

МЕТОДИЧНІ ПІДХОДИ ЩОДО ОРГАНІЗАЦІЇ НАВЧАЛЬНОГО МАТЕРІАЛУ ДИСЦИПЛІНИ «МОДЕЛІ І ЗАСОБИ РОБОТИ З ВЕЛИКИМИ

Розвиток інформаційних технологій призводить до необхідності коригування змісту освіти, насамперед це стосується підготовки ІТ-фахівців у ЗВО. Зокрема, бурхливий розвиток мережових технологій, технологій накопичення, обробки та зберігання даних, поява нових електронних сервісів і пристроїв, сприяє значному зростанню потоків і обсягів даних, що у свою чергу призводить до розробки і впровадження нових технологій, об'єднаних під загальною назвою «Big Data». У системі вищої ІТ-освіти ці процеси також знаходять своє відображення – з'являються нові дисципліни, зокрема, дисципліна «Моделі та засоби роботи з великими обсягами даних (Big Data)». Змістовна складова спеціалізованих професійних знань технологій Big Data висвітлена у достатній кількості відповідної літератури. Однак, аналіз процесу підготовки ІТ-фахівців вказує, зокрема, що подання знань технологій Big Data потребує вдосконалення. Метою цієї статті є висвітлення методичних підходів відбору та подання навчального матеріалу щодо технологій Big Data. Методичні прийоми ґрунтуються на специфіці базових етапів навчання та особливостях дисципліни. Запропоновані методичні підходи відбору та представлення навчального матеріалу дозволяють сформулювати у студентів уявлення про моделі і засоби роботи з великими обсягами даних, а також необхідне теоретичне і практичне підґрунтя для подальшого вивчення цих технологій.

Ключові слова: професійні знання; майбутні фахівці з інформаційних технологій; великі обсяги даних; інформаційні технології; зміст навчання.

Вступ. Бурхливий розвиток інформаційних технологій, зокрема мережових, технологій накопичення, оброблення та зберігання даних, поява нових електронних сервісів, сприяє значному зростанню потоків та обсягів даних. Це призвело до зміни парадигми їхнього накопичення. Самі дані стали цінним ресурсом, який можна застосовувати у різних галузях. Відбувся перехід від парадигми накопичення даних – «ми знаємо, які дані нам потрібні й збираємо саме їх», до – «ми не знаємо, які дані нам потрібні, збираємо усі, що можливо, а потім з'ясуємо їх цінність» [2]. З'являються програмні й апаратні засоби, інформаційні системи, методи роботи з великими обсягами даних. Усе це об'єднано під загальною назвою – «Big Data». У свою чергу в системі вищої ІТ-освіти вводять нові дисципліни, зокрема, дисципліну «Моделі і засоби роботи з великими обсягами даних (Big Data)».

Аналіз досліджень і публікацій. Історично склалося, що поняття «Big Data» має дату появи – 4 вересня 2008 р., коли вийшов спеціальний номер британського журналу Nature, присвячений проблематиці бурхливого зростання глобальних даних та їхньої ролі у

науці [6]. І саме тут уперше звертається увага на зміну відношення до даних, вони розглядаються як ресурс на рівні природних копалин. Згодом починає розроблятися та набувати поширення методологічне підґрунтя технологій Big Data, з'являються відповідні програмні та апаратні засоби [2; 4–7; 11]. Отже у системі вищої ІТ-освіти мають знайти своє відображення ці процеси. За цих умов постає потреба в розробці методики вибору та представлення навчального матеріалу, за допомогою якої можлива ефективна організація як аудиторного (під керівництвом викладача) так і самостійного навчання студентів. Якщо розглядати змістовну складову спеціалізованих професійних знань з технологій Big Data, то тут можна звернути увагу вже на достатню кількість відповідної літератури [1; 3, 8–10; 12]. Нормативні акти, технічний список використаних джерел та дослідження досвіду дає розуміння того, які знання, вміння та навички мають одержати студенти у процесі навчання. Проте, аналіз літератури з проблеми дослідження підготовки ІТ-фахівців вказує, зокрема на те, що представлення знань технологій Big Data потребує вдосконалення.

Метою цієї статті є висвітлення методичних підходів відбору та представлення навчального матеріалу з технологій Big Data на прикладі дисципліни «Моделі і засоби роботи з великими обсягами даних (Big Data)» для спеціальностей галузі знань 12 «Інформаційні технології».

Виклад основного матеріалу. Навчальний процес у ЗВО є теоретичною і практичною підготовкою студентів, що складається із трьох базових етапів: набування нових знань, закріплення отриманих знань (формування умінь і навичок) та контролю знань. На першому етапі студентів пропонується інформація, що містить нові знання, і створюються умови для найбільш ефективного її сприйняття. Мета другого етапу навчального процесу полягає в поглибленні отриманих знань з погляду підвищення їхньої стійкості та розвитку відносин між ними та іншими знаннями. Третій етап навчального процесу передбачає перевірку набутих знань і оцінку їхнього засвоєння з погляду відповідності встановленим вимогам і стандартам.

Розглянемо реалізацію методичних підходів щодо представлення та вибору навчального матеріалу для кожного з етапів навчального процесу на прикладі дисципліни «Моделі і засоби роботи з великими обсягами даних (Big Data)» для студентів спеціальностей галузі знань 12 «Інформаційні технології». Згідно з навчальними планами цю дисципліну викладають на п'ятому курсі вона відноситься до дисциплін циклу професійної та практичної підготовки. Основною метою вивчення дисципліни є формування у студентів теоретичних знань і практичних навичок щодо вибору адекватних засобів технологій Big Data для розв'язання різноманітних задач. Завданнями дисципліни є:

- формування знань з історії появи поняття «Big Data»; проблем використання рішень класу Big Data; провідних технологій, що реалізують Big Data-проекти; концепції NoSQL-систем; провідних моделей даних, що покладені в основу NoSQL-систем (системи «ключ-значення», документні СУБД, системи типу Google BigTable); провідних програмних продуктів, що засновані на NoSQL-системах;

- формування умінь обирати найбільш прийнятні програмні продукти в залежності від специфіки задачі; обирати найбільш ефективні методи та засоби обробки даних в залежності від специфіки задачі; встановлювати та налаштовувати відповідне інформаційне середовище для роботи з великими обсягами даних; застосовувати різні засоби відповідної системи для здійснення CRUD операцій (англ. Create, Read, Update, Delete – створення, зчитування, зміна та видалення) над даними.

Теоретичні знання студенти набувають під час прослуховування лекцій та самостійної роботи з електронним підручником із дисципліни, першоджерелами (нормативними документами) та іншими джерелами інформації. Вивчення теоретичних основ моделей і засобів роботи з великими обсягами даних (Big Data) здійснюється в розрізі наступних тем.

Розділ 1. Загальна характеристика технологій роботи з великими обсягами даних.

Тема 1. 1 Великі обсяги даних (Big Data) : проблеми, технологія, ринок.

Розділ 2. NoSQL системи.

Тема 2. 1 Загальна характеристика NoSQL систем.

Тема 2. 2 Системи «ключ-значення».

Тема 2. 3 Системи типу Google BigTable.

Розділ 3. Документна системи управління базами даних MongoDB.

Тема 3. 1 Документні системи управління базами даних.

Тема 3. 2 Документна система управління базами даних MongoDB.

Формування умінь і навичок, необхідних для практичної діяльності, здійснюється на практичних заняттях у процесі виконання студентами практичних робіт. Тематика робіт, що складає практичну частину, відповідає змісту теоретичного курсу.

- Практична робота № 1. Встановлення MongoDB (встановлення актуальної версії MongoDB, виконання первинних налаштувань середовища MongoDB, перевірка працездатності сервера MongoDB).

- Практична робота № 2. Робота з базами даних MongoDB (знайомство з командами для роботи з БД MongoDB, знайомство з командами для вставки, друку, редагування та видалення документів MongoDB).

- Практична робота № 3. Оболонка Robomongo (встановлення актуальної версії оболонки Robomongo, знайомство з інтерфейсом програмного засобу).

- Практична робота № 4. Типи даних (знайомство з особливостями використання типів даних у MongoDB, знайомство зі значеннями null та undefined, структурою документа MongoDB).

- Практична робота № 5. Пошук документів (імпортування бази даних у систему MongoDB, використання команд для пошуку документів (find, findOne, limit, skip, sort та count), використання курсора у MongoDB).

- Практична робота № 6. Використання модифікаторів під час пошуку (використання регулярних виразів для пошуку інформації, використання модифікаторів умов (\$lt, \$lte, \$gt, \$gte, \$ne), використання модифікаторів логічних умов (\$or, \$nor (not or), \$and, \$not), використання модифікатора залишку від ділення (\$mod), використання модифікатора типу даних (\$type), наявності (\$exists), пошуку значень (\$in, \$nin)).

- Практична робота № 7. Пошук у списках та словниках (пошук документів за значеннями у списках, пошук у вкладених документах за значеннями у списках, пошук за допомогою умов користувача (\$where)).

- Практична робота № 8. Розширена обробка документів (пошук документів з використанням модифікаторів групування (\$project, \$limit, \$sort, \$skip, \$match, \$group, \$unwind, \$addToSet)).

- Практична робота № 9. Керування колекціями та базами даних (оновлення документів з використанням модифікаторів оновлення (\$set, \$unset), створення документів командою update, оновлення списків з використанням модифікаторів списків (\$push, \$pull, \$addToSet, \$each), внесення змін у вкладені списки з використанням оператора позиції – \$, друкування оновлених даних за допомогою команди findAndModify).

– Практична робота № 10. Створення та використання індексів (порівняння продуктивності виконання запитів з використанням індексу і без нього, використання простих, складових, унікальних індексів).

Специфіка викладання зазначеної дисципліни полягає у тому, що її вивчають на 5 курсі. Студенти вже мають достатню базу теоретичних та практичних знань, зокрема вони вже знайомі з мовами програмування; форматами представлення даних json, bson, scv тощо; мають уявлення про бази даних та провідні принципи управління даними. Також студенти вже мають досвід творчого пошуку та самостійного опанування навчальним матеріалом. Отже навчальний матеріал має бути спрямований у першу чергу на розуміння особливостей роботи з великими обсягами даних. Так наприклад, при визначенні поняття Big Data слід звернути увагу на провідні ознаки цих технологій: volume – обсяг; velocity – швидкість оброблення; variety – різноманіття та неструктурованість даних [11]. Не можна казати, що ці ознаки у тому чи іншому вигляді не притаманні технологіям обробки даних, що не віднесені до класу Big Data. Так під «великим обсягом даних» розуміється не скільки певний розмір даних, а якісно нові можливості одержання нової корисної інформації шляхом пошуку закономірностей, що раніше було неможливе. Під різноманіттям та неструктурованістю даних мається на увазі незалежність обробки даних від їх моделі, тобто відхід від принципів логічної цілісності даних теорії реляційних баз даних. Надалі ці особливості знаходять своє відображення у NoSQL-системах, їх організації та моделях даних. Також варто звернути увагу на те, що підхід щодо розкриття поняття Big Data крізь ознаки «тріох V» не розкриває це поняття повною мірою, а лише торкається окремих його аспектів. Усі ці ознаки є якісними. Постають питання, коли, наприклад, обсяг даних великий, а коли він ще недостатньо великий, чим це виміряти, що взяти за точку відліку. Яку швидкість оброблення вважати швидкою тощо. Інший підхід – це розгляд Big Data як технології. Тут слід виокремити три напрями за завданнями, що вирішуються: збір, первісне оброблення та зберігання даних для їхнього подальшого використання; структурування розрізненого контенту: текстового, графічного, відео, аудіо, звукового тощо; бізнес-аналітика на великих обсягах даних. Для кожного напрямку існують свої вимоги та ознаки технологій цього класу. За умови обмеженості часу усі три напрями у цьому курсі докладно не розглядаються. Як подальше поглиблене вивчення моделей і засобів роботи з великими обсягами даних ці напрями можуть розглядатися у окремих дисциплінах.

Теоретичні положення ілюструються на прикладі опанування роботою з системою MongoDB, як приклад реалізації технологій збору, первісного оброблення та зберігання даних для їхнього подальшого використання. Практичні роботи спрямовані на формування знань встановлення, первісного налаштування системи та виконання CRUD операцій. Кожна практична робота містить завдання з наведеною технологією виконання та завдання для самостійної роботи. Завдання для самостійного виконання складені у від-

повідності до основних завдань роботи. А також є завдання, що не можна виконати цілком лише засобами, що розглядаються у роботі. Так, наприклад, у практичній роботі № 7 «Пошук у списках та словниках» зокрема розглядається обмеження виводу результатів пошуку за допомогою певних модифікаторів. Це зручно використовувати для пошуку першого, останнього, передостаннього і т. ін. документа, що відповідає умовам пошуку. Студентам пропонується створити базу даних, яка міститиме інформацію про успішність студентів з дисципліни «Моделі і засоби роботи з великими обсягами даних (Big Data)». База даних зберігає інформацію про практичні роботи: номер роботи (ключ «_id») та тему (ключ «title»), а також успішність студентів за роботою: ПБ студента (ключ «name») та оцінку (ключ «mark»). Нижче наведено приклад одного документа:

```
{ «_id» : 1. 0, «title» : «Установка MongoDB»,
«marks» : [
  { «name» : «Іванов А. Б. », «mark» : 4. 0 },
  { «name» : «Петров О. Б. », «mark» : 5. 0 },
  { «name» : «Попов О. Л. », «mark» : 5. 0 },
  { «name» : «Рябініна А. Д. »},
  { «name» : «Сидоров Г. Г. », «mark» : 1. 0 } ]
}
```

Ця база даних має особливість – є студенти з оцінками, а є без оцінок, тобто ключ «mark» відсутній. Студентам пропонується знайти передостаннього студента, який здав роботу, наприклад, з номером 1 (ключ «_id»). Якщо у запиті використовувати тільки модифікатор пропуску документів \$slice, що розглядається у роботі, то результатом буде: { «name» : «Рябініна А. Д.»}. Однак такий результат це не зовсім відповідає завданню. І дійсно, Рябініна А. Д. є передостаннім студентом у цьому списку, проте вона не є такою, хто здав роботу. Виконати це завдання тільки засобами, що розглядаються у роботі, цілком правильно неможливо. Отже на прикладі таких завдань студенти знайомляться з обмеженнями та особливостями інструментарію MongoDB.

Окрім описаних вище завдань, практичний курс також містить творчі завдання, що сприяють набуттю практичного досвіду роботи з реальними базами даних. Так студентам пропонується база даних, що містить корпоративну переписку. База даних складається з 496 документів обсягом 541 Мбайт. Необхідно знайти відповіді на питання: «Для кожного дня тижня порахуйте, скільки листів було відправлено на адресу ebass@enron.com?», «Проаналізуйте листування Shanna Husser і Eric Bass. Скільки листів кожен з них відправив іншому?», «Скільки чоловік відправляють листи самі собі (тобто адреса відправника і одержувача співпадає)?» тощо. Особливостями бази даних є те, що імена відправників можуть бути написані по різному, наприклад «Shanna Husser» або «Husser, Shanna»; є листи адресовані одному одержувачу, а є розсилки; дата представлена текстовою строчкою тощо. Все це спонукає до більш глибокого дослідження самих даних і вибору відповідних методів та засобів пошуку. Дисципліна завершується тестуванням.

Також слід зазначити, що за умови обмеження часу у цьому практичному курсі не розглядається адміністрування, більш глибоке налаштування та забезпечення безпеки системи MongoDB. Все це також може бути змістом іншої дисципліни.

Висновки. На прикладі дисципліни «Моделі і засоби роботи з великими обсягами даних (Big Data)» для спеціальностей галузі знань 12 «Інформаційні технології» проілюстровано особливості та специфіку

знань, а також методичні підходи щодо формування навчального матеріалу, що враховуватимуть зазначене. Запропоновані методичні підходи відбору та представлення навчального матеріалу дозволяють надати студентам уявлення про моделі та засоби роботи з великими обсягами даних, сформувані необхідне теоретичне і практичне підґрунтя для подальшого вивчення цих технологій.

Список використаних джерел

1. Бэнкер К. MongoDB в действии / К. Бэнкер; пер. с англ. Слинкина А. А. – М. : ДМК Пресс, 2012. – 394с.
2. Найдич А. Big Data : проблема, технология, рынок [Электронный ресурс] / А. Найдич // КомпьютерПресс. – 01.2012. – Режим доступа : <http://compress.ru/article.aspx?id=22725>. Мова рос. – Дата останнього доступу : 15. 03. 2018. – Назва з екрана.
3. Редмонд Э. Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL / Э. Редмонд, Джим. Р. Уилсон ; пер. с англ. Слинкина А. А. – М. : ДМК Пресс, 2013. – 384с.
4. Реймер Д. Gartner : Big Data больше не существует! [Электронный ресурс]. – Режим доступа : <http://denreymer.com/gartner-end-of-big-data>. Мова рос. – Дата останнього доступу : 15.03.2018. – Назва з екрана.
5. Савчук И. Big Data – технология, рождающая новый тип бизнеса [Электронный ресурс] / И. Савчук // БИТ. – № 3 (36). – 2014. – Режим доступа : <http://bit.samag.ru/archive/article/1352>. Мова рос. – Дата останнього доступу : 15.03.2018. – Назва з екрана.
6. Big Data [Электронный ресурс] / Nature. – V. 455. – N. 7209. – pp 1–136. – Режим доступа : <http://www.nature.com/nature/journal/v455/n7209/index.html>. Мова англ. – Дата останнього доступу : 15.03.2018. – Назва з екрана.
7. BIG DATA 2016 : Форум умных решений [Электронный ресурс] / Издательство «Открытые системы». – 07.04.2016. – Режим доступа : <https://www.osp.ru/news/articles/2016/14/13048945>. Мова рос. – Дата останнього доступу : 15.03.2018. – Назва з екрана.
8. Brad Dayley. Node. js, MongoDB and Angular Web Development / Brad Dayley, Brendan Dayley, Caleb Dayley. – Publisher : Addison-Wesley Professional, 2017. – 640 p.
9. Cyrus Dasadia. MongoDB Administrator’s Guide / Cyrus Dasadia. – Publisher : Packt Publishing, 2017. – 429 p.
10. Deepak Vohra. Pro MongoDB Development / Deepak Vohra. – Publisher : Apress, 2017. – 493 p.
11. Gartner Says Solving «Big Data» Challenge Involves More Than Just Managing Volumes of Data [Электронный ресурс] // Gartner. – STAMFORD. – Conn. – June 27, 2011. – Режим доступа : <http://www.gartner.com/newsroom/id/1731916>. Мова англ. – Дата останнього доступу : 15.03.2018. – Назва з екрана.
12. Seguin K. The Little Redis Book [Электронный ресурс] : – Режим доступа : <http://openmymind.net/2012/1/23/The-Little-Redis-Book>. Мова англ. – Дата останнього доступу : 19.10.2015. – Назва з екрана.

Е. А. Кирей,

канд. пед. наук, доцент кафедры,

Л. В. Солобута,

канд. техн. наук доцент кафедры,

ЧНУ ім. Петра Могили, г. Николаев, Україна

МЕТОДИЧЕСКИЕ ПОДХОДЫ К ОРГАНИЗАЦИИ УЧЕБНОГО МАТЕРИАЛА ДИСЦИПЛИНЫ «МОДЕЛИ И СРЕДСТВА РАБОТЫ С БОЛЬШИМИ ОБЪЕМАМИ ДАННЫХ (BIG DATA)»

Развитие информационных технологий приводит к необходимости корректировки содержания образования, прежде всего это касается подготовки ИТ-специалистов в ВУЗах. В частности, бурное развитие сетевых технологий, технологий накопления, обработки и хранения данных, появление новых электронных сервисов и устройств, способствует значительному росту потоков и объемов данных, что в свою очередь приводит к разработке и внедрению новых технологий, объединенных под общим названием «Big Data». В системе высшего ИТ-образования эти процессы также находят свое отражение – появляются новые дисциплины, в частности, дисциплина «Моделі і засоби роботи з великими обсягами даних (Big Data)». Содержательная составляющая специализированных профессиональных знаний технологий Big Data освещена в достаточном количестве в соответствующей литературе. Однако, анализ процесса подготовки ИТ-специалистов указывает, в частности на то, что представление знаний технологий Big Data требует совершенствования. Целью этой статьи является освещение методических подходов отбора и представления учебного материала, касающегося технологий Big Data. Методические приемы основываются на специфике базовых этапов обучения и особенностях дисциплины. Предложенные методические подходы отбора и представления учебного материала позволяют сформировать у студентов представление о моделях и средствах работы с большими объемами данных, а также необходимые теоретические и практические знания для дальнейшего изучения этих технологий.

Ключевые слова: профессиональные знания; будущие специалисты по информационным технологиям; большие объемы данных; информационные технологии; содержание обучения.

E. A. Kirey, L. V. Solobuto,
Petro Mohyla Black Sea National University, Mykolaiv, Ukraine

**METHODOLOGICAL APPROACHES TO ORGANIZATION OF EDUCATIONAL MATERIALS
OF THE DISCIPLINE «MODELS AND MEANS OF WORKING WITH LARGE DATA VOLUMES (BIG DATA)»**

Development of informational technologies leads to the necessity of adjusting the content of education, and this first of all concerns preparation of IT specialists at universities. In particular, boisterous development of network technologies and the technologies of data collection, processing, and storage contribute to considerable increase of data streams and volumes, and in its turn this results in development and implementation of new technologies united under an umbrella name «Big Data».

These processes are also reflected within the higher IT education system: new disciplines appear, in particular the discipline «Models and means of working with big volumes of data (Big Data)». The content component of specialized professional knowledge of Big Data technologies is sufficiently covered in corresponding literature. However, analysis of the process of preparation of IT specialists especially indicates that provision of knowledge about the above-mentioned technologies needs to be augmented.

The goal of this article is to cover methodological approaches to choice and presenting of educational materials about the Big Data technologies. The methodological techniques are based on the peculiarities of the basic stages of learning and the specifics of the discipline. The students already possess a sufficient base of theoretical and practical knowledge, the experience of creative search and of independent learning of study materials. This way the study materials should be primarily aimed at understanding of the peculiarities of work with big volumes of data. Theoretical points are illustrated with an example of working with a system that belongs to the Big Data class, a Document Databases MongoDB. Suggested methodological approaches of educational materials choice and presenting allow to shape the students' understanding of the models and means of working with big volumes of data, as well as the necessary theoretical and practical knowledge for further studying of these technologies.

Key words: professional knowledge; future information technology specialist; Bid Data; information technologies; learning content.

Рецензент: Фісун М. Т., д-р техн. наук, професор.

© Кірей К. О., Солобучо Л. В., 2018

Дата надходження статті до редколегії 22.04.2018