

Jakub Radoszewski Curriculum vitae

data urodzenia 29.08.1984
miejsce urodzenia Poznań
e-mail j.radoszewski@mimuw.edu.pl
www <http://www.mimuw.edu.pl/~jrad>

Wykształcenie

2008-2012 doktor nauk matematycznych w zakresie informatyki
temat rozprawy: „Algorytmiczne i kombinatoryczne problemy związane ze zliczaniem powtórzeń w słowach” (ang. *Algorithmic and Combinatorial Problems Related to Enumeration of Repetitions in Words*)
promotor: prof. dr hab. Wojciech Rytter
Uniwersytet Warszawski, Wydział Matematyki, Informatyki i Mechaniki

2003-2008 magister informatyki
temat pracy magisterskiej: „Generowanie minimalnych leksykograficznie ciągów de Bruijna za pomocą słów Lyndona”
promotor: prof. dr hab. Wojciech Rytter
Uniwersytet Warszawski, Wydział Matematyki, Informatyki i Mechaniki

2003-2007 licencjat z matematyki
temat pracy licencjackiej: „Optymalna aproksymacja ciągów liczbowych ciągami monotonicznymi”
promotor: dr hab. Leszek Plaskota
Uniwersytet Warszawski, Wydział Matematyki, Informatyki i Mechaniki

Doświadczenie zawodowe

2013-obecnie adiunkt
Instytut Informatyki, Uniwersytet Warszawski

2012-2013 asystent
Instytut Informatyki, Uniwersytet Warszawski

07-09.2011 praktyki programistyczne
Facebook (Palo Alto, Stany Zjednoczone)

03-06.2010 wymiana Erasmus dla doktorantów
King’s College London

2007-2008 asystent stażysta
Instytut Informatyki, Uniwersytet Warszawski

07-09.2007 praktyki programistyczne
Google (Mountain View, Stany Zjednoczone)

07-09.2006 praktyki programistyczne
NVIDIA (Santa Clara, Stany Zjednoczone)

Wyróżnienia

2013, 2014	stypendium START Fundacji na rzecz Nauki Polskiej
2013	brązowy krzyż zasługi za działalność w Olimpiadzie Informatycznej
2013	nagroda II stopnia Rektora Uniwersytetu Warszawskiego za osiągnięcia naukowe
2012	wyróżnienie rozprawy doktorskiej
2011	stypendium naukowe w ramach projektu „Nowoczesny Uniwersytet”
2008	magister informatyki z wyróżnieniem, Uniwersytet Warszawski
2007	stypendium TP SA „Polskie Talenty”
2006, 2007	stypendium Ministra Nauki i Szkolnictwa Wyższego

Zainteresowania naukowe

Główny nurt moich badań naukowych dotyczy algorytmów tekstowych i kombinatoryki tekstów, ze szczególnym uwzględnieniem różnych typów powtórzeń w tekstach oraz nieklasycznych modeli tekstów. Poniżej opisałem wybrane wyniki uzyskane w pracy naukowej, poczynając od tych, które umieściłem w rozprawie doktorskiej.

Swoje badania prowadzę w ramach macierzystego zespołu na Uniwersytecie Warszawskim, który utrzymuje współpracę z King’s College London. Praca w stałym zespole owocuje ciągłym strumieniem publikacji na najlepszych konferencjach tematycznych, a także na czołowych konferencjach poświęconych informatyce teoretycznej. Wiele z powstałych prac ustanawia najlepsze znane wyniki dla poszczególnych problemów, jednak w naszym zespole nie wahamy się tworzyć publikacji wskazujących nowe kierunki badań. Moje publikacje przedstawiają zazwyczaj konstrukcje efektywnych, wielomianowych algorytmów natury kombinatorycznej.

Powtórzenia w słowach

Powtórzenia są jednym z centralnych zagadnień kombinatoryki i algorytmiki tekstów. W rozprawie doktorskiej rozpatrywałem powtórzenia będące potęgami słów (głównie kwadraty i sześciany, czyli podwójne i potrójne powtórzenia) oraz tzw. maksymalne powtórzenia i maksymalne powtórzenia sześciennne (ang. *runs* oraz *cubic runs*), występujące jako spójne fragmenty słowa. Maksymalne powtórzenia przedstawiają zwartą reprezentację struktury wszystkich powtórzeń w słowie.

Kluczową własnością kombinatoryczną wymienionych typów powtórzeń w słowach jest to, że ich maksymalna liczba w słowie o zadanej długości jest liniowa. Część kombinatoryczna mojej rozprawy doktorskiej dotyczyła konkretnych oszacowań w tych zależnościach liniowych i wpisywała się w istniejący od kilkunastu lat nurt badań na ten temat. W rozprawie znalazły się ulepszone oszacowania dla powtórzeń silnie okresowych, tj. sześcianów i maksymalnych powtórzeń sześciennych (opublikowane także w pracach [7,8]). Ponadto w rozprawie umieściłem oszacowania na maksymalną sumę wykładników maksymalnych powtórzeń – dolne $2,035n$ i górne $4,1n$, przy czym n to długość słowa (opublikowane też jako [10]). Pierwsze z tych oszacowań pozwoliło obalić hipotezę Kolpakova i Kucherova z 1999 roku, iż suma wykładników maksymalnych powtórzeń nie przekracza $2n$; drugie natomiast poprawiło najlepsze znane wówczas oszacowanie górne. Niektóre z podanych wyników zostały uzyskane z pomocą eksperymentów komputerowych.

W rozprawie pojawił się też wynik algorytmiczny dotyczący powtórzeń: liniowy algorytm pozwalający wyznaczać wszystkie kwadraty (i, ogólniej, potęgi) w słowie oraz tzw. okresy lokalne

słowa na podstawie struktury maksymalnych powtórzeń (także w pracy [2]). Co prawda wcześniej znane były liniowe algorytmy rozwiązujące podane problemy, jednak były one znacznie bardziej skomplikowane. W późniejszej publikacji [19] wraz z zespołem zawarłem inne efektywne algorytmy dotyczące wyznaczania powtórzeń w słowie.

Quasi-powtórzenia

Oprócz klasycznych powtórzeń w słowach zajmuję się także quasi-powtórzeniami, w których dopuszcza się, że poszczególne wystąpienia powtórzeń mogą się nakładać. Quasi-powtórzenia są ważnym uogólnieniem powtórzeń, gdyż pozwalają dostrzec powtarzające się struktury w słowie, gdy nie spełniają one definicji klasycznych powtórzeń. Najczęściej rozważanymi typami quasi-powtórzeń są szablon i quasi-szablon. Szablonem słowa nazywamy jego spójny fragment, którego wystąpieniami możemy pokryć całe słowo, natomiast quasi-szablon to szablon, którego dwa skrajne wystąpienia mogą wystawać poza obręb słowa.

Wraz z zespołem przygotowałem kilka znaczących publikacji w tej dziedzinie, wśród których najważniejszym wynikiem jest liniowy algorytm wyszukiwania wszystkich quasi-szablonów w słowie (praca [21]). Poprzedni algorytm rozwiązujący ten problem, pochodzący z roku 1996, działał w czasie $O(n \log n)$. Zestaw istotnych narzędzi potrzebnych do uzyskania wyniku z pracy [21] opublikowaliśmy wcześniej w pracy [8]. W naszych najnowszych pracach wprowadziliśmy ogólniejsze pojęcia *częściowych* szablonów i quasi-szablonów słowa, które muszą pokrywać zadaną liczbę pozycji w słowie. Problemy te należą do szerszej dziedziny tzw. wyszukiwania przybliżonych wzorców (ang. *approximate pattern matching*). Okazało się, że było możliwe zaproponowanie efektywnych algorytmów (o liniowo-logarytmicznej złożoności czasowej) wyznaczających te uogólnione quasi-powtórzenia. Algorytmy te przedstawiliśmy, odpowiednio, w pracach [4] i [14].

Nieklasyczne modele tekstów

Wraz z zespołem zajmowałem się algorytmiką słów w modelu przemiennym, w szczególności wyznaczaniem okresów abelowych słów i indeksowaniem w modelu abelowym, które stanowią naturalne uogólnienia ważnych zagadnień klasycznej algorytmiki tekstów. Okres abelowy odpowiada podziałowi słowa na równej długości fragmenty równoważne w sensie abelowym. W zależności od tego, czy dopuszczamy, aby skrajne fragmenty podziału były krótsze, otrzymuje się trzy warianty tego problemu. We wszystkich tych wariantach mój zespół podał najszybsze znane obecnie algorytmy (prace [6] i [16]). W problemie indeksowania mamy dany tekst i odpowiadamy na zapytania o występowanie w nim różnych wzorców. W modelu abelowym polega to na znajdowaniu spójnych fragmentów tekstu równoważnych abelowo danemu wzorcowi. Wcześniejsze wyniki dotyczące indeksowania abelowego działały jedynie dla słów binarnych. W pracy [15] mój zespół podał pierwszy indeks abelowy dla słów nad dowolnym stałym alfabetem, w którym złożoność czasowa zapytania jest silnie podliniowa, a rozmiar struktury jest silnie podkwadratowy. Jest to jak dotychczas jedyny nietrywialny wynik dotyczący indeksowania abelowego nad alfabetem większym od binarnego.

W innym modelu nieklasycznym – modelu wyszukiwania kształtów (ang. *order-preserving matching*) – dwa słowa określa się jako równe, jeśli istnieje rosnąca bijekcja przekształcająca litery jednego z nich w litery drugiego. Jest to szczególnie przypadek NP-trudnego w ogólności zagadnienia wyszukiwania wzorców permutacyjnych. Badania mojego zespołu zaowocowały liniowym algorytmem wyszukiwania wzorca w tym modelu nad wielomianowym alfabetem (praca [5]) oraz indeksującą strukturą danych o czasie konstrukcji $O(n \log \log n)$, pozwalającą na wyszukiwanie

wzorców w tekście w czasie proporcjonalnym do długości wzorca (praca [18]). Niewiele później pojawiły się publikacje innych autorów poszerzające te wyniki.

Innym kierunkiem badań, zainicjowanym przez nasz zespół, jest wyszukiwanie kwadratów w etykietowanych drzewach (czyli wyszukiwanie ścieżek prostych o etykietach tworzącej kwadrat). Nasze badania pokazały, że własności kwadratów w drzewach są zaskakująco odmienne od własności kwadratów w zwykłych słowach. W szczególności udało nam się wykazać ściśle ponadliniowe i ściśle podkwadratowe, dokładne oszacowanie asymptotyczne $\Theta(n^{4/3})$ na maksymalną liczbę różnych kwadratów w drzewie (w pracy [20]) oraz przedstawić algorytm o złożoności $O(n \log^2 n)$ wyznaczania zwartej reprezentacji wszystkich kwadratów w drzewie (praca [1]).

Publikacje w czasopismach

1. T. Kociumaka, J. Pachocki, J. Radoszewski, W. Rytter, T. Waleń, *Efficient counting of square substrings in a tree*, Theoretical Computer Science 544, s. 60-73, 2014.
Wersja konferencyjna: Algorithms and Computation – 22rd International Symposium (ISAAC 2012), s. 207-216, 2012.
2. M. Crochemore, C. S. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, *Extracting Powers and Periods in a String from Its Runs Structure*, Theoretical Computer Science 521, s. 29-41, 2014.
Wersja konferencyjna: String Processing and Information Retrieval – 17th International Symposium (SPIRE 2010), s. 258-269, 2010.
3. M. Crochemore, C. S. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, K. Stencel, T. Waleń, *New simple efficient algorithms computing powers and runs in strings*, Discrete Applied Mathematics 163(3), s. 258-267, 2014.
4. T. Kociumaka, S. P. Pissis, J. Radoszewski, W. Rytter, T. Waleń, *Fast Algorithm for Partial Covers in Words*, przyjęta do Algorithmica, 2014.
Wersja konferencyjna: Combinatorial Pattern Matching – 24th Annual Symposium (CPM 2013), s. 177-188, 2013.
5. M. Kubica, T. Kulczyński, J. Radoszewski, W. Rytter, T. Waleń, *A linear time algorithm for consecutive permutation pattern matching*, Information Processing Letters 113(12), s. 430-433, 2013.
6. M. Crochemore, C. S. Iliopoulos, T. Kociumaka, M. Kubica, J. Pachocki, J. Radoszewski, W. Rytter, W. Tyczyński, T. Waleń, *A note on efficient computation of all Abelian periods in a string*, Information Processing Letters 113(3), s. 74-77, 2013.
7. M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, *On the maximum number of cubic subwords in a word*, European Journal of Combinatorics 34(1), s. 27-37, 2013.
Wersja konferencyjna: International Workshop on Combinatorial Algorithms (IWOCA 2009), s. 345-355, 2009.

8. M. Crochemore, C. S. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, *The maximal number of cubic runs in a word*, Journal of Computer and System Sciences 78(6), s. 1828-1836, 2012.
Wersja konferencyjna: Language and Automata Theory and Applications (LATA 2010), s. 227-238, 2010.
9. M. Christou, M. Crochemore, C. S. Iliopoulos, M. Kubica, S. P. Pissis, J. Radoszewski, W. Rytter, B. Szreder, T. Waleń, *Efficient Seeds Computation Revisited*, Theoretical Computer Science 483, s. 171-181, 2012.
Wersja konferencyjna: Combinatorial Pattern Matching – 22nd Annual Symposium (CPM 2011), s. 350-363, 2011.
10. M. Crochemore, M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, *On the maximal sum of exponents of runs in a string*, Journal of Discrete Algorithms 14, s. 29-36, 2012.
Wersja konferencyjna: International Workshop on Combinatorial Algorithms (IWOCA 2010), s. 10-19, 2010.
11. J. Radoszewski, W. Rytter, *On the structure of compacted subword graphs of Thue-Morse words and their applications*, Journal of Discrete Algorithms 11, s. 15-24, 2012.

Publikacje konferencyjne

12. T. Kociumaka, J. W. Pachocki, J. Radoszewski, W. Rytter, T. Waleń, *On the String Consensus Problem and the Manhattan Sequence Consensus Problem*, praca zaakceptowana na konferencję String Processing and Information Retrieval – 21st International Symposium (SPIRE 2014), 2014.
13. T. Kociumaka, J. Radoszewski, W. Rytter, *Computing k -th Lyndon Word and Decoding Lexicographically Minimal de Bruijn Sequence*, Combinatorial Pattern Matching – 25th Annual Symposium (CPM 2014), s. 202-211, 2014.
14. T. Kociumaka, S. P. Pissis, J. Radoszewski, W. Rytter, T. Waleń, *Efficient Algorithms for Shortest Partial Seeds in Words*, Combinatorial Pattern Matching – 25th Annual Symposium (CPM 2014), s. 192-201, 2014.
15. T. Kociumaka, J. Radoszewski, W. Rytter, *Efficient Indexes for Jumbled Pattern Matching with Constant-Sized Alphabet*, 21st Annual European Symposium on Algorithms (ESA 2013), s. 625-636, 2013.
16. T. Kociumaka, J. Radoszewski, W. Rytter, *Fast Algorithms for Abelian Periods in Words and Greatest Common Divisor Queries*, 30th International Symposium on Theoretical Aspects of Computer Science (STACS 2013), s. 245-256, 2013.
17. T. Kociumaka, J. Radoszewski, W. Rytter, T. Waleń, *Linear-Time Version of Holub's Algorithm for Morphic Imprimitivity Testing*, Language and Automata Theory and Applications (LATA 2013), s. 383-394, 2013.

18. M. Crochemore, C. S. Iliopoulos, T. Kociumaka, M. Kubica, A. Langiu, S. P. Pissis, J. Radoszewski, W. Rytter, T. Waleń, *Order-Preserving Incomplete Suffix Trees and Order-Preserving Indexes*, String Processing and Information Retrieval – 20th International Symposium (SPIRE 2013), s. 84-95, 2013.
19. T. Kociumaka, J. Radoszewski, W. Rytter, T. Waleń, *Efficient Data Structures for the Factor Periodicity Problem*, String Processing and Information Retrieval – 19th International Symposium (SPIRE 2012), s. 284-294, 2012.
20. M. Crochemore, C. S. Iliopoulos, T. Kociumaka, M. Kubica, J. Radoszewski, W. Rytter, W. Tyczyński, T. Waleń, *The Maximum Number of Squares in a Tree*, Combinatorial Pattern Matching – 23rd Annual Symposium (CPM 2012), s. 27-40, 2012.
21. T. Kociumaka, M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, *A linear time algorithm for seeds computation*, Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2012), s. 1095-1112, 2012.
22. J. Radoszewski, W. Rytter, *Hamiltonian Paths in the Square of a Tree*, Algorithms and Computation – 22nd International Symposium (ISAAC 2011), s. 90-99, 2011.
23. M. Cygan, M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, *Polynomial-Time Approximation Algorithms for Weighted LCS Problem*, Combinatorial Pattern Matching – 22nd Annual Symposium (CPM 2011), s. 455-466, 2011.
24. M. Crochemore, M. Cygan, C. S. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, *Algorithms for Three Versions of the Shortest Common Superstring Problem*, Combinatorial Pattern Matching, 21st Annual Symposium (CPM 2010), s. 299-309, 2010.
25. J. Radoszewski, W. Rytter, *Efficient Testing of Equivalence of Words in a Free Idempotent Semigroup*, Theory and Practice of Computer Science, 36th Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2010), s. 663-671, 2010.

Publikacje w czasopismach popularyzatorskich

26. J. Radoszewski, *More Algorithms without Programming*, Olympiads in Informatics 8, s. 157-168, 2014.
27. J. Pachocki, J. Radoszewski, *Where to Use and How not to Use Polynomial String Hashing*, Olympiads in Informatics 7, s. 90-100, 2013.
28. T. Kulczyński, J. Łącki, J. Radoszewski, *Stimulating Students' Creativity with Tasks Solved Using Precomputation and Visualization*, Olympiads in Informatics 5, s. 71-81, 2011.
29. M. Kubica, J. Radoszewski, *Algorithms without Programming*, Olympiads in Informatics 4, s. 52-66, 2010.
30. K. Diks, M. Kubica, J. Radoszewski, K. Stencel, *A Proposal for a Task Preparation Process*, Olympiads in Informatics 2, s. 64-74, 2008.