

УДК 004.421

Бойко Сергій Васильович

Магістр кафедри програмного забезпечення автоматизованих систем, *ORCID: 0000-0002-4034-4612*
Черкаський державний технологічний університет, Черкаси

Трегубенко Ірина Борисівна

Кандидат технічних наук, доцент, викладач кафедри програмного забезпечення автоматизованих систем
Черкаський державний технологічний університет, Черкаси

АДАПТИВНИЙ ПРОЦЕС ВИБОРУ АЛГОРИТМІВ АРХІВАЦІЇ ГЕТЕРОГЕННИХ ДАНИХ НА ОСНОВІ МЕТОДУ ДЕЛЬФІ

Анотація Розглянуто метод Дельфі як ключовий у виборі способів архівації даних. На основі Дельфійського методу розроблено експертну анкету для визначення коефіцієнта ущільненості даних. Проведено дослідження ефективності різних методів архівації даних залежно від формату даних, які стискаються. Задля підвищення ефективності архівування гетерогенних даних великого обсягу запропоновано використовувати метод адаптивного стиснення даних, який інтелектуально обирає метод або комплекс методів ущільнення залежно від формату даних.

Ключові слова: метод Дельфі; архівація даних; методи архівації; гетерогенні дані; надлишковість; адаптивний метод

Вступ

На сьогодні актуальною залишається проблема архівування даних на електронних носіях. Сучасні технології дозволяють зберігати великі обсяги даних без архівування. При передачі декількох файлів витрачається більше часу, ніж одного архіву, який може містити декілька файлів одночасно.

Характерною особливістю сучасних сховищ даних є їх гетерогенність за форматами файлів, їх надлишковості та традиційне кодування. Ступінь надлишковості даних залежить від типу даних. Наприклад, для відеоданих ступінь надлишковості в декілька разів більша, ніж для графічних даних, а ступінь надлишковості графічних даних, у свою чергу, більша за ступінь надлишковості текстових даних. Системами кодування можуть бути звичайні мови спілкування, які є ні чим іншим, як системами кодування понять та ідей для висловлення думок.

Існує багато практичних методів стиснення даних, але всі вони базуються на трьох теоретичних процесах зменшення надлишковості даних: зміна вмісту даних, зміна структури даних, одночасна зміна структури та вмісту даних. Найвні методи та алгоритми стиснення даних переважно враховують статистику символів у повідомленні [1-3] або будуються на основі словника [4; 5]. Якщо при стисненні даних відбувається зміна їх вмісту, то метод стиснення є незворотнім. Методи стиснення з регульованими втратами інформації забезпечують значно більший ступінь стиснення. Також, якщо при стисненні даних відбувається тільки зміна структури даних, то метод стиснення є зворотнім.

Останні дослідження і публікації

Дослідник Jo Wainer в статті розглядає метод Дельфі для розробки контенту для національного обстеження сільських лікарів загальної практики, метою якого було встановити спільні досліджені результати сільських лікарів загальної практики з використанням методу Дельфі [6].

Дослідник John Nworie розглядає метод Дельфі для навчальних технологій, його основні переваги, які можуть принести користь для навчання. Також автор зазначає, що даний метод можна використовувати для інших галузей [7].

Дослідники Абашидзе Теймураз Лерьевич та Федотова Елена Леонидовна розглядають використання методу Дельфі для оцінки якості підготовки випускників ВНЗ. Метою даного методу є отримання узгодженої інформації високого ступеня достовірності в процесі анонімного обміну думками між учасниками групи експертів для прийняття рішення про підготовку випускників ВНЗ [8].

На основі викладеного можна припустити, що даний метод може бути використаний в інженерній сфері діяльності, зокрема для обробки великої кількості гетерогенних даних, щоб значно зменшити час витрачений на вибір методу архівації.

Мета статті

На основі викладеного можна сформулювати завдання дослідження, яке полягає в розробці методу архівації на основі методу Дельфі, що може бути використаний для уточнення ефективності методу архівації гетерогенних даних.

Нехай G є множина гетерогенних даних $G = \{G_1, G_2, \dots, G_n\}$ з підмножинами різних типів даних, наприклад G_1 – підмножина відеорозширень даних; G_2 – підмножина текстових розширень даних. Також M є множина методів архівації $M = \{M_1, M_2, \dots, M_n\}$. На основі методу Дельфі та результатів анкетування визначити, який з алгоритмів множини M може бути використаний як кращий спосіб для архівації для певного типу даних множини G .

Виклад основного матеріалу дослідження

Побудуємо спосіб на основі методу Дельфі. Даний метод має широке застосування на етапах формулювання проблеми і оцінки різних способів її вирішення [9].

Відомо, що використання колективних знань веде до можливості знаходження сильних рішень, проте в процесі обміну думками між учасниками може позначитися вплив авторитету колег і все зведеться до появи популярних або авторитетних відповідей.

Метод Дельфі дозволяє вирішити це діалектичне протиріччя. Для цього прямі дискусії експертів замінюються індивідуальними опитуваннями. Зібрані варіанти відповідей піддаються статистичній обробці. Отримані узагальнені відповіді передаються кожному експерту шляхом особистого спілкування, або за звичайною, або електронною поштою з проханням переглянути і уточнити свою думку, якщо він вважатиме необхідним. Ця процедура може повторюватися кілька разів.

Сам же метод Дельфі – інструмент, що дозволяє врахувати незалежну думку всіх учасників групи експертів з обговорюваного питання шляхом послідовного об'єднання ідей, висновків і пропозицій аби дійти згоди. Метод заснований на багаторазових анонімних групових інтерв'ю.

План роботи методу полягає у такому:

1. Сформулювати робочу групу для збирання та узагальнення думок експертів.
2. Сформулювати експертну групу з фахівців, які володіють питаннями з обговорюваної теми.
3. Підготувати анкету, вказавши в ній поставлену проблему, уточнюючі питання.
4. Формулювання повинні бути чіткими та однозначно трактовані, припускати однозначні відповіді.
5. Провести опитування експертів згідно з методикою, яка передбачає за необхідності повторення процедури. Отримані відповіді служать основою для формулювання запитань для наступного етапу.

6. Узагальнити експертні висновки і видати рекомендації з поставленої проблеми.

Також враховується, що метод Дельфі – це системний спосіб узагальнення оцінок експертів; метод найбільш застосовний, якщо до роботи залучаються експерти, компетентні не по всій проблемі, а по її різних складових. Графічне представлення методу Дельфі наведено на рис. 1 [10, 42 с.].

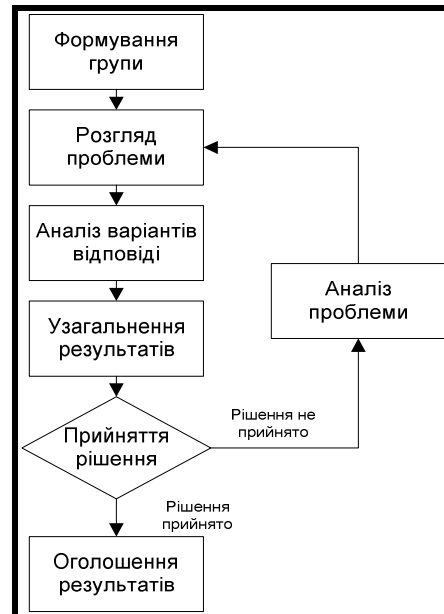


Рисунок 1 – Графічне представлення методу Дельфі

Для дослідження обрано такі методи архівації:

– LZMA (англ. Lempel-Ziv-Markov chain-Algorithm) – метод стиснення даних, що розробляється з 2001 року. Використовується в архіваторі 7-Zip для створення стиснутих архівів у форматі 7z. Метод заснований на схемі стиснення даних за словником, подібної до використаної в LZ77, і забезпечує високий коефіцієнт стиснення, а також дозволяє використовувати словники різного розміру [11].

– PPM (англ. Prediction by Partial Matching – пророкування щодо часткового збігу) – адаптивний статистичний метод стиснення даних без втрат, заснований на контекстному моделюванні та прогнозі. Модель PPM використовує контекст – безліч символів в стилі потоці, що передують даному, щоб передбачати значення символу на основі статистичних даних. Сама модель PPM лише пророкує значення символу, безпосереднє стиснення здійснюється методами ентропійного кодування, як наприклад, метод Хаффмана, арифметичне кодування [12].

– bzip2 – метод реалізації, перетворення Барроуза-Уйлера. bzip2 використовує Перетворення Барроуза-Уйлера для перетворення послідовностей багаторазово чергуються символи в рядках

однакових символів, потім застосовує перетворення MTF, і в кінці кодування Хаффмана. Блоки в bzip2 мають однаковий розмір в стислому потоці. Розмір блоку можна вибрати за допомогою аргументу командного рядка, і він позначається в стислому тексті довільно обраної послідовністю бітів поданням числа π . Попередник bzip2, bzip, використовував арифметичне кодування після сортування блоку. Від нього відмовилися через патентні обмеження.

– Deflate – це метод стиснення без втрат, який використовує комбінацію методу LZ77 і методу Гоффмана. Спочатку він був описаний Філом Кацом для 2-ї версії своєї утиліти для створення архівів PKZIP, який згодом був визначений в RFC 1951 [13].

Сформулюємо анкету у вигляді табл. 1 для опитування експертів на основі множини G та M. Дана експертна таблиця містить такі дані: m_1, m_2, \dots, m_n – алгоритми ущільнення даних; G_1, G_2, \dots, G_n – дані різних форматів; Σ – загальна сума балів виставлена кожним експертом для конкретного алгоритму ущільнення даних.

Кожний з експертів проводить дослідження архівації гетерогенних даних, після чого він заносить отримані результати до таблиці та виставляє бали кожному методу архівації по шкалі від 1 до 10.

Проведено дослідження взаємозв'язку різних методів стиснення (LZMA, PPMd, Deflate, Bzip2) файлів різного розміру та формату, які мають широке застосування. Початкові розміри та тип даних файлів наведені в табл. 2.

Як видно з графіків для файлів JPG (рис. 2, а) методи LZMA, PPMd, Deflate, Bzip2 стискають файли практично з однаковим коефіцієнтом. Для формату PNG (рис. 2, б) – методи LZMA, PPMd,

Deflate показали однаковий розмір архіву, що й самі файли, а архіви методу Bzip2 значно перевищували початкові розміри. Це говорить, що метод Bzip2 неефективний для даного формату. Для відео файлів формату AVI (рис. 2, в) методи показали, що метод LZMA значно ефективніший, ніж інші. Але при проведенні експерименту методи PPMd, Deflate, Bzip2 показали значно менший час стиснення. Для текстового формату TXT (рис. 2, г) метод PPMd показав кращі результати, інші ж показали практично однаковий результат. Для DOC (рис. 2, д) формату кращим є метод LZMA. Для аудіофайлів у форматі MP3 (рис. 2, е) кращими є LZMA та Bzip2.

Проведемо дослідження коефіцієнта залежності ущільнення для одного файлу розміром 1,8 МБ для методів LZMA, PPMd, Deflate, Bzip2, графік якого наведено на рис. 3.

Проведене дослідження залежності ущільнення для одного файлу розміром 1,8 МБ показало, що: методи LZMA та Deflate для файлу формату IPG мають коефіцієнт ущільнення рівний 1; для PNG формату результати ідентичні; формат AVI показав, що кращими є методи LZMA, Deflate Bzip2, а PPMd показав коефіцієнт ущільнення близький до 1; TXT формат кращим є метод LZMA з коефіцієнтом ущільнення 0,02; DOC формат – LZMA з коефіцієнт ущільнення 0,64; для MP3 формату малих розмірів всі методи показали коефіцієнт ущільнення 1. Таким чином, існує залежність величини розміру файлу, чим більший розмір тим краще відбувається архівація файлів.

Результати експертної оцінки вищеописаних дослідів наведено у табл. 3. Експерт виставив оцінку кожному з методів ущільнення.

На основі експертних оцінок було побудовано адаптивний метод вибору алгоритмів архівації (рис. 4).

Таблиця 1 – Загальна анкета опитування експертів

Метод множини M \ Формати даних множини G	Формати підмножини G ₁				Формати підмножини G ₂				...				Формати підмножини G _n			
	DOC	TXT	...	Σ	MP3	MIDI	...	Σ	JPG	PNG	...	Σ
m_1																
m_2																
...																
m_n																

Таблиця 2 – Початкові розміри файлів

Формат	JPG, кб	PNG, кб	AVI, мб	TXT, кб	DOC, кб	MP3, мб
1	358	8	202,6	4	439	9,43
2	175	9	178,8	11	42	13,3
3	60	6	286,4	17	38	7,89
4	5003	22	240,8	136	1858	6,95

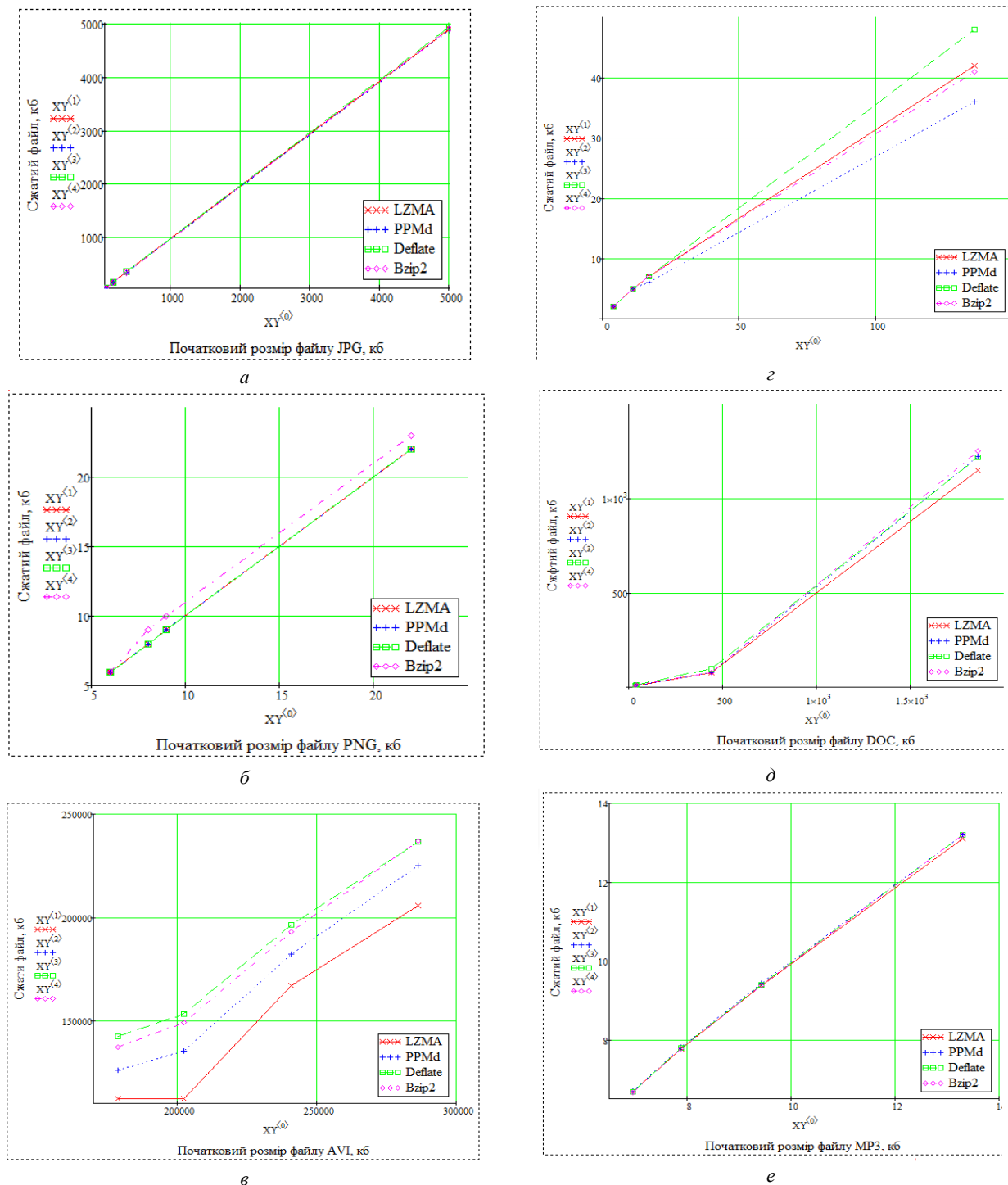


Рисунок 2 – Графік залежності ущільнення гетерогенних даних: а – стиснення JPG файлів; б – стиснення PNG файлів; в – стиснення AVI файлів; г – стиснення TXT файлів; д – стиснення DOC файлів; е – стиснення MP3 файлів

Таблиця 3 – Експертна оцінка методів ущільнення

Метод множини М \Формати даних множини G	Графічні формати			Текстові формати			Аудіоформати		Відеоформати	
	IPG	PNG	Σ	DOC	TXT	Σ	AVI	Σ	MP3	Σ
LZMA	6.5	6.5	13	8	10	18	8	8	8.5	8.5
PPMd	8	7.5	15.5	7	6.5	13.5	6	6	8	8
Deflate	6.5	6.5	13	6	3	9	4.5	4.5	8	8
Bzip2	8	7	15	5.5	6.5	12	4.5	4.5	8	8

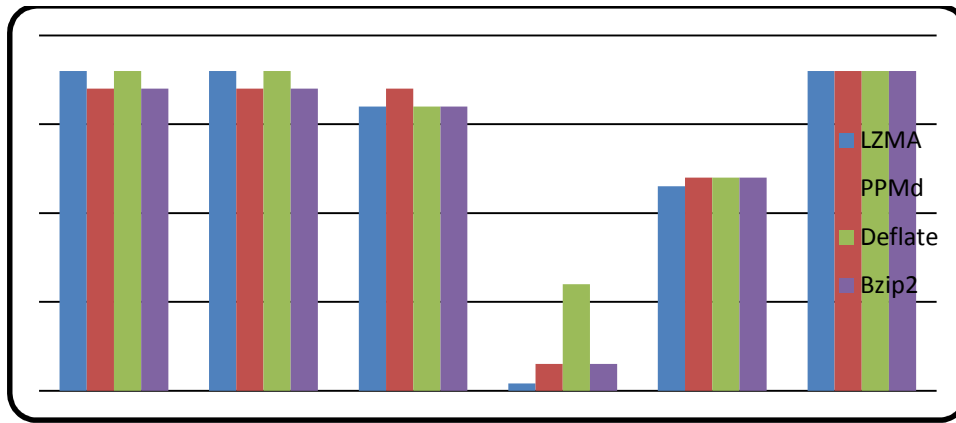


Рисунок 3 – Коефіцієнт залежності ущільнення для одного файлу розміром 1,8 МБ

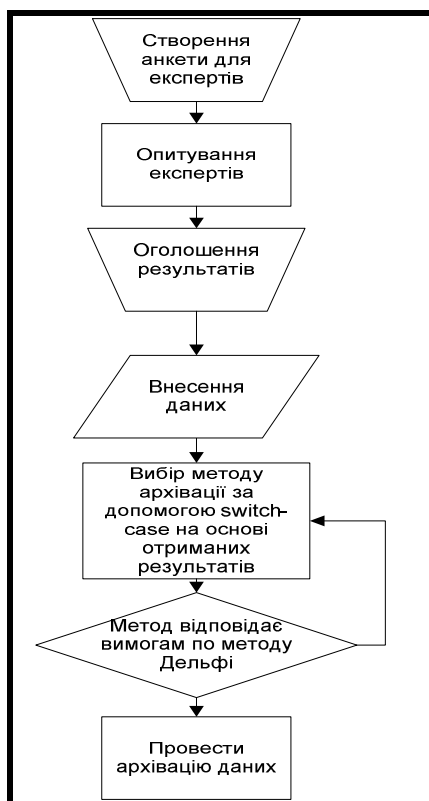


Рисунок 4 – Адаптивний метод вибору алгоритмів архівації

На початку адаптивного методу вибору алгоритмів архівації створюють анкету для проведення опитування. Після створення анкети до опитування залучаються експерти, які без впливу інших заповнюють анкету. Оголошують результати анкетування та порівнюють відповіді кожного з експертів. На основі даних, що були отримані в процесі анкетування, обирається метод для певних типів даних, який є найбільш ефективний. Якщо метод визнано більшістю експертів ефективним, то він обирається для архівації поточного типу даних.

У подальших дослідженнях планується використати нормалізацію для зведення кількох критеріїв архівації даних до єдиного формату.

Висновки та подальші дослідження

Розглянуто основний принцип роботи методу дельфійського оракула як ключовий у виборі способів архівації даних. На основі дельфійського методу розроблено експертну анкету, для визначення коефіцієнта ущільненості даних. Проведено дослідження ефективності різних методів архівації даних залежно від формату даних, які стискаються. Розглянуто базові методи зменшення надлишковості даних, а саме: зміна вмісту даних, зміна структури даних, одночасна зміна структури та вмісту. Задля підвищення ефективності архівування гетерогенних даних великого обсягу запропоновано використовувати метод адаптивного стиснення даних, який інтелектуально обирає метод або комплекс методів ущільнення залежно від формату даних.

Список літератури

1. Huffman D.A. A method for the construction of minimum redundancy codes [Текст] / Huffman D.A. // Proceedings of the Institute of Electrical and Radio Engineers. - 1952.
2. Cormack G.V. Data compression using dynamic Markov modeling [Текст] / Cormack G.V., Horspool R.N. // Comput. J. - 1987.
3. Storer J. A. Data Compression: Methods and Theory [Текст] / Storer J. A. //Computes Science Press. - 1988.

4. Ziv J. Compression of individual sequences via variable-rate coding [Текст] / Ziv J., Lempel A. // IEEE Trans. Inf. Theory. - 1978.
5. Welch T.A. A Technique for High Performance Data Compression [Текст] / Welch T.A. // Computer/ - 1984
6. Jo Wainer, Master of Arts «Cooperative Research with Rural Doctors» [Текст] / - 2004.
7. John Nworie Kent. – State University «Using the Delphi Technique in Educational Technology Research». [Текст] / TehcTrends - 2011
8. Абашидзе Т.Л. та Федотова Е.Л. «Оценка качества подготовки выпускников вуза на основе метода Дельфы». [Текст] / Институт Государственного управления, права и инновационных технологий (ИГУППИТ). Интернет-журнал «НАУКОВЕДЕНИЕ» №1 2013
9. Сайт «Управление и Оптимизация Производственного Предприятия» [Электронный ресурс]. – Режим доступа : <http://productm.ru/methods-of-searching-for-new-ideas/methods-of-control/the-delphi-technique/>
10. Сборник методов поиска новых идей и решений управления качеством [Текст] /сост. В.В. Ефимов. – Ульяновск: УлГТУ, 2011. – 194 с.
11. Сайт «Википедия – LZMA» [Электронный ресурс]. – Режим доступа : <http://ru.wikipedia.org/wiki/LZMA>
12. J.G. Cleary, W.J. Teahan, and I.H. Witten, Unbounded length contexts for PPM, [Текст] / In J.A. Storer and M. Cohn, editors, Proceedings DCC-95, IEEE Computer Society Press, 1995.
13. Сайт «Википедия – Deflate» [Электронный ресурс]. – Режим доступа : <http://ru.wikipedia.org/wiki/Deflate>

Стаття надійшла до редколегії 23.12.2014

Рецензент: д-р техн. наук, проф. С.М. Первунінський, Черкаський державний технологічний університет, Черкаси.

Бойко Сергей Васильевич

Магистр кафедры программного обеспечения автоматизированных систем, ORCID: 0000-0002-4034-4612

Черкасский государственный технологический университет, Черкассы

Трегубенко Ирина Борисовна

Кандидат технических наук, доцент, преподаватель кафедры программного обеспечения автоматизированных систем

Черкасский государственный технологический университет, Черкассы

АДАПТИВНЫЙ ПРОЦЕСС ВЫБОРА АЛГОРИТМОВ АРХИВАЦИИ ГЕТЕРОГЕННЫХ ДАННЫХ НА ОСНОВЕ МЕТОДА ДЕЛЬФИ

Аннотация. Рассмотрен метод Дельфи как ключевой в выборе способов архивации данных. На основе Дельфийского метода разработана экспертная анкета для определения коэффициента уплотнённости данных. Проведено исследование эффективности различных методов архивации данных в зависимости от формата данных, которые сжимаются. Для повышения эффективности архивирования гетерогенных данных большого объема предложено использовать метод адаптивного сжатия данных, который интеллектуально выбирает метод или комплекс методов уплотнения в зависимости от формата данных.

Ключевые слова: метод Дельфи; архивация данных; методы архивации; гетерогенные данные; избыточность; адаптивный метод

Sergey Boyko

Master of the Department of automated software systems, ORCID: 0000-0002-4034-4612

Cherkasy State Technological University, Cherkasy

Tregubenko Irina

Kandedat of technical sciences, lecturer in software systems

Cherkasy State Technological University, Cherkasy

THE ADAPTIVE PROCESS SELECTION ALGORITHMS OF ARCHIVING HETEROGENEOUS DATA BASED ON THE DELPHI METHOD

Abstract. The Delphi method is investigated as key in selecting methods of data archiving. On the basis of the Delphi method it is developed expertise questionnaire for determining the coefficient of data compression. The investigation the effectiveness of different methods of archiving data was provided depending on format of the data which is compressed. It was conducted a study of the relationship of different compression methods (LZMA, PPMd, Deflate, Bzip2) of files with various size and formats that are widely used. Also it was conducted a study of dependency of heterogeneous data for one file size of 1.8 MB for methods LZMA, PPMd, Deflate, Bzip2. We have considered the basic methods of reducing redundancy of data, such as content changing data, changing data structure, the simultaneous change in the structure or content. In order to improve efficiency archiving large volume of heterogeneous data it is suggested to use adaptive method of data compression which intelligently selects the method or set of methods for compression depending on the format of data.

Key words: Delphi method; archiving; methods of archiving; heterogeneous data; redundancy; adaptive method

References

1. Huffman, D.A. (1952). *A method for the construction of minimum redundancy*. *Proceedings of the Institute of Electrical and Radio Engineers*. - 1952.
 2. Cormack, G.V. & Horspool, R.N. (1987). *Data compression using dynamic Markov modeling*: *Comput. J.*
 3. Storer, J. A. *Data Compression: Methods and Theory*: *Computes Science Press*.
 4. Ziv, J. & Lempel, A. (1978). *Compression of individual sequences via variable-rate coding*: *IEEE Trans. Inf. Theory*.
 5. Welch, T. A. (1984). *A Technique for High Performance Data Compression* : *Computer*.
 6. Jo Wainer (2004). *Master of Arts «Cooperative Research with Rural Doctors»*.
 7. John Nworie Kent (2011). *State University «Using the Delphi Technique in Educational Technology Research»*: *TehcTrends*.
 8. Abashidze, T.L. & Fedotova, E.L. (2013.) *Assessment of the quality of graduates based on the Delphi method*. *Institute of Public Administration, Law and innovative technologies (IGUPIT): Internet magazine "Science of Science» №1*
 9. *Delphi method [electronic source]*. – Access mode : \www/ URL: <http://productm.ru/methods-of-searching-for-new-ideas/methods-of-control/the-delphi-technique/> – 22.01.2014
 10. Efimov, V.V (2011) *Collection of methods to search for new ideas and solutions* *Quality Management*. Ulyanovsk, Ulyanovsk State Technical University, 194.
 11. LZMA – *Wikipedia [electronic source]*. – Access mode : \www/ URL: <http://ru.wikipedia.org/wiki/LZMA> – 17.04.2014
 12. Cleary, J.G., Teahan, W.J & Witten, I.H. *Unbounded length contexts for PPM*. *Proceedings DCC-95, IEEE Computer Society Press, 1995*.
- Deflate – Wikipedia [electronic source]*. – Access mode : \www/ URL: <http://ru.wikipedia.org/wiki/Deflate> 18.05.2014
-

Посилання на публікацію

- APA Sergey Boyko, & Tregubenko Irina (2015). *The adaptive process selection algorithms of archiving heterogeneous data based on the delphi method*. *Management of Development of Complex Systems, Issue 21, P. 163 – 169 [in Ukrainian]*.
- ГОСТ Бойко С.В. *Адаптивний процес вибору алгоритмів архівації гетерогенних даних на основі методу Дельфі [Текст]* / С.В. Бойко, І.Б. Трегубенко // *Управління розвитком складних систем*. – 2015. - № 21. – С. 163 - 169.