May 17, 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](https://cenlib.tau.ac.il/Digital-Humanities) 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](https://www.usf.edu/arts-sciences/departments/religious-studies/people/faculty/gil-ben-herut.aspx)

[Click here for my resume and publication list](https://hcommons.org/members/gilbenherut/)

# 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 Asian languages. \*\*Perhaps change this to a project-based learning\*\*The course is built around lesson units, each dedicated to one technological topic or area, including: API and JSON, OCR for Indic scripts, text analysis, fundamentals of programming in Python, 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 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

**Beginnings**

**\*\* Attention! This lesson is held at Rosenberg Building, Room 104**

Part 1: Introductions

Instructor / Oded—lab facilitator / students

Part 2: Online Links

Sanskrit Dictionaries and Other Online Tools

Lesson #2: 3/6

**Processing Sanskrit Texts**

Part 1: Digitization

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

Tokenization and Unsandhi (Python)

Part 2: Tagging

TEI and XML / CoNLL-U

Voyant / AntConc

Lesson #3: 6/6

**Annotation of Sanskrit Texts**

Part 1: General

Working with Catma and Recogito

Part 2: Led by Avigail Penn

A case study of a Sanskrit text with Catma

Lesson #4: 10/6

**Computer and Internet Literacy**

Part 1: Computer

Microsoft Word / writing with South Asian languages

Bibliography managers

Power editors and VimMotions

Part 2: Internet

Communication over the internet

Omeka / WordPress / Wix / GitHub / API, JSON

Lesson #5: 13/6

**DH and AI**

Part 1: Distant Reading. Close Reading

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

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

Part 2: AI

NLP and [nanoGPT](https://github.com/karpathy/nanoGPT)

Techniques for Digital Textual Analysis

Lesson #6: 17/6

**Open Source and “Bottom-up” Programming**

Part 1: History of Open Source

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

\* Eric S. Raymond, *The Cathedral & The Bazaar*, pp. 21−63Programming in Python

Part 2: digitalRoses

\* Mary Rader, [*Read, Hot and Digitized: More is less? Less is more? Minimal computing in South Asian Lexicography*](https://texlibris.lib.utexas.edu/2022/05/read-hot-and-digitized-more-is-less-less-is-more-minimal-computing-in-south-asian-lexicography/)

Lesson #7: 20/6

**Close Reading with Computers**

Part 1: Discussion

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

Part 2: Techniques

Practicing techniques from the book

Lesson #8: 24/6

**Working at the lab on individual projects**

Lesson #9: 27/6

**Student Project Presentations**