

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И РАЗВИТИЕ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ КОМПЕТЕНЦИЙ БАКАЛАВРОВ

Пиотровская К.Р.¹, Тербушева Е.А.²

¹ГОУ ВПО «Российский государственный педагогический университет им. А.И. Герцена», Санкт-Петербург, e-mail: krp62@mail.ru;

²Санкт-Петербургский государственный университет, Санкт-Петербург, e-mail: ekatherina88@mail.ru

Рассматриваются современные требования к формированию научно-исследовательских компетенций студентов исходя из государственных стратегий и образовательных стандартов нового поколения. Анализ требований и обзор научной литературы показывает, что исследовательская компетентность предполагает, в том числе и навыки по обработке и анализу экспериментальных данных с помощью современных методов и информационных технологий. С этой целью в программы обучения предлагается встраивать курсы по интеллектуальному анализу данных, методы которого активно используются в мировом научном сообществе для получения конкурентоспособных результатов. Описаны и обсуждены отечественные и зарубежные подходы к построению подобного курса. Сделан вывод о том, что наличие обязательного курса по интеллектуальному анализу данных поможет развитию научно-исследовательских компетенций у обучающихся различных специализаций.

Ключевые слова: научно-исследовательская компетентность, анализ данных, современные исследования, интеллектуальный анализ данных, ИАД, преподавание интеллектуального анализа данных.

DATA MINING AND THE DEVELOPMENT OF STUDENTS RESEARCH COMPETENCE

Piotrovskaya K.R.¹, Terbusheva E.A.²

¹Russian pedagogical State University. A.I. Herzen, Saint-Petersburg, e-mail: krp62@mail.ru;

²Saint Petersburg University, Saint-Petersburg, e-mail: ekatherina88@mail.ru

The article deals with the modern requirements to formation of research competencies of students on the basis of the nation policies and educational standards of new generation. The analysis of requirements shows that research competence assumes including skills for processing and analysis of experimental data with modern methods and information technologies. For this purpose it is offered to build a course on data mining which techniques are widely used in the world scientific community to achieve competitive results. The article provides an overview of the approaches of Russian and foreign teachers for the introduction of such a course, as well as examples of activities to involve students in the experimental studies.

Keywords: Research competence, data analysis, modern research, data mining, DM, teaching of data mining.

Согласно Программе фундаментальных научных исследований, в РФ на долгосрочный период (2013–2020 годы) [11] необходимым условием для обеспечения качества образования мирового уровня и достижения глобальной конкурентоспособности отечественных университетов среди ведущих мировых научно-образовательных центров требует развития у учащихся научно-исследовательских компетенций нового уровня. Описанные в методической литературе различные подходы к определению понятия научно-исследовательской компетентности (НИК) можно условно разделить на три группы [7]:

- совокупность знаний и умений, необходимых для осуществления исследовательской деятельности, например, с анализом и оценкой научного материала;

- интегральное личностное качество, выражающееся в готовности и способности самостоятельно осваивать и получать системы новых знаний в результате переноса смыслового контекста деятельности от функционального к преобразовательному, базируясь на имеющихся знаниях, умениях, навыках и способах деятельности;
- интеграция знаний, представлений, опыта, программ действий, системы ценностей и отношений, дополненных личностными качествами, которые затем выявляются в исследовательской компетентности нового качества.

Придерживаясь третьей точки зрения, которая объединяет первые два подхода, мы выделяем пять компонентов НИК: личностный, социальный, организационный, отраслевой, ценностно-мотивационный, а под научно-исследовательской компетентностью будем понимать личностное качество, которое формируется в процессе научно-исследовательской деятельности, которое выражается в личностной способности к творческой, деятельностной организации труда в области экспериментальных, конструктивных и композиционных задач, подкрепленной владением специальными знаниями, умениями и навыками.

Обобщенные требования к формированию исследовательских компетенций у обучающихся в обновленных образовательных стандартах (ФГОС 3+) высшего образования для уровня бакалавриата показывают, что одним из видов деятельности, к которому могут готовиться выпускники различных направлений, является научно-исследовательская деятельность. Профессиональные компетенции, соответствующие данному виду деятельности, как правило, включают в себя способности формулировать цели и задачи исследований, собирать, обрабатывать и анализировать данные, интерпретировать и визуализировать результаты эксперимента, использовать математические методы обработки, анализа и синтеза результатов, необходимых для формирования выводов по соответствующим научным профессиональным исследованиям.

Сущность ИАД и его связь с научно-исследовательской компетенцией

Современные информационные технологии позволяют эффективно собирать и накапливать большой объем разнородных экспериментальных данных, грамотный и всесторонний анализ которых является необходимым для проведения полноценного исследования. Стандартные методы математической и статистической обработки данных зачастую не позволяют обнаружить существующие в данных нетривиальные и заранее непредсказуемые закономерности, для выявления которых разрабатываются и используются методы интеллектуального анализа данных. Интеллектуальный анализ данных (ИАД) широко применяется в мировом сообществе при проведении исследований во многих областях, например, в генетике, биоинформатике, астрономии, медицине, экономике,

образовании, информатике и других. Чтобы проводить эффективные эксперименты, получая конкурентоспособные и доказательные результаты, специалисты должны обладать знаниями и навыками в области интеллектуального анализа данных.

В процессе ИАД выделяются несколько этапов [1]: 1) понимание и формулировка задачи анализа; 2) подготовка данных для автоматизированного анализа; 3) применение методов интеллектуального анализа данных и построение моделей; 4) проверка построенных моделей; 5) интерпретация моделей человеком. Как можно заметить, этапы интеллектуального анализа данных схожи с этапами выполнения любой исследовательской работы (формулирование темы, цели и задач исследования, теоретические исследования, экспериментальные исследования, анализ и оформление научных исследований, внедрение и проверка эффективности результатов проведенного исследования [8]). Отличие заключается в том, что окончательная оценка эффективности полученных результатов выходит за рамки непосредственно самого процесса ИАД и возможна после их практического применения. В таблице 1 продемонстрировано, каким образом изучение интеллектуального анализа данных может способствовать развитию научно-исследовательской компетентности:

Таблица 1

Влияние ИАД на развитие НИК

Компонент НИК		Стимуляция развития НИК средствами ИАД
Личностный	Критическое мышление	Выбор и оценка построенных моделей, сравнение различных алгоритмов, сравнение своей реализации выбранного метода с чужими результатами.
	Самостоятельность, саморазвитие	Постоянное самообразование в области ИАД в связи с быстрым развитием данной области.
	Усердность, настойчивость, ориентация на результат	Вариативность в подборе параметров алгоритмов, применение различных методов ИАД с целью обнаружения скрытых знаний и информационных извлечений из сырых данных
Организационный	Анализ данных (методы мат. статистики, современные методы и информационные технологии обработки данных)	Применение конкретных методов и алгоритмов интеллектуального анализа данных. Освоение современных компьютерных ИАД-инструментов (как правило, содержат в себе и возможности статистической обработки данных)
Ценностно-мотивационный		Получение новых знаний, скрытых от непосредственного наблюдения и трудно предсказуемых в объеме данных, с помощью методов ИАД. Анализ примеров успешного применения ИАД в различных областях знаний и производства (в бизнесе, медицине, образовании и др.)

Отраслевой	Изучение ИАД для ИТ-специалистов может включать программирование алгоритмов, анализ их сложности – т.е. развивать навыки, необходимые для осуществления научно-исследовательской деятельности в ИТ сфере.
Социальный	Работа с научной литературой, включая иноязычную. Выступление с докладом по результатам эксперимента. Работа в коллективном эксперименте.

Обзор практики преподавания ИАД в западной и российской высшей школе

Поскольку спрос на специалистов, обладающих знаниями и опытом в области ИАД, неуклонно растёт, то в университетах всего мира активно внедряются курсы по интеллектуальному анализу данных, а в последнее время также наблюдается тенденция к формированию целых кафедр, занимающихся подготовкой специалистов в области анализа данных. Некоторое время подготовка таких специалистов считалась прерогативой вузов и факультетов с ИТ-направленностью, но сегодня все больше на рынке труда ценятся междисциплинарные специалисты, обладающие одновременно, как знаниями в определенной области науки (генетика, медицина, социология, лингвистика и т.д.), так и знаниями по современным способам анализа данных. В то же время, несмотря на бурное развитие этого направления, практика обучения методам ИАД все еще не сложилась, и нет каких-либо устоявшихся канонов преподавания данной дисциплины в высшей школе.

Проведенный нами обзор отечественной литературы по данной тематике показал, что преподаватели внедряют методы интеллектуального анализа данных в программу самыми различными способами. В статье [4] авторы отмечают схожесть определения этапов обработки данных с помощью информационных технологий и этапов анализа данных (сбор данных, их хранение, обработка и представление), заключая таким образом, что задача анализа данных максимально соответствует критерию комплексного подхода к обучению информационных технологий и поэтому предлагают ввести темы, связанные с интеллектуальным анализом данных, в курс «Информационные технологии». При подготовке специалистов в ИТ-области предлагается формировать практические навыки в области ИАД в рамках курсового проектирования, причем с использованием командного метода организации работы, где каждому студенту будет выделен определенный этап исследования или же выполнение одного крупного проекта целой группой студентов [10]. Для будущих инженеров направления «Информатика и ВТ» в рамках учебной работы с массивами данных описывается методика работы с алгоритмами по извлечению данных из социальных сетей [5]. В Санкт-Петербургском политехническом университете при подготовке магистров ИТ-направлений был разработан и опробован курс

«Интеллектуальный анализ данных средствами MS SQL Server 2008» [9].

Все больше попыток внедрения тем, связанных с ИАД, в учебные программы наблюдаются и для специальностей, напрямую не связанных с ИТ и прикладной математикой. Освоение данных методов рассматривается как повышение уровня профессиональной культуры. Опыт внедрения тем интеллектуального анализа данных в курсы «Методы анализа социологической информации», «Методы компьютерной обработки социологической информации: OCA, SPSS», «Методы многомерного анализа социологической информации», для студентов социологов показал, что большинство студентов, ориентированных после окончания университета работать по специальности, интересуются возможностями технологии ИАД и согласны прилагать усилия, чтобы разобраться в ее тонкостях, несмотря на сложность теоретических и прикладных аспектов интеллектуальных вычислений [6]. В ходе научно-исследовательской работы магистрантов направления подготовки «Экономика» предлагалось самостоятельное изучение методов интеллектуального анализа для работы с наборами данных, характеризующих экономические объекты. Студенты сумели в процессе исследования освоить среду R, как соответствующий инструментарий для ИАД [2]. На медицинском факультете СПбГУ кафедра организации здравоохранения обеспечивает курс по медицинской информатике, где помимо общих вопросов применения информационных средств и методов в различных областях физиологии, медицины и здравоохранения, программа дисциплины охватывает и интеллектуальный анализ данных. Студенты учатся формулировать задачи по поиску, сбору, обработке, анализу и представлению медицинской информации, выбирать адекватные средства и методы для их решения, работать с системами автоматизированного анализа электрофизиологических сигналов, медицинскими экспертными системами и базами данных, а также системами, реализующими технологии интеллектуального анализа данных [6]. В курсах «Квантитативная лингвистика и новые информационные технологии» и «Сетевой инструментарий лингвиста» РГПУ им. А.И. Герцена студенты Института иностранных языков и филологического университета знакомятся с открытыми сервисами AntConc, и Intellectual Text Mining Suite, а также проводят собственные мини лингвистические исследования [3].

Анализ зарубежной литературы показывает, что дисциплины по ИАД активно вводятся в программы обучения уже более десяти лет. Почти ни одна программа для подготовки специалистов в области информационных технологий не обходится без данного курса, выработаны различные методики преподавания интеллектуального анализа данных, имеющие перед собой, и различные цели обучения (см. рис. 1).

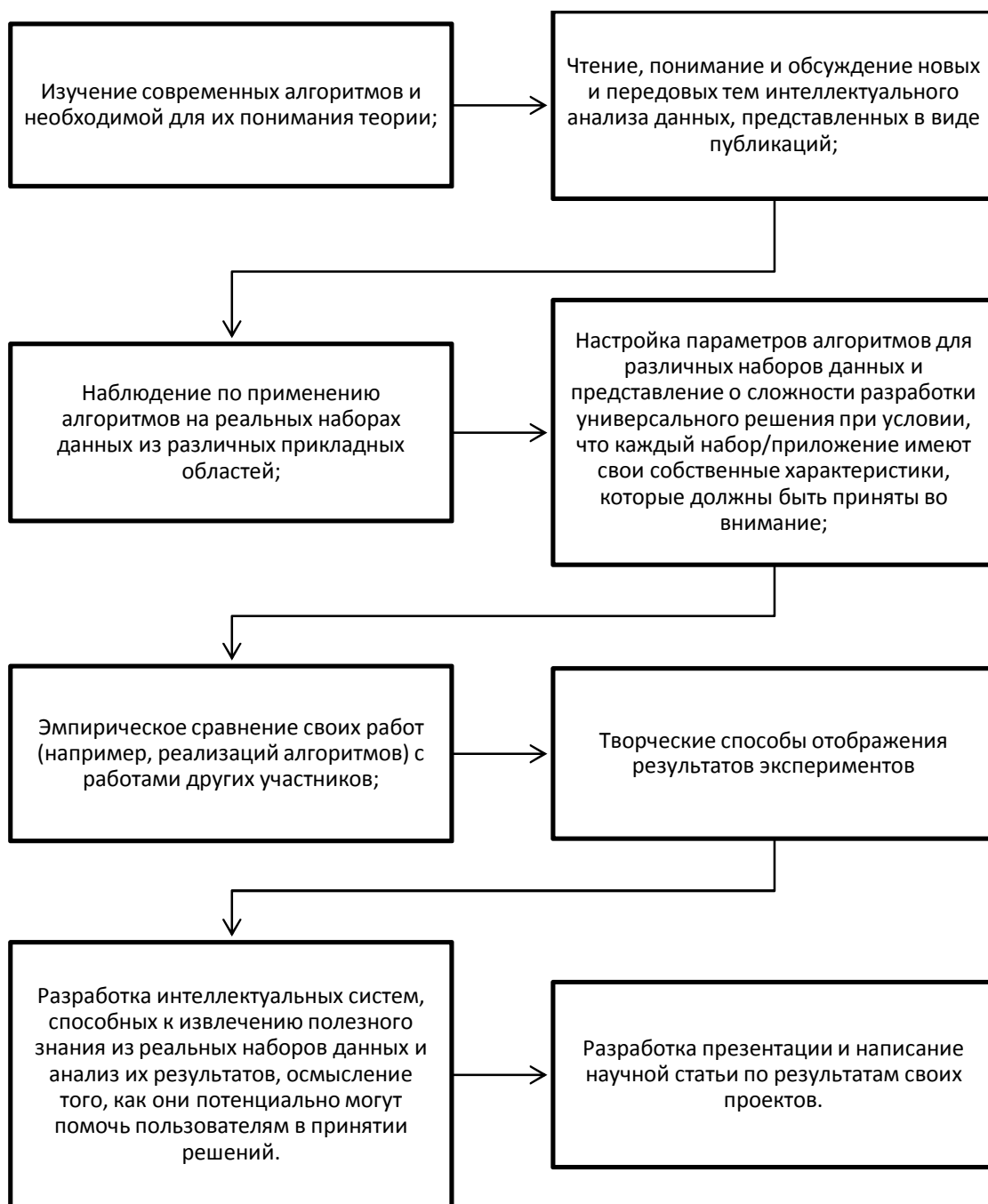


Рис. 1. Индивидуальные и групповые студенческие активности для формирования НИК средствами ИАД



Рис. 2. Подходы к организации работы в ИАД студентов в соответствии с учебной целью

Показателен опыт разработки и проведения курса по ИАД с целью вовлечения студентов в экспериментальные вычислительные исследования, которые описывает Имад Рахаль, американский ученый и профессор университета Святого Джона (США) [16]. Студентам предлагались спланированные индивидуальные и групповые активности, показанные на рисунке 1. Этот опыт оказался полезным при разработке инновационных методик преподавания обсуждаемой дисциплины в соответствии с поставленными учебными целями и представлен рисунке 2.

Заключение

Таким образом, наличие обязательного курса по интеллектуальному анализу данных поможет развитию научно-исследовательских компетенций у обучающихся различных специализаций. Для приобретения навыков в области анализа данных для не ИТ-специалистов рекомендуется сосредоточиться на демонстрации возможностей интеллектуального анализа данных с помощью существующих инструментов (в различных практиках использовались Weka, SAS Enterprise Miner, надстройки Excel, Matlab и др.). А учитывая сложность материала для непрофильных специалистов, знакомство с интеллектуальным анализом данных необходимо осуществлять сразу после изучения обязательного курса математической статистики, причем, для повышения ценностно-мотивационного компонента НИК следует демонстрировать и выполнять анализ на наборах данных из профессиональной области.

Список литературы

1. Баргесян А.А., Куприянов М.С., Холод И.И., Тесс М.Д., Елизаров С.И. Анализ данных и процессов: учеб. пособие. – 3 изд. – СПб.: БХВ-Петербург, 2009.
2. Беляев В.В. Использование методов интеллектуального анализа данных при обучении магистров направления подготовки «Экономика» // Современное образование: содержание, технологии, качество: Материалы XXII международной научно-методической конференции. – СПб.: Изд-во: СПбГУ «ЛЭТИ» им. В.И. Ульянова (Ленина), 2016. – С. 21-23.
3. Беляева Л.Н., Пиотровская К.Р. Сетевой инструментарий лингвиста. Материалы для учебно-методического сопровождения дисциплины. Часть 1. Санкт-Петербург: ООО «Книжный дом», 2014. – 48 с.
4. Гущинский А.Г., Гальченко М.И., Анашкина М.И. Data Mining и статистические методы при оценивании качества курсов повышения квалификации // Научное обеспечение развития АПК в условиях реформирования: Сборник научных трудов по материалам Международной научно-практической конференции профессорско-преподавательского

состава. Министерство сельского хозяйства Российской Федерации, Санкт-Петербургский государственный аграрный университет. – 2014. – С. 323-326.

5. Жуков Н.Н. О содержании обучения будущих инженеров-программистов Дата-майнингу // Ученые записки Института социальных и гуманитарных знаний. – 2015. – № 1. – С. 202-208.

6. Кислова О.Н. Преподавание интеллектуального анализа данных студентам-социологам //Материалы всероссийской социологической конференции «Образование и общество», 20–22 октября 2009. – М.: Изд-во «Институт социологии РАН». – С. 71-76.

7. Комарова Ю.А. Научно-исследовательская компетентность специалистов: функционально-содержательное описание // Известия Российского государственного педагогического университета им. А.И. Герцена. – 2008. – № 68. – С. 69-77

8. Лудченко А.А., Лудченко Я.А., Примак Т.А. Основы научных исследований: учеб. пособие. 2 изд. – К.: О-во «Знание», 2001.

9. Нестеров С.А. Об обучении технологиям интеллектуального анализа данных в рамках курса «Корпоративные системы баз данных» //Преподавание информационных технологий в Российской Федерации / Материалы XIV открытой Всероссийской конференции. – М.: Общество с ограниченной ответственностью «1С-Паблишинг», 2016. – С. 78-79.

10. Смирнов А.В. Формирование практических навыков в области DataMining с помощью курсового проектирования студентов // Современное образование: содержание, технологии, качество. – 2015. – № 2. – С. 169-170.

11. Государственная программа «Информационное общество» (2011–2020 годы). Электронный ресурс: <http://minsvyaz.ru/ru/activity/programs/1/#section-docs>.

12. David A. Dickey. Teaching Data Mining in a University Environment // SAS Users Group International 30, 2005.

13. Giudici P., Figini S. Applied Data Mining for Business and Industry. 2-nd Edition. John Wiley & Sons, 2009.

14. Han J., Kamber M. Data Mining: Concepts and Techniques, 2nd ed. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor, 2006.

15. Maiorana F. A teaching experience on a data mining module //Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, Wrocław, Poland, 9–12 Sept. 2012.

16. Rahal I. Undergraduate research experiences in data mining //Proceedings of the 39th SIGCSE technical symposium on Computer science education, Portland, OR, USA. ACM SIGCSE Bulletin, 2008, vol. 40, № 1, pp. 461-465.

17. Saquer J. A data mining course for computer science and non-computer science students // Journal of Computing Sciences in Colleges, 2007, vol. 22, № 4, pp. 109-114.
18. Tan P., Steinbach M., Kumar V. Introduction to data mining. Boston: Pearson Education, 2006.