

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
 ВЫСШЕГО ОБРАЗОВАНИЯ
 САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
 ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА
 (САМАРСКИЙ УНИВЕРСИТЕТ)

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
 ВЫСШЕГО ОБРАЗОВАНИЯ
 САМАРСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ
 МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
 (ФГБОУ ВО СамГМУ Минздрава России)

СОГЛАСОВАНО

И.о. директора института дополнительного
 образования


 _____ (Аржанов А.П.)
 " " _____ 2018г.

УТВЕРЖДАЮ

Проректор по учебной работе


 _____ (Гаврилов В.А.)
 " " _____ 2018г.

СОГЛАСОВАНО

Директор Института инновационного развития
 ФГБОУ ВО СамГМУ Минздрава России


 _____ (Колсанов А.В.)
 " " _____ 2018г.

УТВЕРЖДАЮ

Директор института профессионального
 образования, проректор по лечебной работе
 ФГБОУ ВО СамГМУ Минздрава России


 _____ (Корымасов Е.А.)
 " " _____ 2018г.

ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА
 ДОПОЛНИТЕЛЬНОГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
 (ПОВЫШЕНИЯ КВАЛИФИКАЦИИ)

«Методы и технологии обработки большого объёма данных в медицине»

(Big Data in Medicine)

**СО СРОКОМ ОСВОЕНИЯ
 72 АКАДЕМИЧЕСКИХ ЧАСА ПО СПЕЦИАЛЬНОСТИ
 31.08.09 РЕНТГЕНОЛОГИЯ**

Самара
 2018

ОБЩИЕ ПОЛОЖЕНИЯ

Цель подготовки по программе:

Формирование профессиональных компетенций в области обработки и анализа сверхбольшого объёма медицинских данных.

Использование технологий обработки и анализа больших данных позволяет не только расширить возможности текущих источников данных, но и сформулировать принципиально новые подходы к решению комплексных междисциплинарных фундаментальных и прикладных задач в медицине.

В настоящее время большинство методов диагностики в медицинских компьютерных системах основаны на визуальном наблюдении и оценивании диагностических объектов на изображениях. Такой подход не использует в полной мере все возможности, предоставляемые современной компьютерной техникой. Информационные технологии поддержки принятия решения врачом-диагностом, позволяют повысить качество и информативность медицинских исследований, помогают решать задачи высокоточной диагностики на ранних стадиях заболеваний и способствуют развитию отечественной научной базы для разработки современного программного обеспечения. Системы интеллектуального анализа большого объёма данных позволят перейти на качественно новый уровень принимаемых решений и обеспечат решающие конкурентные преимущества за счёт анализа всех доступных данных и прогнозирования последствий принимаемых решений. По оценкам ведущих мировых компаний потребность в специалистах данного профиля составляет несколько сот тысяч человек.

Программа предполагает формирование у обучающихся основ теоретических знаний в области сбора, хранения, обработки, поиска, анализа, визуализации данных сверхбольшого объёма; теоретических знаний и практических навыков в области использования программной инфраструктуры распределённых вычислений Hadoop; теоретических знаний об основных архитектурах нереляционных и графовых баз данных; основ теоретических знаний и практических навыков в области разработки распределённых систем переработки данных сверхбольшого объёма с использованием парадигмы MapReduce и технологий потоковой обработки данных. Также обучающийся получит базовые знания о технологиях обработки BigData и технологиях машинного обучения Deep Learning в области обработки медицинских данных, возможностях нейронных сетей по автоматическому распознаванию медицинских изображений. Получит практические навыки по созданию нейронных сетей для обработки медицинских изображений на базе TensorFlow, Caffe, NVidia Digits.

С целью овладения указанными профессиональными компетенциями обучающийся в ходе освоения программы должен:

приобрести практический опыт (приобрести навыки выполнения трудовых действий):

- развертывания и использования программной инфраструктуры распределенных вычислений Hadoop;
- разработки программ распределённой обработки данных с использованием парадигмы MapReduce;
- разработки и использования пакетов построения нейронных сетей TensorFlow, Caffe, Nvidia Digits для обработки медицинских данных.

приобрести умения:

- применять концепцию MapReduce для распределённой обработки данных;
- применять инструментальные средства проектирования и разработки программ распределённой обработки данных с использованием парадигмы MapReduce;
- применять технологии Deep Learning для обработки медицинских данных и изображений.

получить знания:

- теоретические знания о проблематике и применимости современных информационных технологий интеллектуального анализа данных сверхбольшого объёма;
- основы теоретических знаний в области сбора, хранения, обработки, поиска, анализа, визуализации данных сверхбольшого объёма;
- теоретические знания в области использования программной инфраструктуры распределенных вычислений Hadoop;
- основы теоретических знаний в области разработки распределённых систем переработки данных сверхбольшого объёма с использованием парадигмы MapReduce;
- основы теоретических знаний в области машинного обучения и обработки медицинских данных и изображений.

Требования к поступающим на программу:

Для удовлетворительного освоения программы обучающиеся должны иметь базовые знания по технологиям обработки данных, цифровой обработке сигналов, архитектуре распределённых программных систем, иметь навыки программирования на языке JavaSE, желательно иметь базовый опыт работы с SQL-совместимыми СУБД

Срок обучения: 72 часа

Форма обучения: очная

Планируемые результаты обучения:

Обучение по программе предполагает освоение или совершенствование в рамках изучаемых модулей профессиональных компетенций:

Категория работника	Профессиональные компетенции (ПК) (образовательный результат)
Инженер	Интеллектуальный анализ больших массивов неструктурированной медицинской информации с помощью информационных технологий распределённой обработки сверхбольших объёмов данных.

УЧЕБНО - ТЕМАТИЧЕСКИЙ ПЛАН

№	Наименование разделов, дисциплин, тем, модулей	Всего, час.	В том числе			Форма контроля
			Лекции	Практические занятия	Индивидуальные консультации	
1	Модуль 1 <i>Понятие, проблемы и вызовы больших данных</i>	4	4			Опрос
1.1	Определения больших данных, технологии и архитектуры	2	2			
1.2	Новая методология работы с данными. Примеры инновационного использования технологий больших данных	2	2			
2	Модуль 2 <i>Основы программной инфраструктуры распределённых вычислений Hadoop</i>	20	10	10		Контрольное задание
2.1	Парадигма распределённых вычислений MapReduce	2	2			
2.2	Основные определения Hadoop и распределённая файловая система HDFS	4	2	2		
2.3	Основы разработки MapReduce-приложений в среде Hadoop	4	2	2		
2.4	Базовые алгоритмы и шаблоны проектирования (паттерны) в парадигме MapReduce	4	2	2		
2.5	Основные технологии потоковой обработки данных	6	2	4		
3	Модуль 3 <i>Анализ данных на языке программирования R</i>	12	6	6		Контрольное задание
3.1	Основные методы интеллектуального анализа данных на языке программирования R	4	2	2		
3.2	Анализ временных рядов	4	2	2		
3.3	Интеграция с базами данных и технологией Hadoop	4	2	2		

№	Наименование разделов, дисциплин, тем, модулей	Всего, час.	В том числе			Форма контроля
			Лекции	Практические занятия	Индивидуальные консультации	
4	<i>Модуль 5 Графовые модели представления, хранения и обработки данных</i>	14	8	6		Контрольное задание
4.1	Графовая модель данных	2	2			
4.2	Основы вычислений на графах в парадигме MapReduce	4	2	2		
4.3	Графовая база данных Neo4j	4	2	2		
4.4	Основы аналитики на графах	4	2	2		
5	<i>Модуль 5 Обработка медицинских данных и изображений с использованием технологий машинного обучения</i>	18	10	8		Контрольное задание
5.1	Основные методики обработки медицинских данных	2	2			
5.2	Использование Caffe для обработки медицинских изображений	4	2	2		
5.3	Варианты практического использования Nvidia Digits для медицинских данных	4	2	2		
5.4	Варианты построения нейронных сетей на базе TensorFlow в области медицины	8	4	4		
	Итоговая аттестация	4			4	Итоговое тестирование
	Итого	72				

УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

Материально-технические условия:

Материально-технические условия реализации программы	Обеспеченность реализации программы собственными материально техническими условиями (указать наименование, год выпуска используемого оборудования)
<p>Наличие кабинетов:</p> <ul style="list-style-type: none"> – аудитория для проведения лекционных занятий, имеющая необходимое количество посадочных мест и оснащённая оборудованием для проведения презентаций (ноутбук, проектор); – аудитория с персональными компьютерами для проведения лабораторных (практических) занятий, имеющая необходимое количество рабочих мест, оснащенных необходимым системным и программным обеспечением. 	<p>Лекционные демонстрации с помощью мультимедиапроектора, персональные компьютеры (не ниже 2012 года)</p>
<p>Наличие лабораторий (указать каких):</p> <ul style="list-style-type: none"> - лаборатория обработки данных сверхбольшого объёма (Big Data) 	<p>Программно-аппаратный комплекс для обработки структурированных и неструктурированных данных сверхбольшого объёма в составе:</p> <ul style="list-style-type: none"> – программно-аппаратный комплекс (подсистема) хранения и аналитического анализа структурированных данных IBM Puredata for Analytics (Netezza) с объёмом дискового пространства без учета сжатия не менее 96ТБайт (с учетом 4-х кратного сжатия данных); – комплекс серверов IBM System X для подсистемы распределённого хранения и аналитической обработки неструктурированных данных, в т.ч. сервер управления IBM x3630 M4 (два процессора Intel Xeon Processor E5-2450v2; 96 Гбайт памяти; 2 диска по 600ГБ) и четыре сервера обработки данных IBM x3630 M4 (два процессора Intel Xeon Processor E5-2450v2; 96 Гбайт оперативной памяти; 8ТВ дисковой памяти).
<p>Наличие технических средств обучения</p>	<p>Мультимедиапроектор.</p>

Информационное обеспечение образовательного процесса:

а) основная литература

1. Том Уайт. Hadoop. Подробное руководство – СПб.: Питер, 2013. – 672 с.
2. Чак Лэм Hadoop в действии / Перев. с англ. – М.: ДМК Пресс, 2012.

б) дополнительная литература

1. Doherty J, Noiro LA, Mayfield J, Ramiah S, Huang C, Dunagan WC, Bailey TC. Implementing GermWatcher, an enterprise infection control application. AMIA Annu Symp Proc.2006:209.
2. Fukushima Kunihiko Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. // Biological Cybernetics 1980.
3. Gardner RM, Pryor TA, Warner HR. The HELP hospital information system: update 1998. Int J Med Inf. 1999 June; 54(3):169.
4. Glenn Edwards, Paul Compton, Ron Malor, Ashwin Srinivasan, Leslie Lazarus. Peirs: A pathologist maintained expert system for the interpretation of chemical pathology reports. Pathology. 1993, Vol. 25, No.1, pp. 27-34.
5. Большие Данные в медицине и здравоохранении (Интервью Билла Маккормика, менеджера по продуктам компании InterSystems). Автор: Ольга Ревякина. Дата публикации 21.04.2014, Источник: Издательство «Открытые Системы» [Электрон, ресурс].
URL: <http://www.osp.ru/medit/2014/04/13040834.html>.
6. Leading Advances in the Utilization of Big Data in the Healthcare Industry. White Paper Intel Health & Life Sciences [Электрон, ресурс]. - URL: <http://www.intel.ru/content/dam/www/public/us/en/documents/white-papers/big-data-healthcare-tokyo-paper.pdf>.
7. Monyiko I, Chui M., Brown B. etc., Big Data: The Next Frontier for Innovation, Competition, and Productivity/ McKinsey Global Institute, 2011.
8. Big Data - технология, рождающая новый тип бизнеса // Бизнес & информационные технологии. — 2014 — Выпуск № 3 (36) [Электронный документ]. — URL: [/http://bit.samag.ru/archive/article/1352# maintitle](http://bit.samag.ru/archive/article/1352# maintitle).

в) программное обеспечение и Интернет-ресурсы

пакеты прикладных программ

1. IBM Infosphere BigInsights – программное обеспечение распределённой обработки неструктурированных данных.
2. Apache Hadoop – программная инфраструктура распределённой обработки неструктурированных данных.
3. Apache Spark – система распределённой обработки потоковых данных.
4. R – система программирования на языке R.
5. TensorFlow — открытая программная библиотека для машинного обучения.
6. Caffe – среда для глубинного обучения.
7. NVidia Digits – среда глубинного обучения на GPU.

Интернет-ресурсы

1. <http://www.citforum.ru/> (сайт аналитической информации).
2. <http://www.ibm.com/developerworks/ru/> (ресурс для разработчиков и IT-профессионалов).
3. <http://www.apache.org> (сайт группы проектов с открытым кодом).
4. <http://caffe.berkeleyvision.org> (Caffe - deep learning framework).
5. <https://developer.nvidia.com/digits> (NVidia Digits сайт для разработчиков).
6. <https://www.tensorflow.org/> (Сайт проекта с открытым кодом).

Составители программы

Д.т.н., профессор
кафедры технической кибернетики

Куприянов А.В.

Д.т.н., профессор
кафедры технической кибернетики

Попов С.Б.

К.м.н, доцент
кафедры оперативной хирургии
и клинической анатомии с курсом
инновационных технологий СамГМУ

Воронин А.С.

Заведующий лабораторией систем анализа и
обработки медицинских данных
отдела высокопроизводительных вычислений и
технологий искусственного интеллекта в медицине
центра прорывных исследований в медицине СамГМУ

Гусев В.Н.