



Апанович Зинаида Владимировна

МЕТОДЫ ВИЗУАЛИЗАЦИИ ИНФОРМАЦИИ – НАУКОЕМКОЕ НАПРАВЛЕНИЕ СОВРЕМЕННЫХ ИТ

Аннотация

В последние годы возникли и стремительно развиваются такие научные направления как визуализация информации и визуальная аналитика. В отличие от европейских и американских университетов, в программу обучения которых методы визуализации информации вошли достаточно прочно [11], в России подобные курсы почти полностью отсутствуют. В течение последних пяти лет автором этой статьи читается курс, посвященный методам визуализации информации на основе графовых моделей, в Новосибирском государственном университете. Данная статья дает общее представление о созданном курсе.

Ключевые слова: методы визуализации информации, дерево, диаграммы связей вершин и методы заполнения пространства, граф, визуализация планарных графов, силовые алгоритмы размещения неориентированных графов, поуровневое размещение ориентированных графов.

1. ВВЕДЕНИЕ

В последние годы возникли и стремительно развиваются такие научные направления как визуализация информации и визуальная аналитика. Количество теоретических исследований в этих областях возрастает, благодаря быстро расширяющемуся спектру промышленных приложений. Теоретическую основу методов визуализации информации и визуальной аналитики составляют методы рисования и визуализации графов. Эта область является одной из наиболее наукоемких областей современных информационных технологий. Для работы в области визуализации информации, помимо знакомства с теорией графов, необходимо достаточно глубокое знакомство с компьютерной геометрией, компьютерной

графикой и методами вычислений. Именно наукоемкость методов визуализации информации может дать несомненные преимущества на мировом рынке информационных технологий российским программистам, традиционно имеющим хорошую математическую подготовку. В отличие от европейских и американских университетов, в программу обучения которых методы визуализации информации вошли достаточно прочно [11], в России подобные курсы почти полностью отсутствуют.

Одной из основных трудностей в преподавании подобного курса в России является практически полное отсутствие как русскоязычной литературы, касающейся алгоритмов визуализации графов и структур данных, используемых для их эффективной реализации, так и учебных курсов, посвященных этой важнейшей области. В течение последних пяти лет в Новосибирс-

ком государственном университете автором этой статьи читается курс, посвященный методам визуализации информации на основе графовых моделей. В этом курсе рассматриваются теоретические основы построения изображений деревьев и графов, а также дается большое количество примеров реальных приложений, использующих изученные методы для визуализации информации. Данная статья дает общее представление о созданном курсе.

2. ПРИМЕРЫ ТИПИЧНЫХ ПРИЛОЖЕНИЙ, ИСПОЛЬЗУЮЩИХ МЕТОДЫ ВИЗУАЛИЗАЦИИ ГРАФОВ

Алгоритмы визуализации графов имеют огромное количество областей применения. Большинство людей сталкивались с файловой иерархией на компьютере. Эта иерархия может быть изображена в виде дерева, которое является частным случаем графа. Другие знакомые типы графов включают иерархии, представимые организационными диаграммами, и таксономии, которые отражают отношения между разными видами. В биологии и в химии методы визуализации графов применяются к деревьям эволюции, филогенетическим деревьям, молекулярным картам, генетическим картам, визуализации генов сетей, графов экспрессии генов и сигнальных путей. Их области применения включают также графы потоков данных, потоков управления и вызовов подпрограмм, используемые при разработке программного обеспечения, и системы реального времени, представленные сетями Петри.

Визуализация ER-диаграмм применяется в системах управления базами данных, PERT-диаграмм – в системах управления проектами. Весьма популярными в бизнес-приложениях являются визуализации CRM-диаграмм, изображающих отношения с клиентами. Методы визуализации графов необходимы при анализе данных разведки, включая бизнес-разведку. При разработке средств вычислительной техники визуализируют схемы приборов и диаграммы ко-

нечных автоматов. Среди новых приложений следует отметить методы изображения знаний, в частности, онтологии, тезаурусы и семантические карты, а также визуализации для решения задач сетевой безопасности. Красивые изображения графов и сетей используются на обложках книг, журналов и как «сырье» для всевозможных произведений искусства.

Наукоемкие продукты, использующие методы визуализации информации, широко распространяются на мировом рынке в течение последних 10–15 лет. Имеются фирмы, поставляющие библиотеки и программные комплексы, ориентированные на визуализацию графов общего назначения (Tom Sawyer software, ILOG software, Algorithmic Solutions Software GmbH, уWorks), а также системы визуализации программного обеспечения (Imagix Corporation, Absint и др.). Вместе с тем появляется все больше фирм, специализирующихся на визуализации бизнес-информации, необходимой аналитикам различных предприятий и ориентированной на профиль тех или иных предприятий (Enterprise Solutions). Одной из старейших компаний этого направления является фирма Inxight Software, Inc, поставляющая средства визуализации информации для финансовых и биологических фирм. Наконец, совсем недавно появились компании, поставляющие на рынок продукты, использующие новые методы из области визуализации информации, так называемые методы визуальной аналитики. Среди растущего семейства инструментов визуальной аналитики процветает программное обеспечение фирмы HiveGroup, использующее визуализацию иерархических данных на основе Карты Дерева (Treemap) и поставляемое организациям, которым требуется ежедневный мониторинг сложной деятельности с участием тысяч продуктов, проектов и продавцов. Компания Advanced Visual Systems (AVS) поставляет на рынок продукты, позволяющие топ-менеджерам корпораций получать единую картину их бизнеса, а ведущая корпорация в области сервис-ориен-

тированной архитектуры TIBCO приобрела недавно компанию Spotfire вместе с ее платформой Enterprise Analytics.

3. ЦЕЛИ И СОДЕРЖАНИЕ КУРСА ПО ВИЗУАЛИЗАЦИИ ИНФОРМАЦИИ НА ОСНОВЕ ГРАФОВЫХ МОДЕЛЕЙ

Основной целью данного курса является обучение студентов существующим методам визуализации, дающее им понимание сильных и слабых сторон различных методов. Они должны научиться осознанно выбирать и применять подходящие методы визуализации в их практической деятельности, критиковать, предлагать улучшения и создавать новые способы визуализации для новых приложений и новых типов данных. Курс состоит из следующих разделов:

1. Введение в методы и средства визуального анализа на основе графовых моделей.

2. Методы построения статических изображений деревьев для анализа иерархической информации, теоретические оценки.

3. Методы визуализации и навигации для графов и деревьев на основе техники «Фокус+контекст».

4. Вспомогательные методы. Двусвязные графы, построение st-нумерации для двусвязных графов.

5. Методы построения изображения планарного графа и области их применения.

6. Информация, представимая с помощью ориентированных графов, и методы их визуализации.

7. Информация, представимая с помощью неориентированных графов, и методы визуализации, основанные на физических аналогиях.

8. Иерархические методы визуализации и навигации для графов.

9. Задача кластеризации графа.

Вводная лекция курса начинается с примеров визуализации графов, таких как генеалогические деревья или квадраты оппозиции, при помощи которых обучались логике студенты средних веков. Затем демонстрируются примеры современных

научных и коммерческих приложений визуализации информации, таких как изображения структуры организаций, карты рынка, деревьев эволюции, социальных сетей, лингвистических онтологий, транспортных схем, фотоархивов, цифровых библиотек и других. Затем дается математическая формулировка задачи построения графа, рассматриваются эстетические критерии, используемые при оценке качества изображения, и основные стили при построении изображений графов, а также проблемы, возникающие при работе с реальными приложениями.

Раздел, посвященный методам построения статических изображений деревьев для анализа иерархической информации, является весьма обширным и занимает 5 лекций. Сначала рассматриваются специфические эстетические критерии, используемые при визуализации деревьев, связанные с необходимостью изображения иерархий, а также объясняется разница между изображениями типа диаграммы связей вершин и методами заполнения пространства.

Методы построения изображений деревьев, относящихся к классу диаграмм связей вершин, рассматриваются на примере нескольких алгоритмов. Примеры этих изображений показаны на рис. 1. Сначала подробно рассматривается алгоритм Рейнгольда-Тилфорда построения поуровневого изображения бинарных деревьев (см. рис. 1а) и алгоритм Валкера для деревьев произвольной степени, а также структуры данных, необходимые для реализации поуровневого алгоритма за время $O(n)$. Далее рассматриваются несколько методов построения изображений деревьев, используемых для определения теоретических границ оценок площади и коэффициента формы (см. рис. 1б, в), дается понятие алгоритмов построения изображения на основе выбора пути в дереве и на основе разделителей. Рассмотрение радиального алгоритма (см. рис. 1г) построения изображений свободных деревьев сопровождается демонстрацией приложений, DiskTree, MoireTree и GnutellaVision, использующих этот алгоритм. Знакомство с методами построения диаграмм связей вершин

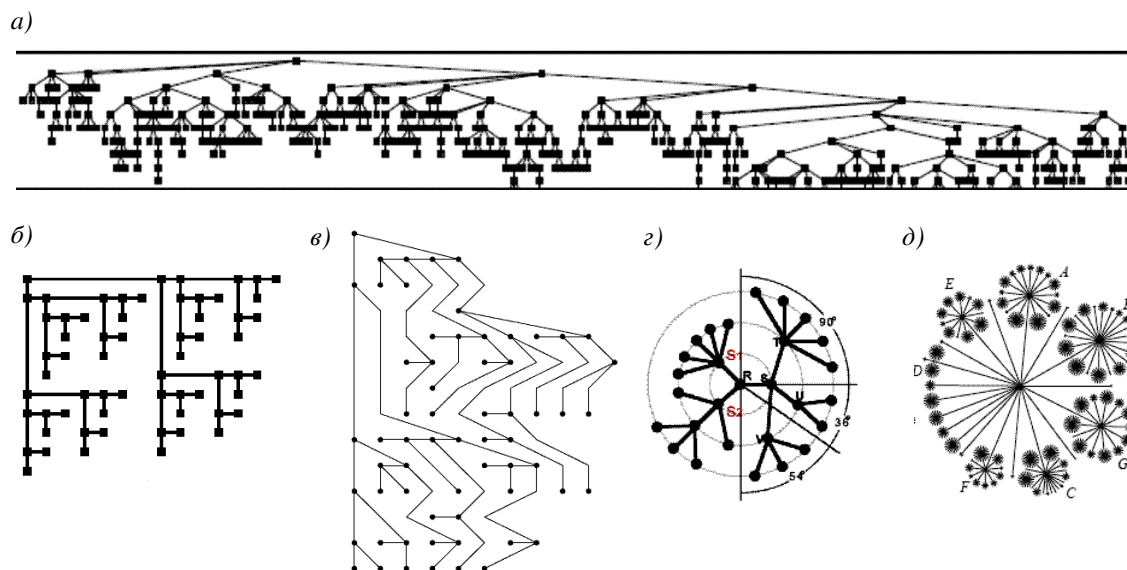


Рис. 1. Примеры изображений деревьев, относящиеся к классу «Диаграммы связей вершин»

завершают несколько вариантов круговых изображений деревьев (см. рис. 1 д).

Рассмотрение методов заполнения пространства начинается с группы наиболее известных алгоритмов этой группы, объединенных общим названием «карта дерева» (Тгеетар), которые выводят на новый уровень возможности визуализации иерархической информации, позволяя изображать не только структурные свойства, но и численные атрибуты различных иерархий. Рассматриваются три алгоритма этой группы: метод продольно-поперечного сечения (см. рис. 2 а), quadriфицирующий (см. рис. 2 б) и полосковый алгоритмы (см. рис. 2 в). Рассматриваются также различные схемы применения этих алгоритмов в случае глубоких иерархий, а также радиальные методы заполнения пространства и карты дерева на основе диаграмм Вороного.

Затем демонстрируется использование этих алгоритмов в конкретных приложениях бизнес-аналитики. Отдельная лекция курса посвящается вопросам построения интерактивных изображений. Наибольшее внимание в

ней уделяется методам, специфическим для работы с информацией. В ней, например, рассматривается методика навигации по дереву на основе функции степени интереса (см. рис. 3) и демонстрируется реализация этой методики на примере приложений DOI-tree и Space-Tree. Также большое внимание уделяется методам геометрического и семантического искажения, представляющим группу методов «Фокус+контекст». В частности, рассматривается построение искажения «Рыбий глаз» на основе функций нелинейного увеличения, включая методы построения интерактивных изображений на основе гиперболической геометрии и знакомство с приложением StarTree. Методы кусочно-линейного увеличения рассматриваются на примере модели «Резино-

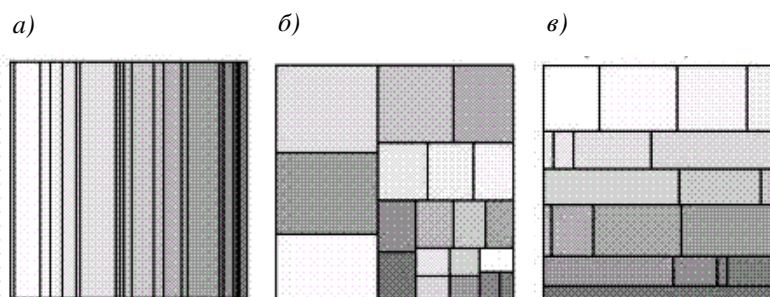


Рис. 2. Примеры работы трех основных алгоритмов построения карты дерева

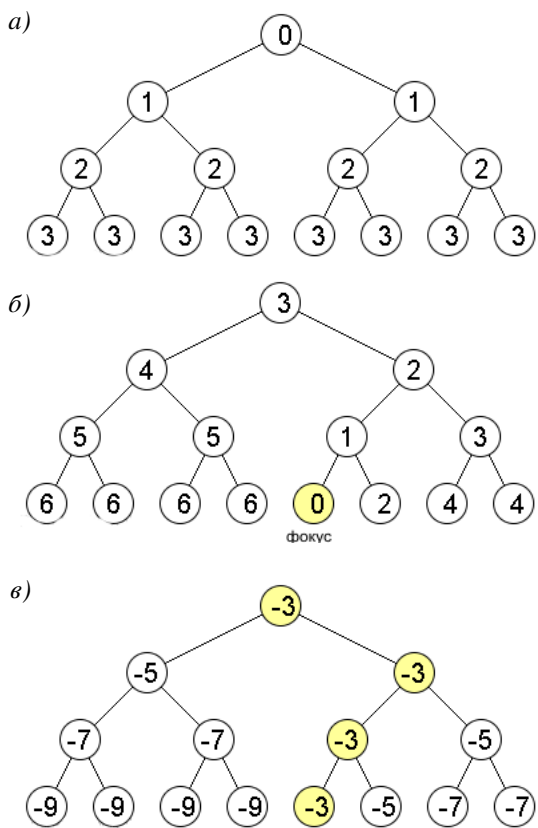


Рис. 3. Подсчет Функции степени интереса. Для каждой вершины показано: а) расстояние до корня, б) расстояние до фокуса, в) значение функции СИ

вый жгут». Демонстрируется применение этой модели для сравнения структуры филогенетических деревьев, а также применение модели «Резиновый жгут» в сочетании с радиальным алгоритмом для визуализации лингвистической онтологии.

Весьма существенное место в данном курсе занимает раздел, посвященный методам визуализации планарных и разреженных графов. Знакомство с этими метода-

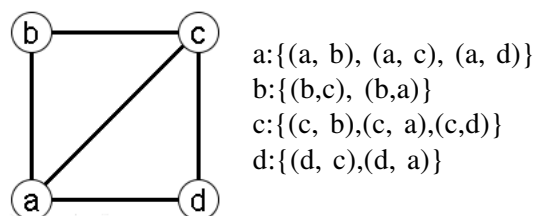


Рис 4. Планарный граф и его комбинаторная укладка

ми требует достаточно глубоких знаний в области теории графов, поэтому пришлось включить в курс две лекции, посвященные основным свойствам планарных графов, алгоритму проверки планарности графов на основе st-нумерации и построению комбинаторной укладки графа, определяемой как циклический порядок ребер, инцидентных каждой вершине (см. рис. 4). Также рассматривается алгоритм построения канонического упорядочения вершин максимальных планарных графов. Затем описывается алгоритм построения прямолинейного сетчатого изображения для планарных графов на основе канонического упорядочения, дается оценка площади получаемого изображения и оценка сложности алгоритма.

При построении ортогональных изображений графов возможно использование двух различных стратегий. Во-первых, существует эвристическая процедура линейной сложности, использующая вспомогательные изображения, называемые обзорными. Для знакомства с этой стратегией вводится понятие сильного и слабого обзорного представления и дается алгоритм построения сильного обзорного представления триангулированного планарного графа на основе канонического упорядочения вершин. Затем описывается эвристическая процедура построения ортогонального изображения на основе обзорного представления и эвристическая процедура минимизации сгибов полученного изображения.

Вторая стратегия позволяет строить ортогональные изображения графов с минимальным количеством сгибов. Эта процедура использует комбинаторную укладку графа, полученную в результате теста на планарность, и строит по этой укладке ортогональное изображение. Центральной частью этой стратегии является алгоритм построения сети по комбинаторной укладке и сведение задачи минимизации сгибов к поиску потока минимальной стоимости в построенной сети (см. рис. 5). Результатом применения обеих стратегий является построение ортогонального представления, к которому может быть применена процедура

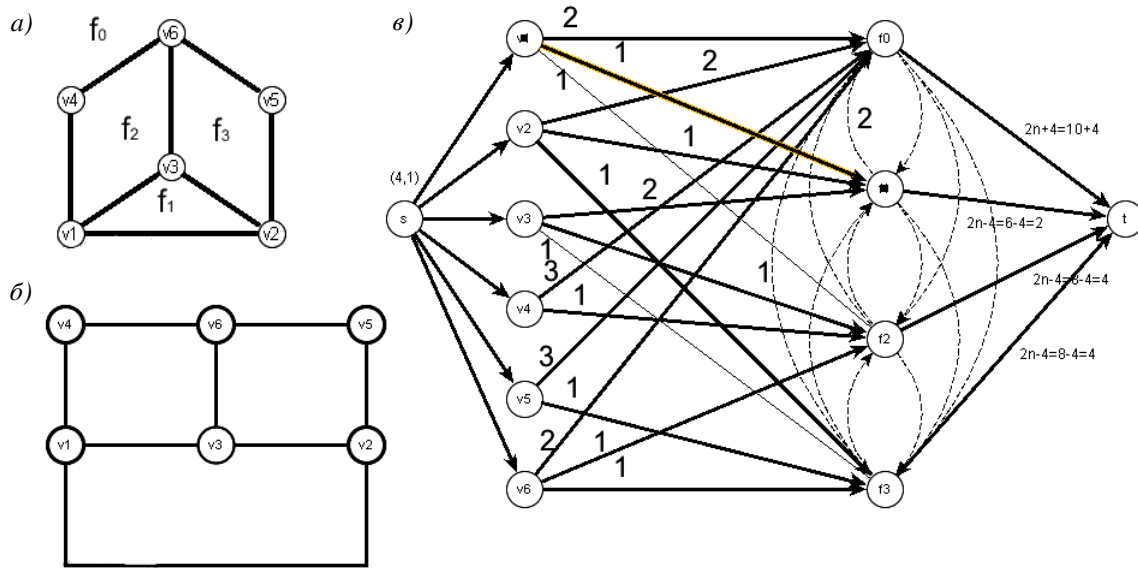


Рис 5: а) прямолинейная укладка планарного графа, б) ортогональное изображение с минимальным количеством сгибов, соответствующее потоку минимальной стоимости в построенной сети, в) потоковая сеть, соответствующая укладке планарного графа и значения потока минимальной стоимости

компактизации для вычисления окончательных размеров ортогонального изображения графа.

Методы визуализации неориентированных графов, основанные на физических аналогиях, являются самым распространенным инструментом при визуализации информации. Две лекции курса посвящены наиболее известной группе алгоритмов, используемых для построения изображений неориентированных графов (см. рис. 6а). К этой группе относятся «пружинные алгоритмы», алгоритмы, имитирующие действие сил гравитации и магнитные силы, а также алгоритмы, основанные на минимизации энергии. Классическими представителями этой группы алгоритмов являются алгоритм Фрюхтермана-Рейнгольда и алгоритм Камады-Кавая. Рассматривается также проблема повышения эффективности силовых алгоритмов размещения за счет применения многоуровневых методов на примере метода Барнеса-Хата.

Также весьма подробно на протяжении трех лекций рассматривается методика построения поуровневого изображения ориентированных графов (см. рис. 6б). Рассматриваются основные этапы стратегии Сугиямы: разбиение множества вершин на слои, упорядочение вершин на одном уровне с целью минимизации количества пересечений ребер, вычисление координат каждой вершины и проблема изображения ребер графа. Еще одна лекция посвящена различным иерархи-

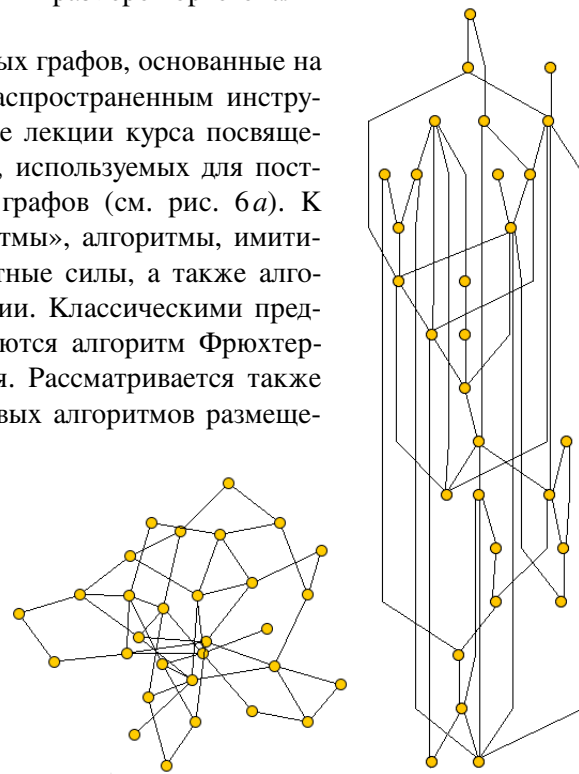


Рис. 6: а) пример размещения неориентированного графа, получаемого при помощи силового алгоритма; б) пример размещения ориентированного графа, получаемого при помощи метода Сугиямы

ческим моделям представления графов. Вводится понятие составного графа, подробно описывается методика построения изображения составных графов, на примере системы VCG, а также демонстрируется альтернативный подход к построению изображения составных графов на примере программы D-Abductor.

Рассмотрение иерархических методов невозможно без рассмотрения методов кластеризации графов, поэтому одна лекция курса посвящена этой теме. Рассматриваются целевые функции, используемые при разбиении графа и эвристические методы их оптимизации, такие как алгоритм Kernighan-Lin, алгоритм Fiducchia – Mattheyses и методы многоуровневого разбиения вершин графа на примере подсистемы hMetis.

Помимо представления теоретического материала, весьма полезны обсуждения, инициированные следующими вопросами:

1. Можно ли модифицировать алгоритм (указание конкретного алгоритма), с тем чтобы появилась возможность визуального анализа новых свойств изображаемой информации? Если да, то как именно?

2. Какие недостатки Вы видите в изученном алгоритме (изображениях, создаваемых при помощи изученного алгоритма) и как Вы бы предложили устранять эти недостатки?

Литература

1. Апанович З.В. Средства для работы с графами большого объема: построение и оптимизация компоновочных планов. // Системная информатика. Вып. 10: Методы и модели современного программирования. Новосибирск: Изд-во СО РАН, 2006. С. 7–58.

2. Апанович З.В. Программа курса «Комбинаторные алгоритмы анализа и синтеза графовой информации» // Сборник учебно-методических материалов по программированию и информатике НГУ, 2007. С. 26-32.

3. Апанович З.В. Методы интерактивной визуализации информации // Проблемы управления и моделирования в сложных системах: Тр. X Межд. конф. (Самара, 23–25 июля 2008 г.) С. 478–489.

4. Апанович З.В. Методы навигации при визуализации графов // Вестник НГУ. Т. 6. Вып. 3. 2008. С. 35–47 .

5. Апанович З.В. Методы заполнения пространства и их применение для визуализации информации и бизнес-аналитики // Проблемы управления и моделирования в сложных системах: Тр. XI Межд. конф. (Самара, 22–24 июня 2009 г.) С. 563–572.

6. Евстигнеев В.А., Касьянов В.Н. Графы в программировании: обработка, визуализация и применение. СПб.: БХВ. 2003.

Следует заметить, что одной из важнейших компонент обучения европейских и американских студентов методам визуализации информации является их участие в соревнованиях, которые ежегодно организуются в рамках таких авторитетных конференций, как IEEE VAST.

4. ЗАКЛЮЧЕНИЕ

Курс «Методы и средства визуализации информации на основе графовых моделей» входит в качестве специальной дисциплины в магистерскую программу «Технология разработки программных систем» по направлению «Информатика и вычислительная техника» в НИУ НГУ. При создании данного курса использовались материалы курсов по визуализации графов и визуализации информации, читаемых в европейских университетах и университетах США.

Опыт преподавания данного курса в Новосибирском государственном университете показал его полезность при подготовке специалистов, работающих в области ИТ, в таких областях как искусственный интеллект и молекулярная биология. В настоящий момент автором этого курса разработан электронный учебник «Методы и средства визуализации информации на основе графовых моделей», а также подготовлена к публикации первая часть учебного пособия по данному курсу.

7. *Apanovich Z.V., Bulyonkova A.A., Bulyonkov M.A., Emelianov P.G., Filatkina N.N.* Using Floorplans for Software Visualization // Bulletin of NCC Issue 24. 2006. P. 27–44.

Abstract

Scientific directions such as information visualization and a visual analytics have arisen quite recently. The quantity of theoretical researches in these areas increases, thanks to the quickly extending spectrum of industrial applications. The theoretical background of the information visualization and visual analytics methods are graph visualization methods. Unlike the European and American universities, which programs include courses on information visualization methods, similar courses in Russia are almost completely absent. One of the basic difficulties in teaching such a course in Russia is almost the total absence of the Russian literature, devoted to this subject. During the last five years the author of this article gives a course devoted to methods of the information visualization based on the graph models, in the Novosibirsk State University. Theoretical foundations of the tree and graph drawing construction are considered in this course. A considerable quantity of examples of real applications is given. This article describes the main topics of the course.

Keywords: information visualization methods, tree, node link diagrams, space filling methods, graph, planar graph visualization, force-directed placement algorithms for undirected graphs, layered drawing of directed graphs.

*Апанович Зинаида Владимировна,
кандидат физико-математических
наук, старший научный сотрудник
Института систем информатики
им. А.П. Ершова СО РАН и
преподаватель Новосибирского
государственного университета
(НИУ НГУ),
apanovich@iis.nsk.su*



Наши авторы, 2010.
Our authors, 2010.