

Винахід відноситься до області ефективного кодування інформації і може бути використаний в комп'ютерних вузлах зв'язку та мережі Інтернет.

Існує пристрій стискання текстових повідомлень методом LZW [Чернега В.С. Сжатие информации в компьютерных сетях / В.С. Чернега - Севастополь: СевГТУ, 1997. - 214 с.], що містить у собі кодову таблицю, кодовий лічильник, блок формування хеш-функції, схему порівняння і префіксийний регістр. Однак, ця схема не враховує інформаційно-статистичні властивості мови текстового повідомлення. Тому на початковому етапі кодування, коли йде побудова кодової таблиці рядків змінної довжини, з виходу пристрою в канал передачі надходять одиночні символи, закодовані кодовими словами довжиною більш 8 біт, що приводить до розширення переданого повідомлення. В міру заповнення кодової таблиці підрядками тексту, що стискається, і кодуванню двох і більш символів (підрядків) однією кодовою комбінацією відбувається зменшення обсягів переданої в канал інформації за рахунок збільшення коефіцієнта стиску.

За прототип узятий пристрій кодування рядків змінної довжини [Welch T.A. A technique for high-performanse data compression // IEEE Computer. - 1984. - Vol.17. -N6. - P. 8-19.]. Даний пристрій містить регістр вхідних символів, до якого підключені блок формування хеш-функції і регістр формування коду під, що утвориться із символів стисливого тексту. Вихід регістра формування коду підрядка підключений до блоку формування хеш-функції, що прискорює процедуру пошуку рядка в таблиці рядків змінної довжини; пристрій порівняння, що здійснює пошук у таблиці рядків перемінної довжини підрядку, що надійшов у вхідний регістр із входу пристрою; і таблиця рядків перемінної довжини, що містить підрядки, що знайдені пристроєм у тексті. Причому вихід регістра формування коду підрядка є виходом пристрою. До блоку вибору адреси таблиці рядків підключені початковий лічильник і блок формування хеш-функції. До виходу блоку вибору адреси таблиці рядків підключена таблиця рядків перемінної довжини, до якої у свою чергу підключений пристрій порівняння регістра формування коду підрядка і кодовий лічильник.

В основу винаходу поставлена задача збільшення ступеня стискання і зменшення часу стискання текстової інформації, особливо на початковому етапі ефективного кодування тексту. Задача вирішується наступним шляхом. Текст пропонується розділяти на різні типи: розмовна мова, белетристика, ділові тексти і т.д. В зв'язку з тим, що для кожного типу тексту однієї і тієї ж мови існує своя статистика підрядків, яку можна заздалегідь зібрати шляхом обробки декількох текстів одного типу, тобто в кожній таблиці знаходяться підрядки з найбільшою імовірністю для кожного типу тексту. Використання цих таблиць дозволяє уникнути побудови не ефективних кодових комбінацій на початковому етапі кодування. Для реалізації цього методу в пристрій стиску текстової інформації вводяться блок оцінки типу тексту, блок таблиць кодування типу тексту, реалізований за допомогою блока оперативних запам'ятовуючих пристроїв (ОЗП), рядків змінної довжини, що відповідають різним типам тексту, а також комутатор таблиць. На кресленні (Фіг.) зображена структурна схема даного пристрою стиску. Пристрій містить блок оцінки типу тексту 1, вхідний регістр символів 2, генератор тактових імпульсів (ГТІ) 3, префіксийний регістр 4, вихідний регістр 5, пристрій керування 6, комутатор таблиць 7, блок ОЗП 8, блок порівняння 9, блок кодових лічильників 10.

Перші входи блоку 1 і блоку 2 з'єднані між собою паралельно, і точка їх з'єднання є входом пристрою. Вихід блоку 1 з'єднаний із входом блоку 7, виходи якого з'єднані з блоком 8 і блоком 6. Виходи блоку 6 з'єднані з другими входами блоків 2, 4 і 10, а відповідні входи з виходами блоків 3, 9 і 7. Виходи блоку 2 з'єднані відповідно з входами блоків 4 і 9. Виходи блоку 10 з'єднані з блоком 8, виходи якого з'єднані з блоком 9. Відповідні входи блоку 9 з'єднані з блоками 1 і 6. Виходи з блоку 4, є відповідно входами блоків 9 і 5. Вихід блоку 5 є виходом із пристрою стискання.

Пристрій працює наступним чином.

Символи, що кодують, надходять послідовно у вхідний регістр символів 2. Потім символ із блоку 2 переноситься шляхом зсуву в префіксийний регістр 4, а у вхідний регістр 2 заноситься наступний символ вхідного потоку. Кодова комбінація, що знаходиться в блоці 4 є префікс, який разом із символом вхідного регістра утворюють рядок. Блок оцінки типу тексту 1 здійснює послідовний пошук рядка в кожній з ОЗП блока таблиць. Таблиця, у якій виявлене найбільше число підрядків, що співпадають із підрядками, які кодуються (в експериментальній моделі, що була створена, через 8 тис. символів) призначається поточною. Індекс рядка таблиці є кодом рядка і записується у вихідний регістр. На початковому етапі кодування використовується узагальнена для всіх типів тексту таблиця.

Код знайденого рядка призначається новим префіксом і заноситься в префіксийний регістр 4. Потім у вхідний регістр 2 вводиться наступний символ потоку, що з'єднується з префіксом і знову здійснюється пошук співпадаючого підрядка. Така операція продовжується до тих пір, поки не буде відзначено, що знову створений рядок відсутній в кожній з таблиць. Одночасно з цим виконується підрахунок кількості підрядків в одиницю часу. Якщо частота знову утвореного рядка перевищує граничну, то вона заноситься в поточну кодову таблицю в позицію першого вільного рядка.

Кодова комбінація з вихідного регістра 5 видається на вихід пристрою. Символ із вхідного регістра 2 пересовується в префіксийний регістр і процедура кодування починається спочатку.

Після надходження на вхід компресора N символів, яких досить для надійної оцінки класу тексту, пошук порівнянного рядка в ОЗП таблиць, що кодують, здійснюється не шляхом послідовного їхнього перегляду, а в таблиці, що відповідає даному класу тексту.

Таким чином, значно скорочується час пошуку порівнянного рядка і тим самим підвищується швидкість компресора.

У процесі кодування символів, що надходять, оцінка класу тексту може змінюватися. Тим самим здійснюється зміна черговості перегляду ОЗП таблиць, що кодують, на предмет наявності співпадаючого підрядка.

Пристрій керування, блок оцінки тексту, вхідний і вихідний регістри реалізовані на базі сигнального мікропроцесора ГТІ - на базі кварцового генератора тактових імпульсів. Блок ОЗП — на базі мікросхем пам'яті.

Таким чином, застосування адаптивного пристрою стиску текстових повідомлень для систем передачі і збереження текстової інформації дозволяє збільшити ступінь стискання даних за рахунок введення блоку оцінки тексту та імовірнісних таблиць рядків змінної довжини для різних типів тексту. Це дозволяє скоротити час

передачі текстових даних по каналах зв'язку, а також зменшити обсяги пам'яті, що займаються текстами при збереженні на магнітних і оптичних носіях.

