Ігор Завадський

доктор фіз.-мат. наук, професор кафедри математичної інформатики

Теорія та практика адаптивного стиснення даних

Класичний підхід до задачі стиснення даних полягає в побудові оптимального коду для наперед заданого розподілу частот символів. Основною ідеєю адаптивного стиснення є врахування статистичних характеристик файлу по мірі його опрацювання. Переваги такого підходу очевидні: по-перше, файл опрацьовується в один прохід; по-друге, пристосування до локальних змін частот символів може підвищити ефективність стиснення. Добре відомі і класичні адаптивні алгоритми, стискальна ефективність яких близька до оптимальної, такі як алгоритми Віттера і Фоллера-Галахера-Кнута. Єдиним їхнім недоліком є низька швидкодія. Останнім часом були запропоновані адаптивні алгоритми, що теоретично є оптимальними як за ємнісними, так і за часовими характеристиками за певних обмежень на розмір алфавіту.

Однак це то випадок, коли теорія сильно розходиться з практикою. У доповіді ми наведемо результати експериментів, що демонструють справжню ефективність різних рішень у практичних кейсах стиснення даних. Також буде запропоновано низку алгоритмів, що поєднують майже оптимальні ємнісні показники із відмінними часовими характеристиками стиснення природно-мовних текстів і програмного коду.