

June 2, 2024

## **Fundamentals of Applied Digital Research in South Asian Languages**

The Lester and Sally Entin Faculty of Humanities - East Asian Studies  
Tel Aviv University  
**June 2024**

Instructor: Gil Ben-Herut, University of South Florida

### **Course Details, Schedule, and Location**

Course Number 0687-2572-01

Mode of Instruction: Seminar

Semester: 2023/2

Credit Hours: 4

Mondays and Thursdays between 14:00 and 18:00

Dates: 30/5, 3/6, 6/6, 10/6, 13/6, 17/6, 20/6, 24/6, 27/6

Rosenberg Building, Room 104, and at the [DH computer lab](#) of the TAU Library

### **Instructor Details**

Gil Ben-Herut, PhD

Associate Professor

Department of Religious Studies

University of South Florida

[gilb@usf.edu](mailto:gilb@usf.edu)

[Click here for my faculty page](#)

[Click here for my resume and publication list](#)

### **Course Description**

This graduate seminar introduces and teaches basic techniques and skills in applied digital research to advanced students with at least two years of training in classical or modern South Asian languages. The course is built around lesson units, each dedicated to one technological topic or area, including: digitization and tokenization, text annotation and analysis, online environments and computing literacy (GitHub, APIs, and JSON, WordPress/Omeka/Wix), and more. Each lesson includes a discussion (with relevant secondary literature) and a session in a computer lab with hands-on training and assignments. The course also includes individual meetings with each student for developing a personal exploratory DH project during the course that will be submitted as the seminar's final assignment (רפרט).

### **Course Format**

- A 4-academic hour seminar structure, with two weekly meetings over four weeks (total of 9 lessons).
- Short frontal lectures and student computer workspaces for assignments and exercises during class.
- The course will be delivered in English and Hebrew

### **Required Background for Students**

Advanced undergraduate (Bachelor's degree) and graduate students with basic training in South Asian languages who are interested in acquiring skills and understanding in how to do digital research.

## Lesson Plan

---

### Lesson #1: 30/5

#### Part 1: Introductions

Instructor / students

#### Part 2: Online Links

Sanskrit Dictionaries and Other Online Tools

---

### Lesson #2: 3/6

#### Part 1: Digitization of Texts Composed in Indian Languages (Lab)

OCR Collab: Abby FineReader, Adobe Acrobat Pro, Tesseract OCR, Transkribus, Google Vision

Tokenization and Unsandhi

#### Part 2: Tagging Texts (Lab)

Voyant / AntConc

TEI and XML / CoNLL-U

---

### Lesson #3: 6/6

#### Part 1: Computer Literacy (Lab)

Microsoft Word / writing with South Asian languages

Bibliography managers

Power editors and VimMotions

#### Part 2: Distant Reading. Close Reading

\* Moretti on Distant Reading (מאת פרנקו מורטי "השערות על ספרות העולם")

\* Marin Paul Eve, *Close Reading with Computers*, pp. 9–24

---

### Lesson #4: 10/6

#### Part 1: General (Lab)

Working with Catma and Recogito

#### Part 2: Led by Abigail Penn (Lab)

A case study of a Sanskrit text with Catma

---

### Lesson #5: 13/6

#### Part 1: Internet (Lab)

Communication over the internet

Omeka / WordPress / Wix / GitHub / API, JSON

#### Part 2: AI (Lab)

AI, LLM, NLP, and [nanoGPT](#)

Techniques for Digital Textual Analysis

---

### Lesson #6: 17/6

#### Part 1: History and Culture of Open Source

\* Neal Stephenson, *In the Beginning ... Was the Command Line*, pp. 24–40

\* Eric S. Raymond, *The Cathedral & The Bazaar*, pp. 21–63

Programming in Python

#### Part 2: digitalRoses

\* Mary Rader, [Read, Hot and Digitized: More is less? Less is more? Minimal computing in South Asian Lexicography](#)

---

## Lesson #7: 20/6

### Part 1: Discussion (Lab)

\* Martin Paul Even, *Close Reading with Computers*, excerpts: pp. 26–37, 61–81, 99–104, 155–9

### Part 2: Techniques (Lab)

Practicing techniques from the book

---

## Lesson #8: 24/6

### **Working at the lab on individual projects**

---

## Lesson #9: 27/6

### **Student Project Presentations**

---