

УДК 004.6-024

ТЕНДЕНЦИЯ ВЫБОРА ЗАДАНИЙ МАГИСТРАНТАМИ В СРЕДЕ BIG DATA



И.И. Пилецкий

Доцент кафедры информатики БГУИР, кандидат физико-математических наук, старший научный сотрудник

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

E-mail: ianmenski@gmail.com, kozub@bsuir.by



В.Н. Козуб

Аспирант кафедры информатики БГУИР, магистр технических наук, ассистент кафедры информатики БГУИР

И.И. Пилецкий

Кандидат физико-математических наук, доцент, доцент кафедры информатики БГУИР, научный руководитель совместной лаборатории БГУИР — ИВА и Академического центра компетенций технологий IBM. Имеет большой опыт в реализации и разработке промышленных решений в ИТ-области, являлся ведущим разработчиком, системным архитектором, руководителем и научным руководителем нескольких десятков крупных проектов, связанных с разработкой программного обеспечения и баз данных корпоративного уровня. Автор более 80 научных публикаций (в том числе монографий и учебных пособий) в области моделирования, технологии разработки программного обеспечения и баз данных. Член редакционной коллегии журнала «Baltic Journal of Modern Computing». Область научных интересов: технологии и аналитические комплексы анализа Big Data, NPL и ML алгоритмы.

В.Н. Козуб

Аспирант, ассистент каф. информатики БГУИР, имеет 5 публикаций тезисов докладов на научной конференции аспирантов, магистрантов и студентов БГУИР, 4 публикации в материалах международных конференций, 1 методическое пособие, ведет лабораторные занятия по курсам "Архитектурные решения для обработки больших объемов информации", "Модели и методы обработки и анализа больших объемов информации", "Технологическая платформа по управлению большими данными".

Аннотация. В данной работе рассмотрены методы и подходы в организации обучения студентов технологиям Big Data, приведён анализ результатов выполнения магистерских диссертаций и практических заданий, приведена тенденция выбора выполненных работ магистрантами.

Ключевые слова: технологии обучения, hard skills интернет-источники, Big Data, Cloud Computing, Machine Learning, Natural Language Processing.

1. Перспективные вакансии и востребованные навыки

Самые перспективные вакансии по данным LinkedIn [1, 2] являлись следующие смотрите таблицу 1. В этом списке указаны рабочие места с высокой средней заработной платой, хорошими вакансиями и годовым ростом, а также рабочие места, которые, скорее всего, приведут к продвижению по службе или продвижению в организации.

Анализ данных LinkedIn (более 500 миллионов человек) и требований различных компаний (о приеме на работу) позволил определить, какие навыки наиболее трудно было найти на январь 2018 года и получить требуемые в 2019 году.

Таблица 1

Нужны в 2018	Нужны в 2019
1. Cloud and Distributed Computing	Cloud Computing
2. Statistical Analysis and Data Mining	Artificial Intelligence
3. Middleware and Integration Software	Analytical Reasoning
4. Web Architecture and Development Framework	People Management
5. User Interface Design	UX Design
6. Software Revision Control Systems	Mobile Application Development
7. Data Presentation	Sales Leadership
8. SEO/SEM Marketing	Translation
9. Mobile Development	Audio Production
10. Network and Information Security	Natural Language Processing

Дополнительно представители LinkedIn измеряли спрос путем **анализа навыков, указанных в профилях самых высокооплачиваемых специалистов**. В выборку при этом включили только те города, где живет не меньше 100 тыс. человек. По данным LinkedIn, сейчас в мире существует порядка 50 тыс. профессиональных навыков.

Самые востребованные так называемые hard skills:

- облачные вычисления,
- искусственный интеллект,
- принятие решений на основе данных,
- управление людьми,
- UX-дизайн.

Если проследить тенденцию нескольких последних лет, то три первые строчки всегда были среди самых востребованных навыков [3].

По данным сайта Парка высоких технологий (на 31.01.19), в ПВТ зарегистрировано 1367 компаний, которые также нуждаются в IT специалистах (428 вакансий) [4]. Если несколько лет назад, требования к специалистам были традиционные для разработчиков и сопровождения ПО, то в настоящее время наметилась тенденция нехватки высоко квалифицированных специалистов, знающих платформу, архитектуру и аналитические методы обработки неструктурированных и мультимедийных данных, знающие технологии IOT и связанных с ними практических решений.

Так, появились вакансии для специалистов, разрабатывающие мобильные приложения на основе компьютерного зрения, знающие технологии разработки ПО на основе искусственного интеллекта и компьютерного зрения, технологии идентификации лиц, жестов и объектов, технологии управления программными роботами, применения AI-технологий, разработки приложений на PaaS-платформе и др. Просматривается общая тенденция к комплексным знаниям для специалистов в некоторой области. Хорошему специалисту предлагается достойная зарплата.

Одними из наиболее перспективными направлениями в настоящее время и недалеком будущем являются: когнитивные системы, приложения и сервисы, NPL и ML –аналитика, (системы типа Watson + Internet of Things) – Интернет вещей как сеть взаимодействия M2M (Machine-to-Machine) позволяет создать безлюдное производство, применение беспилотных машин и роботов в армии и на производстве, «Разумные» дома и города, виртуальные глобальные организации, магазины без товаров, логистические предприятия без подвижного состава и т.д. [3].

В IT-сфере появилась новая проблема как подготовить очень высоко квалифицированных специалистов, которые обладали современными навыками и адаптации образовательных университетских программ к современным требованиям. Высоко квалифицированного специалиста трудно подготовить за время обучения по программам первой ступени

обучения, обязательно нужна целенаправленная подготовка в магистратуре, нужна база и время для подготовки высококвалифицированного специалиста.

Но, современные требования к специалистам должны отражаться в учебных программах, в основных требованиях:

- в программах должен содержать современный контент в IT;
- образовательные программы должно помогать студентам в карьерном плане;
- новые подходы к образованию, которые связаны с современными технологическими возможностями к обучению, должны использоваться студентами.

2. База для подготовки специалистов

2.1. Одно из направлений обучения направление (на кафедре информатики БГУИР факультета КСиС) связано с *обучением студентов в совместных лабораториях, так в лаборатории БГУИР-ИВА* для обучения работы на платформе Big Data обучение проводится на облачной платформе ЦОД БГУИР и платформе IBM Bluemix [5, 6].

Обучение в совместной лаборатории строится на основе регулярно уточняющихся программ, подготовленных с использованием материалов, подготовленных специалистами IBM и преподавателями БГУИР. Здесь важным является, то что в данной лаборатории можно проводить экспериментальные работы в интересах перспективных работ в организации.

Инфраструктура лаборатории используется при обучении магистрантов по специальности «Обработка больших объёмов информации» и для проведения спецкурса для отобранной группы студентов 2-3 курсов, проводятся 50 часовые ознакомительные спецкурсы в области Big Data. Отбор проводится совместно с организацией ИВА Минск.

Спецкурс для студентов по Big Data строится по тем же принципам, что и обучение в магистратуре, однако с учётом одного существенного фактора: ограничение периода обучения – 50 часов и за один месяц. В начале курса лекций студентам предлагается придумать проект (варианты предлагаются, но не навязываются), над которым они будут работать, и собрать команду из числа слушателей курса, обычно образуется команд 4-5 для выполнения проектов. Каждая команда работает над конкретным проектом. На каждое занятие команда составляют план своей работы, а по итогам каждого занятия преподаватель проверяет и корректирует работу команды. В ходе занятий каждый член команды работает над своей частью задач, преподаватель отвечает на вопросы команд, вместе с ними формирует список задач, следит за его выполнением и помогает командам находить нужную информацию. На последнем занятии каждая команда презентует результаты своей работы в виде прототипа приложения. Приглашённые эксперты от компании и преподаватели оценивают результаты работы команд.

К сожалению, в настоящее время, ИТ организация живет сегодняшним днем, видит подготовку специалистов нужных сегодня для тушения пожара в организации, ресурсов для ведения перспективных и исследовательских проектов почти не выделяется. Для более глубокой подготовки нужен научно-исследовательский проект и на весь учебный год, который должен иметь ресурсы.

2.2. Главным направлением подготовки специалистов в области Big data, связано с *обучением студентов в магистратуре* кафедры информатики по специальности «Обработка больших объёмов информации». Базой является образовательный стандарт для второй ступени образования (магистратуры) ОСВО 1-40 81 04-2015 «Обработка больших объёмов информации», который был утвержден Министерством образования Республики Беларусь, подготовленный сотрудниками БГУИР.

В этом стандарте предлагается на основе трех *дисциплин государственных компонент*: «Технологическая платформа по управлению большими данными», «Архитектурные решения для обработки больших объёмов информации», «Модели и методы обработки и

анализа больших объёмов информации» разработать университетам свои программы в зависимости от требований свои программы подготовки специалистов, которые при необходимости можно адаптировать со временем к современным требованиям [3].

В БГУИР разработаны три учебные программы: «Технологическая платформа по управлению большими данными», «Архитектурные решения для обработки больших объёмов информации», «Модели и методы обработки и анализа больших объёмов информации» – по которым три года ведется подготовка магистров по специальности 1-40 81 04 «Обработка больших объёмов информации». По данным программам на кафедре информатики обучаются, в общем, семь групп магистрантов очной и заочной формы обучения. Часть из них выбирают диссертационные работы по Big data, практические задания должны выбирать и выполнять все.

В процессе лекционных занятий рассматриваются файловые системы: HDFS и GPFS (General Parallel File System), модель MapReduce & Hadoop 1 и MapReduce & Hadoop 2, платформа Apache Hadoop, Apache Spark, Apache Storm, Cloudera Hadoop. Платформа IBM Big-Insights, IBM Bluemix (Watson), Microsoft Azure, Cloudera Enterprise строительные блоки. Соотношения разных платформ для обработки больших данных. Языки разработки программ для анализа структурированных и не структурированных данных больших объёмов. Облачные платформы для обработки данных, сравнение технологий обработки данных: традиционная, IaaS, PaaS, SaaS. Облачные решения для анализа данных: IBM Bluemix, IBM Watson, MS Azure. Доступ к сервисам. Языки программирования Pig, Jaql, Hive, SQL, R.

Хранение и доступ к данным, NoSQL базы данных. Архитектура NoSQL баз данных, построенных по принципу «ключ-значение» и колоночных баз данных. Характеристика баз данных типа Riak, Amazon Dynamo, Redis, HBase. Документо-ориентированные базы данных типа MongoDB, CouchDB и графо-ориентированные базы данных типа Neo4J.

Методы, средства и аналитические алгоритмы анализа структурированных и неструктурированных данных, описательная, прогнозная и директивная аналитика, методы нахождения паттернов и аномалий, модели и методы визуального представления данных, примеры графической интерпретации больших объёмов информации.

Поэтому общая тематика этих программ позволяет охватить большинство аспектов «Обработки больших объёмов данных».

2.3. Важным аспектом в подготовке специалистов являются выбор и выполнение магистрантами задач для выполнения практических заданий, и выбор и выполнение исследований при написании диссертаций в магистратуре и аспирантуре.

Тематика многих диссертационных исследований включала значимые практический результаты. Ниже, приведены некоторые направления работ:

- проблемы построения ИТ платформы для реализации приложений работающих с Big data (это работы по созданию ЦОД БГУИР и виртуальной лаборатории кафедры информатики [4,5], работа по созданию ИТ платформы для крупного предприятия);
- разработка методов и алгоритмов анализа поведения пользователей сайта с применением платформ потоковой аналитики;
- работы по созданию аналитического приложения, использующего NPL-алгоритмы для фильтрации данных и векторные ML-алгоритмы для анализа имиджа бренда, и определение тональности упоминаний, на основе текстовых данных, получаемых из социальных сетей (проведен анализ более 400 млн. твитов);
- выполняются работы по прогнозированию цены с использованием нейронных сетей, планируется создать модель, которая будет предлагать рекомендуемую цену для товара;
- разработка мультимедийного ИОТ решения с использованием облачных технологий для управления удаленным музыкальным центром;
- система сбора и анализа данных для ИОТ – электросчетчика;

- обработка GPS данных с помощью скрытых Марковских моделей с целью точной корректировки карт OpenStreetMap;
- применение алгоритмов нейронных сетей для анализа различных фотоизображений (поиск человека в социальной сети по фотографии, проходная на предприятии, автостоянка – определение номера машины и поиск места на стоянке);
- разработку и реализацию искусственного интеллекта в компьютерных играх на основе нейронных сетей;
- нужно отметить ряд работ, связанных с финансовой деятельностью, в которых для поиска оптимальных решений были применены алгоритмы нейронных сетей, так, для определения кредитоспособности клиента был применен алгоритм нахождения весов нейронной сети с применением генетического алгоритма, для прогнозирования цен на фондовой бирже был применен алгоритм нейронных вычислений с глубоким обучением;
- система распознавания отпечатков пальцев;
- NPL- алгоритмы облачной платформы IBM Bluemix и ML- алгоритмов для кластеризации новостей из СМИ, по ключевым словам, и определения тональности статей с использованием вероятностных алгоритмов (PLSA, LDA);
- разработка промышленной системы определения рейтинга влиятельности блогера социальной сети на основе применения NPL- алгоритмов и вероятностных алгоритмов LDA и HDP для анализа блогов (алгоритмы кластеризации и классификации постов блогера по тематикам);
- разработано приложение для заказов такси – аналог UBER;
- выполнена исследовательская работа по применению некоторых методов предиктивной аналитики с использованием облачного сервиса для предиктивной аналитики больших объёмов данных – IBM Watson Analytics;
- проводятся исследования методов и средства анализа мультимедийной информации и методов и средств аналитических решений в сфере транспорта, в сфере электронной коммерции и использование мобильных приложений;
- разработано и внедрено крупное приложение «Оптимизация размещения поликлиник для пожилых людей» для министерства здравоохранения Сингапура или сторонних организаций. Данное приложение учитывает транспортную сеть города, дороги, их загруженность, расстояния, время проезда, наличие других учреждений. Данное решение пригодно для выбора места для размещения другого объекта и в других городах.

Резюмируя приведенные выше направления исследований можно сделать выводы, что многие работы выполнялись в тренде мировых требований для перспективных вакансий и востребованным навыкам, это Cloud Computing, Artificial Intelligence, Statistical Analysis and Data Mining, Web Architecture and Development Framework, Natural Language Processing.

2.4. Так как охватить такую обширную область, как Big Data, даже в рамках годового обучения невозможно, после первой лекции студентам предлагается выбрать курсовой проект в каком-то конкретном направлении и заниматься его реализацией на протяжении всего периода обучения. Этим курсовым проектом мог быть проект связанной с Big data.

При выполнении практических заданий магистрантам был предоставлен выбор задач, которые должны были реализованы с помощью тех средств, которые они могли сами выбрать. Преподаватели формулировали возможные постановки задач, возможные подходы к их решению, а магистранты сами принимали решение, что делать и как. По окончании курса каждый магистрант защищает результаты своей практической работы. Многие выполненные практические задания были выполнены в русле тем магистерских диссертаций, но другие были выполнены только для сдачи практических заданий, например, работы, связанные с чат-ботами. Но, если к чат-бот добавить когнитивную аналитику, то можно заменить многие существующие профессии.

Поэтому важно оценить их выбор с точки зрения современных тенденций к требованиям к навыкам.

Ниже в таблице 1, приведена тематика практических заданий.

Таблица 1

Тематика	Содержание тем
Анализ изображений и видео	Распознавание объектов на изображении, распознавание движущихся объектов по данным с камер, автомобильных номеров в непрерывном видеопотоке, распознавание человеческих эмоций, распознавание дорожных знаков, выделение людей на изображении, распознавание вин, распознавание лиц, распознавание пород собак, распознавание языка жестов по видеопотоку, распознавание жестов на мобильном устройстве, распознавания пожаров по видеопотоку, распознавание текста на изображениях, распознавание меток на изображениях дерматологических патологий, анализ методов обнаружения ошибок фотограмметрии по цифровой модели рельефа, распознавание рукописных цифр, детекция и классификация игральные карты в видеопотоке в режиме реального времени, детектирование лиц на мобильном устройстве, распознавание автомобилей на видео в режиме реального времени.
Совмещение двух и более объектов, картин, изображений, природных явлений	Повышение качества изображения, распознавание изображений и их ведение, наложение 3D объектов, преобразование изображения с целью соответствия стилю художника, колорирование фото и получение ряд других эффектов типа deep fake.
Чат-боты	Слежение за онлайн человеком «вконтакте» через «телеграм бота», получение прогноза погоды, персональные оповещения. Telegram-бот для мониторинга квартир на рынке арендного жилья. Telegram-бот "Погода". Telegram-бот для рекомендации музыки. Использование ботов в социальных сетях. Анализ публикаций в социальной сети Twitter. Топ-10 тем о которых пишет пользователь в Twitter. Telegram bot for Travis CI. Telegram чат-бот. Telegram-бот для интернет-магазина Wildberries. Чат бот на основе BRNN. Погодный чат-бот для Telegram. Telegram-бот для перевода текста. Чат-бот для Telegram Статистический telegram bot
Предиктивная аналитика	Рекомендация книг, рекомендация фильмов, прогнозирование болезней, подбор аналогов лекарств, рекомендация выбора карьеры по анализу личности, выявление хобби человека по анализу его твиттера, прогнозирование числа поездок на велосипеде в зависимости от температуры, оценка успешности атаки баскетбольной команды в момент броска, оценка рисков в логистике. Прогнозирование стоимости аренды недвижимости. Прогнозирование рейтинга ресторана. Прогнозирование ухода сотрудника из компании. Прогнозирование успешности предложений потребительского кредитования. Прогнозирование валютного курса на основании истории изменения курсов.
Интернет-вещей	Навигация внутри помещений, сбор и анализ биометрических данных, управление мультимедиа устройствами, управление кондиционером, отслеживание автомобиля, сбор параметров с датчиков, получение и анализ кардиограммы
Классификация и кластеризация	Классификация отпечатков пальца. Классификация научных документов. Классификация действий водителей. Классификация текстов с помощью нейронной сети. Кластеризация пациентов K-means алгоритмом K-means clustering algorithm. Определение тематики статей из RSS лент новостных порталов и блогов

Тематика	Содержание тем
Облачные вычисления	Сбор и анализ биометрических данных, приём и обработка заявок, организация вычислительного кластера
Анализ данных в соцсетях	Сбор информации о людях, поиск человека по фото, анализ тональности сообщений медиа ресурсов, анализ тональности текстов
NLP	Выделения текста из видео, выделение текста из аудио, перевод текстов, верификация человека по голосу, перевод текста в речь
Тематика других направлений	Анализ задержек авиарейсов. Анализ возможности обработки отзывов о товарах на китайском языке с использованием bag-of-words малых размеров. Интеллектуальная игра "Мафия" на основе технологии блокчейн. Создание рекомендаций на основе внешнего вида продукта. Анализ данных криптовалютных бирж. Анализ обработки заявок в банковской системе клиентских заявок. Построение модели валютного курса на основе текущих биржевых и внебиржевых показателей. Прогнозирование цен акций на фондовом рынке с помощью нейронных сетей. Рекомендация фильмов пользователю при помощи коллаборативной фильтрации. Распознавание объектов, нарисованных человеком, на мобильном устройстве. Моделирование загруженности перекрестка. Поиск маршрута между остановками общественного транспорта. Распознавание атрибутов объекта по его изображению Методы Data Mining. Приложение для прогнозирования необходимого количества денежных средств для загрузки в устройства самообслуживания. Использование акустических отпечатков для поиска аудиозаписей. Масштабируемое приложение для контроля доступа к офисным помещениям.

Анализируя тематику выполненных практических заданий можно сделать вывод, большинство магистрантов хорошо понимают тенденции востребованных навыков и понимают свою востребованность, ибо выполнение ими заданий потребовало навыки и знания современных ИТ платформ, сервисов и готовых библиотек.

2.5. При проведении занятий магистрантам предоставляются нужные информационные ресурсы по темам проектов: книги, журналы, статьи. Также магистранты активно используют интернет для нахождения нужной информации.

Магистрантам предлагается дополнительно пройти один или несколько курсов на интересующую их тематику на одной из популярных обучающих платформ: **cognitiveclass.ai**, **coursera.org**, **Microsoft Azure**, **IBM Bluemix**. Это набор базовых курсов в различных областях Big Data. Так же для сотрудников и студентов кафедры доступны обучающие курсы по программе IBM Academic Initiative. По программе Academic Initiative компания IBM предлагает разнообразие тематических учебных материалов как в виде статей и книг, так и в виде on-line занятий и видеоматериалов на You Tube. Магистрантам предлагается пройти интересующие их курсы по желанию.

На базе кафедры информатики создан и функционирует вычислительный кластер (ЦОД). Он состоит из четырёх серверов, объединенных в единый вычислительный центр с использованием технологии VMware vSphere. Магистранты и студенты получают доступ к кластеру для запуска своих проектов на его мощностях, так как многие проекты требуют достаточно большого количества ресурсов и не могут выполняться на персональных компьютерах в разумный срок.

Студентам и магистрантам также предлагается использовать возможности облачных сервисов для выполнения своих проектов. Такие сервисы, как правило, предоставляют ограниченный набор базовых возможностей бесплатно.

Кроме того, кафедра сотрудничает с московским представительством компании IBM, благодаря чему студенты имеют возможность получить расширенный аккаунт в облачном сервисе IBM Bluemix и использовать многие возможности данного сервиса, в частности, алгоритмы предиктивной аналитики IBM Watson. Ибо, наряду с уже существующей базой, для организации и проведения занятий требуются промышленные технологии, так как решить проблему качественного образовательного процесса силами университетов невозможно.

Так, для проведения занятий применялись и выполнения заданий использовались в первую очередь облачные платформы и технологии такие как:

- облачная платформа Microsoft Azure, Amazon Web Services, IBM Bluemix;
- когнитивные сервисы IBM Watson;
- популярные БД MongoDB, Amazon Dynamo, Redis, CouchDB, Neo4J;
- open source платформы Apache Hadoop, Cloudera Hadoop Distribution Platform, Apache Spark.

В качестве ресурсов для образования могут быть использованы и другие Интернет ресурсы, как платные, так и бесплатные.

3. Резюме

Выбор магистрантами практических заданий и тем диссертационных работ по Big data показывает их понимание перспективных вакансий в мире ИТ. Образовательный процесс должен регулярно корректироваться в соответствии с мировыми тенденциями и проводиться в тренде востребованным навыкам и перспективным вакансиям.

Данный процесс не может быть реализован только силами сотрудников университетов, в ИТ-сфере появилась новая проблема: постоянной адаптации и трансформация образовательного процесса к современным требованиям.

Наряду с уже существующей базой, для организации и проведения занятий требуются *индустриальные технологии* [3], так как решить проблему подготовки специалистов к современным требованиям индустрии ИТ только силами университетов крайне сложно. Необходимо сотрудничество в области образования с ведущими мировыми компаниями, такими как, IBM, Microsoft, SAP др., которые предоставляют специальные образовательные программы и предлагают свои ресурсы в учебных целях бесплатно.

Проблемы подготовки специалистов высокой квалификации есть и остаются длительное время нерешенными, нужны высококвалифицированные профессионалы, которых трудно удержать в университете. Глубокое понимание аналитических алгоритмов обработки больших данных требуют доскональных математических знаний, т.к. данные алгоритмы по стандартной схеме «старт - стоп» не работают, эти алгоритмы нужно обучать, нужно знать математическую основу NPL- алгоритмов и ML- алгоритмов, нужно уметь оценить достоверность полученных результатов, нужно выбрать и обучить тот алгоритм, который наиболее подходит для решения данной задачи.

Как и во всем университетском мире, нам нужны совместные научно-образовательные проекты с нашими ведущими ИТ предприятиями, реализация которых позволит создать в университетах творческие команды, работающие на перспективу ИТ, а не решать текущие проблемы в организации.

Литература

- [1] The Skills Need Most in 2018 – And The Courses to Get Them [Электронный ресурс] / LinkedIn The Learning blog. – Режим доступа: <https://learning.linkedin.com/blog/top-skills/the-skills-companies-need-most-in-2018--and-the-courses-to-get-t>. – Дата доступа: 30.01.2019.

[2] The Skills Companies Need Most in 2019 – And And How to Learn Them [Электронный ресурс] / LinkedIn The Learning blog. – Режим доступа: <https://learning.linkedin.com/blog/top-skills/the-skills-companies-need-most-in-2019-and-how-to-learn-them>. – Дата доступа: 30.01.2019.

[3] Пилецкий, И.И. BIG DATA. Трансформация магистерских программ/ И.И. Пилецкий, А. Е. Лещев, В. Н. Козуб //BIG DATA and Advanced Analytics: collection of materials of the third inter-national scientific and practical conference. (Minsk, Belarus, May 3 – 4, 2017) / editorial board: M. Batura [etc.]. – Minsk, BSUIR, 2017. – 350 с. /с. 159-164.

[4] ИТ в Беларуси. Публикации компании вакансии зарплаты календарь [Электронный ресурс] /– Режим доступа: <https://jobs.dev.by/>– Дата доступа: 30.01.2019.

[5] Пилецкий, И. И. Облачная платформа IBM Bluemix для тренинга по технологиям Big Data / И. И. Пилецкий, А. Е. Лещев, В. Н. Козуб // BIG DATA and Advanced Analytics. Использование BIG DATA для оптимизации бизнеса и информационных технологий : сборник материалов II международной научно-практической конференции; Минск, 15-17 июня 2016 г. / редкол.: М. П. Батура [и др.]. – Минск: БГУИР, 2016. – С. 146-153.

[6] И.И. Пилецкий и др. Виртуальная ИТ среда БГУИР для исследования Big Data и VCL, с. 21-32, BIG DATA and Predictive Analytics. Использование BIG DATA для оптимизации бизнеса и информационных технологий: сборник материалов междунар. науч.-практ. конф. / редкол.: М.П. Батура [и др.]. – Минск: БГУИР, 2015. – 220 с.

SYSTEM FOR COMPLEX ANALYSIS OF DATA FROM INTERNET SOURCES

I.I. PILETSKI, PhD

*Associate Professor of Informatics
Department of the BSUIR*

V.N. KOZUB

*Postgraduate student of the BSUIR,
Master of Technical Sciences, Assistant of In-
formatics Department the BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
E-mail:ianmenski@gmail.com, kozub@bsuir.by*

Abstract. In this paper, the methods and approaches in the organization of students' training in Big Data technologies are considered, the analysis of the results of master's theses and practical tasks is given, the tendency of the choice of work performed by undergraduates is given.

Keywords: learning technologies, hard skills, internet sources, Big Data, Cloud Computing, Machine Learning, Natural Language Processing.