Teaching Big Data by Three Levels of Projects

Jianjun Yang
Department of Computer Science and Information Systems
University of North Georgia
jianjun.yang@ung.edu

Ju Shen
Department of Computer Science
University of Dayton
jshen1@udayton.edu

Abstract

"Big Data" is a new topic and it is very hot nowadays. However, it is difficult to teach Big Data effectively by regular lecture. In this paper, we present a unique way to teach students Big Data by developing three levels of projects from easy to difficult. The three levels projects are initializing project, designing project, and comprehensive projects. They are developed to involve students in Big Data, train students' skills to analyze concrete problems of Big Data, and develop students' creative abilities and their abilities to solve real setting problems.

Keywords: Big Data, initializing, designing, comprehensive, projects

1. Introduction

Numerous technological innovations are driving the dramatic increase in data and data gathering. Data are being collected at unprecedented scale in many areas, such as networking [11] [12], image processing [13] [14] [15], virtualization, scientific computation, and algorithms. The huge data nowadays are called Big Data. Big data are all-encompassing terms for any collection of data sets so large and complex that it becomes difficult to process them using traditional data processing applications [1]. They are used in a wide variety of applications, such as traffic patterns, purchasing behaviors, online video, and real-time inventory management. Big Data have become a recent area of strategic investment for IT organizations [2].

Big Data are critical to students' current study and future career; hence many schools are training Big Data to students. However, it is extremely difficult to teach them because: First, manipulating data sets often requires massively parallel software running on tens, hundreds, or even thousands of servers. Second, there is no specific Big Data course in most schools. Many instructors met a lot of challenges when they teach Big Data to students. The challenges on teaching and learning Big Data include analysis, capture, search, sharing, storage, transfer, visualization, and privacy violations.

In this article, we present a unique way to teach Big Data in networking and image processing courses. We develop three levels' projects to students via three phases. They are initializing project, designing project, and comprehensive projects. Initializing projects ask students to test basic concepts and check the results. They are normally interesting projects with good visual effect, such as the simulator for networking. This kind of project aims to introduce Big Data and attract student learning interest. Designing projects ask students to analyze a given topic and then to come up with their own solutions. This type is to train student analytical ability. The third type is comprehensive project, which has challenge topics and may cover multiple course materials. We ask students to analyze the questions in depth, propose their approaches, test the approaches and improve them. This type is to develop student's creative habit and help student solve real setting problems.

Our projects are conducted in two areas. The first one is web/mobile application in networking course, in which initializing projects and designing project are provided. Student learned the concepts of Big Data and how to simplify Big Data from complex networking to a visual computer program. Then we set comprehensive projects in the course of image processing. We used novel image based rendering algorithm with user intervention to generate realistic 3D virtual world in this phase. Students' learning outcomes are significant.

2. Teaching Big Data Based on Three levels of Projects

The literature highlights the importance of hands-on activities in the teaching of technologies [3]. Hence we trained students Big Data by Projects. We assigned three levels of projects in three phases from easy to difficult.

2.1 Phase 1: Introduce Concepts of Big Data with Initializing Projects.

Big Data are critical in Computer Science not only because they are emerging technology, but also they are fundamental for students' future career. Some Computer Science scholars have generally gravitated toward introducing easy content under the assumption that the students would be more receptive to it. It is not true. If the goal of teaching Big Data is just to introduce the basic concepts, it would be an easy task by simplifying the course. However, this could make students, especially those in computer science major, get bored easily with those trivial and superficial contents. Moreover, this teaching strategy prevents students from grasping the fundamentals concretely. To instil the joy of Big Data to students, we demonstrated interesting cases by assigning initializing projects to stimulate students' learning interest.

When we taught Computer Network courses, we studied the characteristics of wireless devices including laptops, iPads, iPhones and Android Phones. In order to consistently create an enthusiastic learning environment and facilitate student's success, we applied simulator as a tool to introduce and simplify Big Data in networking. In particular, we set initializing projects to introduce Big Data to students. We taught students to conduct simulations, which are acts of imitating the behavior of a physical or abstract system, such as an event, a situation, or a process that does or could exist [4]. Some scholars [5] consider simulations as a perfect educational technique that creates learning by reproducing all or part of an event or situation. Theoretically, simulations could be created for any number of topics, courses, or programs in education. Some more popular simulations are offered in various academic programs including business, health care, and transportation. Technology advances allow individuals to design self-placed simulations in their classrooms with limitless options. We designed initializing projects via simulators for mobile networking in our networking courses.

In the projects, we taught students to simplify big network data by simulators. Figure 1 is the real network topology and figure 2 is the graphical interface of the designed simulator. When we introduced Big Data, we presented the scenario of connected network devices illustrated in figure 1. Since many modern and popular devices are used in the scenario, it makes the class compelling and retain students' attentions. High volume data are demonstrated from different aspects, such as their structures and transmissions among the network devices including their structure, transmission, and representation under the network devices. Then we introduced how to retrieve the critical content from the Big Data, such as IP addresses, locations, the resource capabilities [6]. Afterwards, we instructed students to practice manipulating Big Data through hands-on projects. Students were guided to allocate the resources to the mobile devices by solving linear equations. We pinpointed areas where students can add virtual computers based on the properties of the heterogeneous devices, in order to increase the number of equations. This is the way to simplify the big network data. Students later implemented the equations by programming and the results are displayed in the simulator. This provided undergraduate students a unique opportunity to use experimental technologies to be adaptively involved in learning complicated Big Data problems and understanding the abstract concepts.

Figure 1 Network Scenario

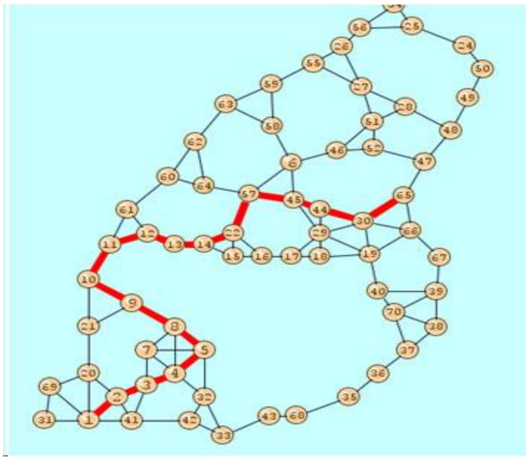

Figure 2 The interface of simulator

2.2 Phase 2: Train Students' Analysis Abilities with Designing Projects

Marc Prensky[7] created a powerful summary when he said games offer fun, play, rules, goals, interactivity, outcomes, feedback, conflict, opposition, problem solving, structure, flow, motivation, and pleasure. With such a list of benefits, it is a good idea to use smart phone for teaching in the classroom.

We taught the environment of Android development. Then we assigned projects to ask students to design Apps for Big Data. App Inventor for Android is a new visual programming platform for creating mobile applications (Apps) for smart phones. It was developed at Google Labs by a team led by MIT. Students do not write code to developed apps in App Inventor. Instead, they designed tools to visualize the App by using block based GUI to directly control the App's behaviors through interlocking components. App Inventor aims to develop intuitive tools that facilitate novices to program in an enjoyable manner. App Inventor lets students create apps for smart phones. Given the popularity and ubiquity of mobile phones among today's generation of students, App Inventor seems to hold a great potential for attracting a new generation of students to problem-solving thinking to handle Big Data.

Students found App Inventor very accessible and they learned how to develop Apps of their own design quickly. Though the App looks simple, it actually incorporates a large amount of data with different formats (e.g. images, sounds, labels, etc.), and involves considerable control logics. Hence App is able to let students focus on problem solving on handling the big data rather than coding syntax. We asked students to design some

very interesting App projects. For example, we assigned the students to develop an interactive map of the attractions in Paris. When an attraction is clicked, its corresponding information will be displayed. Figure 3 shows the logic design of the App. Figure 4 shows the interfaces on a virtual cell phone.
App is a good tool to develop students' problem-solving ability since it is not only easy to follow and reproduce already written apps, but also straight forward to develop completely new Apps based on the principles acquired through the tutorials and demonstrations. Students progressed quickly from writing "Hello Kitty" to develop Apps using database, interactive maps, client server communication, and other advanced concepts. Thus they know how to manipulate Big Data, even when they encounter problems. Students were able to apply their programming skills to new types of problems including databases, client-server communication, images processing and algorithms.

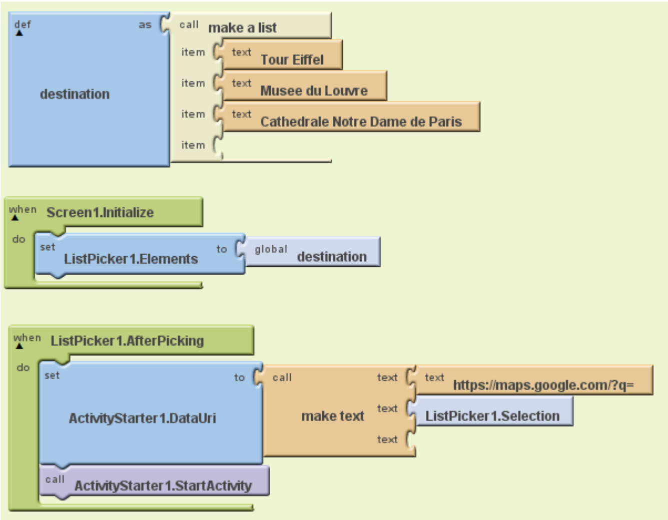

Figure 3 Logic blocks

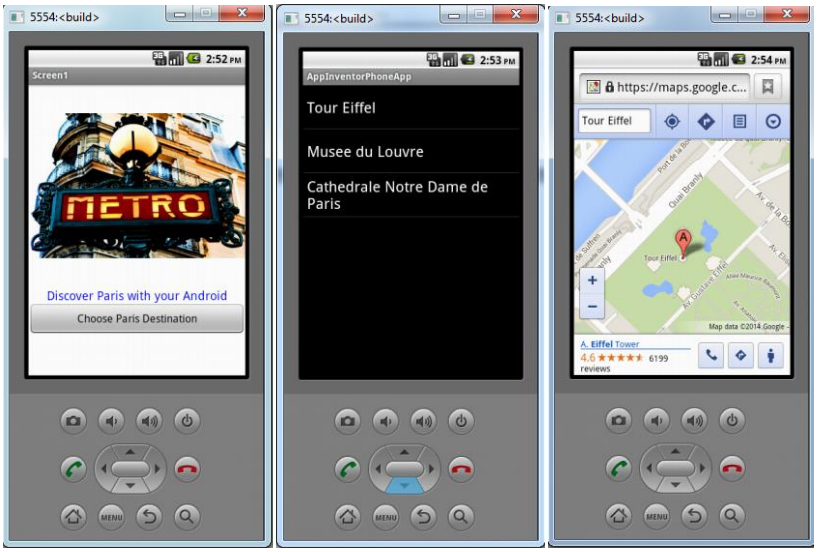

Figure 4 Interfaces on a Virtual Cell Phone

2.3 Phase 3: Develop students' creative abilities and help students develop research abilities by comprehensive projects.

This is the third phase to train students Big Data. We proposed an interactive system to operate on big visual data that supports online picture sharing or virtual 3D world navigation when we taught Interactive Media. Students got involved of the whole process of system development, such as coding, online image editing, and 3D model designing.
With the explosive growth of Internet and web-based cameras, billions of photographs are uploaded to the internet every day. The massive collections of imagery have inspired a wave of different applications on such

large visual data. Part of the excitement in these areas is due to the facts that images are easy to take nowadays everywhere from our daily devices, like cell phones, tablets, and the efficient online access via WiFi or any phone network. Imagine building a virtual 3D world by taking the advantage of these large online images, such as the Google street view databases or the Flickr image collection. This system can provide virtual environment and immersive experience that allows users to walk freely in a re-constructed virtual world and view the scene from any arbitrary perspectives. In addition to its virtual reality value, as a photo warehouse, such system can also support large visual information. For example, for a travelling resort, people often take many pictures during the trips. However, sometimes the taken pictures may be less than satisfactory, such as the background scene is not fully captured or occluded by some objects. Some photo editing tools are available to improve the images. However, it could be a pain to modify the picture directly without any extra information, which often introduces noticeable artifacts. Things can become much easier, if there are additional available pictures taken from the same location at similar time. In such a way, travellers can share their experiences and enrich their photo collections from the large visual data.

We assigned students with comprehensive projects, which are on image retrieval, localization and reconstructing 3D geometry from a large, unordered collection of online images on landmarks and cities[8][9][10]. We asked students to analyze the questions in depth, propose their approaches, test the approaches and improve them. Because students have experience from the first two phases, we asked students to use image feature descriptors, such as SIFT or SURF, as the cue to identify similar images for clustering. Then based on the detected feature correspondences across multiple views, the scene geometry can be approximated estimated. The use of real photos not only supports realistic image synthesis with little user intervention, but raises the important issue of controlling and altering the representations. The students were really interested in the projects and happy to present their work to the instructors. Many results have demonstrated that, through training, students have developed the ability to use tools to render realistic view of novel images efficiently and accurately.

The projects of this phase present an integrated research and educational program with two goals. The first goal of the phase is to produce new technologies on intuitive and interactive pictorial editing tools that allow undergraduates to manipulate and alter large visual data directly in high dimensions or temporal domain. The second goal of the phase is to expose the cutting edge technologies in Big Data processing, especially for visual data clustering and reconstruction to undergraduates, which can stimulate student interests in the related fields and promote their pursuit of careers. This phase is not only undergraduate oriented as many available software tools can be used straight away, such as the image matching APIs, 3D transformation tools, but also requires students to explore the core techniques and develop novel solutions on efficiently manipulating large visual data. During the phase, students had the chance to learn those well-established algorithms and state-of-art Big Data technologies in image matching, 3D graphics, and data visualization. Figure 5 is an example that shows the process to reduce the Big Data to represent a tree to much smaller data that represents the outline of the tree.

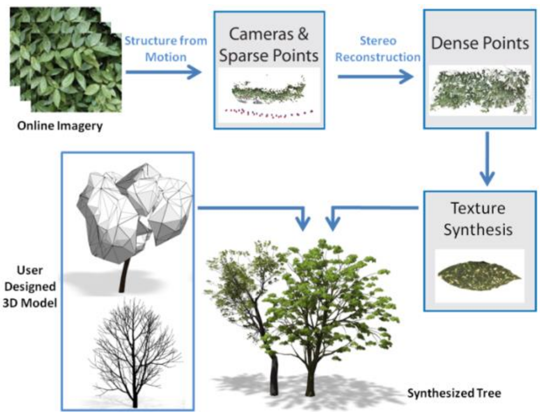

Figure 5 User Editing for 3D Scene Reconstruction: A Big Data application of 3D Tree Synthesis

3. Conclusion

Big Data are very important yet very difficult to teach for students. In this paper, we proposed an effective way to teach Big Data to students. We do not merely mechanically introduce the concepts of Big Data. Instead, we use concrete projects to illustrates Big Data to students gradually through three phases. We assigned students relevant projects to train their skills on handling Big Data, train students' abilities to analyze questions of Big Data, and develop students' creative abilities and their abilities to solve real setting problems.

.

4. References

[1] Big Data, http://en.wikipedia.org/wiki/Big_data, 2015

[2] Introduction to Big Data: Infrastructure and Networking Considerations, Juniper Networks, Inc, 2012

[3] Curto, K.. and Bayer, T, Writing and speaking to learn biology: An intesection of critical thinking and communication skills, Bioscene, 31(4), pp. 11-19, 2005

[4] Damassa, D. A., and Sitko, T. D., Simulation technologies in higher education: Uses, trends, and implications. ECAR Research Bulletin 3. Boulder, CO: Educause Center for Applied Research, 2010

[5] Maran, N. J., and Glavin, R. J., Low- to high-fidelity simulation—a continuum of medical education? Medical Education, 37(1), 22-28, 2003

[6] Yang, J., and Fei, Z., Broadcasting with Prediction and Selective Forwarding in Vehicular Networks , International Journal of Distributed Sensor Networks, 2013.

[7] Prensky, M., Digital game-based learning. Chicago, IL: McGraw-Hill, 2001

[8] Irschara, A., Zach, C., Frahm, J.M., and Bischof, H., From structure-from-motion point clouds to fast location recognition. IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2009

[9] Li, Y., Snavely, N., and Huttenlocher, D.P., Location Recognition using Prioritized Feature Matching, 2010

[10] Hays, J., and Efros, A.A., Estimating geographic information from a single image", IEEE conference on Computer Vision and Pattern Recognition, 2008

[11] J. Yang and Z. Fei, Bipartite Graph Based Dynamic Spectrum Allocation for Wireless Mesh Networks, ICDCS Workshops, 2008

[12] J. Yang and Z. Fei, HDAR: Hole detection and adaptive geographic routing for ad hoc networks, Computer Communications and Networks (ICCCN), Proceedings of 19th International Conference. IEEE, 2010.

[13] J. Shen, P. Su and S. Cheung, Virtual Mirror Rendering with Stationary RGB-D Cameras and Stored 3D Background, IEEE Transactions on Image Processing, vol. 22, issue 9, pp. 1-16.

[14] Ju Shen and Wai-tian Tan, Image-based indoor place-finder using image to plane matching, Multimedia and Expo (ICME), 2013 IEEE International Conference on.

[15] J. Shen and S. Cheung Layer Depth Denoising and Completion for Structured-Light RGB-D Cameras, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013), Portland, USA 2013.