

Представление знаний в информационных системах с учетом свойств наноразмерных объектов и материалов¹

В.А. Шахнов, А.Е. Аверьянихин, А.И. Власов, Л.В. Журавлева, Л.А. Зинченко

Аннотация. Рассмотрены подходы к представлению знаний в информационных системах с учетом свойств наноразмерных объектов и материалов. Отличительной особенностью предложенных подходов является использование онтологий, концептуальных карт и когнитивных технологий при представлении знаний. Обсуждаются особенности реализации предложенных подходов на примере реализованной базы знаний по технологиям микро- и наносистем.

Ключевые слова: знания, нанотехнологическая информатика, онтологии.

Введение

Экспоненциальный рост объема информации, полученной в области нанотехнологий, требует нового представления накопленных знаний.

Визуальные методы представления знаний в информационных системах являются мощными инструментами в условиях все возрастающего потока информации [1, 2].

При решении этой задачи одной из наиболее сложных проблем является формализация представления знаний в информационных системах с учетом фундаментальных свойств наноразмерных объектов и материалов. Это объясняется тем, что исследования в области нанотехнологий характеризуются разнородностью данных и знаний, мультязычностью литературных источников научной информации, междисциплинарностью, необходимостью привлечения экспертов из различных областей исследований.

В простейших моделях представления знаний в информационных системах использовались определения, которые затем систематизировались как словари некоторой предметной области. Примером использования такой модели представления знаний является словарь нанотехнологических и связанных с нанотехнологиями терминов [3, 4], включающих в настоящее время 592 термина. Для поиска информации используется предметный указатель. В электронной версии связи между понятиями реализованы с использованием механизма гиперссылок. В этом словаре отсутствует визуализация иерархии связей, что затрудняет проведение аналитических исследований с помощью этого словаря.

Универсальная десятичная классификация (УДК) является более совершенной системой систематизации знаний с использованием цифрового десятичного кода элементов. Однако соотношения между понятием и соответствующим десятичным кодом не несут какой-либо

¹ Работа поддержана Министерством образования и науки РФ (Государственный контракт 16.647.12.2049 от 26 августа 2011) и РФФИ (грант 13-07-00073а).

визуальной и когнитивной информации и требуют от исследователя долгой и кропотливой работы в библиотечном каталоге. Особенно усложняется задача поиска и анализа информации для области нанотехнологий, требующей применения нескольких УДК.

Для каталогизации и классификации патентных документов используется система МПК (международная патентная классификация). Каждому патентному документу присваивается индекс, являющийся комбинацией букв и цифр, каждая из которых несет определенную смысловую нагрузку. В последней редакции МПК также допускается наличие электронного слоя, содержащего различную визуальную информацию. К недостаткам этого подхода следует отнести отсутствие информации о результатах исследований, которые были выполнены без последующего патентования, что значительно сужает объем доступной для поиска информации.

В последние годы в области инженерии знаний и глобальных компьютерных сетей предложены принципиально новые подходы к представлению, хранению и обработке знаний в глобальных компьютерных сетях. Такие модели хранения знаний, как, например, онтологии, концептуальные карты, семантическая паутина (англ. Semantic Web) и др. [5, 6], включают в себя, помимо традиционных определений, информацию о связях между различными объектами, понятиями, определениями. Это позволяет перейти на принципиально новый уровень информационного обеспечения исследуемой предметной области и значительно повысить эффективность представления, поиска и анализа информации за счет визуализации связей между различными понятиями и использованием визуальной аналитики.

Целью данной статьи является обсуждение особенностей реализации базы знаний по технологиям микро- и наносистем для семантической поддержки методов формализации и сравнения научно-технологических решений в области технологий микро- и наносистем. Для структурирования различных понятий в этой предметной области были выбраны онтологический уровень представления знаний и концептуальные карты. Для наполнения базы знаний были использованы различные источники, в том числе [7]. Предло-

женные подходы по формализации знаний в области технологий микро- и наносистем могут быть расширены для применения в других областях нанотехнологий.

1. Технологии микро- и наносистем

Под технологиями микро- и наносистем понимается комплекс организационных мер, операций и приемов, направленных на создание, изготовление, обслуживание, ремонт, эксплуатацию и/или утилизацию функционально законченных сложных многоуровневых опто-электромеханических приборов, устройств и систем, использующих элементы и блоки, имеющие микро- и/или наноразмерные цепи, созданные на базе микро- и нанотехнологий.

Для определения предметной области был проведен опрос экспертов с помощью технологии экспертных панелей по структуре и составу направлений исследований и разработок в области технологий микро- и наносистем. Согласно результатам опроса к области технологий микро- и наносистем относятся, например:

- различного рода информационные микро- и наносистемы, включающие системы локации и связи, системы сбора, обработки и передачи данных, датчики и сенсоры, а также устройства мехатроники на их основе;
- биомедицинские приборы и устройства, включающие приборы и устройства как диагностики, так и локальной доставки лекарств и терапии;
- технологическое оборудование для производства наноструктур, наноматериалов и наносистем;
- измерительное и аналитическое оборудование для диагностики и анализа наноструктур, наноматериалов и наносистем;
- программные комплексы и технологии моделирования и проектирования наноструктур, наноматериалов, приборов, устройств и систем на их основе;
- изделия машиностроения и приборостроения, имеющие размерные цепи с величинами порядка единиц - десятков нанометров.

На Рис.1 приведена разработанная в ходе систематизации существующих результатов в области технологий микро- и наносистем мен-

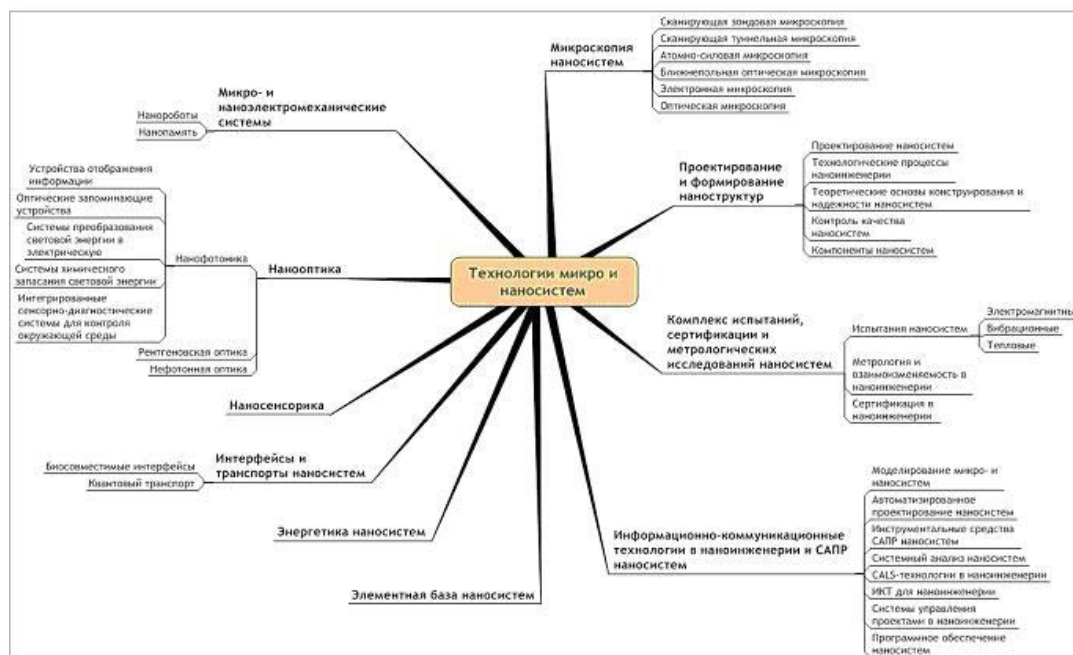


Рис. 1. Ментальная карта направления «Технологии микро- и наносистем»

тальная карта. Она позволяет наглядно представить существующую совокупность результатов научных исследований в указанной области в виде иерархической структуры. Однако эта карта не позволяет формализовать основные компоненты информационной системы. В связи с этим была необходима разработка способа представления знаний для указанной предметной области.

2. Подходы к представлению знаний

Базы знаний [5, 8] подразделяются на две большие группы.

1. Машино-ориентированные базы знаний. В этом случае знания хранятся в формате, удобном для последующей машинной обработки. Такие базы знаний используются в интеллектуальных системах, в частности, в экспертных системах [8]. Базы знаний этого класса также используются при построении семантической паутины, следующим шагом в развитии глобальных компьютерных сетей [9, 10].

2. Базы знаний, ориентированные на использование человеком (в основном в информационных системах, ориентированных на обмен информацией между различными пользователями).

Для создания баз знаний существуют различные подходы [5]. Способ представления знаний в информационной системе определяется выбранной моделью представления знаний. Необходимо отметить, что выполнить грамотный выбор модели возможно только после рассмотренных выше этапов получения и структурирования знаний.

В работах [11, 12] была предложена следующая классификация уровней представления знаний.

Самым низшим уровнем представления знаний являются примитивы, относящиеся к уровню физической реализации, например, указатели в языке С. На основе этих примитивов возможно только создание структур данных при отсутствии семантических связей. Однако наличие большого количества связей между понятиями в области нанотехнологий, и в частности, в области технологий микро- и наносистем не позволяет выбрать эту модель представления знаний для решения поставленных задач.

Логический уровень представления знаний, базирующийся на использовании, например, предикатов, обеспечивает возможность формальной семантики взаимосвязей между объектами с учетом свойств наноразмерных объектов

и материалов. Однако этот уровень представления знаний дает только общие, нейтральные механизмы и инженер по знаниям должен найти подходящее представление знаний. Учитывая бурное развитие нанотехнологий, этот подход пока не представляется пригодным для реализации в связи с постоянно меняющимся наполнением содержания.

Эпистемологический уровень позволяет обеспечить адекватное представление примитивов для передачи знаний. Концептуальные связи дополняются эпистемологическими связями, что обеспечивает лучшую формализацию знаний по сравнению с логическим уровнем. Однако выбор заранее оговоренной структуры представления знаний делает этот подход мало эффективным для практического применения в области нанотехнологий.

Концептуальный уровень представления знаний отличается определенной когнитивной интерпретацией примитивов, представляющих собой некоторые тематические роли или элементарные действия. Однако выделение указанных выше примитивов является процессом субъективным, поэтому невозможно разработать какие-либо практические подходы к формализации знаний в области нанотехнологий, базируясь на этом подходе.

Лингвистический уровень представления знаний наиболее сложен для формализации. В настоящее время в области нанотехнологий отсутствуют информационные системы, позволяющие общаться с ЭВМ на естественном языке.

В последние годы для представления знаний развивается онтологический подход. Он отличается переходом к использованию значения конкретного понятия. Под моделью онтологии понимается упорядоченная тройка вида [5]

$$O = \langle A, B, C \rangle \quad (1)$$

где A - множество понятий (сущностей, концептов), используемых в предметной области $\{A_1, A_2, \dots, A_n\}$;

B - множество связей между понятиями;

C - множество функций интерпретации, заданных на множествах A и B .

Интерпретация позволяет обеспечить однозначное и корректное толкование знания в формализованном виде.

При построении онтологии для выбора множества понятий используются только понятия, обладающие свойствами, относящимися в философии к сортирующим [13]. При этом понятия, обладающие только характеризующими свойствами, не могут быть включены в множество понятий A . К сортирующим относятся все свойства, которые могут быть каким-либо образом численно определены (например, измерены, взвешены и т. д.) [12].

Сортирующие свойства, в свою очередь, подразделяются на строгие, полустрогие и нестрогие [8]. К строгим свойствам относятся те, которые обязательны для исследуемого понятия и изменения этого свойства приводят к изменению самого объекта. Примером строгого сортирующего свойства является понятие размерности системы (например, микросистема, наносистема).

При построении онтологии каждый концепт должен обладать строгим свойством.

Строгие свойства подразделяются на типичные и квазитипичные [8]. К типичным свойствам могут быть отнесены те свойства, которые позволяют четко идентифицировать объект. Примером такой классификации является выделение понятия *акселерометр ADXL202E* из множества микросистем. Дальнейшая идентификация этого объекта по дате производства как дополнительного условия классификации позволяет четко идентифицировать объект из всего множества A . Однако, например, свойство *сенсор* не позволяет четко идентифицировать объект из всего множества A , поэтому оно относится к квазитипичным.

Для задания внутренней и внешней структуры понятия используются бинарные свойства [8].

Бинарное свойство рассматривается как атрибут для некоторого объекта A_i , если некоторое число Y является атрибутом объекта A_i .

В зависимости от степени завершенности исследований в предметной области рассматривают два возможных подхода к построению онтологии [14]. Подход, получивший название «предположение закрытого мира» (Closed World Assumption (CWA)), подразумевает, что все возможные понятия уже включены в онтологию и дальнейшее расширение онтологии невозможно. Противоположный подход, полу-

чивший название «предположение открытого мира» (Open World Assumption (OWA)), обеспечивает возможность дополнения онтологии новыми понятиями.

Для перехода к онтологиям используются концептуальные карты, интеллект-карты и таблицы решений [5].

Концептуальная карта (концепт-карта, Concept Map (CMap)) является инструментом для организации и представления знаний. С математической точки зрения концептуальная карта представляет собой граф, при этом каждой вершине графа поставлено в соответствие выбранное понятие из множества A , входящего в кортеж (1), а ребра графа несут в себе информацию о множестве связей, входящих в множество B . При этом информация может быть представлена с помощью следующих выражений: «известен как...», «является», «приводит к...», «требуется для...», «вносит вклад в...». Подобное представление визуализирует логическую структуру рассматриваемого объекта.

При построении и использовании концептуальных карт привлекаются все возможности человеческого мозга [2, 15]: левое полушарие (логическое мышление) и правое полушарие (пространственно-образное мышление), что позволяет обучающимся и взаимодействующим специалистам быстро и точно усваивать представленные картами знания во всей их полноте.

После построения ментальной модели предметной области в виде концептуальной карты выполняется переход к построению базы знаний в виде семантической сети, фрейма и т.д. [5].

3. Особенности представления знаний в базе знаний по технологиям микро- и наносистем

При реализации базы знаний для представления знаний о свойствах наноразмерных объектов и материалов был выбран онтологический уровень представления знаний. Для представления выбранной модели исследуемой предметной области были выбраны концептуальные карты.

Область технологий микро- и наносистем постоянно развивается, что отражается в постоянном расширении и динамичном изменении

понятийной системы. Для формирования базы знаний в области технологий микро- и наносистем было использовано предположение открытого мира (OWA). Это позволило обеспечить в дальнейшем возможность дополнения разработанной онтологии новыми понятиями.

Для построения концептуальных карт были проанализированы и систематизированы различные данные, в том числе [7].

Концептуальные карты были разработаны с использованием специализированного программного комплекса ИМС Cmap Tools [6], предназначенного для создания баз знаний на основе концептуальных карт.

В качестве сортирующего строгого свойства понятий были выбраны физические эффекты, определяющие энергетические процессы в искомом устройстве. Это объясняется тем, что энергия как скалярная физическая величина является единой мерой различных форм материи и единой мерой перехода материи из одних форм в другие. Использование энергетических характеристик позволяет охватить всю предметную область технологий микро- и наносистем и обеспечить построение онтологической модели, имеющей строгий физический смысл.

Учитывая гетерогенность знаний в области нанотехнологий, для повышения эффективности работы с базой знаний для представления свойств понятия были использованы элементы когнитивных технологий.

Когнитивные технологии [16, 17] являются основой для творческого компонента всех форм познавательной активности. Каждый из видов когнитивных структур обеспечивает активную форму упорядочивания поступающей информации - ее идентификацию, хранение, селекцию по релевантным признакам и внутренний контроль.

Для заполнения базы знаний в области технологий микро- и наносистем был использован шаблон, пример заполнения которого приведен на Рис. 2. Для описания свойств понятия с использованием когнитивных технологий была выбрана следующая структура представления информации о выбранном концепте (понятии).

Каждое понятие для визуальной концептуализации включало в себя три графических элемента: визуальный образ понятия, метафора понятия и инверсия понятия (Рис. 2).

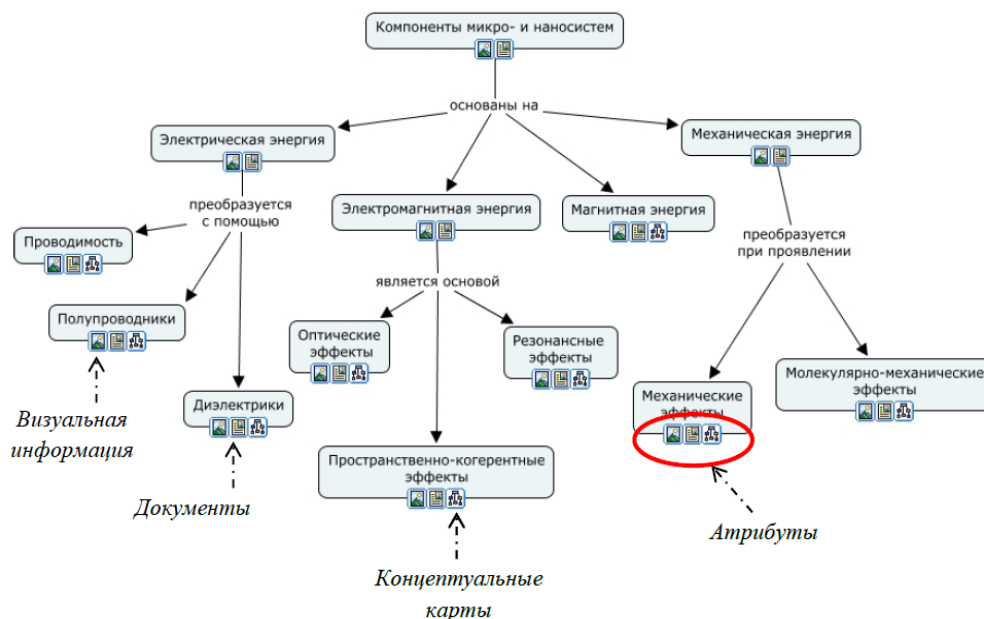


Рис. 3. Концептуальная карта выбранной предметной области

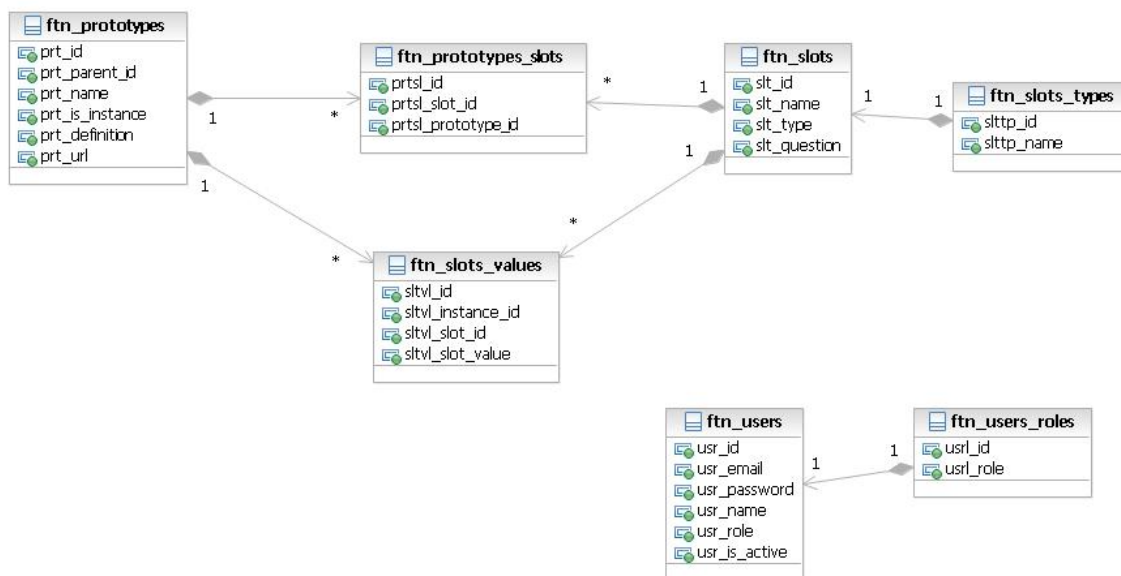


Рис. 4. Логическая модель представления знаний в базе знаний в области технологий микро- и наносистем

Использование онтологического уровня представления знаний при разработке базы знаний в области технологий микро- и наносистем [18] позволило реализовать информационную систему, отличающуюся достаточно большой гибкостью по сравнению с системами, базирующимися на теории формальных систем,

и с информационными системами – словарями [4], и в тоже время обладающей достаточной физически корректной внутренней структурой для структуризации и формализации исследуемой предметной области нанотехнологий.

Выделение понятий (концептов) онтологии и связей между ними позволило структуриро-

вать знания в предметной области технологий микро- и наносистем. Применение концептуальных карт позволило перейти от хаотического массива информации в области технологий микро- и наносистем к упорядоченному представлению знаний и визуализации связей между понятиями.

Использование элементов когнитивных технологий дало возможность разработать уникальное программное обеспечение, которое может быть использовано как для хранения информации, так и для поиска новой информации с целью дальнейшего развития технологий микро- и наносистем.

Необходимо отметить, что предложенные в статье подходы могут быть легко адаптированы для решения различных задач в области нанотехнологий.

Литература

- Hey T, et. al. eds. (2009) The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research. 284 pp.
- Шахнов В.А., Зинченко Л.А. Нанотехнологическая информатика – направление развития информационных технологий // Информационные технологии и вычислительные системы. 2012. №3. С. 84-92.
- Словарь нанотехнологических и связанных с нанотехнологиями терминов. М. Физматлит, 2010 г.
- <http://thesaurus.rusnano.com/>
- Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. - С-Пб.: Питер, 2000.
- <http://cmap.ihmc.us>
- Библиотека «Наноинженерия». Под ред. Шахнова В.А. В 17 кн. М.: МГТУ им. Н.Э. Баумана, 2011.
- Статические и динамические экспертные системы: Учеб. пособие/ Попов Э.В., Фоминых И.Б., Кисель Е.Б., Шапот М.Д. - М.: Финансы и статистика, 1996.
- Dacosta M. C., et al. The Semantic Web : A Guide to the Future of XML, Web Services, and Knowledge Management . John Wiley, 2003.
- <http://www.w3.org/>
- Brachman R.J. On the Epistemological Status of Semantic Networks. In N.V. Findler (Ed.), Associative Networks: Representation and Use of Knowledge by Computers. Academic Press, 1979.
- Guarino N. The Ontological Level: Revisiting 30 Years of Knowledge Representation. Conceptual Modelling: Foundations and Applications. Essays in Honor of John Mylopoulos, Springer Verlag 2009, pp. 52-67.
- Grandy R.E. Sortals. In Zalta E.N. (ed.) The Stanford Encyclopedia of Philosophy. 2007.
- Sengupta K., Krisnadhi A.A., Hitzler P. Local Closed World Semantics: Grounded Circumscription for OWL. In ISWC, volume 7031 of LNCS. Springer, 2011. pp. 617–632.
- Graudina V., Grundspenkis J. Concept map generation from OWL ontologies. Concept Mapping: Connecting Educators. Proc. of the Third Int. Conference on Concept Mapping, 2008.
- Величковский Б.М. Когнитивная наука: Основы психологии познания: - М.:Издательский центр «Академия», в 2 т., 2006.
- Власов А.И. Гексагональная понятийная модель визуального представления сложных производственных систем// Вестник МГТУ им. Н.Э. Баумана, сер. Приборостроение. Специальный выпуск №5. 2012. С.157-169.
- Власов А.И. и др. Экспертная система по технологиям микро- и наносистем. Свидетельство о государственной регистрации программ для ЭВМ №2012610397 от 10 января 2012 г.

Шахнов Вадим Анатольевич. Заведующий кафедрой в МГТУ им. Н.Э. Баумана. Окончил Московское высшее техническое училище им. Н.Э. Баумана в 1966 году. Доктор технических наук, профессор, член-корреспондент РАН. Автор более 200 печатных работ из них 7 монографий и 14 учебников и учебных пособий. Область научных интересов: информационные технологии, наноинженерия, вычислительная техника. E-mail: shakhnov@mail.ru

Аверьяннин Артур Евгеньевич. Ассистент МГТУ им. Н.Э. Баумана. Окончил МГТУ им. Н.Э. Баумана в 2010 году. Автор 10 печатных работ. Область научных интересов: информационные технологии, наноинженерия, вычислительная техника. E-mail: vlasov@iu4.ru

Власов Андрей Игоревич. Доцент МГТУ им. Н.Э. Баумана. Окончил МГТУ им. Н.Э. Баумана в 1994 году. Кандидат технических наук. Автор более 50 печатных работ. Область научных интересов: информационные технологии, наноинженерия, вычислительная техника. E-mail: vlasov@iu4.ru

Журавлева Людмила Васильевна. Доцент МГТУ им. Н.Э. Баумана. Окончила МАТИ в 1978 году. Кандидат технических наук. Автор более 50 печатных работ. Область научных интересов: технологии приборостроения и новых педагогических методов. E-mail: vlasov@iu4.ru

Зинченко Людмила Анатольевна. Профессор МГТУ им. Н.Э. Баумана. Окончила Таганрогский радиотехнический институт в 1987 году. Доктор технических наук. Автор более 200 печатных работ из них 12 монографий и учебных пособий. Область научных интересов: информационные технологии, САПР, моделирование. E-mail: lzinchenko@bmstu.ru