

зовательно-квалификационной характеристики бакалавров и магистров компьютерных наук.

В течение всех 15 лет, каждую 2-ю среду ноября, на кафедре совместно студентами и преподавателями проводятся мероприятия, посвященные Всемирному Дню ГИС. Всемирный День ГИС – это ежегодное широкомащштабное событие, призванное привлечь внимание и интерес людей во всем мире к географии, а также к практическим и образовательным аспектам технологии географических информационных систем. Обмен информацией и результатами, достигнутыми в процессе проведения этого дня, мировое сообщество осуществляет через сайт <http://www.gisday.com>.

Розглянуто питання організації, відкриття і розвитку кафедри геоінформаційних систем Державного вищого навчального закладу „Національний гірничий університет“. Представлено напрямки і спеціальності навчання студентів на кафедрі, головні досягнення в сфері викладання сучасних навчальних дисциплін, а також у підготовці методичних матеріалів і виконанні наукових розробок. Наведено професійні функції і задачі спеціаліс-

тів, місця проходження навчальної та виробничої практики студентами кафедри.

Ключові слова: геоінформаційні системи і технології, комп'ютерний еколого-економічний моніторинг, геоінформатика, інтелектуальні системи прийняття рішень

The paper considers formation of the Geoinformation Systems Department of the State Higher Educational Institution “National Mining University”, its creation, opening and further development. Authors review education areas and specialties of students of the department, main achievements in the field of teaching state-of-the-art educational disciplines, drafting of learner’s guides and scientific developments. They specify professional functions and tasks of future specialists, places of training and study courses.

Keywords: geoinformation systems and technologies, ecology-economic computer monitoring, geoinformatics, decision-making intelligence systems

Рекомендовано до публікації докт. техн. наук І.О. Садовенком. Дата надходження рукопису 14.03.11

УДК 004.627:004.772

А.В. Качанов

Государственное высшее учебное заведение „Национальный горный университет“, г. Днепропетровск, Украина, e-mail: gis@alex29.dp.ua

СЖАТИЕ ДАННЫХ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ ДЛЯ INTERNET

A.V. Kachanov

State Higher Educational Institution “National Mining University”, Dnipropetrovsk, Ukraine, e-mail: gis@alex29.dp.ua

DATA COMPRESSION FOR INTERNET-GIS

Статья посвящена анализу алгоритмов и подходов к сжатию данных в онлайн-ГИС. В статье выделены основные типы данных в базе геоданных. Классификация проведена по особенностям, критичным для разработки алгоритма компрессии. Рассмотрены основные подходы, алгоритмы и стандарты компрессии, учитывающие специфику данных каждого типа. Предложено решение для сжатия данных, основанное на архитектуре прокси-сервера с раздельным сжатием различных типов данных.

Ключевые слова: онлайн-ГИС, сжатие геоданных, прокси-сервер, XML в ГИС

В сетевых геоинформационных системах (ГИС), как и в любых других сетевых приложениях, важным критерием эффективности работы является время доступа к данным. Несмотря на удешевление и увеличение объемов запоминающих устройств и носителей, мощности каналов в сетевых средах растут не так быстро, как объемы геоданных. В настоящий момент объем типичной пространственной базы данных, включающей изображения земной поверхности со спутника, значительно превосходит пропускную способность канала при передаче за приемлемое время. Поэтому компрессия (сжатие) является актуальной проблемой при обработке и передаче данных в сете-

вых ГИС. Однако необходимо отметить, что в автономных (stand-alone) системах, в отличие от сетевых, время обработки и доступа может быть даже выше при использовании несжатых данных, т.к. отсутствуют затраты времени на декомпрессию при обращении.

Постепенный переход области функционирования геоинформационных систем в сетевую среду и Internet накладывает отпечаток на форматы представления и хранения геоинформации. В автономной ГИС формат базы данных не так существен, т.к. для передачи пространственных данных пользователям, работающим в других ГИС, можно использовать конвертеры форматов. В сетевой же среде работают одновременно ГИС различных производителей и использование единого стандарта на представление и

передачу просторових даних являється обов'язковим для успішного взаємодія поставщиків і користувачів геопросторових даних.

Найбільш ефективно алгоритми стиснення працюють в тому випадку, якщо вони розроблені і застосовуються з урахуванням специфіки оброблюваних даних. Наприклад, алгоритми, застосовувані при стисненні аудіо-даних, не дадуть ефекту або будуть непридатними при стисненні текстових даних.

Можно виділити такі напрями розвитку алгоритмів стиснення: алгоритми загального призначення, стиснення аудіоданих, відеоданих, растрової графічної інформації (зображень), стиснення текстових і бінарних даних. В кожному з цих напрямків можна виділити окремі напрями застосування. Наприклад, сімейство алгоритмів стиснення відеоданих має кілька груп, таких як стиснення для передачі по каналах зв'язу (швидким і повільним), для зберігання і поширення відео, для подальшої обробки. В кожній групі вимоги до стиснення можуть суттєво відрізнятися.

Мають певну специфіку і просторові дані, але розробка спеціальних алгоритмів стиснення для них знаходиться на початковій стадії. Наповнення бази геопросторових даних можна розділити на кілька типів, відрізняються за властивостями, важливими для розробки алгоритму стиснення. До таких типів належать:

- растрові графічні дані;
- векторні дані;
- атрибутивна інформація, зазвичай – текст;
- допоміжні дані – дані, формуються при передачі і отриманні запитів і відповідей в мережних ГІС-застосунках.

В цій статті досліджено стан проблеми стиснення просторових даних применливо до роботи в мережній середовищі, відзначено загальні недоліки методів і виділено перспективи подальшого розвитку даної області.

Растрові графічні дані. Зазвичай це великі масиви зображень, отриманих в результаті дистанційного зондування Землі. Ця область представлена найбільшим кількістю різних алгоритмів стиснення, т.к. тут можливо застосування алгоритмів, розроблених для стиснення зображень. Такі алгоритми застосовуються як без змін, так і з модифікаціями, які враховують відмінності зображень фотографічного якості і растрових, що представляють просторові дані. При стисненні зображень виділяють дві групи алгоритмів: працюючі без втрат (loseless) і з втратами (lossy). При стисненні без втрат і подальшій декомпресії отримують повну копію вихідного зображення. Стиснення з втратами враховує особливості побудови людського зору і видаляє частину „лишньої“ інформації, не впливаючої на сприйняття зображення. Після декомпресії зображення не рівно оригінальному, але при цьому може не мати суттєвих візуальних відмінностей – залежності від алгоритму і ступеня стиснення.

Серед спеціальних форматів найбільш популярні MrSID, розширення GeoTIFF і JPEG2000.

MrSID (Multiresolution Seamless Image Database) [1] – захищений розробкою компанії LizardTech. Поступово стає стандартом де-факто для поширення даних дистанційного зондування Землі. MrSID оснований на вейвлет-перетвореннях. Можливості формату: висока ступінь компресії, підтримка кількох розрешень, вибіркова декомпресія частин зображення, внутрішнє представлення як набір бітових матриць (bitplane). Бітові матриці ділять зображення на частини, представляють ці частини в різних масштабах. Така технологія дозволяє за запитом користувача витягати з стисненого зображення тільки необхідну інформацію – за положенням і за розрешенням. Стиснення може виконуватися з втратами і без втрат. До недоліків можна віднести захищеність формату і платність ліцензування технології.

Формат GeoTIFF [2] – розширення формату TIFF (Tagged-Image File Format) [3]. TIFF – один з найбільш популярних растрових форматів. Використовується для зберігання, передачі, відображення растрових зображень. Підтримує компресію різними алгоритмами, розбиття на ділянки, різні кольорові простори, багатосторонність, включення додаткової інформації (тегов), що і використовується для зберігання просторових метаданих. Формат GeoTIFF формалізує зберігання метаданих для географічно-прив'язаних зображень, використовуючи підходящі теги і структури TIFF. В частині зберігання растра не відрізняється від TIFF. Інформація про розширенні тегах GeoTIFF є відкритою і може бути вільно використана.

Аналогічні доработки пов'язані і з стандартом JPEG2000 [4]. Формат стиснення зображень JPEG2000 оснований на технології вейвлет-перетворень. Формат дозволяє включати додаткові дані в файл, що і використовується в його розширенні. Комітет OGC (Open Geospatial Consortium Inc.) розробив специфікації GMLJP2 [5], які вказують, як використовувати мову GML разом з зображеннями в форматі JPEG2000. Мова GML (Geography Markup Language) – оснований на XML граматику для кодування просторових даних: географічних об'єктів, покриттів, спостережень, топології, геометрії, проекцій і т.д. Специфікації GMLJP2 включають вказівки по використанню і впровадженню об'єктів GML в файли JPEG2000, схеми застосування GML для підтримки покриттів в файлах JPEG2000. Зв'язка GML і JPEG2000 в певній ступені подібна розширенню GeoTIFF.

На даній стадії розвитку розробки не дають суттєвого приросту в ступені стиснення, як, наприклад, в JPEG2000 порівняно з JPEG. Нові формати дають нові можливості і зручності, як JPEG2000 дозволяє проводити декомпресію тільки потрібної області зображення. В основному зусилля дослідників в області стиснення геоізображень направлені на доработку існуючих висококоэф-

фективных алгоритмов с целью добавления к хранимым данным географической информации – координат, проекций и других метаданных.

В области сжатия векторных данных нет стандартов и общепринятых форматов. Разработаны отдельные алгоритмы, дающие возможность сжимать векторную информацию с потерями и без потерь. Самый простой подход [6] использует уменьшение избыточности векторных данных путем аппроксимации полигонов базовыми линиями, которые кодируются без потерь, и остаточными векторами, сжимаемыми с помощью векторного квантования. Более сложные техники используют сжатие с помощью словаря [7]. Словарь создается на основе кластеризации, проводимой по начальным точкам векторных объектов. Центроиды кластеров затем становятся элементами словаря, относительно которых и кодируются остальные векторные данные. Интересной техникой является использование вейвлет-преобразований, которые позволяют достигнуть уровня сжатия географических карт до 50% со среднеквадратичной ошибкой от 0,003 до 0,01° при восстановлении [8].

Сжатие текстовой информации (атрибутивных данных). Для сжатия текстовой информации могут использоваться любые алгоритмы компрессии, которые работают без потерь (loseless). Некоторые из них могут давать большую степень сжатия при соблюдении определенных ограничений. Одним из таких ограничений может быть язык текста. Например, текст на английском языке можно представить в 7-битной кодировке, только использование которой дает уже степень компрессии $7/8=0.875$. Некоторые исследователи адаптируют алгоритмы к специальным областям применения. Например, для текста в 2-байтной кодировке Unicode предложен алгоритм быстрого сжатия до 60%, работающий на порядок быстрее обычных алгоритмов [9].

Как и с векторными данными, в случае текстов нет единых стандартов и форматов, алгоритмов, а в геоинформационных системах сжатие текстов практически не применяется.

Сжатие вспомогательных данных. К таким данным относятся бинарные файлы проектов, различные метаданные (например, файлы палитр и проекций) и XML-файлы, сопровождающие передачу геоданных по сети (запросы, ответы). Сжатие бинарных данных обычно не применяется в ГИС при хранении (например, файлы проектов и широко распространенные shape-файлы хранятся в исходном виде и не сжимаются). При передаче по сети таких данных могут использоваться проверенные и общепринятые алгоритмы компрессии общего назначения (zip, gzip, zlib и др.).

Особо выделяются **XML-данные**. Инициативы OGC по принятию единых стандартов геоданных на основе языка XML и его разновидности GML приводят к все более частому использованию подобных структур данных. Но даже в хорошо разработанной схеме (xml-schema) файл GML может быть недопустимо больших размеров. Т.к. GML является подвидом XML, он хранится в текстовом виде и, по сравнению с бинарными

файлами, избыточен и потому медленно обрабатывается. Существует несколько путей решения этой проблемы, но ни один из них не стал общепринятым.

Часто на практике используется обычное сжатие потока XML, например, по методу gzip. Т.к. данные текстовые, то некоторое сжатие достигается, но алгоритм gzip не учитывает особенностей структуры XML-потоков, поэтому не дает максимальных степеней сжатия.

Другим подходом является разработка бинарного формата для хранения и передачи XML (WAP Binary XML) [10]. Этот подход является, в некоторой степени, шагом назад в развитии XML, который изначально создавался как язык, понятный человеку и машине, а в бинарном виде он лишается этого преимущества. Высокие степени сжатия достигаются за счет замены имен тегов и атрибутов на токены – байты из кодовой таблицы. Формат бинарного XML не сжимает строковые данные, которые сохраняются в исходном виде.

Стандарт ASN.1 [11] поддерживает кодирование XML-данных. ASN.1 – формальная запись данных абстрактного типа для передачи их по сети. К стандарту разработаны дополнения: правила кодирования XER – XML encoding rules и отображения из схем XML в ASN.1, которые позволяют переводить текстовый XML-формат в изначально разработанный для передачи двоичных данных ASN.1. Эффективность этого подхода снижается его универсальностью и тем, что он не учитывает структуру XML-данных.

Разработаны и другие приложения для сжатия XML-данных (XMill, XMLZip), но в настоящее время нет единого стандарта и не создано решения, позволяющего проводить компрессию и декомпрессию XML „на лету“, что обязательно для использования этого решения в онлайн-приложениях на основе XML-RPC, SOAP. Особенности сжатия в языке GML, который только недавно стал стандартом, не исследованы. К таким особенностям относится, например, соотношение объемов данных и разметки, ведь GML используется для хранения больших объемов пространственных данных.

Выше выделены основные типы данных в базе геоданных и рассмотрены основные подходы к компрессии, учитывающие специфику каждого типа. В настоящее время разработка и принятие стандартов в этой области находятся на начальном этапе.

Ввиду указанных причин, задача создания специализированного решения, позволяющего эффективно и компактно передавать геоданные по каналам связи, является актуальной. Такое решение должно позволять проводить сжатие и распаковку разнородных данных в едином технологическом цикле, оптимальным для каждого типа данных способом. Оно должно быть прозрачно для пользователя и приложения, что позволит применить его в уже разработанных геоинформационных системах.

Для соответствия всем этим требованиям, можно предложить архитектуру, разрабатываемую по принципу прокси-сервера.

Исследования показали [12], что использование HTTP-прокси серверов для сжатия данных позволяет

економить до 33% канала передачі при роботі з веб-сайтами. В цих дослідженнях використовувались тільки загальні алгоритми стиснення ZLIB і GZIP, використовувались вони тільки для текстових даних. А використання спеціалізованих алгоритмів, оброблюваних різних типів даних, повинно суттєво підвищити ефективність онлайн-ГІС.

В систему „клієнт-сервер“ пропонується додати проміжне ланцюгове звено (прокси), яке займається питаннями перетворення, в даному випадку – компресії і декомпресії. Потіки даних повинні аналізуватися і класифікуватися за типом даних, потім оброблятися з використанням відповідного алгоритму компресії. В мережній моделі OSI такому прокси-серверу відповідає рівень представлення. Наприклад, згаданий вище стандарт ASN.1 реалізований саме на цьому рівні.

Для функціонування онлайн-ГІС необхідно використовувати два прокси – клієнтський і серверний. Клієнтський декомпресуючий прокси-сервер знаходиться „ближче“ до клієнтського застосування, серверний компресуючий – „ближче“ до джерела геопросторових даних і ГІС-серверу. В окремих випадках прокси можуть об'єднуватися з застосуваннями і серверами, т.е. немає необхідності в запуску і налаштуванні клієнтського прокси-сервера, всі його функції виконуються в застосуванні на стороні клієнта, будь то браузер або спеціалізована система (прокси в вигляді „плагінів“, додаткових модулів і подібних прозорих для користувача рішень). В такому випадку, якщо існує декілька потоків даних, наприклад, клієнтське застосування використовується не тільки для перегляду і аналізу даних, але і для додавання нових геопросторових даних в БД сервера, або ведеться одночасна (але прозора для користувача) робота з декількома серверами, може знадобитися додавання декількох прокси для стиснення і розпаковки даних, в тому числі для стиснення на стороні клієнта.

Використання запропонованого прокси-сервера для стиснення потоків геопросторових даних в ГІС для інтернету дозволить значно підвищити швидкість передачі об'ємних геопросторових даних, а диференційоване стиснення (використання оптимальних алгоритмів для різних типів даних) усуне слабкі сторони алгоритмів стиснення загального призначення. Крім того, підхід з використанням прокси-сервера не потребує внесення значущих змін в архітектуру існуючих ГІС.

Список литературы

1. LizardTech. MrSID Technology Primer. [Електронний ресурс]: Режим доступу: http://www.lizardtech.com/files/geo/techinfo/MrSID_Tech_Primer.pdf.
2. Ritter N. GeoTIFF Format Specification. 2000 [Електронний ресурс] / Ritter N., Ruth M.: Режим доступу: http://remotesensing.org/geotiff/spec/geotiff_home.html.
3. Adobe Systems. TIFF Specification, revision 6.0. 1992 [Електронний ресурс]: Режим доступу: <http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf>.

4. Joint Photographic Experts Group. JPEG 2000 Specifications. [Електронний ресурс]: Режим доступу: <http://www.jpeg.org/jpeg2000/>.

5. OGC Reference Model. Open Geospatial Consortium Inc. [Електронний ресурс]: Режим доступу: <http://www.opengis.org>.

6. Akimov A. Reference line approach for vector data compression. / Akimov A., Kolesnikov A., Franti P.; International Conference on Image Processing – 2004. – P. 1891 – 1894

7. Shekhar S. Dictionary design algorithms for vector map compression. / Shekhar S., Huang Y., Djughash J.; Data Compression Conference – 2002. – P. 471.

8. Ioup J.W. Vector Map Data Compression with Wavelets / Ioup J.W., Gendron M.L., Lohrenz M.C.; Journal of Navigation. – 2000, №53. – P. 437–449

9. Studeny P. Unicode Technical Note #31. Fast Compression Algorithm for UNICODE Text. 2007 [Електронний ресурс] / Studeny P.: Режим доступу: <http://unicode.org/notes/tn31/>.

10. WAP Binary XML Content Format. W3C NOTE 24. 1999 [Електронний ресурс]: Режим доступу: <http://www.w3.org/TR/wbxml/>.

11. Abstract Syntax Notation One (ASN.1). Specification of Basic Notation. ITU-T Rec. X.680 (2002) /ISO/IEC 8824-1:2002.

12. Chi-Hung C. Compression proxy server: design and implementation, / Chi-Hung C., Jing D., Yan-Hong L.; in Proceedings of the 2nd conference on USENIX Symposium on Internet Technologies and Systems – Volume 2. 1999, USENIX Association: Boulder, Colorado.

Стаття присвячена аналізу алгоритмів і підходів до стиснення даних в онлайн-ГІС. У статті виділено основні типи даних у базі геопросторових даних. Класифікація проведена за особливостями, критичними для розробки алгоритму компресії. Розглянуто основні підходи, алгоритми та стандарти компресії, що враховують специфіку даних кожного типу. Запропоновано рішення для стиснення даних, засноване на архітектурі прокси-сервера з розділеним стисненням різних типів даних.

Ключові слова: онлайн-ГІС, стиснення геопросторових даних, прокси-сервер, XML у ГІС

Paper is dedicated to analysis of data compression algorithms and approaches in online-GIS. Main geodatabase data types are marked out. They are classified by features, critical for compression algorithms. Main approaches, algorithms and standards of compression for each data type are discussed. Data compression solution based on proxy-server architecture with differential compression is proposed.

Keywords: online-GIS, data compression, proxy-server, XML for GIS

Рекомендовано до публікації докт. техн. наук Б.С. Бусигінім. Дата надходження рукопису 14.03.11