

І.С. СКАРГА-БАНДУРОВА, д-р техн. наук, доц., проф., СНУ
ім. В. Даля, Северодонецьк,
М.О. ГРУШКА, магістр, СНУ ім. В. Даля, Северодонецьк,
Л.В. БАРБАРУК, ст. викл., СНУ ім. В. Даля, Северодонецьк

ПІДХОДИ ДО ЕФЕКТИВНОГО СПРОЩЕННЯ І ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ НАБОРІВ ДАНИХ

Розглянуто проблеми візуалізації послідовностей великих наборів даних. Задачу пошуку оптимального відображення даних сформульовано у вигляді задачі вибору алгоритму спрощення простих полігональних ланцюгів з мінімальною кількістю помилок. Виконано аналіз ефективності алгоритмів Рамера-Дугласа-Пекера, Реумана-Віткама, Опхейма, Ланга, алгоритму виключення за висотою трикутника та алгоритму радіального виключення точок при роботі з великими даними. Наводяться результати порівняння та оцінки ефективності їх використання для аналізу реальних даних, що збираються в системах локомотивної безпеки. Іл.: 2. Табл.: 1. Бібліогр.: 15 назв.

Ключові слова: візуалізація; великі набори даних; алгоритм спрощення; полігональний ланцюг; система локомотивної безпеки.

Вступ. Стан будь-якого технологічного об'єкта, що оснащений датчиками, з яких постійно знімається інформація, може бути описаний послідовностями їх значень, які, у свою чергу, відображаються точками, лініями, кривими або поверхнями. Графічне відображення – дуже простий та дієвий спосіб представлення даних, що дозволяє полегшити їх сприйняття та надає достатньо інформації для аналізу і прийняття рішень [1]. Візуалізація даних має важливе значення для вивчення та розуміння структур та моделей даних, їх залежностей та виявлення незвичних спостережень. Необхідність бачити дані неможливо перебільшити. Однак наявність великої кількості доступних даних суттєво впливає на вибір засобів і технології їх відображення [2]. Використання традиційних методів візуалізації для послідовностей великих наборів різноманітних даних приводить до втрати інформативності й може створити на екрані одну велику пляму, що складається з множини точок, які представляють кожен рядок даних. Зрозуміло, що таке відображення унеможливує подальший аналіз та інтерпретацію даних. Інша проблема полягає в тому, що дані надходять з різних джерел. Хоча вони можуть охоплювати аналогічний діапазон дат, різні набори даних можуть мати різні інтервали між записаними значеннями. Нарешті, поряд з природою даних, їх обсяги та вимоги до швидкості обробки постійно зростають, і це стає класичною проблемою Big Data. Для того, щоб успішно аналізувати такі типи даних, їх необхідно спростити для графічного відображення, що вимагає пошуку

і використання нових підходів та принципів обробки. Це дасть змогу дослідити менший розмірний простір і виявити цікаві структури та незвичні спостереження.

У цій статті головний акцент зроблено на розгляді проблем візуалізації даних у вигляді полігональних ланцюгів та результатах роботи з великими наборами даних.

Огляд проблем спрощення полігональних ланцюгів. Полігональний, або багатокутний ланцюг P в евклідовій площині визначається як упорядкований список вершин $\langle p_1, p_2, \dots, p_n \rangle$, при якому будь-які дві послідовні вершини з'єднуються лінійним сегментом.

У загальному виді проблема спрощення полігональних ланцюгів [3] формулюється наступним чином. Нехай дана полігональна крива $\langle p_1, \dots, p_n \rangle$ і крива $P' = \langle p_{i_1}, \dots, p_{i_k} \rangle$ з $1 = i_1 < \dots < i_k = n$, що спрощує криву P , де $P(i, j)$ це відрізок шляху з p_i до p_j . Для пари індексів $1 \leq i \leq j \leq n$, $\delta_F(p_i p_j, P)$ позначає помилку спрощення сегмента $p_i p_j$ відносно $P(i, j)$. Тоді,

$$\delta_F(P', P) = \max_{1 \leq j \leq k} \delta_F(p_i p_{j+1}, P), \quad (1)$$

де P' є ε -спрощенням P , якщо $\delta_F(P', P) \leq \varepsilon$.

Проблема оптимізації відображення даних може бути сформульована у вигляді задачі вибору алгоритму спрощення простих полігональних ланцюгів, а саме: Для наданої полігональної кривої P , необхідно знайти спрощення P' з мінімальною кількістю вершин. В літературі [3, 4] ця проблема розглядається головним чином як задача ε -або $\#$ -мінімізації.

Проблема ε -мінімізації. Для наданого полігонального ланцюга P і цілого числа $k \leq n$, знайти серед усіх апроксимацій P з не більше ніж k вершинами апроксимацію з мінімальною погрішністю.

Проблема $\#$ -мінімізації. Для наданого полігонального ланцюга P і дійсного числа $\varepsilon \geq 0$, знайти серед усіх ε -апроксимацій P ε -апроксимацію з мінімальною кількістю вершин.

Надалі в роботі розглянуто проблему ε -мінімізації.

Спрощення полігональних ланцюгів. Суть алгоритмів спрощення полігональних ланцюгів полягає у зменшенні кількості точок шляхом видалення тривіальних точок, але без порушення істотної форми вихідної лінії. Перше застосування з 1960 р. алгоритми спрощення полігональних ланцюгів набули у картографії [5, 6] з подальшим поширенням на географічні інформаційні системи, комп'ютерну графіку, візуалізацію часових рядів [7, 8], проблеми

стиснення даних, тощо. Нині більшість навігаційних додатків використовують методи і алгоритми спрощення ліній для зменшення об'єму карт та поліпшення швидкості виконання таких операцій як масштабування та прокрутка. Розроблені й протестовані різноманітні алгоритми спрощення, наприклад, алгоритм Дугласа-Пекера [9], що спрощує лінію на основі вертикального інтервалу; алгоритм Реумана-Віткама [10], що спрощує лінію на основі функцій кутового повороту [11]; алгоритм послідовного спрощення лінії [12], що спрощує лінію на основі області та інші, однак, як зазначається в [13], питання вибору алгоритму та оцінка ефективності залишається відкритою.

Ключовим моментом у використанні алгоритмів спрощення полігональних ланцюгів є вибір порогу, що визначає силу спрощення. Саме від цього параметру залежить кількість точок та вид отриманої лінії.

У різних алгоритмах поріг спрощення визначається по-різному. Найпоширеніші варіації порогу спрощення:

- Відстань від точки до лінії, утвореної сусідніми або крайніми точками.
- Радіус, в межах якого видаляються усі точки
- Номер точки: видаляється кожна N точка.
- Площа, в межах якої залишається лише зазначена кількість точок.

Чим більшим є поріг спрощення тим менше рівень деталізації отриманої лінії та тим менше точок складають отриману лінію. На рис. 1 показано полігональний ланцюг, що складається з 13 вершин: $\{p_1, p_2, \dots, p_{13}\}$ та приклад спрощення при $k = 3$, що містить вершини: $\{p_1, p_6, p_9, p_{13}\}$.

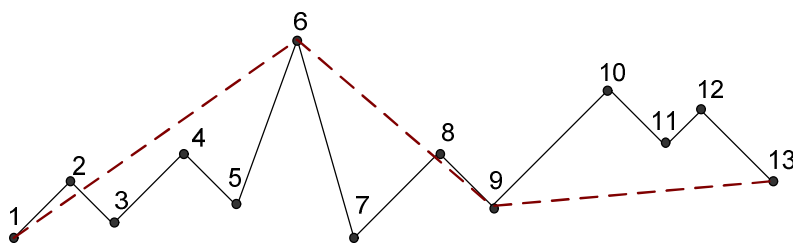


Рис. 1. Приклад полігонального ланцюга та його спрощення

Різноманіття варіацій порогу спрощення призводить до необхідності детального аналізу характеру вхідних даних. Задача математичного обмеження алгоритмів спрощення полігональних ланцюгів є найбільш складною. Недоопрацьоване застосування процесу спрощення ліній, імовірно, не зможе значно зменшити розмір ланцюга.

Натомість надмірна спрощена лінія може призвести до спотворення його відображення. Погано підібраний поріг спрощення може призвести до втрати значущих даних та нашттовхнути людину, що аналізує візуалізовані дані, до хибних висновків. Таким чином, обов'язковими кроками є розрахунок порогу спрощення і вибір показників для визначення якості.

Метою роботи є обґрунтування вибору методів та алгоритмів спрощення полігональних ланцюгів, які лежать в основі системи візуалізації великих наборів технологічних даних. Найкращий алгоритм має забезпечити візуальне відображення даних у різних масштабах зі збереженням їх точності.

Розрахунок сили спрощення полігонального ланцюга. Рівень деталізації спрощеної лінії тісно пов'язаний з порогом спрощення відповідного алгоритму. Поріг асоціюється із силою спрощення. Для вимірювання сили спрощення зазвичай використовують алгоритм знаходження середньої відстані зсуву.

Для отримання знаходження сили спрощення у класичному випадку використовується формула [14]

$$D_{simpl} = SR_p \cdot N, \quad (2)$$

де SR_p – просторова розподільна здатність пікселя; N – мінімальна кількість пікселів, що зможе розрізнити людське око на певній відстані від монітора.

Просторова розподільна здатність являє собою число незалежних пікселів значень на дюйм. Цей показник головним чином впливає на здатність ока людини розрізнити деталі на лінії й обчислюється наступним чином

$$SR_p = \frac{D_p}{PPI \cdot Scale}, \quad (3)$$

де D_p – фізичний розмір пікселя $D_p = \frac{1}{PPI}$, $Scale$ – масштаб PPI – кількість пікселів на дюйм

$$PPI = \frac{\sqrt{H_r^2 + V_r^2}}{D_s}, \quad (4)$$

де D_s – розмір дисплея, $H_r \times V_r$ – розподільна здатність дисплея.

Для реалізації можливості зміни пропорцій графіків відносно один одного до формули (2) додамо співвідношення поточної висоти

компоненти відображення графіків (H) до висоти графіка, для якого виконується обчислення (H_{plot})

$$D_{simpl} = SR_p \cdot N \cdot \frac{H}{H_{plot}}. \quad (5)$$

Оцінка якості спрощення полігональних ланцюгів. Зрозуміло, що спрощення полігонального ланцюга змінює його форму. Чим вище ступінь спрощення, тим більше спрощений ланцюг відрізняється від оригінального. Для оцінки застосовності алгоритмів спрощення до мети роботи необхідно розрахувати позиційні помилки. Спосіб вимірювання похибки спрощення полягає в знаходженні відстаней між вихідними точками, що були спрощені, та спрощеним ланцюгом.

Найбільш часто використовуваними показниками для визначення якості спрощення є [15]:

– максимальне відхилення (максимальна відстань від оригінальної точки, що була спрощена, до спрощеної лінії)

$$\max_{i=1, N} (p_i - p_i^m); \quad (6)$$

– мінімальне відхилення (максимальна відстань від оригінальної точки, що була спрощена, до спрощеної лінії)

$$\min_{i=1, N} (p_i - p_i^m); \quad (7)$$

– середньоквадратичне відхилення

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (p_i - p_i^m)^2}{N}}; \quad (8)$$

– сума відхилень (сума відстаней від оригінальних точок, що були спрощені, до спрощеної лінії)

$$S = \sum_{i=1}^N (p_i - p_i^m). \quad (9)$$

Дослідження алгоритмів спрощення на практичному прикладі. Описані вище положення використовуються для реалізації завдання з формування та візуалізації послідовностей великих наборів даних систем локомотивної безпеки, призначених для забезпечення безпеки руху в поїзній і маневровій роботі.

В системах локомотивної безпеки, що організовані за допомогою сучасних високотехнологічних пристроїв, використовуються датчики й устаткування для автостопного і службового гальмування; обладнання

для приймання та обробки вхідної інформації та формування керуючих команд; операторське обладнання; сервісне та діагностичне обладнання; блоки живлення. Багато вимірювань, такі як, наприклад, значення автоматичної локомотивної сигналізації безперервної дії (АЛСБ), збираються з частотою 50, 100 або 150 Гц. Для аналізу архівних даних, діагностики і визначення причин несправностей та аварійних ситуацій необхідно мати можливість відображення усіх параметрів, в залежності від типу, на чотирьох графіках: (1) дискретні параметри, (2) аналогові параметри, (3) параметри швидкості та обмеження, (4) сигнал світлофору та об'єкти з електронної карти. По осі x має бути відображені дата і час з точністю до секунди та значення лінійної координати у кілометр-пикетах. По осі y – значення відповідного параметра у заданій величині. Необхідно також надати можливість перегляду даних за максимально можливою за тривалістю поїздки, що складає 3 доби.

Для пошуку найкращого алгоритму спрощення та забезпечення якісної візуалізації послідовностей наявних даних було обрано 6 алгоритмів: Дугласа-Рамера-Пекера (ДРП), Реумана-Віткама (Р-В), Опхейма, Ланга, алгоритм виключення за висотою трикутника (ВТ) та алгоритм радіального виключення точок (Рад). Порівняння ефективності алгоритмів спрощення складалося з двох етапів: візуального аналізу та розрахунку похибок. Візуальний аналіз використовувався для оцінки відтворюваності, тобто наскільки спрощений ланцюг передає поведінку (форму) початкового ланцюга. Математичний аналіз дозволив обрати з поміж алгоритмів, що однаково добре пройшли візуальний аналіз, алгоритм з найменшим значенням позиційної похибки.

Для тестування алгоритмів спрощення було обрано значення середньо-квадратичного відхилення сигналу АЛСБ. Результати спрощення надано на рис. 2.

Для порівняння обраних алгоритмів спрощення полігональних ланцюгів виконано обчислення позиційних похибок. Значення похибок обчислювались по формулам (6) – (9) окремо для кожної точки. Результати обчислення похибок спрощення ланцюгів різними алгоритмами надані у табл.

Координати точок на рис. 2: по осі X – дата і час в мілісекундах з початку епохи (1970), по осі Y – значення середньоквадратичного відхилення сигналу АЛСБ. Загальний проміжок – 10 секунд. На верхньому графіку кожного рисунку позначений оригінальний ланцюг. На нижньому графіку кожного рисунку позначено спрощений ланцюг. Початкова кількість точок 15425.

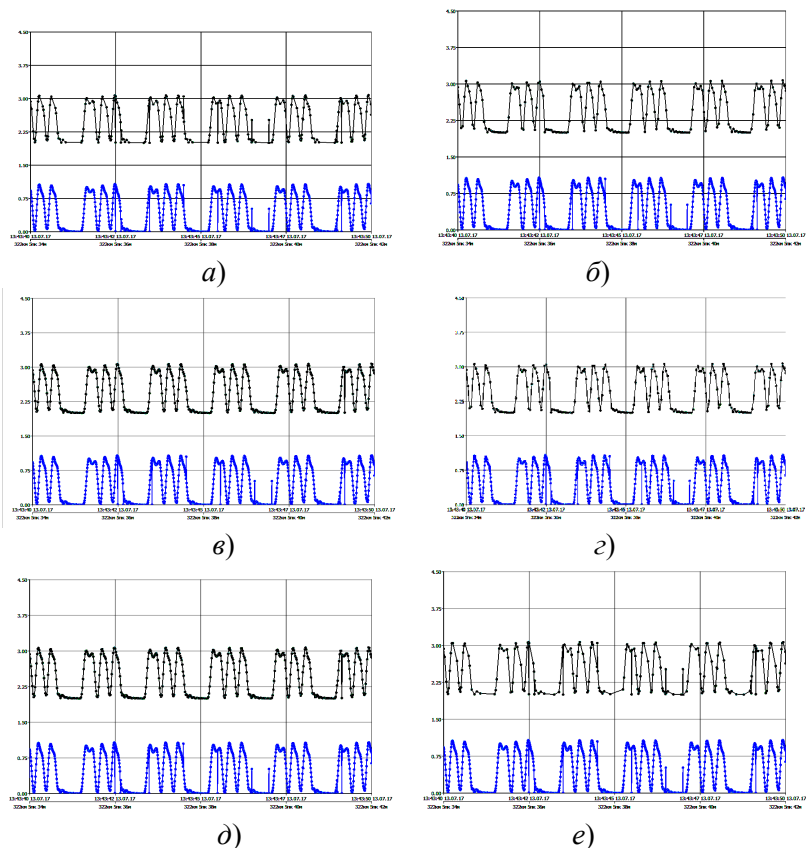


Рис. 2. Результат спрощення алгоритмами (а) Дугласа-Рамера-Пекера, (б) радіального виключення точок, (в) виключення за висотою трикутника, (г) Ланга, (д) Опхейма, (е) Реумана-Віткама

Таблиця

Результати обчислення похибок

Алгоритм	ДРП	Рад	ВТ	Ланг	Опхейм	Р-В
Показник						
Максимальна похибка	0.022	1.062	1.064	1.062	1.064	0.127
Середня похибка	0.005	0.023	0.008	0.023	0.008	0.018
Середньо-квдратична похибка	0.005	0.069	0.064	0.069	0.066	0.019
Сума похибок	82.08	352.09	129.49	356.01	128.89	285.63
Кількість точок після спрощення	3856	3090	7713	3086	7726	2867

В результаті роботи, враховуючи вимоги забезпечення найменшої похибки, найбільш оптимальним алгоритмом для спрощення точок сигналу АЛСБ є алгоритм Дугласа-Рамера-Пекера.

Висновки. Обґрунтовано вибір алгоритму спрощення візуального відображення великих наборів даних для використання в діагностичному програмному комплексі системи локомотивної безпеки. Вибір оптимального алгоритму проведено з урахуванням розрахунків порогу спрощення і показників якості спрощення. Розрахунки проводилися з урахуванням характеристик дисплею. У рамках поставленої задачі, баланс між вимогами до деталізації вихідного зображення та отриманням мінімальної кількості точок був зсунутий у бік зменшення сили спрощення та збільшення деталізації.

Результати роботи дозволили вирішити частину багатоетапного завдання з розробки підсистеми адаптивної візуалізації архівних даних систем локомотивної безпеки. Наступним кроком реалізації проекту є вирішення проблеми пошуку і виділення послідовностей даних, що найменше схожі на інші, або дисонансів часових рядів, рішення задачі виявлення цікавих структур та незвичних спостережень в наборах великих даних.

Список літератури: 1. *Gorodov E.Yu.* Analytical Review of Data Visualization Methods in Application to Big Data / *E.Yu. Gorodov, V.V. Gubarev* // Journal of Electrical and Computing Engineering. – 2013. – Vol. 2013. – P. 201-207. 2. *Hyndman R.J.* Visualization of big time series data [Електронний ресурс] / *R.J. Hyndman*. – Режим доступу: www.robjhyndman.com/seminars/big-time-series/ – 21.10.2017 р. 3. *Agarwal P.K.* Efficient Algorithms for Approximating Polygonal Chains / *P.K. Agarwal, K.R. Varadarajan* // Discrete Computing Geometry. – 2000. – Vol. 23. – P. 273-291. 4. *Buzer L.* Optimal simplification of polygonal chains for subpixel-accurate rendering / *L. Buzer* // Computing Geometry Theory Application. – 2009. – Vol. 42 (1). – P. 45-59. 5. *Buttenfield B.* Treatment of the cartographic line / *B. Buttenfield* // Cartographica. – 1985. – Vol. 22. – P. 1-26. 6. *McMaster R.B.* Automated line generalization / *R.B. McMaster* // Cartographica. – 1987. – Vol. 24 (2). – P. 74-111. 7. *Kumar N.* Time-series bitmaps: a practical visualization tool for working with large time series databases / *N. Kumar, N. Lolla, E. Keogh et al.* // Proceedings of the 5th SIAM International Conference on Data Mining, Newport Beach, California, April 2005. – P. 531-535. 8. *Lin J.* Visually mining and monitoring massive time series / *J. Lin, E. Keogh, S. Lonardi, et al.* // Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2005. – P. 460-469. 9. *Douglas D.H.* Algorithms for the reduction of the number of points required to represent a digitized line or its caricature / *D.H. Douglas, T.K. Peucker* // The Canadian Cartographer. – 1973. – Vol. 10. – №. 2. – P. 112-122. 10. *Reumann K.* Optimizing curve segmentation in computer graphics / *K. Reumann, A.P.M. Witkam* // Proceedings of the International Computing Symposium, Amsterdam, North-Holland Publishing Company, 1974. – P. 467-472. 11. *Rangayyan R.M.* Polygonal approximation of contours based on the turning angle function / *R.M. Rangayyan, D. Guliato, J.D. de Carvalho, S. Santiago* // Journal of Electronic Imaging. – 2008. – Vol. 17 (2). – P. 16-23. 12. *Qingsheng G.* A progressive line simplification algorithm / *G. Qingsheng, C. Brandenberger, L. Hurni* // Geo-Spatial Information Science. – 2002. – Vol. 5 (3). – P. 41-45. 13. *Shi W.* Performance Evaluation of Line Simplification Algorithms

for Vector Generalization / *W. Shi, CK. Cheung* // *The Cartographic Journal*. – 2006. – Vol. 43. – №. 1. – P. 27-44. **14.** *Song J.* A Novel Evaluation Approach for Line Simplification Algorithms towards Vector Map Visualization / *J. Song, R. Miao* // *International Journal of Geo-Informatics*. – 2016. – Vol. 5. – P. 223. **15.** *Ekdemir S.* Efficient Implementation of Polyline Simplification for Large Datasets and Usability Evaluation [Електронний ресурс] / *S. Ekdemir* – Режим доступу: [www. URL: www.diva-portal.org/smash/ get/diva2:444686/FULLTEXT01.pdf](http://www.diva-portal.org/smash/get/diva2:444686/FULLTEXT01.pdf) – 21.10.2017 р.

References:

1. Gorodov, E.Yu. and Gubarev, V.V. (2013), "Analytical Review of Data Visualization Methods in Application to Big Data", *Journal of Electrical and Computing Engineering*, Vol. 2013, pp. 201-207.
2. Hyndman, R.J. (2015), "Visualization of big time series data, available at: URL <https://robjhyndman.com/seminars/big-time-series/> (accessed 21 October 2017)
3. Agarwal, P. K. and Varadarajan, K. R. (2000), "Efficient Algorithms for Approximating Polygonal Chains", *Discrete Computing Geometry*, Vol. 23, pp. 273-291.
4. Buzer, L. (2009), "Optimal simplification of polygonal chains for subpixel-accurate rendering", *Computing Geometry Theory Application*, Vol. 42 (1), pp. 45-59.
5. Buttenfield, B. (1985), "Treatment of the cartographic line", *Cartographica*, Vol. 22, pp. 1-26.
6. McMaster, R.B. (1987), "Automated line generalization", *Cartographica*, Vol. 24 (2), pp. 74-111.
7. Kumar, N., Lolla, N., Keogh, E., et al. (2005), "Time-series bitmaps: a practical visualization tool for working with large time series databases", *Proceedings of the 5th SIAM International Conference on Data Mining, Newport Beach, California, April 2005*, pp 531-535.
8. Lin, J., Keogh, E., Lonardi, S., et al. (2004), "Visually mining and monitoring massive time series", *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 460-469.
9. Douglas, D.H. and Peucker, T.K. (1973), "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature", *The Canadian Cartographer*, Vol 10, No. 2, pp. 112-122.
10. Reumann, K. and Witkam, A.P.M. (1974), "Optimizing curve segmentation in computer graphics", *Proceedings of the International Computing Symposium*, Amsterdam, North-Holland Publishing Company, pp. 467-472.
11. Rangayyan, R.M., Guliato, D., de Carvalho, J.D. and Santiago, S. (2008), "Polygonal approximation of contours based on the turning angle function", *Journal of Electronic Imaging*, Vol. 17 (2), pp. 16-23.
12. Qingsheng, G., Brandenberger, C. and Hurni, L. (2002), "A progressive line simplification algorithm", *Geo-Spatial Information Science*, Vol. 5 (3), pp. 41-45.
13. Shi, W., and Cheung, CK. (2006), "Performance Evaluation of Line Simplification Algorithms for Vector Generalization", *The Cartographic Journal*, Vol. 43 No. 1, pp. 27-44.
14. Song, J. and Miao, R. (2016), "A Novel Evaluation Approach for Line Simplification Algorithms towards Vector Map Visualization", *International Journal of Geo-Informatics*, Vol. 5, pp. 223.

15. Ekdemir, S. "Efficient Implementation of Polyline Simplification for Large Datasets and Usability Evaluation", available at: URL www.diva-portal.org/smash/get/diva2:444686/FULLTEXT01.pdf (accessed 21 October 2017).

Статтю представив д-р техн. наук., проф. НТУ "ХПІ" Леонов С.Ю.

Надійшла (received) 01.12.2017

Skarga-Bandurova Inna, Dr. Sci. Tech., Professor
Volodymyr Dahl East Ukrainian National University
59-a Central Avenue, Severodonetsk, Luhansk region, Ukraine, 93400
Tel.: (064) 522-89-97, e-mail: skarga_bandurova@ukr.net
ORCID ID: 0000-0003-3458-8730

Grushka Mykhailo, Magistr
Volodymyr Dahl East Ukrainian National University
59-a Central Avenue, Severodonetsk, Luhansk region, Ukraine, 93400
Tel.: (064) 522-89-97, e-mail: mikkgru@yahoo.com
ORCID ID: 0000-0003-4145-1425

Barbaruk Lina, Senior Lecturer
Volodymyr Dahl East Ukrainian National University
59-a Central Avenue, Severodonetsk, Luhansk region, Ukraine, 93400
Tel.: (064) 522-89-97, e-mail: barbaruk.angelina@gmail.com
ORCID ID: 0000-0001-6639-3001

УДК 004.922

Підходи до ефективного спрощення та візуалізації великих наборів даних / Скарга-Бандурова І.С., Грушка М.О., Барбарук Л.В. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2017. – № 50 (1271). – С. 55 – 65.

Розглянуто проблеми візуалізації послідовностей великих наборів даних. Задачу пошуку оптимального відображення даних сформульовано у вигляді задачі вибору алгоритму спрощення простих полігональних ланцюгів з мінімальною кількістю помилок. Виконано аналіз ефективності алгоритмів Рамера-Дугласа-Пекера, Реумана-Віткама, Опхейма, Ланга, алгоритму виключення за висотою трикутника та алгоритму радіального виключення точок при роботі з великими даними. Наводяться результати порівняння та оцінки ефективності їх використання для аналізу реальних даних, що збираються в системах локомотивної безпеки. Іл.: 2. Табл.: 1. Бібліогр.: 15 назв.

Ключові слова: візуалізація; великі набори даних; алгоритм спрощення; полігональний ланцюг; система локомотивної безпеки.

УДК 004.922

Подходы к эффективному упрощению и визуализации больших наборов данных / Скарга-Бандурова И.С., Грушка М.О., Барбарук Л.В. // Вестник НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2017. – № 50 (1271). – С. 55 – 65.

Рассмотрены проблемы визуализации последовательностей больших наборов данных. Задача поиска оптимального отображения данных сформулирована в виде задачи выбора алгоритма упрощения простых полигональных цепей с минимальным количеством ошибок. Выполнен анализ эффективности алгоритмов Рамера-Дугласа-Пэкера, Реумана-Виткама, Опхейма, Ланга, алгоритма исключения по высоте треугольника и алгоритма радиального исключения точек при работе с большими данными. Приведены результаты сравнения и оценки эффективности их использования для анализа реальных данных, собираемых в системах локомотивной безопасности. Ил.: 2. Табл.: 1. Библиогр.: 15 назв.

Ключевые слова: визуализация; большие наборы данных; алгоритм упрощения; полигональная цепь; система локомотивной безопасности.

UDC 004.922

Approaches for Efficient Simplification of Large Datasets / Skarga-Bandurova I.S., Grushka M.O., Barbaruk L.V. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2017. – №. 50 (1271). – P. 55 – 65.

The problems of visualization of sequences of large data sets are considered. The task of finding an optimal data mapping is formulated in the form of selecting the best algorithm for simplification of simple polygonal circuits with a minimum number of errors. The analysis of the efficiency of variation of Ramer-Douglas-Packer algorithm, Reumann-Whitcom algorithm, Opheim algorithm, Lang simplification algorithm, the triangle exclusion algorithm and the radial exclusion algorithm for large data sets are performed. The results of comparison and estimation of their efficiency for the analysis of real data collected in systems of locomotive safety are presented. Figs.: 2. Tabl.: 1. Refs.: 15 titles.

Keywords: visualization; large data sets; simplification algorithm; polygonal circuits; system of locomotive safety.