

Information interaction in Applied Geoinformatics

Andrey Alexandrovich Maiorov, Professor, doctor of technical sciences. Sciences, head. Department of Information and Measuring Systems, Academician of the Russian Academy Cosmonautics named after K.E. Tsiolkovsky, Moscow State University of Geodesy and Cartography (MIIGAiK)

The article analyzes the information interaction in Applied Geoinformatics. The article shows the difference between awareness and information exposure. The article shows the difference between the impact of information and information interaction. This article describes three types of information exchange. The article describes the content requirements for the information interaction. These requirements include the following characteristics: Target certainty, temporary alignment, structural correspondence, communications line, the cyclical.

Keywords: applied informatics, communication, information, information influence, the target uncertainty, structural correspondence.

УДК 004.9

**НЕОПРЕДЕЛЕННОСТЬ КАК ФАКТОР НЕОБХОДИМОСТИ
ОБНОВЛЕНИЯ БАЗ ДАННЫХ**

*Василий Тимофеевич Матчин, старший преподаватель,
e-mail: matchin.v@gmail.com,
Московский технологический университет (МИРЭА),
<https://www.mirea.ru>*

Статья исследует новый фактор, вызывающий потребность в обновлении базы данных – информационную неопределенность. Рассмотрены общие вопросы обновления базы данных, которые определяют место информационной неопределенности среди прочих факторов обновления. Исследованы виды и причины информационной неопределенности в базах данных. Предложена структурная схема преобразования моделей от сбора информации до модели базы данных. Схема показывает динамику информационной неопределенности и причины ее появления. Статья дает различие в применении энтропии и негэнтропии при обновлении базы данных.

Ключевые слова: информация; базы данных; обновление базы данных; принципы обновления базы данных; информационная неопределенность; преобразование моделей; информационное соответствие; энтропия; негэнтропия.

Ведение

DOI: 10.21777/2500-2112-2017-2-98-104

В настоящее время важными являются вопросы обновления информации в базах данных [1, 2]. Большинство информационных ресурсов хранится в БД. Актуальность информации и информационных ресурсов является обязательным условием для использования их при поддержке принятия решений и управления [3]. Актуальность информации в БД является обязательным требованием к хранимой информации. Снижение или потеря актуальности служит основой обновления базы данных. Это обуславливает применение процедуры актуализации как основной технологии обновления, которая не решается в полной мере средствами СУБД. Обновление пространственной информации является сложной процедурой. Актуализация считается основной причиной обновления. Однако во многих работах по теории баз данных не принимают во внимание вторую причину – «НЕ-фактор» [4], который состоит в неопределенности информации в базе данных. Наряду с потерей актуальности существует независимый процесс накопления разной неопределенности информации в базе данных. Иногда эту неопределенность упрощенно называют «информационной неопределенностью». Информационная неопределенность [5], является одной из мно-



В.Т. Матчин

гих неопределенностей в базе или банке данных, которая служит основой обновления информации. По этой причине при обновлении используют две дополняющие технологии. Первая технология – это замена информации БД на более актуальную. Вторая технология – это технология замены информации в БД, содержащей много неопределенности, на информацию с меньшим количеством неопределенности. Это приводит к необходимости исследования неопределенности как фактора обновления базы данных.

Особо следует выделить базы данных пространственной информации, или базы пространственных данных (БПД), или базы геоданных (БГД) [1]. В этих базах содержится большее количество факторов неопределенности в сравнении с текстовыми или с фактографическими базами данных. БПД является более широким понятием. Оно включает, например, картографические базы данных или базы данных изображений. Как средство уменьшения неопределенности в базах пространственной информации используют геоданные. Эти данные являются высокоорганизованными, структурированными данными, которые являются системным ресурсом [6], что позволяет решать на их основе системные задачи и проводить системный анализ на основе БГД.

Материалы и методы. В качестве материалов использованы существующие описания и статистика информационного поиска. В качестве методов использовался системный анализ, структурный анализ и информационное моделирование.

Проблемы обновления баз данных. Большой объем данных затрудняет процессы обновления. Выход из сложившейся ситуации видится в создании интегрированного технологического обеспечения для обновления БД и условий, обуславливающих обновление. Процедуру интегрированного обновления баз данных целесообразно проводить поэтапно. На первом этапе необходимо провести исследование информационной потребности в обновлении. На втором этапе необходимо решить языковые задачи семиотического обеспечения процедур обновления. На третьем этапе необходима разработка математического обеспечения технологий обновления. На четвертом этапе целесообразно проводить обобщение и анализ методологических аспектов теории, чтобы сделать принципы обновления функционально переносимыми.

Одной из частных технологий обновления базы данных является информационный поиск [7]. При этом следует качественно различать информационный поиск документа в базе данных и информационный поиск фрагмента информации для обновления базы данных. Информационный поиск как самостоятельная технология ищет документ или файл, обладающий как минимум свойством целостности, как максимум свойством системности. Информационный поиск как вспомогательная технология при обновлении базы данных ищет фрагмент, который может не обладать свойством целостности.

Принцип информационного поиска при обновлении БД состоит в поиске структурированной информации, которая является частью целостного информационного объекта или информационной конструкцией. Информационный поиск при обновлении представляет собой процесс отыскания необходимой информационной конструкции, которую необходимо заменить новой информационной конструкцией. Термин информационная конструкция [8, 9] очень хорошо вписывается в технологию обновления, так как охватывает целостный документ, паттерн и фрагмент документа.

Предметом поиска выступает информационная потребность в обновлении информации, формально выраженная в поисковом запросе. Следует различать информационный поиск и «поиск данных». Поиск может осуществляться с помощью разных методов: циклически, рекурсивно, многоаспектно [10, 11]. Принципиальной позицией автора является то, что термин обновление базы данных следует заменять в большинстве случаев на обновление фрагментов БД. Полностью БД обновляется весьма редко. Чаще обновляют ее фрагменты. Поэтому и говорить в таких случаях следует о фрагментах. Проблемы обновления информации связаны с утерей ценности фрагментов информа-

ции БД, со снижением актуальности фрагментов информации БД, ростом неопределенности фрагментов информации БД. Проблеме неопределенности фрагментов БД в литературе уделяется мало внимания.

Неопределенность в базах данных. В общем случае информационная потребность в обновлении отражает информационную неполноту БД. Информация для обновления должна быть измерена в терминах сокращения неопределенности, то есть должна быть возможность измерить ее с помощью различных математических моделей.

Информационная потребность в обновлении данных определяется как различие между семантикой базы данных и текущим состоянием фактов реального мира. Можно рассматривать семантику баз данных как представление реального мира. Протекающие процессы в реальном мире создают различие между реальными фактами и фактами в базе данных

Семантика баз данных [12] включает содержательную качественную и количественную составляющие. Базы данных содержат также информацию, полученную на основе измерений, такая информация включает приближения, ошибки измерения и упрощения реальности. Многие базы данных содержат результаты большого числа измерений. Например, сканированное цифровое изображение включает миллионы пикселей

Многие базы данных и системы хранения информации при построении основаны на полисемических и синонимических терминах, которые задают двойственность [13]. Все ситуации двойственности также создают неопределенность для анализа содержимого БД. Исключение неопределенности требует использования математических моделей, оперирующих данными БД. Информационная неопределенность варьируется от нечетких наборов данных до стохастических наборов данных. Неопределенность обусловлена не только неточной, но и отсутствующей информацией. Неопределенность включает в себя ошибку как компонент [14].

База данных предоставляет ряд инструментальных средств манипулирования данными, однако не имеет методов контроля правильности их употребления. Это серьезная проблема. Без количественной оценки неопределенности результаты анализа данных часто можно рассматривать только как качественную информацию. Чтобы решить проблему качественного использования данных, требуется разработать алгоритмический подход к управлению информацией БД.

Алгоритмический подход должен поддерживать типы данных, собранных на основе различных моделей и учитывать присущие им качественные и количественные ошибки. Этот подход должен разработать специальные методы контроля неопределенности. Особенно интересна комбинация неопределенности в разных информационных моделях. Различные аспекты неопределенности могут быть измерены объективно, но их важность в значительной степени определена конкретной задачей. Моделирование неопределенности представимо в трех этапах:

- 1) формализация неопределенности [15], связанной с каждым набором данных, позволяющая говорить об объективном характере надежности,
- 2) разработка методов контроля уровня неопределенности в процессе обработки и преобразования данных,
- 3) определение пригодности набора данных для конкретной задачи.

Работы по исследованию неопределенности часто посвящены анализу ошибок, присущих определенным типам структур данных или моделям данных, но проблеме моделирования неопределенности пока уделено недостаточное внимание. На рис. 1 показано преобразование моделей при использовании четырех категорий моделей данных. На этом рисунке представлены четыре качественные модели и информационные отношения [16], переводящие модели одну в другую.

Первая модель является набором эмпирических данных (DS – data set), вторая модель называется формальной информационной моделью (FIM), или первичной инфор-

мационной моделью (FIM), третья модель называется объектной информационной моделью (ИМО), четвертая модель называется моделью базы данных (DBM). Основная функция первой модели – охват эмпирических данных в единую совокупность. Основная функция второй модели – исключение избыточности, установление связей и отношений между данными.

Основная функция модели ИМО – оптимальное описание свойств объекта или объектов, данные о которых находятся в DS, описание внешних и внутренних связей объектов, описание отношений и процессов, в которых участвуют объекты. Основная функция модели DBM – оптимальная адаптация объектных моделей ИМО в структуру базы данных, реализация структурной и информационной согласованности ИМО и DBM, а также структурной согласованности параметров ИМО с требованиями базы данных.

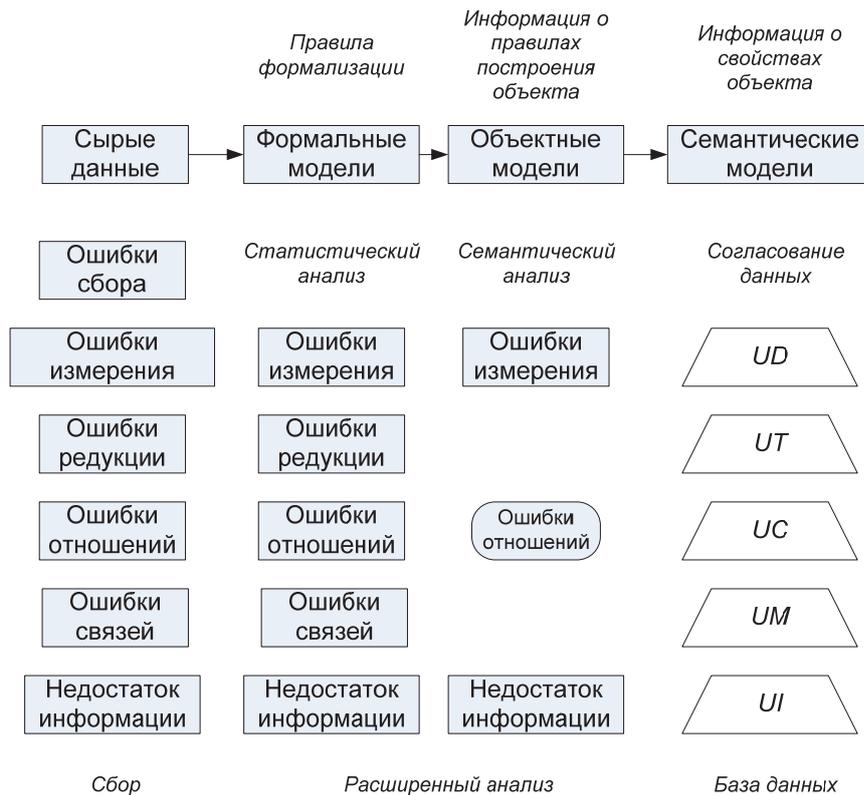


Рис. 1. Преобразование моделей и изменение неопределенности при подготовке информации для базы данных

Под каждой моделью размещены виды факторов, обуславливающих неопределенность. При преобразовании в соответствии с переходом от одной категории модели к другой характеристики неопределенности в данных изменяются. Формальные модели данных создают на основе статистического анализа и принятой формализации. Статистический анализ данных исключает ошибки сбора и уменьшает ошибки измерений. Семантический анализ исключает ошибки редукции и ошибки связей, а также уменьшает ошибки отношений. Согласование данных является рубежным преобразованием, уменьшающим неопределенность.

Описание используемой неопределенности связано с источниками ошибок, возникающих при трансформации моделей. Итоговая неопределенность включает следующие виды: UD – неопределенность значения данных, включая ошибки измерения; UT – временная неопределенность; US – неопределенность согласованности; UM – неопределенность завершенности моделирования (ИМО, DBM); IU – информационная неопределенность. Неопределенности обозначены в виде трапецевидных чисел, что подчеркивает их связь с теорией нечетких множеств.

При этом следует отметить, что исключение одних неопределенностей сопровож-

дается появлением новых неопределенностей, обусловленных неопределенностью моделирования и анализа.

Из перечисленных неопределенностей UD, UT могут рассматриваться применительно к любому набору данных. Информационная неопределенность IU рассматривается применительно к решаемой задаче. Неопределенность завершенности UM рассматриваются применительно к объекту моделирования. Неопределенности согласованности US и завершенности UM рассматриваются применительно к базе данных.

Элементарная модель данных может быть описана тремя параметрами: количественным значением d , семантикой s , временным параметром t . Каждый параметр связан с неопределенностью U и имеет три соответствующих компонента:

$$IM(d, s, t; UD, IU, UT). \quad (1)$$

В отдельных случаях неопределенности в (1) могут быть представлены в виде ортогональных компонент. В других случаях это сделать невозможно. Элементарная модель базы данных может быть описана тремя наборами параметров: количественными значениями Q_v , семантическим набором описаний S_s , временным набором параметров T_s :

$$DBM(Q_v, S_s, T_s; UD, IU, UT, US, UM). \quad (2)$$

Каждая элементарная модель базы данных создана в соответствии с некой концептуальной моделью C_m :

$$DBM(Q_v, S_s, T_s; UD, IU, UT, US, UM) \in C_m. \quad (3)$$

Преобразования, показанные на рис. 1, могут быть описаны системой парадигматических отношений, в которой каждая последующая модель добавляет новое качество:

$$D \rightarrow FM \rightarrow OM \rightarrow SM. \quad (4)$$

Необходимо рассматривать три механизма, связанных с неопределенностью в информационных процессах: 1) механизм анализа, уменьшающий неопределенность; 2) механизм накопления неопределенности (диссипация); 3) механизм создания неопределенности (ошибки вывода, неправильные процедуры анализа). Каждое из преобразований в схеме (4) не только создает новое качество, но и изменяет характер неопределенности:

$$UD \rightarrow UFM \rightarrow UOM \rightarrow USM. \quad (5)$$

Каждая из перечисленных суммарных неопределенностей в выражении (5) содержит неопределенности, показанные на рис. 1. Эти неопределенности расположены в вертикальных столбцах под соответствующей моделью.

Моделирование объекта в базу данных – это преобразование от эмпирических данных до модели базы данных:

$$OOM(Q_v, S_s, T_s) \rightarrow DBM(Q_v, S_s, T_s; UD, IU, UT, US, UM). \quad (6)$$

Выражение (6) означает следующее. Объективная модель объекта (OOM) не содержит ошибок, поэтому левая часть выражения (6) не имеет признаков неопределенности.

Изменение в наборе значений количественных данных Q_v происходит на стадии сбора информации. Оно возникает из-за ошибок измерения и других эффектов. Изменение в наборе значений семантических данных S_s связано с грубым экспертным оцениванием, использованием стереотипов.

Изменение в наборе значений временных данных T_s связано с протеканием процессов, меняющих информационную ситуацию [17], в которой находится объект, и меняющих состояние объекта независимо от управления им. По существу, T_s задает время экспозиции DBM как некоего снимка. Чем быстрее протекают процессы в информационной ситуации, тем больше временная неопределенность UT.

Появление неопределенностей US, UM обусловлено когнитивными факторами

[18], поскольку человек пропускает через себя исходную информацию и на основе когнитивных процедур строит связи и отношения в модели объекта, когда у него нет необходимой информации. То есть информационная неопределенность IU в виде нехватки информации обуславливает неопределенности US, UM.

Дискуссия

Приведенные выражения предлагают некоторую интерпретацию [19] сложных процессов построения модели базы данных для реальных объектов. С повышением уровня сложности используемых данных растет неопределенность, вносимая при анализе и обработке. Ярким примером этого является проблема «больших данных», [20] которая давно существует в области наук о Земле. Одна из трудностей анализа неопределенности состоит в разнообразной связи между разными типами неопределенностей. Например, ошибки в исходных данных могут исказить либо описание самого объекта, либо отнести этот объект другому классу. Анализ работ показывает, что обновление баз данных осуществляется по одному принципу – повышению актуальности информации. По нашему мнению, должен быть введен еще один принцип – принцип уменьшения неопределенности в базе данных и замена информации в БД на информацию, содержащую меньшую неопределенность.

Одним из методов устранения информационной неопределенности может быть интеграция [21]. В частности, при работе с пространственной информацией интеграция пространственных данных, получаемых из различных источников, в единую интегрированную информационную основу создает условия для уменьшения неопределенности.

Заключение

Проведенный анализ показывает, что при обновлении следует употреблять термин «обновление фрагментов базы данных», так как он соответствует сущности реального процесса обновления. Приведенные выражения (1)–(6) служат основой формирования алгоритмов обновления базы данных. Приведенные схемы показывают, что кроме неопределенности, свойственной процессу получения информации об объекте в базе данных, действия по анализу и обработке также вносят неопределенность в модель базы данных и накапливаются на каждом шаге преобразования. Причинами неопределенности могут быть накапливаемая неопределенность и нарушение принципа информационного соответствия. Причиной неопределенности может быть большой объем данных и рост слабой структуризации данных. С этих позиций проблему больших данных можно интерпретировать как проблему роста неопределенности. По нашему мнению, целесообразно введение понятия «принцип обновления базы данных» или «условие необходимости обновления базы данных». В настоящее время используют одни принцип, в то время как их может быть много. Необходимо введение алгоритмов и технологии «контроля неопределенности» в базе данных. Как показывает опыт, при обновлении баз данных применяется информационный поиск [10, 22]. Однако технология этого поиска и поисковые запросы существенно отличаются от обычного поиска документов по образцам. Это ставит задачу разработки специальной технологии информационного поиска для решения задач обновления базы данных. Это ставит дополнительную задачу дифференциации технологий информационного поиска. Еще одна проблема – это правильное применение понятий энтропия и негэнтропия [23]. Энтропия описывает информацию, уменьшающую неопределенность, поэтому она хороша как сигнальная информация для фиксации неопределенности. Обновление на основе энтропии некорректно. Для обновления необходимо применять понятие негэнтропии как содержательной информации. Это связывает проблему неопределенности с энтропией и негэнтропией.

Литература

1. Цветков В. Я., Матчин В. Т. Обновление баз геоданных // Перспективы науки и образования. 2015. № 5. С. 15–20.
2. Цветков В. Я. Обновление банка пространственных данных // Славянский форум. 2016. № 4 (14). С. 282–289.

3. *Коллин К. К.* Природа информации и философские основы информатики // Открытое образование. 2005. № 2. С. 43–51.
4. *Нариньяни А. С.* НЕ-факторы: неточность и недоопределенность – различие и взаимосвязь // Известия РАН. Серия: Теория и системы управления. 2000. № 5. С. 44–56.
5. *Цветков В. Я.* Информационная неопределенность и определенность в науках об информации // Информационные технологии. 2015. № 1. С. 3–7.
6. *Савиных В. П., Цветков В. Я.* Геоданные как системный информационный ресурс // Вестник Российской Академии наук. 2014. Т. 84. № 9. С. 826–829.
7. *Басинов А. А., Демич О. В.* Семантический поиск: проблемы и технологии // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2012. № 1.
8. *Tsvetkov V. Ya.* Information Constructions // European Journal of Technology and Design. 2014. Vol. 5. Iss. 3. P. 147–152.
9. *Дешко И. П.* Информационное конструирование: монография. – М.: МАКС Пресс, 2016. 64 с.
10. *Беззубова Ю. О.* Модели программных агентов в задачах информационного поиска // Славянский форум. 2015. № 2 (8). С. 41–49.
11. *Романов В. П.* Теоретические основы информатики. Информационные структуры и фактографический поиск информации. – М.: РЭА им. Г. В. Плеханова, 1996. 190 с.
12. *Цаленко М. Ш.* Моделирование семантики в базах данных. – М.: ФМЛ, 1989.
13. *Казенников А. О.* Разработка моделей и алгоритмов для комплекса автоматической обработки и анализа потоков новостных сообщений на основе методов компьютерной лингвистики: дис. ... канд. тех. наук. – М.: МИРЭА, 2014. 138 с.
14. *Цветков В. Я.* Информационно-измерительные системы и технологии в геоинформатике. – М.: МАКС Пресс, 2016. 94 с.
15. *Никифоров В. О., Слута О. В., Ушаков А. В.* Интеллектуальное управление в условиях неопределенности. – СПб: СПбГУ ИТМО, 2011.
16. *Tsvetkov V. Ya.* Information Relations // Modeling of Artificial Intelligence. 2015. Vol. 8. Iss. 4. P. 252–260.
17. *Цветков В. Я.* Применение темпоральной логики для обновления информационных конструкций // Славянский форум. 2015. № 1 (7). С. 286–292.
18. *Rozenberg I. N.* Information reception in information and cognitive systems // European Journal of Technology and Design. 2015. Vol. 10. Iss. 4. P. 140–148.
19. *Чехарин Е. Е.* Методы и алгоритмы информационной интерпретации // Образовательные ресурсы и технологии. 2016. № 5 (17). С. 39–49.
20. *Чехарин Е. Е.* Большие данные: большие проблемы // Перспективы науки и образования. 2016. № 3. С. 7–11.
21. *Коваленко Н. И.* Интеграция геоинформатики и логистики // Перспективы науки и образования. 2014. № 6. С. 26–30.
22. *Розенберг И. Н.* Комплексность информационного поиска // Образовательные ресурсы и технологии. 2017. № 1 (18). С. 41–49.
23. *Лотоцкий В. Л.* Энтропия и неэнтропия // Перспективы науки и образования. 2017. № 1 (25). С. 20–23.

Uncertainty as a factor in the need to update databases

Vasily Timofeevich Matchin, Senior Lecturer, Moscow Technological University (MIREA).

The article explores a new factor that causes the need to update the database – information uncertainty. The article explores general issues of updating the database, which determine the place of information uncertainty among other update factors. The article explores the types and causes of information uncertainty in databases. The author suggested a structural scheme for transforming models from information collection to a database model. The diagram shows the dynamics of information uncertainty and the reasons for its occurrence. The article gives a difference in the use of entropy and negentropy when updating a database.

Keywords: Information, databases, updating of the database, database update principles, information uncertainty, model transformation, information compliance, entropy, negentropy.