

УДК 528.9

**ВЫЯВЛЕНИЕ ОШИБОК В ПРОСТРАНСТВЕННОМ КОМПОНЕНТЕ  
ГЕОДАНЫХ**

К.В. Хурцилава

*Институт проблем математических машин и систем НАН Украины*

e-mail: 1524nsurena\_580@rambler.ru

Как указано в ([1], с. 175 - 183) сложность географического мира делает практически возможным его полное цифровое отображение только в ограниченной области. Хотя существуют некоторые исключения (мы можем, например, создать совершенное цифровое отображение широты экватора, или прямой линии на земной поверхности), во всем остальном будут различия между содержимым базы данных и явлениями ими представленными. Многие термины используются для разъяснения этих различий, в зависимости от контекста. Различия могут существовать вследствие ошибок в измерениях, в то время как термин "неопределенность" кажется, более подходящим если цифровое представление просто неполное. В целом, можно просто сослаться на «качество» представления.

Если качество данных важное свойство почти всех географических данных, тогда оно должно влиять на решения, применяемые к этим данным. В общем, чем хуже качество данных, тем хуже решения. Плохие решения могут иметь тяжелые последствия, такие как больница, построенная в неудачном месте, или школа, по небрежности построенная около заброшенного хранилища опасных отходов. Географические данные часто используют для регуляторных целей, или для решения споров: хранители таких данных, безусловно, несут ответственность, если данные окажутся ошибочными.

Несмотря на, как представляется, ясные аргументы в пользу подробно разработанного толкования качества данных в ГИС, и несмотря на достаточные исследования соответствующих методов, многие ГИС методы продолжают применяться, так, как будто данные были совершенны. Результаты ГИС-анализа - в форме ли таблиц, карт или демонстраций - редко показывают оценку достоверности или другие индикаторы эффективности качества данных. В частности, такие положения унаследованы из картографической практики, впоследствии это затрудняло определение качества изображенной на карте информации. В частности, они возможно отражая общую тенденцию - доверять компьютерам больше, чем следовало - верили этому потому, что числа и карты, как по волшебству, появлялись из цифровых черных ящиков, они должны непременно быть надежными.

Существуют следующие причины возрастания, в последнее время, интереса к качеству данных:

1. Геоданные производят не только государственные организации, но и частные предприятия. Частные компании не согласуются с известными стандартами.
2. Увеличение применения ГИС как инструмента поддержки принятия решений. Ущерб от низкого качества данных существенно возрос.
3. Рост доверия к чужим данным.

Географические наблюдения описывают феномены с пространственными, временными и тематическими компонентами. Пространство - важнейший элемент этой тройки. Однако, времени не уделяется достаточно внимания. Время важно для понимания географических явлений как событий. Географические явления не только пространственные, а тематические. Мы можем рассматривать пространство (пространство-время) как рамку, с которой соотносится тема. Качество данных нужно рассматривать для пространственных, временных и тематических компонентов.

Модель сущность-атрибут-значение как концептуальная основа для большинства реализаций баз данных:

'сущность' - феномен реального мира;

'атрибут' - свойство феномена;

'значение' - качественное или количественное значение свойства.

Ошибка - как различие между записанным и оригинальным значением отдельного атрибута. Точность обратная ошибке. Это определение до некоторой степени ограничено. Географические данные это не тема с добавкой пространства и времени. Изменение пространства или времени влечет изменение темы и так далее.

Определение ошибки, данное выше предполагает, что существует, и может наблюдаться некая объективная реальность с которой могут быть сопоставлены полученные значения.

Истина может быть ненаблюдаемой, как для исторических данных. Наблюдение может быть непрактичным (например, слишком дорогим). Истин может быть несколько, так как сущности в базе геоданных это скорее абстракции, чем феномены реального мира.

К счастью, объективная действительность не должна быть ясно сформулированной, чтобы выполнить оценку точности. Это вызвано тем, что геопространственные данные всегда собираются при помощи модели, которая определяет, явно или неявно, необходимый уровень абстракции и обобщения относительно реальных явлений.

Временная точность не получила много внимания в литературе (напр. в классификации [15] временная компонента никак не обозначена), так же, как с самим временем не имеют дело явно в обычных геопространственных моделях данных.

Меры тематической точности ('точности атрибутов') меняются с изменением измерительной шкалы.

Логическая непротиворечивость ([16], с. 111) связана с внутренней непротиворечивостью структуры данных, с топологическим представлением данных, что означает наличие исчерпывающего списка взаимоотношений между связными геометрическими представлениями данных без измерения хранимых координат пространственных объектов. Она обычно заключается в ответах на вопросы: замкнуты ли полигоны, нет ли полигонов без меток или с несколькими метками, есть ли узлы на всех пересечениях дуг. Логические противоречия могут быть вызваны проблемами согласования информации и географических границ при совмещении данных из разных источников.

Полнота – указывает на связи между объектами в базе данных и “абстрактной вселенной” всех подобных объектов. Выбранные критерии, определения и другие картографические правила, используемые для создания базы данных, это определяющие факторы полноты. Существуют два вида полноты, “полнота данных” – указывающая на соответствии данных базы данных и её модели, а также “полнота модели” указывающая на соответствии структуры базы данных и “абстрактной вселенной”, для которой создана база данных.

Остановимся подробнее на логической непротиворечивости, а именно, средствах применяемых в нескольких распространенных геоинформационных системах для проверки топологии.

### **Топология и проверка геометрии.**

Ниже кратко рассмотрим средства проверки топологии и выявления ошибок в геометриях в нескольких известных ГИС.

### **ArcGis 10.2** (см. [2]).

Содержит инструмент проверки геометрии и инструмент восстановления геометрии. Эти средства работают хорошо. Топология в шейп-файлах – только простые соотношения. Топологические правила в базах геоданных – более тридцати правил (не ниже лицензии ArcEditor). Очень дорогой (ориентировочно 230000 гривен). Цена зависит от количества лицензий и функциональности.

### **ГИС Карта 2011.**

Топологические средства, как в режиме редактирования шейп-файлов ArcGis. В [3] упомянуто о согласованном сглаживании при генерализации.

В [4] описан комплекс автоматизированного контроля качества и исправления цифровых карт.

### **Quantum GIS.**

Имеет топологические правила. Можно использовать не только для баз геоданных но и для файлов. Правил, пока, гораздо меньше чем в ArcGis 10.2. Нет правила о включении одного полигона в другой, что, в настоящее время, препятствует использованию этой топологии для геоданных имеющих иерархический характер.

О плагине Quantum GIS, проверяющем топологию смотреть в [5] и [6].

### **GRASS GIS.**

Свободно распространяемая система с открытым исходным кодом. При исправлении серьезных ошибок в геометрии в Quantum GIS применяется GRASS GIS, поставляющийся вместе с Quantum GIS. В [7], [8] и [9] описаны некоторые способы исправления ошибок в геометрических данных с помощью GRASS GIS. Долгое время GRASS GIS разрабатывалась с интерфейсом командной строки. В настоящее время для GRASS GIS существуют несколько вариантов визуального пользовательского интерфейса.

### **POST GIS** ([10], [11]).

Возможна работа совместно с Quantum GIS. Имеет средства исправления ошибок в геометриях. Например, применяется федеральной полицией Бразилии [12] для больших объемов неверных геоданных. ([13] с. 269)

Есть инструмент для импорта шейп-файлов ([13] с. 177). Бесплатная система с открытым исходным кодом.

### **SAGA GIS** [14].

Маленькая, бесплатная (имеется даже версия, не требующая установки) геоинформационная система, ориентированная на автоматизированный геоанализ, имеет средства для работы с PostGIS.

ArcGis 10.1 обладает довольно развитыми средствами топологического редактирования и проверки геометрий, но очень дорог. ГИС “Карта” 2011 имеет более примитивные средства топологического редактирования, но существенно дешевле. Однако его средства контроля качества, по нашим впечатлениям, ориентированы на издание тематических карт.

Quantum GIS бесплатен, однако средства топологического редактирования и проверки геоданных пока недостаточны для многих задач. Вместе с тем топологические правила могут быть установлены и для шейп-файлов, что иногда бывает нужно. Из-за этого Quantum GIS применяется вместе с PostGIS и GRASS GIS.

GRASS GIS, розроблювана вже около 30 лет геоинформационная система с открытым исходным кодом.

PostGIS пространственное расширение объектно-реляционной базы данных PostgreSQL, системы с открытым исходным кодом. Для достоверных выводов требуется дальнейшее изучение её возможностей. Правда, обнаруженные нами отзывы о возможностях PostGIS по выявлению и исправлению ошибок в геоданных весьма благоприятны для этой системы.

Ощущается необходимость в недорогой, или вовсе бесплатной, но вместе с тем функциональной системе проверки и исправления топологии и геометрий с удобным пользовательским интерфейсом. В соответствии с распространенной сейчас клиент-серверной архитектурой функции проверки топологии и корректности геометрии следует возлагать на базу геоданных, однако множество пользователей до сих пор применяют файловые базы (например, на основе коллекций шейп-файлов (см. [17])). Такие пользователи также должны иметь возможность для проверки качества данных, что затрудняется большим разнообразием форматов хранения геоданных в файлах.

#### Список литературы

1. New Developments in Geographical Information Systems: Principles, Techniques, Management and Applications P. A. Longley, D. W. Goodchild, M. F., Maguire, D. J., Rhind 1999 1101 pages. Wiley
2. <http://resources.arcgis.com/ru/help/main/10.2/index.html#/006200000003000000>
3. <http://www.gisinfo.ru/newspages-news-1625-0>
4. <http://gistoolkit.ru/download/doc/sxfcorrectdoc.pdf>
5. <https://www.youtube.com/watch?v=huhkTZkoKC8>
6. [http://docs.qgis.org/2.0/en/docs/user\\_manual/plugins/plugins\\_topology\\_checker.html](http://docs.qgis.org/2.0/en/docs/user_manual/plugins/plugins_topology_checker.html)
7. <https://faunaliagis.wordpress.com/2013/08/14/bad-bad-polygon-fixing-invalid-geometries-with-quantum-gis/#print>
8. <http://www.northrivergeographic.com/archives/invalid-geometry-qgis>
9. [http://grasswiki.osgeo.org/wiki/Vector\\_topology\\_cleaning](http://grasswiki.osgeo.org/wiki/Vector_topology_cleaning)
10. <http://www.postgis.org/>
11. <http://www.slideshare.net/maximdubinin/postgrespostgis>
12. <https://vimeo.com/106220883>
13. PostGIS in Action Regina O. Obe, Leo S. Hsu 425 pages Manning Publications (April, 2011)
14. <http://www.saga-gis.org/en/index.html>
15. [http://www.miigaik.ru/nauka/dissertacionyy\\_soviet/zasedaniya/20131023163850-4688.pdf](http://www.miigaik.ru/nauka/dissertacionyy_soviet/zasedaniya/20131023163850-4688.pdf)
16. Геоинформатика: Учеб. для студ. вузов / Е.Г.Капралов, А. В. Кошкарев, В. С. Тикунов и др.; Под ред. В. С. Тикунова. — М.: Издательский центр «Академия», 2005. — 480 с.
17. Хурцилава К. В. Підсистема внесення змін до (актуалізації) змісту бази даних лісовпорядкування /К. В. Хурцилава, Б. О. Білецький // Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-прак. конф. з міжнар. участю. «СППР'2012». — Київ: СП «Інтертехнодрук», 2012. — С. 153 – 157.