

Автором публікації «Порівняльний аналіз розподілених суфіксних дерев і традиційних методів управління даними» є Садовий Я. С. Державний університет «Житомирська політехніка», м. Житомир, Україна.

В публікації досліджується ефективність різних методів побудови суфіксних дерев у порівнянні з традиційними структурами даних, такими як префіксні дерева та перетворення Берроуза-Вілера (BWT). Основною метою дослідження є визначення найефективнішого алгоритму для роботи з великими обсягами текстових даних та оптимізація ресурсів пам'яті й часу обчислень.

Дослідження розглядає три основні підходи до побудови суфіксного дерева:

Наївний метод, який ітеративно додає кожен суфікс у структуру, але має квадратичну складність $O(n^2)$, що робить його непридатним для великих даних.

Алгоритм Укконена, який забезпечує побудову дерева в лінійний час $O(n)$, використовуючи оптимізацію вставки суфіксів.

Алгоритм Partition and Write Only Top Down (PWOTD), який використовує попереднє розділення вхідних суфіксів, що дозволяє значно зменшити вимоги до пам'яті та ефективно будувати дерево.

Також проведено порівняння суфіксних дерев з традиційними методами управління даними, включаючи префіксні дерева, які ефективні для зберігання префіксів, та перетворення Берроуза-Вілера, яке використовується для стиснення даних.

Аналіз показує, що:

Суфіксні дерева ефективніші за традиційні методи у задачах пошуку та обробки рядкових даних.

PWOTD є найкращим методом серед розглянутих алгоритмів для побудови суфіксного дерева, оскільки зменшує використання пам'яті за рахунок розбиття суфіксів, хоча це збільшує обчислювальні витрати.

Префіксні дерева мають переваги у зберіганні даних із загальними префіксами, проте поступаються суфіксним деревам у швидкості пошуку підрядків.

Перетворення Берроуза-Вілера ефективно для стиснення, але не придатне для онлайн-обробки тексту.

Ключові інсайти:

PWOTD забезпечує найкращий баланс між швидкістю та використанням пам'яті.

Цей алгоритм дозволяє будувати суфіксне дерево для великих рядків, розбиваючи їх на підмножини, що зменшує навантаження на оперативну пам'ять.

Важливо враховувати збільшення обчислювальних витрат на етапі розділення.

Алгоритм Укконена є найефективнішим у плані швидкості, якщо ресурс пам'яті не є критичним.

Він забезпечує лінійну складність, що робить його оптимальним для роботи з текстовими даними в реальному часі.

Цей алгоритм корисний для задач, що вимагають швидкого оновлення дерева без необхідності його повної реконструкції.

Суфіксні дерева значно перевершують традиційні методи у завданнях обробки тексту.

На відміну від префіксних дерев, вони дозволяють швидко знаходити підрядки, що особливо важливо для біоінформатики та пошуку схожих фрагментів тексту.

Їхня динамічна модифікація робить їх придатними для реального застосування в інформаційних системах.

Висновок:

Публікація надає детальний огляд методів побудови суфіксних дерев та їх порівняння з традиційними методами управління даними. Алгоритм PWOTD показав

найкращі результати щодо оптимізації використання пам'яті, а алгоритм Укконена залишається найефективнішим для швидкої обробки великих текстових масивів.

Отримані висновки є цінними для розробки ефективних алгоритмів обробки тексту, пошуку інформації та роботи з великими наборами даних. Майбутні дослідження можуть зосередитися на подальшому покращенні ефективності PWOTD та його адаптації до різних застосувань, таких як обробка природної мови та аналіз біологічних послідовностей.