

Abstract:

Although data cleansing and preparation are significant tasks in many real-world data projects, they are rarely found in project assignments in IS database courses. This paper describes a pilot study of a relatively open-ended project assignment in a graduate database course. The project required the students to cleanse and prepare five datasets on educational statistics from United Nations Data before storing them in relations that they designed. To gauge the level of students' prior knowledge on data preparation, the instructor deliberately provided no prior lecture on the topic. A follow-up assignment was a PHP/MySQL Web database application to display educational statistics for a user-specified country. Submitted works and post assignment surveys were studied and analyzed. The result indicated that both assignments were well received and generally beneficial. Although our students appeared not to be well trained in data preparation in their undergraduate studies, they were able to learn quickly enough to produce acceptable products. This approach also appeared to encourage more creativity and better diversity in students' database designs. Our experience suggested that while it was not difficult to identify interesting realworld datasets of appropriate complexity, the instructors will need to put in extra effort on project evaluation. We believe that this kind of assignment can be adapted in many ways to satisfy different educational objectives and it fits well in a well-rounded IS curriculum. Thus, the goal of the paper is to foster interests in real-world data cleansing projects in database courses with a well-examined case study.

Citation:

Yue, K., A Realistic Data Cleansing and Preparation Project, Journal of Information Systems Education (JISE), Volume 23, Number 2, 2012, pp205-216.