

## **КОНЦЕПЦИЯ**

**Образовательной программы магистратуры «Интеллектуальный анализ данных»  
направление 010400.68 «Прикладная математика и информатика»,**

НИУ ВШЭ – Нижний Новгород

Нижний Новгород  
2014

## 1. Общая характеристика

- Основная образовательная программа
- Магистерская
- Очная форма обучения
- 25 бюджетных мест
- Академическая и практико-ориентированная
- Языки преподавания: русский и английский
- Руководители и разработчики программы:

Программа «Интеллектуальный анализ данных», руководитель всей программы Бацын Михаил Владимирович (mbatsyn@hse.ru)

Специализации:

- Анализ данных и задачи оптимизации (руководитель Бацын Михаил Владимирович, mbatsyn@hse.ru)
- Компьютерная лингвистика (руководитель Сибирцева Вера Григорьевна, vsibirtseva@hse.ru)

## 2. Актуальность, цели и задачи.

Актуальность магистерской программы «Интеллектуальный анализ данных», включающей две специализации: «Анализ данных и задачи оптимизации» и «Компьютерная лингвистика» обусловлена рядом взаимосвязанных факторов:

- большим числом прикладных задач, возникающих в реальной практике, которые относятся к задачам анализа данных и задачам оптимизации.
- развитием вычислительных средств, позволяющих накапливать и обрабатывать большие массивы информации и эффективно решать сложные прикладные задачи.
- запросами современного информационного общества, в котором нарастает значимость экономики знаний, развивается рынок высоких технологий и инновационных производств.
- необходимостью овладения новыми инструментами и новыми компетенциями в управлении информационными потоками и анализе информации.
- функционированием российских университетов в глобальной образовательной среде, развитием в ведущих европейских и американских вузах в рамках магистерских программ в сфере “applied linguistics”.

Основные цели и задачи магистерской программы – обеспечить выпускников фундаментальными знаниями во многих областях современной прикладной математики, включая анализ данных, машинное обучение, модели и алгоритмы решения задач оптимизации, теорию игр, технологии анализа сетевых структур, современную теорию вероятностей, математическую статистику, задачи обработки языковых данных, прикладные задачи компьютерной и корпусной лингвистики. Выпускники магистратуры обладают необходимыми знаниями и способны разрабатывать новые математические модели и эффективные алгоритмы для решения сложных задач, возникающих в реальной

жизни. Выпускники программы будут востребованы во всех развитых странах и регионах мира.

В нашу эру информации сотрудники со знаниями в области анализа данных и алгоритмов оптимизации востребованы во многих сферах деятельности. В сфере информационно-коммуникационных технологий, как одной из наиболее быстро развивающихся областей в наше время, зарплаты сотрудников со знаниями в области анализа данных и алгоритмов оптимизации в разы превышают зарплаты рядовых разработчиков программного обеспечения. Наиболее заинтересованы в таких сотрудниках современные крупные компании, такие как Yandex, Google, Intel, Kaspersky Lab, Microsoft, Apple, Samsung и др.

Анализ российского в целом и рынка труда провинции в частности (как наиболее полно отражающего современное состояние IT-рынка) демонстрирует потребность в специалистах в области компьютерной лингвистики, в особенности ее технической составляющей. Ср. рынок Европы и США, в университетах которых существуют многочисленные программы «Теоретическая компьютерная лингвистика и математическая лингвистика» (Theoretical Computational Linguistics and Mathematical Linguistics), «Прикладная компьютерная лингвистика» (Applied Computational Linguistics), «Текстовые технологии» (Text Technology), «Формальная семантика», «Когнитивистика». Кроме того, отсутствие аналогичных предложений на образовательных рынках в крупных регионах, соседних с Нижегородской областью, может привлечь в магистратуру иногородних бакалавров, желающих вернуться на работу в родной город.

Выпускники специализации «Компьютерная лингвистика» найдут применение в таких сферах как моделирование и аналитика коммуникативных ситуаций, мониторинг массивов информации с использованием современных компьютерных инструментов лингвистического исследования, разработка и применение интеллектуальных программных систем поиска и анализа текстовых данных в сети Интернет.

**3. Целевой аудиторией** программы являются студенты, имеющие, в первую очередь, степень бакалавра по направлениям «Прикладная математика и информатика», «Математика и компьютерные науки», «Бизнес-информатика», «Фундаментальная и прикладная лингвистика». Планируемое число студентов – 25. К основным критериям набора студентов следует отнести высокий средний балл по техническим дисциплинам в бакалавриате, высокий уровень базовых знаний в области математических и компьютерных наук, желание решать сложные прикладные задачи, применяя современные математические и алгоритмические подходы.

**4. Международный и отечественный опыт в избранной сфере, особенности проекта в свете этого опыта.**

Международный и отечественный опыт в данной области связан с деятельностью лаборатории алгоритмов и технологий анализа сетевых структур. Высокий уровень исследований сотрудников лаборатории позволяет привлекать студентов магистратуры к исследовательским и практическим проектам лаборатории, знакомя их с потенциальной научной деятельностью выпускников магистратуры.

**5. Характеристика сегмента рынка образовательных услуг, основные конкуренты, сравнительные преимущества ОП.**

Подготовка магистров по направлениям близким к направлению «прикладная математика и информатика» ведется в нескольких ВУЗах Нижнего Новгорода.

Основными конкурентами являются ННГУ им. Н.И.Лобачевского и НГТУ им. Р.Е. Алексеева.

В ННГУ на факультете ВМК открыт прием на магистерские программы

- Прикладная математика и информатика
- Прикладная информатика
- Фундаментальная информатика и информационные технологии

Факультет ВМК традиционно имеет высокую репутацию в области подготовки специалистов по прикладной математике и информационным технологиям. Вместе с тем, магистерские программы ННГУ пока еще не получили должного развития: на сайте факультета ВМК сложно найти информацию об учебных планах и программах, о трудоустройстве выпускников и о специализациях магистерских программ, не выделены конкретные актуальные научные проблемы, с которыми будут работать студенты магистратуры. Это создает конкурентное преимущество программе ВШЭ, которым необходимо эффективно воспользоваться.

В НГТУ имеется программа подготовки магистров по направлению «прикладная математика и информатика». В качестве направлений научных исследований указаны: развитие новых математических моделей в механике, геофизике, физике плазмы, астрофизике, экономике; проблемы информатики, включая исследования в области создания параллельных алгоритмов; теоретическая и прикладная математика, включая теорию аппроксимаций, исследование некорректных задач; динамика нелинейных волновых процессов в сплошных средах. Это существенно отличает направленность программы от программы ВШЭ, которая ориентирована на современные компьютерные науки и приложения.

## **6. «Портрет выпускника»**

Выпускник специализации «Анализ данных и задачи оптимизации» имеет глубокие фундаментальные и практические знания во многих областях современной прикладной математики, включая анализ данных, машинное обучение, модели и алгоритмы решения задач оптимизации, теорию игр, технологии анализа сетевых структур, современную теорию вероятностей и математическую статистику. Выпускники магистратуры обладают необходимыми знаниями и способны разрабатывать новые математические модели и эффективные алгоритмы для решения сложных задач, возникающих в реальной жизни. Обладая глубокими знаниями в области анализа данных и алгоритмов оптимизации, выпускники востребованы во многих сферах деятельности, включая прикладные научные исследования лабораторий при крупных современных компаниях, информационно-коммуникационные технологии, бизнес-аналитику, транспортную и складскую логистику и др.

Конкурентоспособность выпускников магистратуры специализации «Компьютерная лингвистика» на рынке труда связана прежде всего с широким набором компетенций, приобретаемых в ходе освоения образовательной программы. Выпускник специализации «Компьютерная лингвистика» обладает:

*Общекультурными компетенциями*, связанными с углублением культуры мышления и когнитивных способностей к обобщению и анализу информации; владением английским языком на уровне, необходимом для профессионального общения.

*Профессиональными компетенциями*, которые позволят магистру успешно осуществлять разработку, внедрение и поддержку проектов, созданных с помощью

информационных технологий в компаниях, ориентированных на создание и применение специализированных продуктов и систем (тестирование лингвистических компьютерных технологий, создание электронных библиотек, архивов, баз данных);

**Научно-исследовательскими компетенциями,** формирующими навыки аналитической работы со сложными текстами различной знаковой природы; создание, интерпретацию и обработку текстов широкой тематики и различной степени сложности;

**Проектными и производственно-практическими компетенциями,** необходимыми для успешной командной работы по анализу, поиску информации и созданию баз данных.

**Институциональными компетенциями,** характеризующими способности выпускника-магистра применять на практике теоретические знания, необходимые для внедрения лингвистических технологий в электронные информационные системы, системы машинного перевода.

**Социально-личностными компетенциями,** призванными подготовить выпускника к эффективной публичной деятельности, в том числе и в качестве преподавателя университета.

## 7. Структура учебного плана.

Магистерские программа составлена в соответствии с Оригинальным образовательным стандартом НИУ ВШЭ по направлению «Прикладная математика и информатика» и формируется из дисциплин программы, научно-исследовательских семинаров и научно-исследовательской работы. Программа предполагает двухгодичное обучение. При этом второй семестр второго года обучения посвящен научно-исследовательской работе студентов, включающей, как результат, подготовку и защиту магистерской диссертации.

Программа ориентирована на подготовку выпускников, способных:

- создавать математические модели для задач оптимизации и анализа данных, возникающих в реальной практике, на основе построенных моделей разрабатывать эффективные алгоритмы для решения таких задач, реализовывать разработанные алгоритмы в виде программного обеспечения, применяемого в реальной деятельности.
- анализировать существующие подходы в любых сферах человеческой деятельности на основе имеющихся данных, строить математические модели, делать выводы об эффективности анализируемой деятельности, предлагать решения, повышающие эффективность.
- применять на практике знания в области автоматической обработки текста и программирования;
- организовывать и проводить различные исследования с целью разработки и создания инновационных электронных языковых ресурсов (корпусов текстов, словарей, фонетических, лексических, терминологических баз данных);
- осуществлять мониторинг массивов информации, разрабатывать краткосрочные и долгосрочные прогнозы в различных предметных областях с использованием современных компьютерных инструментов лингвистического исследования.

Основные курсы магистерской программы представлены дисциплинами «Современные методы анализа данных», «Современные методы принятия решений», «

Стохастические модели», «Дискретные модели и сложность алгоритмов», «Дискретная оптимизация и исследование операций», «Модели корпусной лингвистики», «Технологии работы с большими массивами данных», «Прикладные задачи анализа сетевых структур», «Машинное обучение», «Методы и алгоритмы распознавания речи», «Автоматическая обработка текстов», «Дополнительные главы методов оптимизации».

Для выпускников других вузов предусмотрена возможность слушать адаптационные курсы. Адаптационные курсы включают «Статистические методы обработки данных» и «Иностранный язык как инструмент научной работы».

Программа подготовки магистров включает курсовую работу, научно-исследовательскую практику и магистерскую диссертацию. Программу подготовки магистров обеих специализаций «Анализ данных и задачи оптимизации» и «Компьютерная лингвистика» дополняют научно-исследовательские семинары «Современные задачи прикладной математики и компьютерной лингвистики» и «Методы интеллектуального анализа данных», непосредственно посвященные современным задачам анализа данных, исследования операций и компьютерной лингвистики.

### **8. Концепция научно-исследовательского семинара.**

Научный семинар является одной из форм работы со студентами, которой в магистерской программе отводится приоритетное место. Главная цель научного семинара состоит в выработке у студентов умений и навыков исследовательской деятельности, необходимых для подготовки магистерской диссертации. В ходе работы на семинарах происходит непосредственная передача от преподавателя к студенту опыта научно-исследовательской работы.

Семинар «Методы интеллектуального анализа данных» охватывает наиболее важные темы анализа данных, включающие методы и алгоритмы классификации, задачи распознавания изображений и речи, существующее программное обеспечение, необходимое для решения различных задач анализа данных. Семинар «Современные задачи прикладной математики и компьютерной лингвистики» рассматривает как классические, так и прикладные задачи оптимизации, возникающие в реальной практике, включая сложные задачи маршрутизации транспорта с множеством ограничений, задачи оптимизации в складской логистике, задачи планирования производства и др. Магистранты также знакомятся с современными математическими методами и компьютерными алгоритмами анализа и синтеза речи. На материале конкретных проектов, осуществляемых учеными НИУ ВШЭ – Нижний Новгород и приглашенными специалистами, рассматриваются современные методы и средства компьютерной лингвистики. Ряд тем в семинаре обсуждаются на английском языке. Предполагается участие в работе семинара приглашенных специалистов из университетов Европы и США, что способствует интернационализации магистерской программы.

### **9. Характеристика кадрового потенциала**

Основные курсы, научно-исследовательские семинары и научно-исследовательские проекты ведут доктора физико-математических и технических наук: Калягин В.А. (ординарный профессор), Чистяков В.В. (ординарный профессор), Колданов А.П., Малышев Д.С., Крылов В.В., доктора филологических и политических наук: Зусман В.Г., Бухаров В.М., Цветкова М.В., Романова Т.В., Гронская Н.Э., а также кандидаты физико-математических наук: Карпов Н.В., Асеева Н.В., Бацын М.В., и кандидаты филологических наук: Сибирцева В.Г., Ляшевская О.Н., Бонч-Осмоловская А.А.

Курс «Машинное обучение» читает Лысенков И.Д., профессионально занимающийся в компании ItSeez задачами компьютерного зрения и машинного обучения, победитель конкурса «Интернет-математика: Яндекс.Карты» (см. статью с описанием алгоритма победителей: <http://habrahabr.ru/company/itseez/blog/158645/>), выпускник Школы Анализа Данных Яндекса.

Курс «Прикладные задачи анализа сетевых структур» читает Пономаренко А.А., сотрудник лаборатории алгоритмов и технологий анализа сетевых структур, профессионально занимающийся исследованиями в области сетевых структур для оптимальной организации поиска в больших массивах данных.

Все преподаватели активно занимаются научной деятельностью и имеют публикации в лучших российских и зарубежных журналах с высоким импакт-фактором. Каждый преподаватель руководит курсовыми и выпускными работами бакалавров и магистров, активно привлекая студентов к научной и проектной деятельности.