

УДК004.414.23

В.К. Гулаков, Е.О. Трубаков

## **ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ РАЗНЫХ ПОДХОДОВ К ПОСТРОЕНИЮ ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ТРАЕКТОРИЙ ДВИЖЕНИЯ МОБИЛЬНЫХ ОБЪЕКТОВ**

Рассмотрены общие подходы к пространственно-временному математическому моделированию траекторий движения мобильных объектов. Исследована эффективность применения этих подходов в системах мониторинга и быстродействие их основных алгоритмов. Даны рекомендации по выбору подхода к моделированию траекторий движения при определенных начальных условиях.

Ключевые слова: пространственно-временные математические модели, пространственно-временные структуры данных, системы мониторинга, мобильные объекты, методы доступа.

Текущий этап информатизации общества подразумевает обработку больших объемов информации. В частном случае можно говорить о временной информации. Это связано с тем, что большинство данных имеют тенденцию к изменению с течением времени. В данной статье рассмотрим один из видов временных данных – пространственно-временные данные.

Под термином «пространственно-временные данные» понимаются некоторые пространственные данные (координаты) объектов, изменяющих своё положение (например, автотранспорт). Ручная обработка такого рода данных затруднительна в связи с их объемами и тенденцией изменения. Для автоматизации процессов обработки были разработаны геоинформационные системы, или системы мониторинга мобильных объектов. Данные системы собирают, хранят, обрабатывают и визуализируют данные за приемлемое время. Но с развитием технологий и всё большим их внедрением (например, мобильных навигаторов) требования к эффективности таких систем и скорости обработки данных возрастают. Поэтому актуальность развития данного направления по-прежнему высока.

Сфер применения систем мониторинга мобильных объектов множество: мониторинг автотранспорта, мониторинг движения водного, воздушного и железнодорожного транспорта, отслеживание миграций животных, перемещения мобильных объектов (сотовые телефоны и пр.) и многое другое.

В каждой из сфер применения может быть множество уникальных для этой области запросов. Например, определение скорости сближения самолета, находящегося в той же плоскости и двигающегося навстречу. Для систем мониторинга воздушного движения данный запрос является актуальным и достаточно часто встречается на практике. Если же рассматривать сотовую связь, то необходимость в подобном запросе отпадает. Однако возникает необходимость определения точек скопления объектов и количества объектов, покидающих зону покрытия вышки. Данные запросы необходимы для повышения качества связи.

Несмотря на разнообразие пространственно-временных запросов, их можно разделить на три основные группы. Первая группа – это пространственные запросы. В данном случае временной параметр фиксируется в определенном значении и обрабатываются пространственные запросы (поиск по области, поиск соседей и др.)[2].

Ко второй группе относят временные запросы, т.е. запросы, в которых пространственная составляющая играет второстепенную роль. Примером таких запросов может быть поиск маршрута движения объекта в заданный интервал времени или же поиск положения объекта в определенный момент (случай, когда начальное и конечное значения интервала времени совпадают).

К третьей группе относятся смешанные запросы, т.е. запросы, в которых указываются условия как для пространственных, так и для временных параметров. К данной группе относятся топологические (направлены на траекторию движения) и навигационные запросы (направлены на сопутствующие параметры: скорость движения, направление и т.д.).

Для решения этих и многих других задач используют пространственно-временные системы мониторинга, которые состоят из трех основных частей. Первая часть - аппаратная. Это некоторое устройство позиционирования (ГЛОНАСС, GPS и др.), набор датчиков для отслеживания необходимых параметров (положение ключа в замке зажигания, уровень топлива и др.) и устройство передачи информации на сервер (GPRS, SMS или каким-либо другим способом). На серверной стороне находится устройство приема этой информации и две другие части.

Вторая часть – математическая модель. Так как данных на сервер может приходиться очень много, то их нужно эффективно хранить и обрабатывать. Именно этой части и посвящена данная статья.

Третья часть – это интерфейс (программное обеспечение), с помощью которого происходит взаимодействие пользователя с данными (обработка, визуализация и другие операции).

При выборе математической модели для системы мониторинга необходимо проанализировать область применения и выявить как можно большее количество особенностей работы в ней. Учет этих особенностей позволит повысить эффективность работы системы. Например, учет отсутствия произвольного движения объектов в пространстве (имеется возможность передвижения только по фиксированным сетям (автомобили, поезда и др.)) значительно повысит производительность. Определение круга типов запросов и выявление основных также повышает производительность. Таким образом, существующие математические модели пространственно-временных данных узкоспециализированы. Они могут снижать эффективность работы системы при применении в похожих областях или могут быть неприменимы вообще.

Существующие пространственно-временные модели можно разбить на группы, например по временному интервалу существования объекта. В первую группу входят структуры, сохраняющие только прошлое положение объектов. С их помощью можно собирать статистические данные, создавать различного рода отчеты и анализировать ситуации на основе траекторий движения объектов. Во вторую группу входят структуры, отслеживающие только текущее положение объектов. Благодаря таким структурам получают актуальную информацию о положении всех объектов. В третью группу входят структуры мониторинга текущего положения мобильных объектов с возможностью прогнозирования будущего изменения параметров. На основе этих структур получают актуальную информацию на текущий момент и прогнозируют с некоторой вероятностью будущее изменение параметров (как пространственного положения, так и значений дополнительных динамических параметров). Четвертую группу составляют структуры, поддерживающие работу со всей временной осью (траекторию изменения, текущее значение и прогнозирование будущего). На данный момент структуры этой группы являются наиболее актуальными, так как они совмещают в себе возможности остальных групп. При разработке структур данной группы также стоит учитывать тот факт, что они содержат в себе и все недостатки структур предыдущих групп (разработчики сталкиваются со всеми их проблемами)[1].

При индексировании данных по всей временной оси, как и при индексировании только прошлого состояния объектов, обрабатывается траектория движения. Рассмотрим методы организации структур данных, используемые при математическом моделировании траекторий движения мобильных объектов.

Первые пространственно-временные структуры данных основывались на хорошо зарекомендовавших себя многомерных деревьях с небольшими изменениями (например, R-

дерево [5]). Изначально эти структуры решали большинство задач, но со временем развивались сопутствующие технологии и, как следствие, эффективность этих структур снижалась. Поэтому возникает задача об увеличении эффективности работы модели и разработке новых пространственно-временных структур данных. Однако методы организации информации остались теми же. На момент развития данного направления хорошо зарекомендовали себя пространственные структуры, поэтому они и использовались для индексирования нового вида данных путем добавления измерения времени к пространственным осям. Далее рассмотрим более подробно этот метод индексирования пространственно-временных данных.

Существует два подхода к добавлению измерения времени. Один заключается в разделении данных на отдельные группы и их внесении в отдельные структуры (пространственные данные – в одну структуру, а временные – в другую) (рис. 1).

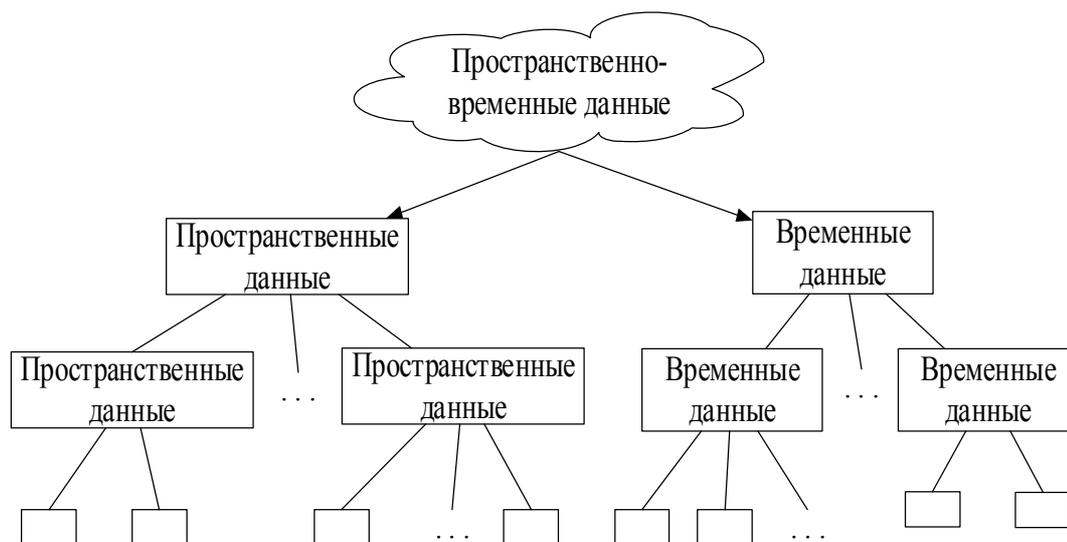


Рис.1. Схема индексирования двумя структурами данных

Второй метод организации данных состоит в устранении различия между пространственной и временной составляющими и индексировании в одной структуре путем добавления времени как дополнительной оси (рис. 2).

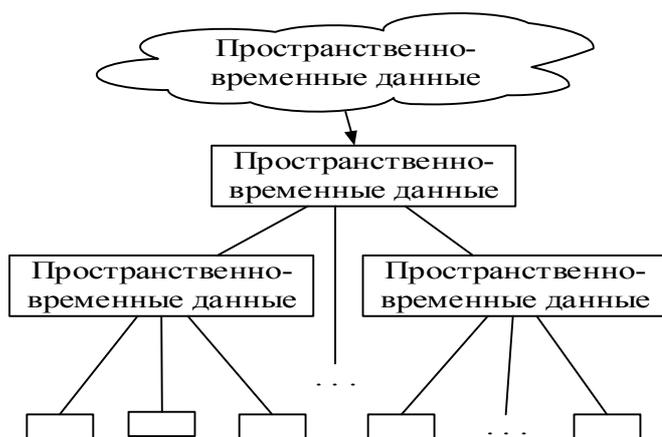


Рис.2. Схема индексирования одной структурой данных

В обоих подходах к организации математической пространственно-временной модели данных имеются как достоинства применения на практике, так и недостатки. При этом положительные стороны одного метода являются узкими местами другого и наоборот. Так, в системах первой моделью данных, в которых преобладают пространственно-временные типы запросов, необходимо обращаться к двум структурам данных (к дереву с пространственной информацией и дереву с временной информацией), что является достаточно дорогостоящей операцией.

Суть обработки запросов состоит в том, чтобы сначала произвести выборку объектов, удовлетворяющих одному из параметров (например, пространственному). Но

результатирующий набор данных будет содержать большое количество объектов, не подходящих под условия второго параметра. Поэтому на следующем этапе обработки запроса делают выборку для второго параметра данных (временного). Результирующая выборка также будет содержать большое количество объектов, не удовлетворяющих условиям по первому параметру, которые необходимо отсекают. Таким образом, получаются две выборки, содержащие как правильный ответ, так и излишние объекты. Для устранения неточности результата производят выборку из полученных двух наборов тех объектов, которые находятся как в первом наборе, так и во втором. В итоге только после третьего шага получают нужный набор объектов. При использовании второго подхода после первого шага обработки запроса получают объекты, удовлетворяющие условиям сразу двух параметров. Данный набор объектов и будет итоговым для ответа на запрос, поскольку шаг с объединением результирующих наборов попросту отсутствует.

Если рассматривать запросы вида срезов, т.е. только пространственные или только временные (второй параметр остается фиксированным), то из логики построения индекса видно, что более эффективным окажется метод организации данных с двумя самостоятельными структурами. Это связано с тем, что при поиске результирующего набора данных, удовлетворяющих исходному условию, необходимо проверять меньшее количество измерений.

Если сравнивать эти методы организации данных с простым перебором, то, безусловно, они более эффективны. Именно поэтому они используются для индексирования пространственно-временных данных.

Таким образом, очевидно, что в сферах применения пространственно-временных структур данных с преобладающим количеством смешанных запросов наиболее целесообразно использовать метод построения с одной структурой, а в случае преобладания срезов – метод с двумя самостоятельными структурами.

Однако на практике пространственно-временные системы мониторинга объектов используют как один тип запросов, так и другой. Поэтому в данной статье будем сравнивать указанные подходы по другим критериям, например по объёму дискового пространства, необходимому для хранения данных (поскольку данных может быть много и в итоге индекс будет занимать достаточно большой объём, что снижает эффективность работы системы и увеличивает стоимость оборудования, необходимого для работы). Вторым критерием сравнения стоит выбрать количество обращений к диску. Это связано с тем, что именно эта операция считается наиболее дорогостоящей при работе с индексами. В качестве третьего критерия сравнения выберем время вставки нового объекта в индекс, поскольку пространственно-временные данные постоянно обновляются и операция вставки является частой, что естественно увеличивает общее время работы системы.

Для проведения эмпирической части работы использовались три вида наборов данных, моделирующих основные виды распределения объектов в пространстве. Первый вид набора данных имеет равномерное распределение, которое моделирует общие случаи расположения объектов в пространстве и их изменение с течением времени.

Считается, что величина  $X$  равномерно распределена на интервале  $(a, b)$ , если все её возможные значения находятся на этом интервале и плотность распределения вероятностей постоянна:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{при } x \in (a, b), \\ 0 & \text{при } x \notin (a, b). \end{cases}$$

Для случайной величины  $X$ , равномерно распределенной в интервале  $(a, b)$ , вероятность попадания в любой интервал  $(x_1, x_2)$ , лежащий внутри интервала  $(a, b)$ , равна

$$P\{x_1 < X < x_2\} = \frac{x_2 - x_1}{b - a},$$

Второй вид набора данных имеет нормальное распределение. Это распределение моделирует скопление объектов в районе какой-либо точки и уменьшение их при отдалении от центра. Примером такого распределения объектов на практике может быть город (сам город как центр скопления объектов и его окрестности как отдаление от него).

Считается, что величина  $X$  нормально распределена, если плотность распределения её вероятностей определяется зависимостью

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

где  $m=M(x)$  – математическое ожидание (условно соответствует центру города);  $\sigma = \sqrt{D(x)}$  – дисперсия (условно показывает размер кластера).

Третий набор данных имеет кластерное распределение. На практике этот набор данных представляет собой несколько центров скопления объектов, с отдалением от каждого из которых снижается количество объектов. В жизни такое можно встретить, рассматривая несколько городов и объекты, находящиеся в них. По сути, это распределение данных представляет собой набор нормальных распределений [4].

При проведении опытов для моделирования в качестве пространственной структуры использовалось R-дерево. Это связано с тем, что данная структура хорошо себя зарекомендовала и на её основе построено большое количество пространственно-временных моделей данных.

Одним из ключевых моментов использования R-дерева является выполнение основных операций с данными без переиндексирования. Также имеются алгоритмы для параллелизации обработки поисковых запросов и операций добавления-удаления.

В данной структуре каждый листовой узел состоит из элементов, имеющих вид

$$[MBR, \text{идентификатор\_записи}],$$

где *идентификатор\_записи* ссылается на запись в БД, а MBR – это n-мерный прямоугольник, который является минимальным охватывающим прямоугольником для пространственного объекта (со сторонами, параллельными осям координат). Обычно MBR задают в виде

$$MBR = \{I_1, I_2, \dots, I_n\},$$

где  $n$  – число размерностей, а  $I_i$  – интервал с закрытыми концами  $[a, b]$ , который характеризует размер объекта по соответствующей оси координат  $i$ . Кроме этого, листовые вершины имеют ссылки на сами географические объекты.

Внутренние узлы дерева содержат элементы, имеющие сходную структуру:

$$[MBR, \text{ссылка\_на\_потомка}],$$

где *ссылка\_на\_потомка* – это адрес узла низшего уровня в R-дерево (дочернего по отношению к данному), все записи внутри которого покрываются прямоугольником MBR.

Для определения эмпирической трудоемкости основного алгоритма работы (вставки данных) был реализован ряд программ в виде консольных приложений. Для экспериментов использовался наиболее часто встречаемый случай пространственно-временных данных – трехмерные данные (два пространственных измерения и одно временное). После многократного запуска приложений на различных ЭВМ и с разными начальными условиями были получены экспериментальные данные, обработанные статистическими методами. Каждый из опытов был проведен тысячу раз. Это необходимо из-за невозможности абсо-

лютно точного измерения времени работы приложения. Только при большом количестве выборок и проводимых опытов можно говорить о приемлемой точности измерения [3]. Результаты опытов в абсолютных величинах получились различными (в зависимости от ЭВМ, на которой они проводились, а также начального и поискового множества вершин). Однако относительная разница между подходами к моделированию не зависит от перечисленных факторов, поэтому результаты опытов можно считать адекватными.

Структуры данных сравнивались по времени вставки нового объекта, количеству обращений к диску при поиске места вставки новых данных и размеру занимаемого дискового пространства.

На рис. 3 представлены результаты одной из серий опытов. Были построены структуры, содержащие 30 000 вершин, и измерено время поиска места вставки для каждого случая новых данных.

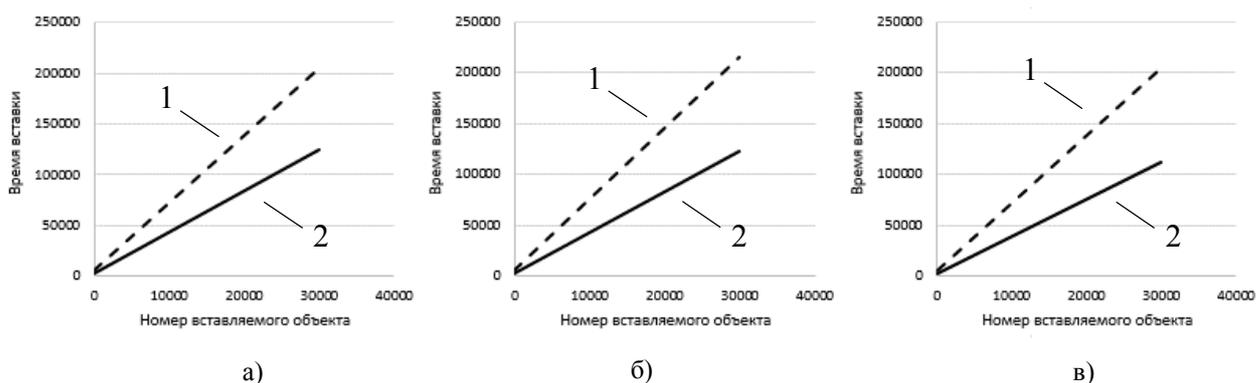


Рис. 3. Время вставки новых данных: а – кластерное распределение; б – нормальное распределение; в – равномерное распределение; 1- структура с двумя деревьями; 2 – структура с одним деревом

На рис. 3 представлены результаты вычислений для кластерного, нормального и равномерного распределений. Как видно из рисунка, для всех трех типов наборов данных время вставки новых трехмерных данных (два измерения пространственных и одно временное) в структуре с одним деревом меньше. Разница времени вставки увеличивается линейно с количеством вставленных данных, и она составляет 47, 46 и 42 процента от времени вставки в структуру с двумя деревьями для равномерного, нормального и кластерного распределений данных соответственно.

На рис. 4 представлены результаты второй серии опытов. Были построены структуры, содержащие 30 000 вершин, и подсчитано количество обращений к диску при вставке новых данных. Представлены результаты вычислений для кластерного, нормального и равномерного распределений. Как видно из рисунка, для всех трех типов наборов данных количество обращений к диску в структуре с одним деревом меньше. Разница количества обращений увеличивается логарифмически с количеством вставленных данных, и она составляет 47, 48 и 47 процентов от количества обращений к диску в структуре с двумя деревьями для равномерного, нормального и кластерного распределений данных соответственно.

На рис. 5 представлены результаты третьей серии опытов. Были построены структуры, содержащие 30 000 вершин, и подсчитан размер занимаемого дискового пространства.

На рис. 5 представлены результаты вычислений для кластерного, нормального и равномерного распределений. Видно, что для всех трех типов наборов данных размер дискового пространства, занимаемого структурой с одним деревом, меньше. Разница размеров индексов увеличивается линейно с количеством вставленных данных, и она составляет 12, 13 и 11 процентов от дискового пространства, занимаемого структурой с двумя

деревьями, для равномерного, нормального и кластерного распределений данных соответственно.

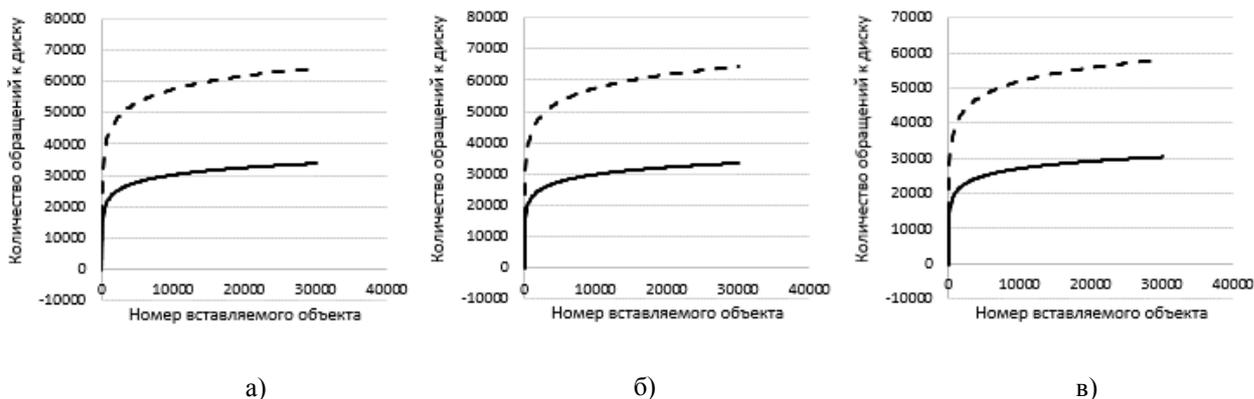


Рис. 4. Количество обращений к диску: а – кластерное распределение; б – нормальное распределение; в - равномерное распределение; 1- структура с двумя деревьями; 2 - структура с одним деревом

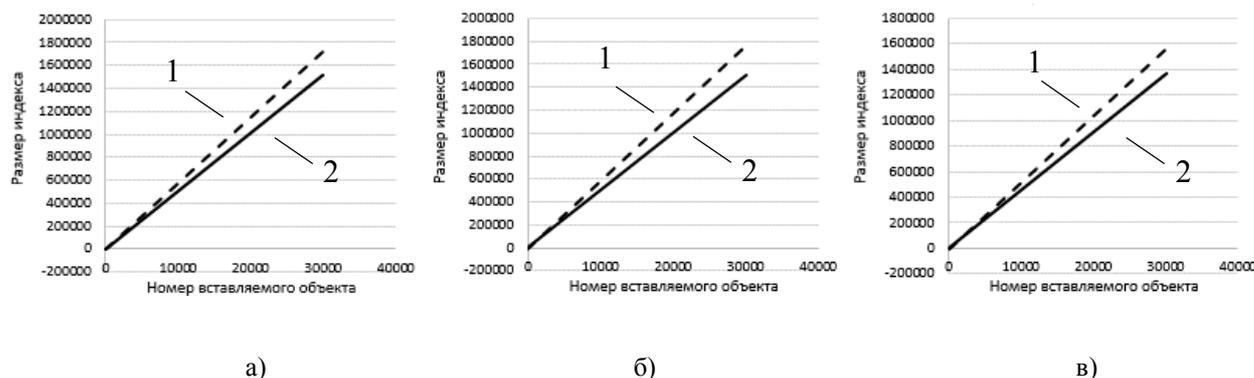


Рис. 5. Размер индекса: а – кластерное распределение; б – нормальное распределение; в - равномерное распределение; 1- структура с двумя деревьями, 2-структура с одним деревом

В данной работе была исследована эффективность применения двух различных подходов к организации исторической информации в пространственно-временных структурах данных. Первый подход подразумевает использование двух самостоятельных пространственных структур (одно дерево – для пространственной информации, второе – для временной), а второй подход заключается в объединении пространственного и временного изменений в одной структуре. В ходе исследований было выяснено, что в сферах применения пространственно-временных структур данных с преобладанием срезов (пространственных или временных) наиболее удобно использовать организацию с двумя древовидными структурами данных, в противном случае – структуру с одним деревом. Эмпирическим путем на основе статистических данных было выяснено, что структура с одним деревом занимает меньше места на жестком диске и быстрее вставляет новые данные при меньшем количестве обращений к диску. Таким образом, организация с одной структурой данных позволяет за одно и тоже время выполнить больше операций, чем структуры с двумя деревьями.

#### СПИСОК ЛИТЕРАТУРЫ

1. Гулаков, В.К. Пространственно-временные структуры данных: монография / В.К.Гулаков, А.О.Трубаков, Е.О.Трубаков. - Брянск: БГТУ, 2013.-215с.

2. Гулаков, В.К. Многомерные структуры данных: монография / В.К. Гулаков, А.О. Трубаков. – Брянск: БГТУ, 2010. – 387 с.
3. Новицкий, П.В. Оценка погрешностей результатов измерений / П.В. Новицкий, И.А. Зограф. – Л.: Энергоатомиздат, 1991. – 303 с.
4. Колемаев, В.А. Теория вероятностей и математическая статистика: учебник / В.А. Колемаев, В.Н. Калинина; под ред. В.А. Колемаева. – М.:ИНФРА-М, 2001. – 302 с.
5. Guttman, A. R-trees: A dynamic index structure for spatial searching / A. Guttman // ACM SIGMOD International Conference on Management of Data. – 1984. – P. 47-54.

Материал поступил в редколлегию 10.02.14.