

Wbrew pozorom nie chodzi o szukanie błędów w szkolnym dyktandzie, ale o efektywną analizę ciągu znaków, takich jak np. kod DNA. Dr Jakub Radoszewski z Instytutu Informatyki Wydziału Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego pracuje nad stworzeniem algorytmów automatycznego przeszukiwania i przetwarzania tekstów z błędami i niejednoznacznościami, które znajdą zastosowanie w biologii obliczeniowej.

„W informatyce teoretycznej znanych jest bardzo wiele efektywnych algorytmów przetwarzania tekstów idealnych, czyli takich, w których nie dopuszczamy możliwości, że niektóre symbole zostały zniekształcone w procesie uzyskiwania tekstu bądź jego przekazywania. Liczba znanych algorytmów poświęconych tekstom z błędami lub z niejednoznacznościami jest znacznie mniejsza, a właśnie takie algorytmy są niezbędne w biologii obliczeniowej, z uwagi na naturę i pochodzenie analizowanych danych” – mówi dr Jakub Radoszewski, laureat drugiego konkursu w programie HOMING realizowanego przez Fundację na rzecz Nauki Polskiej w ramach Programu Operacyjnego Inteligentny Rozwój.

Efektom jego pracy będą pomysły na narzędzia i programy – „cegielki informatyczne” – które będą mogły być wykorzystane w tworzeniu większych systemów służących do analizy niedoskonałych, niedookreślonych czy zniekształconych danych. „W ramach projektu będziemy się także zajmować wyszukiwaniem podobieństw między tekstami zawierającymi zniekształcenia, znajdowaniem regularności w takich tekstach oraz próbą analizy danych powstałych w wyniku sekwencjonowania DNA. W tym przypadku informatyka pełni zatem rolę służebną wobec biologii obliczeniowej i innych dziedzin, które mają do czynienia z takimi właśnie danymi” – wyjaśnia dr Radoszewski.

Projekt będzie wykonywany we współpracy z uczonymi z King's College London w Wielkiej Brytanii, którzy będą służyli wsparciem w zakresie zastosowań na styku informatyki i biologii obliczeniowej oraz z badaczką z Paryża z dużym doświadczeniem w zakresie teoretycznej analizy metod przeszukiwania tekstów.