For this class of monoids we provide a sequentialisation construction for transducers and appropriately generalise the notion of Twinning Property. We provide a construction to test the Twinning Property for transducers over the considered class of monoids and prove that it is a necessary and sufficient condition for the sequentialisation construction to terminate.

2 Група Г

1. Статия: Stefan Gerdjikov, Alexander Wolff, Decomposing simple polygon into pseudo-triangle and convex polygons, Computational Geometry Theory and Applications, vol. 41, 1-2, pp. 21 – 30, ISSN: 0925-7721, 2008

Резюме: В тази статия разглеждаме проблема за разбиването на прост многоъгълник на многоъгълници с върхове измежду върховете на дадения. Позволяваме два типа многоъгълници да участват в разбиването: псевдо-триъгълници и изпъкнали многоъгълници. Така полученото разбиване наричаме ПТ-изпъкнало. Интерес за нас ще представляват минималните разбивания, тоест разбивания на изходния многоъгълник на възможно най-малък брой многоъгълници от разглеждания вид. Допускането на два типа многоъгълници в разбиването има потенциалната възможност да намали сложността на резултатното разбиване (сравнено с разбиване, в което участват само изпъкнали многоъгълници, например). Проблемът за разбиването на прост многоъгълник на най-малък брой изпъкнали многоъгълници вече е разглеждан. Ние над-

²Тоест такива, за които равенството $a^k = b^k$ за някое положително k винаги влече, че $a = b$.

граждаме схемата на динамично програмиране от алгоритъма на Keil и Snoeyink за този проблем за случая, когато са допустими както изпъкнали многоъгълници, така и псевдо-триъгълници. Нашият алгоритъм намира минимално ПТ-разбиване на прост многоъгълник с n върха за $O(n^3)$ време и памет.

Abstract: In this paper we consider the problem of decomposing a simple polygon into subpolygons that exclusively use vertices of the given polygon. We allow two types of subpolygons: pseudo-triangles and convex polygons. We call the resulting decomposition PT-convex. We are interested in minimum decompositions, i.e., in decomposing the input polygon into the least number of subpolygons. Allowing subpolygons of one of two types has the potential to reduce the complexity of the resulting decomposition considerably. The problem of decomposing a simple polygon into the least number of convex polygons has been considered. We extend a dynamic-programming algorithm of Keil and Snoeyink for that problem to the case that both convex polygons and pseudo-triangles are allowed. Our algorithm determines such a decomposition in $O(n^3)$ time and space, where n is the number of the vertices of the polygon.

2. Статия: Stefan Gerdjikov, José Ramon González de Mendivil, Conditions for the existence of maximal factorizations, Fuzzy Sets and Systems, vol. 397, pp. 186 – 196, 2020, ISSN: 0165-0114, doi:10.1016/j.fss.2019.07.006

Резюме: Обобщаването на класически автоматни алгоритми за обикновени претеглени автомати или автомати преобразуващи думи за автомати с подлежаща по-обща алгебрична структура е значим интересен както от практическа, така и от теоретична гледна точка. Разбира се, обобщаването на класическите алгоритми налага ограничения върху подлежащата структура. В тази връзка свойството за *максимална факторизация* се оказва достатъчно за много практически проблеми, например минимизация и канонизация. Наскоро, аксиоматичен подход за описанието на свойства на моноидите, показва, че моноидите с *най-общ изравнител* (mge-моноиди) дават алтернатива за постигането на подобни резултати. В настоящата статия изучаваме теоретичната връзка между моноиди, които допускат максимална факторизация и mge-моноидите. Описваме необходими условия за съществуването на максимална факторизация и посочваме необходими условия за това mge-моноид да допуска максимална факторизация.

Abstract: Extending classical algorithms for ordinary weighted or string-to-string automata to automata with underlying more general algebraic structures is of significant practical and theoretical interest. However, the generalization of classical algorithms sets certain assumptions on the underlying structure. In this respect the maximal factorization turns out to be a sufficient condition for many practical problems, e.g. minimization and canonization. Recently, an axiomatic approach on monoid structures suggested that monoids with most general equalizer (mge-monoids) provide an alternative framework to achieve similar results. In this paper, we study the fundamental relation between monoids admitting a maximal factorization and mge-monoids. We describe necessary conditions for the existence of a maximal factorization and provide sufficient conditions for an mge-monoid to admit a maximal factorization.

3. Статия: Sebastian Wandelt, Dong Deng, Stefan Gerdjikov, Shashwat Mishra, Petar Mitankin, Manish Patil, Enrico Siragusa, Alexander Tiskin, Wei Wang, Jiaying Wang, State-of-the-art in string similarity search and join, ACM SIGMOD Record, vol. 43, 1, pp. 64 – 76, 2014, ISSN: 0163-5808, doi:10.1145/2627692.2627706

Резюме: Търсенето на близки стрингове и негови варианти е фундаментален проблем

с много приложения в области като интеграция на данни, качество на данни, компютърна лингвистика или биоинформатика. През последните десетилетия бяха разработени множество от методи за неговото решаване. Тъй като резултатите се публикуват в различни области без много интердисциплинарни обсъждания, статиите използват различни множества от данни и често изучават не основния проблем, а негова лека модификация, а броят на предложените методи надхвърля възможностите на всяка една отделна група, е много трудно да се получи обзор върху съществуващите най-добри алгоритми. В настоящата статия, докладваме резултатите за може би най-голямото представително съпоставяне на съществуващите алгоритми, някога правено в тази област. За да избегнем недостига на ресурси, организирахме представителното изследване като международно състезание, работна среща в рамките на EDBT/ICDT 2013. Отбори от различни области и от различни краища на света разработиха и настроиха своите програми за две строго дефинирани задачи. Всички алгоритми бяха оценени от външна група на две различни машини. Общо, сравнихме 14 различни програми върху две задачи за търсене на стрингове (k -приближено на търсене на даден стринг и всички k -близки стрингове в дадено (крайно) множество), използвайки множества от данни с нарастващ размер и различни характеристики от две различни области. Сравнихме програмите основно по тяхното бързодействие, но също даваме резултатите за ресурсите, които те изискват като памет, време за индексирание, ефекта от различния брой заявки, а също и способността на програмите да се възползват от паралелизъм на многоядрен процесор. Резултатите бяха осреднени от няколко независими пускания на програмите и потвърдени на втора, хардуерно различна система. Интересно наблюдение е, че различните области могат и би трябвало да учат повече една от друга, тъй като трите най-добре представили се отбора са от областите на компютърната лингвистика, бази от данни и биоинформатика, съответно.

Abstract: String similarity search and its variants are fundamental problems with many applications in areas such as data integration, data quality, computational linguistics, or bioinformatics. A plethora of methods have been developed over the last decades. Obtaining an overview of the state-of-the-art in this field is difficult, as results are published in various domains without much cross-talk, papers use different data sets and often study subtle variations of the core problems, and the sheer number of proposed methods exceeds the capacity of a single research group. In this paper, we report on the results of the probably largest benchmark ever performed in this field. To overcome the resource bottleneck, we organized the benchmark as an international competition, a workshop at EDBT/ICDT 2013. Various teams from different fields and from all over the world developed or tuned programs for two crisply defined problems. All algorithms were evaluated by an external group on two machines. Altogether, we compared 14 different programs on two string matching problems (k -approximate search and k -approximate join) using data sets of increasing sizes and with different characteristics from two different domains. We compare programs primarily by wall clock time, but also provide results on memory usage, indexing time, batch query effects and scalability in terms of CPU cores. Results were averaged over several runs and confirmed on a second, different hardware platform. A particularly interesting observation is that disciplines can and should learn more from each other, with the three best teams rooting in computational linguistics, databases, and bioinformatics, respectively.

4. Статия: Petar Mitankin, Stefan Gerdjikov, Stoyan Mihov, An approach to unsupervised historical text normalization, Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, pp. 29 – 34, 2014, ISBN:978-1-4503-2588-2, doi:10.11452595188.2595191
Резюме: Представяме нов подход за неинформирана корекция на текст. Нашият под-

ход се основава на автоматично извличане на шаблони от исторически вариации, при които се анализира структурата на думите в историческия корпус и се сравнява с тази на думите в съвременен речник. Стъпвайки върху така извлечените шаблони за вариации основният генератор на кандидати, REBELS, генерира думи за корекция дори извън съвременния речник. За корекцията на цели изречения, моделът се допълва с езиков модел като двата модела се комбинират по лог-линеен начин. Качеството на нашия неинформиран метод са сравнени емпирично с еталонна система за нормализация на текст, която разполага с примери за обучение. Експериментите показват, че нашата система постига 81.79% точност на ниво думи при нормализацията исторически текстове на английски език от 17-и век.

Abstract: We present a novel approach to unsupervised noisy text correction. Our approach is based on automatic extraction of historical variation patterns by analysing the structure of the words from a historical corpus and comparing it with the structure of the contemporary dictionary. Based on the extracted variation patterns the core candidate generator, REBELS, produces correction candidates even outside the modern dictionary. Further, the sentence correction is complemented with a modern language model combined in a log-linear model. The quality of our unsupervised approach is empirically compared against a supervised system competitive with the state-of-the-art supervised text normalisation systems. The experiments show that our system delivers 81.79% normalisation accuracy of 17th century English historical texts in a fully unsupervised setup.

5. Статия: Andrey Sariev, Vladislav Nenchev, Stefan Gerdjikov, Petar Mitankin, Hristo Ganchev, Stoyan Mihov, Tinko Tinchev, Flexible Noisy Text Correction, Proceedings of 11th IAPR International Workshop On Document Analysis Systems, pp. 31 – 35, 2014, ISBN: 978-1-4799-3243-6, doi:10.1109/DAS.2014.12

Резюме: Представяме нов общ и езиковонезависим подход за корекция на текстове, разработен и имплементиран в рамките на проекта CULTURA. Накратко описваме основния генератор на кандидати на думи за корекция, REBELS, цялостната идея на системата, нейната ефективна имплементация, основаваща се на функционални автомати и нейното непосредствено приложение. Качеството на цялата система е оценено в различни тестови установки, при които се мени езикът и произходът на грешки/шум.

Abstract: We present a new general and language independent approach to the noisy text correction problem developed and implemented in the framework of the CULTURA project. We briefly describe the core candidate generator, REBELS, the complete system concept, its efficient implementation based on functional automata and its immediate applications. The quality of the whole system is empirically established in different experimental settings where language and noise sources are varied.