

# COURSE APPROVAL/CHANGE DOCUMENT

(See back of form for instructions)

Submit

1. ADDITION  REVISION  TERMINATION

2. IF REVISION: denote changes (i.e. Title only; Title, CIP and Description, etc.):

3. COURSE NUMBER EG 316

4. COURSE TITLE DATA CLEANING & LINKAGE

5. IF REVISION: Previous Course No. \_\_\_\_\_ Previous Title \_\_\_\_\_

6. FOR ADDITIONS AND REVISIONS -  
FIRST TERM/YEAR TO BE OFFERED:

Fall  Spring  Summer  Term 2019

7. FOR TERMINATIONS ONLY -  
LAST TERM/YEAR TO BE OFFERED:

Fall  Spring  Summer  Term \_\_\_\_\_

8. COLLEGE: Coll of Sci, Tech, Engr & Math

9. DEPARTMENT NAME: Engineering & Technology

10. CIP CODE (Classification of Instructional Program / US Bureau of Labor Statistics): 14.2701

11. FIXED CREDIT HOURS: YES  NO

3 Total Credit Hours

3 Lec Contact Hours

\_\_\_\_ Lab Contact Hours

\_\_\_\_ Other Contact Hours

12. VARIABLE CREDIT HOURS: YES  NO

\_\_\_\_ Min Total Credit Hours \_\_\_\_ Max Total Credit Hours

\_\_\_\_ Min Lec Contact Hours \_\_\_\_ Max Lec Contact Hours

\_\_\_\_ Min Lab Contact Hours \_\_\_\_ Max Lab Contact Hours

\_\_\_\_ Min Other Contact Hours \_\_\_\_ Max Other Contact Hours

13. CAN THIS COURSE BE TAKEN FOR ADDITIONAL CREDIT: YES  NO  If YES, total number of times course can be taken \_\_\_\_\_

14. MAXIMUM ENROLLMENT ALLOWED FOR COURSE: 15 Justification of maximum enrollment:

Course will be computer based and frequent student-teacher interaction is needed (so small class size will facilitate the learning environment)

15. CLASS SCHEDULE TYPE/ FACULTY WORKLOAD: Choose appropriate schedule type: L - Lecture

Faculty Workload: 3 Class schedule type justification:

No labs involved

16. COURSE LEVEL: Undergraduate

17. GRADE TYPE: Standard Grade

18. DEVELOPMENTAL COURSE: YES  NO

19. CROSS-LISTED COURSE:

YES  WITH \_\_\_\_\_ NO

20. SPECIAL COURSE FEE? (Must be Board approved)

YES  Amount \$ \_\_\_\_\_ NO

21. Required faculty qualifications to teach this course:

PhD or MS degree with data analysis experience

22. UNIVERSITY STUDIES COURSE: YES  NO

If yes, please select one general education category:

If yes, please select up to three general education learning goals that reflect the priorities for student learning in the course. Please rank these in priority order, i.e. 1,2,3 by inserting the numbers/rankings into the boxes:

- \_\_\_ General Education Learning Goal 1: Ethical Reasoning
- \_\_\_ General Education Learning Goal 2: Global Learning
- \_\_\_ General Education Learning Goal 3: Information Literacy
- \_\_\_ General Education Learning Goal 4: Written Communication
- \_\_\_ General Education Learning Goal 5: Oral Communication
- \_\_\_ General Education Learning Goal 6: Critical Thinking
- \_\_\_ General Education Learning Goal 7: Quantitative Literacy

If the proposed new or revised course is a General Education/University Studies course, please provide a short rationale why this course should be considered as a general education course.

Attach the following:

- a) If applicable, memos from Department Chair(s) in affected department(s) stating support or that issues/conflicts are resolved.
- b) Memo from Library Dean assessing available and needed library resources and services.
- c) Class syllabus using the syllabus template: [Syllabus Template Link](#)

COURSE APPROVAL SIGNATURES

Department Chairperson

**Sophia Scott**

Digitally signed by Sophia Scott  
Date: 2019.04.09 10:42:47  
-05'00'

Dean of Kent Library

**Barbara C.  
Glackin**

Digitally signed by Barbara C.  
Glackin  
Date: 2019.05.03 13:31:01 -05'00'

College Council

**Brad Deken**

Digitally signed by Brad Deken  
Date: 2019.05.10 09:29:45  
-05'00'

Educator Preparation Committee

University Studies Council

Graduate Council

To obtain the next signature, save the pdf to your desktop and then email the form as an attachment to the next individual for signing. When submitting the form, the email must come from your Southeast email account.

Registrar's Office Use Only

SCACRSE \_\_\_\_\_ Bulletin \_\_\_\_\_ Degree Audit \_\_\_\_\_ SHATATR \_\_\_\_\_

**COURSE APPROVAL DOCUMENT**  
Southeast Missouri State University

Department: Engineering and Technology

Course No.: EG316

Title of Course: Data Cleaning & Linkage

Date: 2/28/19

Please check:  New  
Revision

I. Catalog Description (3 Credit Hours):

This course is designed to cover the theory and practice of data cleaning & merging process via an advanced analytic tool. (3)

II. Prerequisites: MA115 or MA116

III. Purposes or Objectives of the Course

Upon completion of this course, the students should be able to:

- A. Recognize common data quality issues and solutions
- B. Normalize and standardize data for optimal data usability
- C. Develop regular expressions for matching, parsing, tokenizing, replacing and transforming data
- D. Use probabilistic and deterministic record linkage tools to merge data
- E. Perform exploratory data profiling on industry-type data

IV. Course Learning Outcomes:

- A. Students will be able to perform detailed data munging and exploratory analysis.
- B. Students will be able to create a data profiling tool to identify the concepts while providing descriptive statistics to help make basic sense of the data and help assess the data quality.
- C. Students will be able to build a workflow to merge records from two or more sequential files.

V. Names of Faculty Qualified to Teach the Proposed Course:

- A. Dr. Serkan Varol
- B. Dr. Khaled Bawaneh
- C. Dr. Jingjing Tong
- D. Dr. Shaojun Wang

VI. Course Content or Outline (Indicate number of class hours per unit or section):


Topics	Hours
Basics of Metadata & unique identifiers	6
Basics of ETL operations (extract-transform-load) & Lookup tables	3
Data types and functions	6
Data cleansing and standardization	8
Midterm Exam	1
Data wrangling operations (transformation, joining, filtering etc)	6
Record Linkage Process	3
Deterministic and probabilistic record linkage models	6
Business Data Profiling & Exploratory Data Analysis	6
Final Exam	

45

Signature: 

Date: 3/2/19

Chair

Signature: 

Date: 3/6/19

Dean

## Class Syllabus Template

- (1) Course Number: **EG316**
- (2) Course Title: **Data Cleaning & Linkage**
- (3) Catalog Description: **This course is designed to cover the theory and practice of data cleaning & merging process via an advanced analytic tool.**
- (4) Prerequisites: **MA115 or MA116**
- (5) Co-requisites:
- (6) Credit Hours: **3**
- (7) Semester: **Fall 2019**
- (8) Class Meeting Time(s), location and format: **FACE TO FACE**
- (9) Instructor:
- (10) Instructor Contact Information:
- (11) Concerns: **Questions, comments or request regarding this course should be taken to the instructor. Unanswered questions or unresolved issues about this class can be directed to Dr. Sophia Scott, Chair, Department of Engineering and Technology ([sscott@semo.edu](mailto:sscott@semo.edu), 573-651-2965)**
- (12) (a) Course Learning Outcomes:
  1. **Students will be able to perform detailed data munging and exploratory analysis**
  2. **Students will be able to create a data profiling tool to identify the concepts while providing descriptive statistics to help make basics sense of the data and help assess the data quality**
  3. **Students will be able to build a workflow to merge records from two or more sequential files**
- (b) Purposes or Objectives of the Course:

**Upon completion of this course, the students should be able to:**

  1. **Recognize common data quality issues and solutions**
  2. **Normalize and standardize data for optimal data usability**
  3. **Develop regular expressions for matching, parsing, tokenizing, replacing and transforming data**
  4. **Use probabilistic and deterministic record linkage tools to merge data**
  5. **Perform exploratory data profiling on industry-type data**
- (13) Course-specific Required Materials:

**"Learning Alteryx: A beginner's guide to using Alteryx for self-service analytics and business intelligence" by Renato Baruti (ISBN: 978-1788392655, Publisher: Pack Publishing)**

(14) Course Content:

Topics	Hours
Basics of Metadata & unique identifiers	6
Basics of ETL operations (extract-transform-load) & Lookup tables	3
Data types and functions	6
Data cleansing and standardization	8
Midterm Exam	1
Data wrangling operations (transformation, joining, filtering etc)	6
Record Linkage Process	3
Deterministic and probabilistic record linkage models	6
Business Data Profiling & Exploratory Data Analysis	6
Final Exam	
Total:	45

(15) Grading Scale and Policies:

Project	20%
In Class Activity	15%
Homework	20%
Mid-term exam	20%
Final exam	25%
<b>Total</b>	<b>100%</b>

**Grading Scale:**  $90\% \leq A \leq 100\%$   
 $80\% \leq B < 90\%$   
 $70\% \leq C < 80\%$   
 $60\% \leq D < 70\%$   
 $F < 60\%$

(16) Final Exam Schedule: **TBD**

(17) Expectations of Students

- A. Attendance and participation are suggested to be prepared for projects and exams.**
- B. Students are expected to spend time outside of class to improve their practical skills relevant to the course material**
- C. Late assignments will not be accepted without prior approval from the instructor. Approval is generally granted in cases of documented emergency or degree-related professional activities.**
- D. Exams cannot be made up without prior approval from the instructor. Approval is generally granted in cases of verifiable/documentated excuses provided to the instructor at least 24 hours prior to the exam time**

- (18) **Academic Honesty** – Southeast Missouri State University expects all students, faculty and staff to operate in an honest and ethical manner. Academic dishonesty is a very serious offense because it undermines the value of your education and the education of others. Students who engage in academic dishonesty face significant penalties. Forms of academic dishonesty include, but are not limited to, plagiarism, cheating, contract cheating, misrepresentation, and other actions you take. Some of these are defined below:
- Plagiarism means passing off someone else’s work as your own, whether it is intentional or unintentional.
  - Cheating includes copying from another person or source of information to meet the requirements of a task.
  - Contract cheating is paying someone else or a company to do your work.
  - Misrepresentation means you are posing as someone else or someone else is posing as you to complete a task.
  - Collusion means working with one or more people to cheat. If you help someone cheat or plagiarize you will face the same penalties.

For more information, visit the Responsible Redhawks Code of Conduct

<http://www.semo.edu/responsibleredhawks/code-of-conduct.html> or the Faculty Handbook Section (D) on Academic Honesty <http://www.semo.edu/facultysenate/handbook/5d.html>

- (19) **Accessibility** – Southeast Missouri State University and Disability Services are committed to making every reasonable educational accommodation for students who identify as people with disabilities. Many services and accommodations which aid a student’s educational experience are available for students with various disabilities. Students are responsible for contacting Disability Services to register and access accommodations. Accommodations are implemented on a case by case basis. For more information, visit <http://www.semo.edu/ds/> or contact Disability Services at 573-651-5927.
- (20) **Civility** – Your university experience is purposely designed to introduce you to new ideas, help you think effectively, develop good communication skills, evaluate information successfully, distinguish among values and make sound judgements. Doing this well requires respectful and courteous discussion among and between students and the instructor. Together, we must create a space where we acknowledge and respect others have different experiences, perspectives and points of view. Disagreements are likely. Mutual respect for one another and a willingness to listen are important. Remember, you are responsible for your behavior and actions. There is a no tolerance policy on bullying or harassment of any kind. Additional information on student conduct may be found at: <http://www.semo.edu/pdf/stuconduct-code-conduct.pdf?ver=1.0> and [http://www.semo.edu/pdf/Conduct\\_Faculty\\_Resource\\_Guide.pdf](http://www.semo.edu/pdf/Conduct_Faculty_Resource_Guide.pdf)
- (21) **Mandatory Reporting** – I will keep information you share with me confidential to the best of my ability, but as a professor I am legally required to share information about sexual misconduct and crimes I learn about to make our campus and community safe for everyone.
- (22) **Student Success** – This course uses SupportNET, Southeast’s student success network, to improve communication between students, faculty and staff on campus. You’ll get emails through SupportNET with information about resources or concerns. Please read these emails—they are sent to help you succeed! You can access SupportNET through your portal, Moodle or directly at [supportnet.semo.edu](http://supportnet.semo.edu) to see any academic alerts, ask for help and to access resources to support your success at Southeast.

\*Definition of ‘blended’ to be added.