

Integrating Data Cleansing With Popular Culture: A Novel SQL Character Data Tutorial

David H. Olsen, Utah State University, USA
Pamela A. Dupin-Bryant, Utah State University, USA

ABSTRACT

Big data and data science have experienced unprecedented growth in recent years. The big data market continues to exhibit strong momentum as countless businesses transform into data-driven companies. From salary surges to incredible growth in the number of positions, data science is one of the hottest areas in the job market. Significant demand and limited supply of professionals with data competencies has greatly affected the hiring market and this demand/supply imbalance will likely continue in the future. A major key in supplying the market with qualified big data professionals, is bridging the gap from traditional Information Systems (IS) learning outcomes to those outcomes requisite in this emerging field. The purpose of this paper is to share an SQL Character Data Tutorial. Utilizing the 5E Instructional Model, this tutorial helps students (a) become familiar with SQL code, (b) learn when and how to use SQL string functions, (c) understand and apply the concept of data cleansing, (d) gain problem solving skills in the context of typical string manipulations, and (e) gain an understanding of typical needs related to string queries. The tutorial utilizes common, recognizable quotes from popular culture to engage students in the learning process and enhance understanding. This tutorial should prove helpful to educators who seek to provide a rigorous, practical, and relevant big data experience in their courses.

Keywords: Databases; Big Data; Data Science; 5E Instructional Model; SQL Server; Teaching/Learning Methods; Character Data; String Functions; Business Intelligence; Business Analytics

INTRODUCTION

Big data and data science have experienced unprecedented growth in the recent past (Gantz & Reinsel, 2012; Nadkarni & Vesset, 2015). Indeed, it is axiomatic to state that big data is integral to all areas of business. From salary surges to incredible growth in the number of positions, big data and data science is one of the hottest areas in all of the job market (Adams, 2015; Krantz & Lee, 2015; RHT, 2015). Positions in big data science are becoming more visible and sought after within successful organizations. A recent report highlights an 8.9% projected increase in salaries for big data positions in 2016 with a salary range of \$129,000-\$183,000 depending upon location and experience (RHT, 2015). Significant demand and limited supply of professionals with data competencies has greatly affected the hiring market and this demand/supply imbalance will likely continue in the future (Dumbill, Libby, Stanton, Mueller, & Farnham, 2013; RHT, 2015).

A major key in supplying the market with qualified big data professionals, is bridging the gap from traditional Information Systems (IS) learning outcomes to those outcomes requisite in this emerging big data area. Big data scientists need a variety of skills sets and knowledge domains. Current literature underscores the importance of educating a new generation of workers in ‘Information Competencies’ (Aasheim, Williams, Rutner, & Gardiner, 2014; Azat, Evgenia, & Tatiana, 2016; Borthick, Schneider, & Viscelli, 2016; Cegielski & Jones-Farmer, 2016; Gorman & Klimberg, 2014). Big data professionals need a deep understanding of relational database implementation and querying. They also need to understand acquisition and cleansing of data. SQL/Query skills and basic analytics are often highlighted as requisite skills for data scientists (Cegielski & Jones-Farmer, 2016; Mamonov, Misra, & Jain, 2015; Wixom, et al., 2014). To fully comprehend the area of querying, data scientists must understand the nature of data and the tools that are available to manipulate and transform the data. One area that must be explored and mastered by data science professionals involves string, character, or textual data.

The purpose of this paper is to share an SQL Character Data Tutorial for IS education. The tutorial was designed to expose students to User-Defined Functions in SQL Server and specifically, string oriented functions that are considered an absolute building block for solving business problems. This practical tutorial was designed as a stand-alone activity that can be employed in any database course to help students learn to employ the data science tools offered in the Microsoft SQL Server Development Environment. Utilizing the 5E Instructional Model (a well-respected learning model), this tutorial helps students (a) become familiar with SQL code, (b) learn when and how to use SQL string functions, (c) understand and apply the concept of data cleansing, (d) gain problem solving skills in the context of typical string manipulations, and (e) gain an understanding of typical needs related to string queries. The tutorial utilizes common, recognizable quotes from movies and television programs to help students learn to use SQL Server String Functions. The use of quotes from popular culture emphasizes “pop culture is a useful tool that can help get students interested and engaged in learning, and even offer resources for better understanding classroom material” (Staff Writers, 2013, para. 1). Students spend a lot of time interacting with popular culture and using simple, recognizable quotes from popular culture as an educational tool helps IS educators make learning more engaging and productive.

Four areas provide a framework for the paper exposition: (a) highlight relevant literature pertaining to big data and data science in both business and educational environments, (b) outline the methods and educational model employed, (c) highlight each phase in the SQL Character Data Tutorial, and (d) provide concluding remarks and suggestions for future research and practice. This paper seeks to highlight a practical approach for helping students gain a deep understanding of relational database implementation and querying. The tutorial identified in this paper should prove helpful to information systems educators who seek to provide a rigorous, practical, and relevant big data experience in their courses.

LITERATURE REVIEW

Over the past several years, the term ‘big data’ has been extensively employed in business to describe vast, diverse, multifaceted, and/or longitudinal datasets; ‘big data’ is considered the driver of business intelligence and analytics (BI&A) (Gantz & Reinsel, 2012; Gorman & Klimberg, 2014). The terms ‘big data’ and ‘data science’ refer to the analytical techniques required to extract useful information from large-scale data. These terms gained prominence as organizations began collecting datasets that far exceeded the capacity of traditional data storage systems and analytical tools (Provost & Fawcett, 2013). Big data technologies are “designed to economically extract value from very large volumes of a wide variety of data by enabling high – velocity capture, discovery, and/or analysis. There are three main characteristics of Big Data: the data itself, the analytics of the data, and the presentation of the results of the analytics” (Gantz & Reinsel, 2012 p. 9).

Big data and data science have experienced unprecedented growth in the recent past (Gantz & Reinsel, 2012; Nadkarni & Vesset, 2015). Big data has become integral to all areas of business in varied industries. Gorman and Klimberg (2015) suggest “business analytics sophistication grows with the degrees of intelligence employed and directly affects a company’s competitive advantage” (p. 331). The big data market continues to exhibit strong momentum as businesses accelerate their transformation into data-driven companies. Driven by widespread industry adoption, the big data market is projected to grow to \$48.6 Billion in 2019 according to a recent IDC Forecast (Nadkarni & Vesset, 2015).

From salary surges to incredible growth in the number of positions, data science is one of the hottest areas in all of the job market (Adams, 2015; Krantz & Lee, 2015; RHT, 2015). “A data scientist is a hybrid role that requires a broad combination of technical and business skills and multidisciplinary knowledge domains” (Wang, 2015). Positions in big data science are becoming more visible and sought after within successful organizations. “Job postings seeking data scientists and business analytics specialists abound these days” (Chen, Chiang, & Storey, 2012, p. 1182). A recent report highlights an 8.9% projected increase in salaries for big data positions in 2016 with a salary range of \$129,000-\$183,000 depending upon location and experience (RHT, 2015). The current Job’s Related Almanac, ranked Data Scientist as 6th out of 200 in their overall ‘best jobs’ ranking. The overall ranking considered variables such as work environment, stress, income, and hiring outlook (Krantz & Lee, 2015). “With the explosion of big data and the need to track it, employers keep on hiring data scientists. But qualified candidates are in short supply” (Adams, 2015, para. 1).

Significant demand and limited supply of professionals with data competencies has greatly affected the hiring market and this demand/supply imbalance will likely continue in the future (Dumbill, et al., 2013; RHT, 2015). Murray (2015) suggests “the demand for the hottest data roles continues to stymie companies, who are demanding that managers become masters of data analysis” (para. 2). A recent study investigated big data impact on company productivity and found the major performance impact is the close connection between investment in big data technologies and labor skills (Bughin, 2016). A recent article in the Wall Street Journal emphasized the importance of updating big data skills and competencies to remain relevant in the information age or face the ‘permanent pink slip’ (Waller, 2014). Cegielski and Jones-Farmer (2016) note “demand for education in the technical and managerial skills related to managing, extracting, analyzing, and communicating information from large data sets is high” (p. 95).

A major key in supplying the market with qualified big data professionals, is bridging the gap from traditional information systems learning outcomes to those outcomes requisite in this emerging big data area. “Higher education needs to quickly respond to the rise of big data and offer adequate relevant educational opportunities” (Aasheim, et al., 2014, p. 2). Over the past several years, many noteworthy reports have stressed the vital relationship between business intelligence/analytics initiatives and academic programs (Aasheim, et al., 2014, Cegielski & Jones-Farmer, 2016; Gorman & Klimberg, 2014; Murray, 2015; Wilder & Ozgur, 2015; Wixom et al., 2014). Universities have been challenged to develop programs that will produce graduates who can apply a variety of tools and techniques to capture, manage, and identify patterns in large amounts of data (Aasheim, et al., 2014).

Many recent studies underscore the relationship between big data and formal education. These studies note current programs and initiatives; suggest strategies for developing curriculum; propose analytical tools to integrate in IS education; highlight requisite knowledge, skills, and abilities; and share cases, methods, activities, and pedagogical resources. Several studies have identified the current of state of BI&A in academia (Aasheim, et al, 2014; Cegielski & Jones-Farmer, 2016; Davis & Woratschek, 2015; Wang, 2015; Wilder & Ozgur, 2015; Wixom, et al, 2014). Since 2007, over 130 institutions have begun offering degrees in Business Analytics or Data Science (Wixom et al., 2014). While academia is making great strides in this area, with the development of a wide variety of relevant programs, these studies suggest that the integration of BI&A initiatives into educational arenas is still immature but evolving rapidly. However, the supply of appropriately skilled graduates still lags behind the demand (Dumbill, et al., 2013).

Many studies highlight the requisite skill sets and knowledge domains of BI&A professionals. Big data professionals need a deep understanding and application of relational database implementation and querying. In order to fully comprehend the area of querying, data scientists must understand the nature of data and the tools that are available to manipulate and transform the data. SQL and basic querying skills were highlighted as a fundamental skill in a variety of studies (Cegielski & Jones-Farmer, 2016; Mamonov, et al., 2015; Wixom, et al., 2014). In a recent study, Cegielski and Jones-Farmer (2016) asked the question “What knowledge, skills, and abilities should be taught in business schools to prepare students for careers as data analytics professionals?” They found through a triangulated mixed-method research approach that “the most important technical prerequisites for entry-level business analytics positions are a high level of mastery of Excel and SQL” (p. 105). Wilder and Ozgur (2015) noted data specialists should be able to understand how data is managed. In addition, Wixom et al. (2014) surveyed practitioners in the field and found the most critical skills are communication, SQL/Query Skills, and basic analytics. Mamonov et al. (2015) took a competence-based perspective and researched companies’ open position announcements; they identified ‘data storage/extraction’ and ‘SQL’ as commonly publicized mandatory technical skills.

In a review of BI&A research, Wang (2015) notes a variety of studies that focus on teaching cases, activities, and pedagogies. Borthick, Schneider and Viscelli (2016) outline a case that helps students develop database skills such as importing spreadsheet data into a database, joining tables, and analyzing data with database querying. They suggest that cases of this nature help students learn to think critically and employ database query expressions to achieve important business objectives. Edgington (2011) shared innovative practices for introducing text analytics in a graduate business school course. A recent study used a simulation and evolutionary lab sequence to teach business intelligence (Pittarese, 2015). Wang and Wang (2013) published their process for designing and delivering a technical module related to business intelligence. Wang (2015) provides a comprehensive list of recent journal

articles and conference proceedings related to teaching cases and learning activities in business intelligence and analytics education.

Current literature underscores the importance of educating a new generation of workers in ‘Information Competencies’ (Aasheim, et al., 2014; Azat, et al., 2016; Borthick, et al., 2016; Cegielski & Jones-Farmer, 2016; Gorman & Klimberg, 2014). “High quality training of the specialists working in the information sphere has become an important driver in successful development of the economy and the society. Expectations of the employers towards employees’ information competencies are changing simultaneously with the public, social and technological progress” (Azat, et al., 2016, p. 79). However, many BI&A programs lack appropriate pedagogical design for developing student professional skills (Wilder & Ozgur, 2015). High-quality big data education must focus on pedagogical innovations that emphasize fostering students’ problem solving, team work, and communication skills (Wilder and Ozgur 2015). “More relevant and practical real-world teaching cases and computer-aided pedagogies...that teach BI&A concepts or techniques should be presented in the future BI&A research” (Wang, 2015, p. 5).

METHODS

The SQL Character Data Tutorial outlined in this paper was designed to expose students to User-Defined Functions in SQL Server. This functional tutorial was designed as a stand-alone activity that can be employed in any database course to help students learn to employ the business intelligence and analytics tools offered in the Microsoft SQL Server Development Environment. Specifically, this tutorial helps students (a) become familiar with SQL code, (b) learn when and how to use SQL string functions, (c) understand and apply the concept of data cleansing, (d) gain problem solving skills in the context of typical string manipulations, and (e) gain an understanding of typical needs related to string queries.

The tutorial utilizes common, recognizable quotes from movies, television programs, and popular culture to help students learn to use SQL Server String Functions. The use of quotes from popular culture as the content students query in the tutorial emphasizes “pop culture is a useful tool that can help get students interested and engaged in learning, and even offer resources for better understanding classroom material” (Staff Writers, 2013, para. 1). According to a recent study, children (ages 8-18) spend approximately 7 and a half hours per day interacting with media including TV, music, computers, video games, and movies (Rideout, Foehr, & Roberts, 2010). Students spend a lot of time interacting with popular culture and using simple quotes from popular culture as an educational tool helps IS educators make learning more engaging and productive.

The popular culture quotes used in this paper are provided as a starting point for the tutorial; however, any number of popular culture quotes may be substituted into the tutorial framework to ensure students are introduced to fundamental SQL character data. The popular culture quotes selected to begin the tutorial should be easily recognizable to the majority of students. As the lesson progresses, the instructor may draw more apropos quotes from students to incorporate into future courses.

The SQL Character Data Tutorial was designed to be taught over a two week period (i.e. 4 class sessions/75-minutes each). Recognizing varied course scheduling and time constraint issues, this tutorial could be reduced/increased to slightly shorter/longer time frames. This tutorial targets students in upper division database courses who have already been introduced to basic database principles and have a strong SQL understanding. While this tutorial has been successfully integrated into various database courses, IS educators could easily adjust the tutorial design to suite specific educational needs and teaching styles.

The 5E Instructional Model was employed as the framework for this tutorial. The 5E Instructional Model is a highly respected, learning and teaching model based on the ideas of educational theorists Herbart, Dewey, and Piaget. The model is based on a constructivist approach to learning in which people ‘construct’ knowledge and meaning through experience; the 5E Instructional Model also incorporates aspects of behaviorist and cognitivist learning theories (Bybee et al., 2006; Bybee, 2009; Jobrack, 2011). “When organized and correctly operated, this 5E learning cycle model enhances not only the students’ achievement but also the permanence of knowledge in various fields of education” (Dorji, Panjaburee, & Srisawasdi, 2015, p. 93).

The 5E Instructional Model is widely used in Science, Technology, Engineering, and Math (STEM) education. More than a quarter million lesson plans have been developed and implemented with this approach and countless research articles highlight the merit of this inquiry based teaching model (Bybee et al., 2006; Dorji et al., 2015; Goldston, Dantzer, Day, & Web, 2013; Jobrack, 2011; Piyayodilokchai et al., 2013). The 5E model has been cited as a useful model for teaching SQL (Dupin-Bryant & Olsen, 2014; Piyayodilokchai et al., 2013). A recent study that focused on Structured Query Language Instructional Units, found the 5E learning cycle model to be more effective than traditional learning models. In this study, students significantly outperformed peers on understanding “key aspects and concepts in SQL, especially regarding basic knowledge of SQL and ability to apply SQL” (Piyayodilokchai et al., 2013, p. 156). Studies in a variety of scientific disciplines show positive trends for student mastery of subject matter and interest when employing the 5E model (Bybee et al., 2006).

The phases of the 5E Instructional Model include: (E1) engagement, (E2) exploration, (E3) explanation, (E4) elaboration, and (E5) evaluation (Bybee, et al., 2006; Bybee, 2009). Each of the distinct phases are supported by research and have been widely applied in diverse educational settings (Bybee, et al., 2006; Bybee, 2009; Jobrack, 2011). Figure 1 summarizes each phase in the 5E Instructional Model (Bybee, 2009, p. 5).

Figure 1. Phases of the 5E Instructional Model

(1) Engagement

- "The teacher assesses...the learners' prior knowledge and helps them become engaged in a new concept through the use of short activities that promote curiosity and elicit prior knowledge."
- "The activity should make connections between past and present learning experiences, expose prior conceptions, and organize students' thinking toward the learning outcomes of current activities."

(2) Exploration

- "Exploration experiences provide students with a common base of activities within which current concepts (i.e., misconceptions), processes, and skills are identified and conceptual change is facilitated."
- "Learners may complete lab activities that help them use prior knowledge to generate new ideas, explore questions and possibilities, and design and conduct a preliminary investigation."

(3) Explanation

- "Focuses students' attention on a particular aspect of their engagement and exploration experiences and provides opportunities to demonstrate their conceptual understanding, process skills, or behaviors."
- "Provides opportunities for teachers to directly introduce a concept, process, or skill."
- "Learners explain their understanding of the concept."
- "An explanation from the teacher or the curriculum may guide them toward a deeper understanding, which is a critical part of this phase."

(4) Elaboration

- "Teachers challenge and extend students' conceptual understanding and skills."
- "Through new experiences, the students develop deeper and broader understanding, more information, and adequate skills."
- "Students apply their understanding of the concept by conducting additional activities."

(5) Evaluation

- "Encourages students to assess their understanding and abilities."
- "Provides opportunities for teachers to evaluate student progress toward achieving the educational objectives."

Source: Summary of the BSCS 5E Instructional Model (Bybee, 2009, p. 5)

The 5E Instructional Model is “grounded in sound educational theory, has a growing base of research to support its effectiveness, and has had a significant impact on science education” (Bybee, et al., 2006, p. 41). The instructional flow of this SQL Character Data Tutorial follows the 5E Instructional Model. The alignment of the tutorial with the 5E Instructional Model provides an educational environment in which students can thrive.

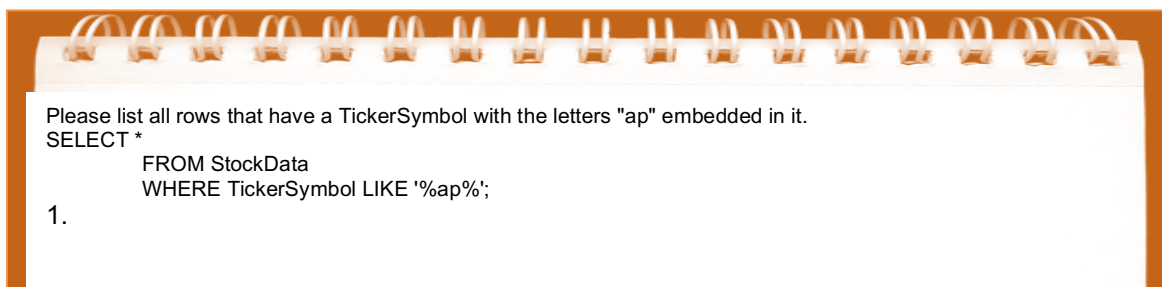
CHARACTER DATA TUTORIAL PHASES

Following the 5E Instructional Model, this SQL Character Data Tutorial design includes sound educational methods and activities for each distinct phase (i.e. Engagement, Exploration, Explanation, Elaboration, and Evaluation). All the components of this tutorial are instructionally aligned with the learning objectives. The following section outlines each of the phases of instruction employed.

Engagement (E1)

At the beginning of the SQL Character Data Tutorial, the instructor should assess the students’ prior knowledge. The instructor could simply ask each student to complete a basic string query. Figure 2 provides an example string query used to assess prior knowledge.

Figure 2. Basic String Query Assessment



The instructor introduces each specific learning objective during the engagement phase. Learning objectives should be defined, distributed, and published for easy student retrieval (e.g. posted on the course Learning Management System). The five objectives for the SQL Character Data Tutorial include:

1. learn when and how to use SQL string functions,
2. understand and apply the concept of data cleansing,
3. gain problem solving skills in the context of typical string manipulations, and
4. gain an understanding of typical needs related to string queries.

The instructor then presents a variety of materials regarding the importance of learning advanced querying techniques for string, character, or textual data to introduce students to this unfamiliar material. Content for this activity can be found online, and/or in course textbooks. A Common SQL String Functions List (see Table 1) and/or links to comprehensive Function Lists may be provided to students to ensure familiarity with common string functions. The standard string manipulation functions in Table 1 were compiled from the Microsoft Developer Network – String Functions page (<https://msdn.microsoft.com/en-us/library/ms181984.aspx>) and an SQL – String Functions online tutorial (<http://www.tutorialspoint.com/sql/pdf/sql-string-functions.pdf>).

In addition, students are asked to find and write down four of their favorite quotes from popular culture, including one quote in each of the following categories: (1) movie, (2) television, (3) music, and (4) inspirational. Students should be allowed a few minutes in class to explore the internet and find the quotes. Students are then asked to share one quote with the class. This activity promotes curiosity amongst students and sets the stage for a structured activity associated with the learning outcomes.

Table 1. Common SQL String Functions List

| Function | Results |
|-----------------------|---|
| ASCII | Returns numeric value of left-most character |
| BIT_LENGTH | Returns length of argument in bits |
| CHAR_LENGTH | Returns number of characters in argument |
| CHAR | Returns the character for each integer passed |
| HEX | Returns a string representation of a hex value |
| INSERT | Inserts a substring at the specified position up to the specified number of characters |
| LEFT | Returns the leftmost number of characters as specified |
| LENGTH | Returns the length of a string in bytes |
| LOWER | Returns the argument in lowercase |
| LPAD | Returns the string argument, left-padded with the specified string |
| LTRIM | Removes leading spaces |
| MID | Returns a substring starting from the specified position |
| OCT | Returns a string representation of the octal argument |
| QUOTE | Escape character |
| REPLACE | Replaces occurrences of a specified string |
| REVERSE | Reverses the characters in a string |
| RIGHT | Returns the specified rightmost number of characters |
| RPAD | Appends string the specified number of times |
| RTRIM | Removes trailing spaces |
| SPACE | Returns a string of the specified number of spaces |
| STRCMP() | Compares two strings |
| SUBSTRING_INDEX() | Returns a substring from a string before the specified number of occurrences of the delimiter |
| SUBSTRING(), SUBSTR() | Returns the substring as specified |
| TRIM() | Removes leading and trailing spaces |
| UCASE() | Synonym for UPPER() |
| UPPER() | Converts to uppercase |

Resources: Microsoft Developer Network – String Functions (<https://msdn.microsoft.com/en-us/library/ms181984.aspx>)
 SQL – String Functions Online Tutorial (<http://www.tutorialspoint.com/sql/pdf/sql-string-functions.pdf>)

Exploration (E2)

The second phase of tutorial helps students become familiar with SQL code and learn when and how to use SQL string functions. During this phase students explore string functions by completing the following activity. Table 2 shows a Sample Popular Culture Quotes Table that the students are asked to query. The popular culture quotes used in this paper are provided as a starting point for the tutorial; however, any number of popular culture quotes may be substituted into the tutorial framework to ensure students are introduced to fundamental SQL string functions. The popular culture quotes selected to begin the tutorial should be easily recognizable to the majority of students. Students should be given a copy of the Popular Culture Quotes (see Table 2) and associated string query instructions (see Figure 3). Students should then be asked to perform each of the string query instructions. The quotes do not need to be part of a table(s). Rather, students can simply paste the quotes into a SELECT statement and leave out the FROM.

Table 2. Sample Popular Culture Quotes Table

| Quote | Popular Culture Reference |
|--|---|
| "No way Bells is bigger than Waimea bro." | <i>Point Break</i> |
| "Either they don't know, don't show, or don't care about what's going on in the hood." | <i>Boyz n the Hood</i> |
| "You see, it's the slow knife, the knife that takes it's time, the knife that waits years without forgetting, then slips quietly between the bones, that's the knife, that cuts deepest. " | <i>The Dark Knight Rises</i> |
| "It's a Sicilian message. It means Luca Brasi sleeps with the fishes." | <i>The Godfather</i> |
| "Yo, Mini-Man, takin' on the jellies. You've got serious thrill issues, dude. Awesome." | <i>Finding Nemo</i> |
| "A census taker once tried to test me. I ate his liver with some fava beans and a nice chianti." | <i>Silence of the Lambs</i> |
| "What was that for? It doesn't matter, it's in the past." | <i>The Lion King</i> |
| "Get busy living or get busy dying." | <i>The Shawshank Redemption</i> |
| "On this night, you have allowed your friends to die for you, rather than face me yourself. There is no greater dishonor." | <i>Harry Potter and the Deathly Hallows: Part 2</i> |
| "That man is playing Galaga! He thought we wouldn't notice, but we did. " | <i>The Avengers</i> |
| Dr. Silberman: I'm sure it feels very real to you. Sarah Connor: On August 29th, 1997, it's gonna feel pretty real to you too. Anybody not wearing 2 million sunblock is gonna have a real bad day. Get it? | <i>Terminator 2</i> |
| "Elliot, if we're gonna go that way, you'll need a bigger knife." | <i>Breaking Bad</i> |
| "It's not personal Sonny, it's strictly business." | <i>The Godfather</i> |

Figure 3. Query Instructions

Use SQL Server string functions to accomplish the following. Possible functions include LEFT, LEN, LOWER, LTRIM, RTRIM, LOWER, SUBSTRING, UPPER.

Note: Declare keyword is used to assign the variable.

1. Return letters 13-19 in the *Point Break* quote.
2. Return the last word in the quote from *Boyz n the Hood*. Put it in all capital letters.
3. Return the word "quietly" from *The Dark Knight Rises* quote.
4. Return "Luca Brasi" from *The Godfather* quote.
5. Return "liver" from the *Silence of the Lambs* quote.
6. Return "doesn't" from the *Lion King* quote.
7. Return the last two letters from *The Shawshank Redemption* quote.
8. Return the last word from the *Harry Potter and the Deathly Hallows: Part 2* quote.
9. Return the word after "playing" regardless of what that word is from *The Avengers* quote.
10. Return the *Terminator 2* quote with the year 2017 replacing 1997.
11. Return "takin' " from the *Finding Nemo* quote.
12. Return the word "bigger" from the *Breaking Bad* quote.
13. Return "it's" from *The Godfather* quote.
14. Create a table called "Stairway to Heaven."
 - a. The first column is the row number and the second column is the lyrics in each row.
 - b. Import the song into the table.
 - c. Use a simple solution (like using Excel) or a scalable solution that would work for any song.
 - d. Do a SELECT * to show the contents of the table.

Explanation (E3)

In the explanation phase, the instructor should give an explanation of querying strings in general and specific string functions in particular to guide students toward a deeper understanding. The instructor will conduct a discussion on how big data professionals need a deep understanding of relational database implementation and querying related to strings. The process of detecting and correcting or removing inaccurate data (also known as “data cleansing” or “scrubbing”) should be explored. Current concrete examples of “Data Cleansing” may help students make connections between the tutorial and real world application. The answers to the popular culture query assignment should be provided and reviewed in detail (See Table 3). Upon completion of this phase, students should now be able to write the basic syntax for SQL string functions and understand and apply the concept of data cleansing.

Table 3. String Query Answers

| Item | Answers |
|------|---|
| 1 | <code>DECLARE @Quote VARCHAR(MAX) = 'No way Bells is bigger than Waimea bro' SELECT SUBSTRING(@Quote, 17, 8) AS [Word];</code> |
| 2 | <code>DECLARE @Quote VARCHAR(MAX) = 'Either they don't know, don't show, or don't care about what's going on in the hood'; -- Basic Method 1 SELECT UPPER(SUBSTRING(@Quote,80,4)) AS [Last Word]; -- Intermediate Method 2 SELECT UPPER(SUBSTRING(@Quote,(LEN(@Quote)-3),4)) AS [Last Word]; -- Complex Method 3 SELECT UPPER(REVERSE(SUBSTRING(REVERSE(@Quote), 1, CHARINDEX(' ', REVERSE(@Quote)) - 1))) AS [Last Word];</code> |
| 3 | <code>DECLARE @Quote VARCHAR(MAX) = 'You see, it's the slow knife, the knife that takes it's time, the knife that waits years without forgetting, then slips quietly between the bones, that's the knife, that cuts deepest.' SELECT SUBSTRING(@Quote,120,8); SELECT SUBSTRING(@Quote,121,7);</code> |
| 4 | <code>DECLARE @Quote VARCHAR(MAX) = 'It's a Sicilian message. It means Luca Brasi sleeps with the fishes' SELECT SUBSTRING(@Quote, 33, 11);</code> |
| 5 | <code>DECLARE @Quote VARCHAR(MAX) = 'A census taker once tried to test me. I ate his liver with some fava beans and a nice chianti.' SELECT SUBSTRING(@Quote, 48, 6); SELECT SUBSTRING(@Quote, 49, 5);</code> |
| 6 | <code>DECLARE @Quote VARCHAR(MAX) = 'What was that for? It doesn't matter, it's in the past' SELECT SUBSTRING(@Quote, 23, 7);</code> |
| 7 | <code>DECLARE @Quote VARCHAR(MAX) = 'Get busy living or get busy dying' SELECT SUBSTRING(@Quote,32,2);</code> |
| 8 | <code>DECLARE @Quote VARCHAR(MAX) = 'On this night, you have allowed your friends to die for you, rather than face me yourself. There is no greater dishonor.' SELECT SUBSTRING(@Quote, 111,9); SELECT SUBSTRING(@Quote, 112,8);</code> |
| 9 | <code>DECLARE @Quote VARCHAR(MAX) = 'That man is playing Galaga! He thought we wouldn't notice, but we did.' SELECT SUBSTRING(@Quote, 20,8); SELECT SUBSTRING(@Quote, 21,7);</code> |
| 10 | <code>DECLARE @Quote VARCHAR(MAX) = 'Dr. Silberman: I'm sure it feels very real to you. Sarah Connor: On August 29th, 1997, it's gonna feel pretty real to you too. Anybody not wearing 2 million sunblock is gonna have a real bad day. Get it?' SELECT STUFF(@Quote, 83, 4,'2017'); SELECT REPLACE(@Quote, 1997, 2017);</code> |

(Table 3 continued on next page)

(Table 3 continued)

| Item | Answers | | | | | | | | | | | | | | | | | | | | | | |
|-----------|--|-----------|--------|---|---|---|--|---|---|---|--|---|--|---|--|---|---|---|--|---|---|----|---|
| 11 | <pre>DECLARE @Quote VARCHAR(MAX)='Yo, Mini-Man, takin" on the jellies. You"ve got serious thrill issues, dude. Awesome.' SELECT SUBSTRING(@Quote,15,6); SELECT SUBSTRING(@Quote,14,7);</pre> | | | | | | | | | | | | | | | | | | | | | | |
| 12 | <pre>DECLARE @Quote VARCHAR(MAX)='Elliot, if we"re gonna go that way, you"ll need a bigger knife.' SELECT SUBSTRING(@Quote,51,6); SELECT SUBSTRING(@Quote,50,7);</pre> | | | | | | | | | | | | | | | | | | | | | | |
| 13 | <pre>DECLARE @Quote VARCHAR(MAX)='It"s not personal Sonny, it"s strictly business.' SELECT SUBSTRING(@Quote,26,4); SELECT SUBSTRING(@Quote,25,5);</pre> | | | | | | | | | | | | | | | | | | | | | | |
| 14 | <p>Prepare the work book “Stairway to Heaven” as attached</p> <div data-bbox="297 661 490 787" data-label="Image"> <p>Stairway to Heaven.xlsx</p> </div> <p>Import the Excel using Import Data option in Task, as you right click on the Database. Then select Source Data and Destination, finally click finish.</p> <div data-bbox="451 877 1247 1486" data-label="Image"> <p>The screenshot shows a SQL query window with the following text: <code>Select * from [dbo].[Stairway to Heaven]</code>. Below the query, there is a 'Results' tab showing a table with two columns: 'RowNumber' and 'Lyrics'. The table contains 10 rows of data, with the first row highlighted.</p> <table border="1"> <thead> <tr> <th>RowNumber</th> <th>Lyrics</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>There's a lady who's sure all that glitters is gold</td> </tr> <tr> <td>2</td> <td>And she's buying a stairway to heaven.</td> </tr> <tr> <td>3</td> <td>When she gets there she knows, if the stores are all c...</td> </tr> <tr> <td>4</td> <td>With a word she can get what she came for.</td> </tr> <tr> <td>5</td> <td>Ooh, ooh, and she's buying a stairway to heaven.</td> </tr> <tr> <td>6</td> <td></td> </tr> <tr> <td>7</td> <td>There's a sign on the wall but she wants to be sure</td> </tr> <tr> <td>8</td> <td>'Cause you know sometimes words have two meanings.</td> </tr> <tr> <td>9</td> <td>In a tree by the brook, there's a songbird who sings,</td> </tr> <tr> <td>10</td> <td>Sometimes all of our thoughts are misgiven.</td> </tr> </tbody> </table> </div> | RowNumber | Lyrics | 1 | There's a lady who's sure all that glitters is gold | 2 | And she's buying a stairway to heaven. | 3 | When she gets there she knows, if the stores are all c... | 4 | With a word she can get what she came for. | 5 | Ooh, ooh, and she's buying a stairway to heaven. | 6 | | 7 | There's a sign on the wall but she wants to be sure | 8 | 'Cause you know sometimes words have two meanings. | 9 | In a tree by the brook, there's a songbird who sings, | 10 | Sometimes all of our thoughts are misgiven. |
| RowNumber | Lyrics | | | | | | | | | | | | | | | | | | | | | | |
| 1 | There's a lady who's sure all that glitters is gold | | | | | | | | | | | | | | | | | | | | | | |
| 2 | And she's buying a stairway to heaven. | | | | | | | | | | | | | | | | | | | | | | |
| 3 | When she gets there she knows, if the stores are all c... | | | | | | | | | | | | | | | | | | | | | | |
| 4 | With a word she can get what she came for. | | | | | | | | | | | | | | | | | | | | | | |
| 5 | Ooh, ooh, and she's buying a stairway to heaven. | | | | | | | | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | There's a sign on the wall but she wants to be sure | | | | | | | | | | | | | | | | | | | | | | |
| 8 | 'Cause you know sometimes words have two meanings. | | | | | | | | | | | | | | | | | | | | | | |
| 9 | In a tree by the brook, there's a songbird who sings, | | | | | | | | | | | | | | | | | | | | | | |
| 10 | Sometimes all of our thoughts are misgiven. | | | | | | | | | | | | | | | | | | | | | | |

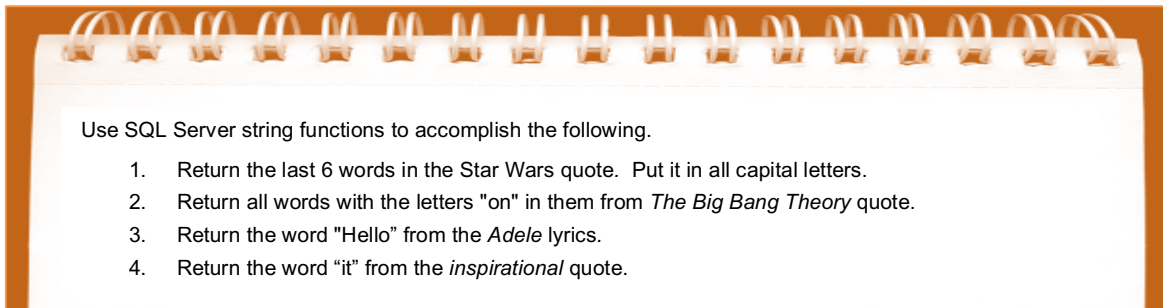
Elaboration (E4)

In the elaboration phase, students apply their understanding of SQL String Functions with their four popular culture quotes they selected at the beginning of the tutorial during the Engagement (E1) phase. Piyayodilokchai, et al., (2013) suggest students would benefit more if SQL teachers could demonstrate material then allow students to practice by themselves. In this elaboration phase, students do just that – they practice importing their quotes (see Table 4) and creating their own questions/instructions (see Figure 4).

Table 4. Student Sample Popular Culture Quotes Table

| Quote | Popular Culture Reference |
|--|-------------------------------------|
| Finn: We'll figure it out. We'll use the Force. Han Solo: That's not how the Force works. | <i>Star Wars--The Force Awakens</i> |
| Sheldon: Why are you crying? Penny: Because I'm stupid. Sheldon: That's no reason to cry. One cries because one is sad. For example, I cry because others are stupid, and that makes me sad. | <i>The Big Bang Theory</i> |
| "Hello, it's me. I was wondering if after all these years you'd like to meet." | <i>Adele</i> |
| "If it is to be, it is up to me." | <i>Unknown</i> |

Figure 4. Student Sample Query Instructions



During this phase, instructors may choose to organize students into small groups (2-3 individuals) to design queries to answer the questions. Table 5 provides common answers to the example Student Sample String Queries. Students are able to discuss and ask questions that arise as they explore the material as a group. A key to success for learning business analytics is to integrate the concept of ‘learning by doing’ such as hands-on projects and trial-and-error experimentation (Chen et al., 2012). During this phase of the Character Data Tutorial, students actively participate in the learning process and develop a deeper understanding of querying techniques through direct, experiential learning. Students get their hands on the ‘data’ and begin finding creative ways of turning it into valuable ‘information.’

Table 5. Student Sample String Query Answers

| Item | Answers |
|------|--|
| 1 | <pre> DECLARE @Quote VARCHAR(MAX) = 'Finn: We'll figure it out. We'll use the Force. Han Solo: That's not how the Force works'; SELECT UPPER(SUBSTRING(@Quote,58,31)) AS [Last 6 Words]; </pre> |
| 2 | <pre> DECLARE @Quote VARCHAR(MAX) = 'Sheldon: Why are you crying?'; @Quote1 VARCHAR(MAX) = 'Penny: Because I'm stupid.'; @Quote2 VARCHAR(MAX) = 'Sheldon: That's no reason to cry. One cries because one is sad. For example, I cry because others are stupid, and that makes me sad.' SELECT SUBSTRING(Col, 1, 7) + ' ' + SUBSTRING(Col,20,7) + SUBSTRING(Col, 34,4) + SUBSTRING(Col, 52,4) as [Words] from (SELECT @Quote AS Col UNION SELECT @Quote1 AS Col UNION Select @Quote2 as Col) T Where Col like '%on%' </pre> |
| 3 | <pre> DECLARE @Quote VARCHAR(MAX) = 'Hello, it's me. I was wondering if after all these years you'd like to meet'; SELECT SUBSTRING(@Quote,(LEN(@Quote)-74),5) AS [Word]; </pre> |
| 4 | <pre> DECLARE @Quote VARCHAR(MAX) = 'If it is to be, it is up to me.'; -- Basic Method 1 SELECT SUBSTRING(@Quote,4,2) AS [Word]; -- Intermediate Method 2 SELECT SUBSTRING(@Quote,(LEN(@Quote)-27),2) AS [Word]; -- Complex Method 3 SELECT REVERSE(SUBSTRING(REVERSE(@Quote), 14, CHARINDEX(' ', REVERSE(@Quote)) - 1)) AS [Word]; </pre> |

Evaluation (E5)

In the last phase of the project, student understanding and skills should be assessed. While informal evaluation should occur in every phase of the SQL Character Data Tutorial, a more formal assessment of student learning should be conducted during this phase. Assessment not only focuses on writing syntactically correct statements, but more importantly on whether or not students can translate a natural language question into a semantically correct SQL expression.

The answers to the four queries (related to the student quotes) are graded by the instructor. Results are evaluated based on the following: (a) *Complete* – the student submitted 4 quotes, 4 sample query questions/instructions, and 4 corresponding SQL queries; and (b) *Correct* – each SQL String query is assessed on whether it is syntactically correct and whether or not it is fully functional (i.e. do the queries extract the data proposed in the student sample query instructions). The instructor returns the assessment to the student and provides feedback related to issues with query semantics. Software systems that automatically assess SQL statements can also be employed. These systems increase the efficiency of grading student submissions and enable students to continuously re-submit their solutions until competency is achieved (Kleiner, Tebbe, & Heine, 2013).

Prior to receiving formal feedback from the instructor, each student is asked to check their own queries (self-assessment) and compare them to one of their peers. The student is then asked to compare their own assessment to the instructor graded work. Each student should be encouraged to make the necessary corrections and resubmit their

queries to the instructor. This phase should help students make adjustments on their progress toward achieving the learning objectives. During this evaluation phase, students benefit from spending time thinking about the semantics of questions and expressing them in SQL.

CONCLUSION

The need for specialists in the big data and analytics area is axiomatic. One needs to only look at organizational hiring needs or jobs forecasts to see the explosive growth in data analytics related positions. Indeed, shortages in many analytics areas are so acute that organizations are simply “going without” or are attempting to make do with “homegrown talent.” One logical solution to the problem relates to Information Systems programs producing more and better graduates with extensive data capabilities.

The talents and skills that IS programs must facilitate to address big data concerns include data cleansing, data visualization, data manipulation, and identifying the story that is being told by the data. One critical notion is providing data analysts with an arsenal of SQL techniques that can be recalled on demand. Within the sphere of SQL capabilities and one that is not always included is the area of character data manipulation. The ability to skillfully manipulate character data is paramount for both data cleansing and querying. The tutorial presented here is both interesting and salient for students gain these skills.

For future research, we are going to test students that have completed this tutorial and determine if their abilities exceed those of students from comparable universities. We also intend to extend the depth and breadth of these modules in order further student capabilities.

AUTHOR BIOGRAPHIES

David Olsen is the Management Information Systems Department Head at Utah State University. He received his Ph.D. in Management Information Systems from The University of Arizona in 1993 and taught at The University of Akron accounting department in accounting information systems for five years. Dr. Olsen joined the MIS department at Utah State University in 1998 and teaches primarily in the database area as well as the MBA strategy and management course. His research interests include database concurrency control, accounting information systems, the integration of SQL, XML and XBRL, and database modeling. His research has been published in journals such as Communications of the ACM, Issues in Accounting Education, and the Journal of Database Management. Dr. Olsen is most pleased with the teaching and mentoring awards he has received. E-mail: david.olsen@usu.edu (contact author)

Pam Dupin-Bryant is a Professor of Management Information Systems in the Jon M. Huntsman School of Business at Utah State University. She earned her Ph.D. at the University of Wyoming and her master's and bachelor's degrees at USU. Throughout her career, Dr. Dupin-Bryant has employed a wide variety of delivery methods and educational strategies to facilitate learning for traditional and nontraditional students. Her primary teaching activities include business applications programming, web design/development, and principles of MIS. Her research and scholarly writings have focused primarily on information systems pedagogy, e-learning/distance education, and business ethics. Dr. Pam Dupin-Bryant has received many awards for her teaching innovations, research, and service. E-mail: pam.dupin-bryant@usu.edu

REFERENCES

- Aasheim, C., Williams, S., Rutner, P., & Gardiner, A. (2014). Big data analytics and data science undergraduate degree programs. Decision Sciences Institute Proceeding. Retrieved from <http://www.decisionsciences.org/Portals/16/Proceedings/AM-2014/files/p747132.pdf>.
- Adams, S. (2015). The 10 toughest jobs to fill in 2016. Forbes/Leadership. Retrieved from <http://www.forbes.com/sites/susanadams/2015/09/24/the-10-toughest-jobs-to-fill-in-2016/#7645d0f04596>
- Azat, S., Evgenia, S., & Tatiana, U. (2016). Advancement of information competencies in training a new generation of economists. In J. Becker, O. Kozyrev, E. Babkin, V. Taratoukhine, & N. Aseeva (Eds.), *Emerging trends in information systems: Recent innovations, results and experiences*. (pp. 79-83). Switzerland: Springer

International.

- Borthick, A. F., Schneider, G. P., & Viscelli, T. R. (2016) Analyzing data for decision making: Integrating spreadsheet modeling and database querying. *Issues in Accounting Education*.
- Bughin, J. (2016). Big data, big bang? *Journal of Big Data*, 3(2), 1-14.
- Bybee, R. W. (2009). The BSCS 5E instructional model and 21st century skills. The National Academies Board on Science Education. Retrieved from http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_080127
- Bybee, R., Taylor, J. A., Gardner, A., Van Scotter, P., Carlson, J., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. Colorado Springs, CO: BSCS.
- Cegielski, C. G., & Jones-Farmer, L. A. (2016). Knowledge, skills, and abilities for entry-level business analytics positions: A multi-method study. *Decision Sciences Journal of Innovative Education*, 14(1), 91-117.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.
- Davis, G. A., & Woratschek, C. R. (2015). Evaluating business intelligence/business analytics software for use in the information systems curriculum. *Information Systems Education Journal*, 13(1), 23-29.
- Dorji, U., Panjaburee, P., & Srisawasdi, N. (2015). A learning cycle approach to developing educational computer game for improving students' learning and awareness in electric energy consumption and conservation. *Educational Technology & Society*, 18(1), 91-105.
- Dumbill, E., Liddy, E. D., Stanton, J., Mueller, K., & Farnham, S. (2013). Educating the next generation of data scientists. *Big Data*, 1(1), 21-27.
- Dupin-Bryant, P. A., & Olsen, D. H. (2014). Business intelligence, analytics and data visualization: A heat map project tutorial. *International Journal of Management & Information Systems*, 18(3), 185-200.
- Edgington, T. M. (2011). Introducing text analytics as a graduate business school course. *Journal of Information Technology Education: Innovations in Practice*, 10, 207-234.
- Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East. EMC Corporation. Retrieved from <https://www.emc-technology.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
- Goldston, M. J., Dantzler, J., Day, J. & Web, B. (2013). A psychometric approach to the development of a 5E lesson plan scoring instrument for inquiry-based teaching. *Journal of Science Teacher Education*, 24(3), 527-551.
- Gorman, M. F., & Klimberg, R. K. (2014). Benchmarking academic programs in business analytics. *Interfaces*, 44(3), 329-341.
- Jobrack, B. (2011). *The 5E instructional model*. Retrieved from https://www.mheonline.com/secondaryscience/pdf/5e_lesson_cycle.pdf
- Kleiner, C., Tebbe, C., & Heine, F. (2013). Automated grading and tutoring of SQL statements to improve student learning. *Proceedings of the 13th Koli Calling International Conference on Computing Education Research, ACM*, 161-168.
- Krantz, L., & Lee, T. (2015). *The jobs rated almanac: The best jobs and how to get them*. iFocus Books: Lake Geneva, WI.
- Mamonov, S., Misra, R., & Jain, R. (2015). Business analytics in practice and in education: A competency-based perspective. *Information Systems Education Journal*, 13(1), 4-13.
- Murray, S. (2015). Future of big data: Rapid growth of business analytics degrees eases skills crunch. Retrieved from <http://www.businessbecause.com/news/future-of-big-data/3556/future-of-big-data-rapid-growth-of-analytics-degrees>
- Nadkarni, A., & Vesset, D. (2015). Worldwide big data technology and services forecast, 2015-2019. International Data Corporation. IDC #259532.
- Pittarese, T. (2015). Teaching business intelligence to computing students through a simulation and evolutionary lab sequence. *Journal of Computing Sciences in Colleges*, 30(5), 13-19.
- Piyayodilokchai, H., Panjaburee, P., Laosinchai, P., Ketpichainarong, W., & Ruenwongsa, P. (2013). A 5E learning cycle approach-based, multimedia-supplemented instructional unit for structured query language. *Educational Technology & Society*, 16 (4), 146-159.
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59.
- Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). Generation M²: Media in the lives of 8- to 18-Year Olds. A Kaiser Family Foundation Study. Retrieved from <files.eric.ed.gov/fulltext/ED527859.pdf>
- RHT. (2015). 2016 salary guide for technology professionals. Robert Half Technology. Retrieved from https://www.roberthalf.com/sites/default/files/Media_Root/images/rht-pdfs/robert_half_technology_2016_salary_guide.pdf
- Staff Writers. (2013, March 5). Cool teachers' guide to pop culture in the classroom [Web log post]. Retrieved from

- <http://www.onlineuniversities.com/blog/2013/03/cool-teachers-guide-pop-culture-classroom/>
- Waller, N. (2014). Get familiar with big data now—or face ‘permanent pink slip’: Demand rises for analytics professionals, data scientists. *The Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/SB10001424052702304819004579489541746990638>
- Wang, Y. (2015). Business intelligence and analytics education: Hermeneutic literature review and future directions in is education. *Proceeding of Twenty-First Americas Conference on Information Systems (AMCIS), Puerto Rico, 2015*.
- Wang, S., & Wang, H. (2013). Design and delivery of technical module for the business intelligence course. *Journal of Information Technology Education: Innovations in Practice* 13, 169-184.
- Wilder, C., & Ozgur, C. (2015). Business analytics curriculum for undergraduate majors. *Interfaces*, 44(3), 329–341.
- Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., Kulkarni, U., Mooney, B., Phillips-Wren, G., & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems*, 34, 1–13.

NOTES