

# **INFO-H 643 Natural Language Processing (NLP) and Text Mining for Biomedical Records and Reports**

**Fall 2017**

**Time and location:** 03:00P-05:40P Tuesday at IT 256

**Total credits:** 3

**Instructor:** Dr. Mathew J. Palakal, PhD

**Co-Instructor:** Dr. Heng-Yi Wu, PhD ([hengwu@iupui.edu](mailto:hengwu@iupui.edu))

**TA:** Venu Kolli ([kollive@iu.edu](mailto:kollive@iu.edu))

**Office Address:** IT 475H

**Office Phone:** 317-278-7689

**Office Hours:** Tuesdays, 1:30pm – 2:30pm

**Email Address:** mpalakal@iupui.edu

**The Mission of IUPUI is to provide for its constituents excellence in**

- **Teaching and Learning**
- **Research, Scholarship, and Creative Activity**
- **Civic Engagement**

**With each of these core activities characterized by**

- **Collaboration within and across disciplines and with the community**
- **A commitment to ensuring diversity, and**
- **Pursuit of best practices**

IUPUI's mission is derived from and aligned with the principal components – Communities of Learning, Responsibilities of Excellence, Accountability and Best Practices – of Indiana University's Strategic Directions Charter.

## **Statement of Values**

IUPUI values the commitment of students to learning; of faculty to the highest standards of teaching, scholarship, and service; and of staff to the highest standards of service. IUPUI recognizes students as partners in learning. IUPUI values the opportunities afforded by its location in Indiana's capital city and is committed to serving the needs of its community. Thus, IUPUI students, faculty, and staff are involved in the community; both to provide educational programs and patient care and to apply learning to community needs through service. As a leader in fostering collaborative relationships, IUPUI values collegiality, cooperation, creativity, innovation, and entrepreneurship as well as honesty, integrity, and support for open inquiry and dissemination of findings. IUPUI is committed to the personal and professional development of its students, faculty, and staff and to continuous improvement of its programs and services.

## **Other Policies**

- Policy on Academic Dishonesty /Integrity  
[http://campuslife.indiana.edu/Code/Part\\_4B.html](http://campuslife.indiana.edu/Code/Part_4B.html)
- Policy regarding late work and make-up exams  
All late work regardless of the excuse will be discounted 50%.

- All students are responsible for reading the Code of Student Rights, Responsibilities and Conduct of IUPUI.  
[http://campuslife.indiana.edu/Code/Part\\_3A.html](http://campuslife.indiana.edu/Code/Part_3A.html)
- Policy on Plagiarism <http://www.indiana.edu/~wts/wts/plagiarism.html>

### **Policy regarding children attending**

“Children are not permitted to attend class with parents, guardians, or childcare providers. This conduct has the effect of unreasonably interfering with an individual’s work or academic performance creating an offensive learning environment.”

“A student must not violate course rules as contained in a course syllabus, which are rationally related to the content of the course or to the enhancement of the learning process in the course.”

[Code of Student Rights, Responsibilities, and Conduct, page 29]

### **1. Textbook**

#### **Text Analytics with Python**

A Practical Real-World Approach to Gaining Actionable Insights from your Data

Sarkar, Dipanjan, Apress, 2016

### **2. Course Objectives:**

This course familiarizes students with the basic analysis and applications of Natural Language Processing and text mining. This course introduces commonly used processes, algorithms, techniques, and software. The assignments are projects will provide hands-on experience with a variety of text mining applications.

### **3. Grading and course evaluation:**

Homework assignments: 40%

Quizzes: 20%

Quizzes will cover the content of the lectures, reading material, and exercises. Quizzes will be held in the room where the lecture section of the class regularly meets for half an hour.

Project: 40%

Project investigation/implementation will require more time commitment than for regular homework.

### **4. Course logistics:**

Canvas, the Indiana University online teaching resource, will be used for this course. Students will be given instructions to use Canvas after they enroll. All communication will be initiated from Canvas, so correctly specifying your email address and setting up the mail forward feature is critical for communication. Use the **Canvas** messaging to communicate with the Instructors and TA so your emails do not get lost.

	<b>Topic(s)</b>	<b>Labs</b>	<b>Due Date</b>
Week 1 8/22/2017	Course introduction and organization Overview, what and why Text mining?	Introduction to Text Mining and Natural Language Processing	
Week 2 8/29/2017	Applications of NLP (General, clinical, Bio surveillance and Public Health) Terminologies, Coding, and Natural Language Processing	NLTK Introduction <b>Assignment #1</b>	9/05/2017
Week 3 9/05/2017	Linguistic Essentials, Morphology, stemming, tokenization	Stemming and Tokenization using NLTK <b>Assignment #2</b>	9/12/2017
Week 4 9/12/2017	Part of speech tagging & intro to graphical models (CRF and HMM)	POS tagger using NLTK <b>Assignment #3</b>	9/19/2017
Week 5 9/19/2017	Named entity recognition	NER using NLTK, <b>Assignment #4</b>	9/26/2017
Week 6 9/26/2017	Regular Expressions and rule based NLP methods	<b>Quiz #1</b> , Regex tutorial, <b>Assignment #5</b>	10/03/2017
Week 7 10/03/2017	Statistical/machine learning methods	Weka Tutorial <b>Assignment #6</b>	10/10/2017
Week 8 10/10/2017	Higher level NLP tasks: Co-reference, Word sense disambiguation, Negation detection	Negex and ConText <b>Assignment #7</b>	10/24/2017
<b>Week 9</b> <b>10/17/2017</b>	<b>Fall Break</b>		
Week 10 10/24/2017	Higher level NLP tasks: Context free grammar Shallow parser Semantic role labeler Dependency parser	Parsers <b>Assignment #8</b>	10/31/2017
Week 11 10/31/2017	Text classification	<b>Project pre-proposal Due</b> <b>Assignment #9</b>	11/07/2017
Week 12 11/07/2017	Text summarization	<b>Project final proposal Due</b> <b>Quiz #2</b> <b>Assignment #10</b>	11/28/2017
Week 13 11/14/2017	Document clustering	Q&A related to the project	
Week 14 11/21/2017	Case study 1: Biomedical text analysis	Q&A related to the project	
Week 15 11/28/2017	Case study 2: HER text analysis	Q&A related to the project	
Week 16 12/05/2017	Case study 3: Sentiment analysis	Q&A related to the project	
Week 17 12/12/2017	Final project presentations	Project Presentations	

## Academic Calendar (Fall 2017) - Important Dates

<b>Fall Term</b>	<b>Day</b>	<b>2017</b>
Classes Begin	M	Aug. 21
Labor Day - no classes	M	Sept. 4
Fall Break	M,T	Oct. 16-17
Thanksgiving Break Begins	W	Nov. 22
Classes Resume	M	Nov. 27
Last Day of Classes	M	Dec. 11
Finals Begin	T	Dec. 12
Finals End	F	Dec. 15