МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ, МОЛОДЕЖИ И СПОРТА УКРАИНЫ

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ

НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ "ХАРЬКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ"

РВУЗ "КРЫМСКИЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ"

ПРОБЛЕМЫ ИНФОРМАТИКИ И МОДЕЛИРОВАНИЯ

ТЕЗИСЫ ДВЕНАДЦАТОЙ МЕЖДУНАРОДНОЙ НАУЧНО-ТЕХНИЧЕСКОЙ КОНФЕРЕНЦИИ

(24 – 30 сентября 2012 года)

Харьков – Ялта

2012

УДК 621.387: 681.327 Проблеми інформатики і моделювання. Тезиси дванадцятої міжнародної науково-технічної конференції. – Харків: НТУ "ХПІ", 2012. – 93 с., російською мовою.

ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ:

- Министерство образования и науки, молодежи и спорта Украины
- Национальная Академия наук Украины
- Петровская Академия наук и искусств РФ
- Институт проблем моделирования в энергетике имени Г.Е. Пухова НАНУ
- Национальный технический университет "ХПИ"
- Национальный аэрокосмический университет "ХАИ"
- Национальный исследовательский университет "Белгородский государственный университет", Россия
- Республиканское высшее учебное заведение "Крымский гуманитарный университет"
- Северо-Кавказский государственный технический университет, Ставрополь, Россия
- Институт радиофизики и электроники НАНУ
- Институт радиотехники и электроники им. В.А. Котельникова Российской Академии наук, Москва, Россия
- Полтавский национальный технический университет им. Ю. Кондратюка
- Харьковский университет Воздушных Сил имени Ивана Кожедуба
- Харьковский национальный университет радиоэлектроники
- Украинская государственная академия железнодорожного транспорта
- Кировоградский национальный технический университет
- Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина, Воронеж

© НТУ "ХПИ", 2012

"НАУКА О ДАННЫХ" КОРПОРАЦИИ EMC – ИННОВАЦИОННЫЙ УЧЕБНЫЙ КУРС ДЛЯ IT-СТУДЕНТОВ

к.ф.-м.н., с.н.с. С.И. Николенко, к.т.н., доц. Г.Н. Смородин,
Академический университет Российской АН, г. Санкт-Петербург

Термин "наука о данных / Data Science" – направление интеллектуального анализа данных, рассматриваемое как последовательное развитие активно используемого в деловой и научной среде термина "добыча данных / Data Mining".

В докладе проведен критический анализ учебного курса "Наука о данных", предлагаемого корпорацией EMC для студентов математических и инженерных специальностей, связанных с компьютерными науками. Учебно-методический комплекс данного курса доступен для всех-вузов партнеров корпорации EMC. Курс предназначен для подготовки специалистов по анализу больших массивов данных и имеет выраженную практическую направленность.

В курсе рассматриваются основные составляющие типичного проекта по анализу больших данных (big data), а также даётся краткое введение в классические статистические методы обработки данных: проверка гипотез, ANOVA, обучение ассоциативных правил, кластеризация методом *k*-средних, анализ временных рядов (ARIMA), основы анализа текстов; даются примеры на языке R. Курс завершается описанием парадигмы параллелизации MapReduce на примере Hadoop и продуктов дочерней компании корпорации EMC – GreenPlum для работы с этой парадигмой (MADlib).

В зависимости от наличия либо отсутствия в учебном плане специализированных дисциплин, курс может быть логично дополнен элементами визуализации и регрессионного анализа как важными элементами анализа данных: визуализация данных (классические методы визуализации и современные языки и библиотеки для визуализации общего характера – Processing, d3; визуализация графов и проект Graphviz), основные методы решения задач регрессии (байесовская регрессия, разные виды регуляризации) и классификации (логистическая регрессия, SVM).

С практической стороны в курсе предлагается рассмотреть новые методы и парадигмы параллелизации, обобщающие и расширяющие связку MapReduce/BigTable: Percolator, Dremel, Pregel, Graphlab.

НАУКОВЕ ВИДАННЯ

ТЕЗИСЫ ДВЕНАДЦАТОЙ МЕЖДУНАРОДНОЙ НАУЧНО-ТЕХНИЧЕСКОЙ КОНФЕРЕНЦИИ "ПИМ-2012"

"ПРОБЛЕМЫ ИНФОРМАТИКИ И МОДЕЛИРОВАНИЯ"

Науковий редактор д.т.н. Дмитрієнко В.Д.
Технічний редактор к.т.н. Леонов С.Ю.

Дизайн та оформлення к.т.н. Гладких Т.В., к.т.н. Заковоротний О.Ю.

Підп. до друку 13.09.2012 р. Формат 60x84 1/16. Папір Copy Paper.
Гарнітура Таймс. Умов. друк. арк. 4,30.
Облік. вид. арк. 4,0. Наклад 120 прим.
Ціна договірна

НТУ "ХПІ", 61002, Харків, вул. Фрунзе, 21

Видавничий центр НТУ "ХПІ"
Свідоцтво ДК № 116 від 10.07.2000 р.

Отпечатано в типографии ООО "Цифра Принт"
на цифровом комплексе Xerox DocuTech 6135
Свидетельство о Государственной регистрации
А01 № 432705 от 03.08.2009 г.
Адрес: г. Харьков, ул. Культуры, 22-Б