Workshops Digital Humanities Institute – Beirut 2015

DHI-Beirut workshops will run every day from Monday 2 March to Friday 6 March for two hours in the morning (10-12 noon) and two hours in afternoon (2-4pm) and they will include both instruction and hands-on sessions. There will also be a set of "lightning" talks, keynotes and social events in the lunch and afternoon hours. So that it is possible for all graduate students to attend, the Departments of English and Arabic will not hold graduate courses in the week of 2-6 March 2015.

Because of the intensive nature of the workshops, _it will only be possible for each participant to attend one workshop this week_. Students who attend the whole workshop and the events of the week will be issued, upon request, a certificate of participation in DHIB.

Priority is given to members of the AUB community, but participants from other foreign and Lebanese institutions may attend as well provided that they register. Workshops are free for all members of the AUB community. Depending on the numbers of outside participants, a nominal fee may be charged to those attending from outside.

Workshops will be offered in English, but many of the participants are fluent in French and Arabic. Access to computer labs will be limited. Participants should plan on bringing their own laptop computers. Instructors will assist participants in downloading any needed tools.

WORKSHOP OFFERINGS (ONLY ONE MAY BE TAKEN PER PERSON)

(1) "Text Processing on the *nix Command Line" (Kamal Bou Mikhael, @kamaLdotAm, AUB) - Jafet Library e-classroom / Fisk Hall 1st floor Education Lab / West Hall Room 310

The command line interface (CLI) found in Unix based systems (*nix) facilitates efficient, innovative, and elegant interaction with the computer that differs considerably from that of the standard click-based graphical user interface. In addition to the standard commands related to browsing of directories and listing of files, many powerful text processing and publishing tools exist in *nix CLI environments. Standard text processing tools allow the search, sort, and manipulation of plain text files. In addition, command line mechanisms allow (1) the output of one command to serve as the input to another and (2) data to be written to or read from files. Since all such input is based on typed text, sequences can be saved as automated scripts for later reuse. Additional related tools are also available on the command line. LaTeX and pdftk are publishing related tools for the creation and manipulation of PDF files. vim is a command line based text editor with an internal set of commands that facilitates powerful text editing. Complimentary to all the above tools are regular expressions (REs), a general mechanism that is used in many contexts of data processing; they can be employed for effective string matching and substitution. These are crucial to the effective operation of the above tools. This workshop will acquaint the student with the tools and mechanisms described above. The skills learned will be useful for general use and applicable to projects that involve the analysis and annotation of textual data. Use cases will be presented from the instructor's everyday experience as well as from current linguistic projects he is embarking on. Students will come away familiar with skills to process text in a precise and efficient manner. They will also acquire the knowledge needed to explore the other command line tools that they may find useful in the future.

(2) "Introduction to Python for the Humanities" (Loa Aude and Tracey El Hajj @TraceyDH_Leb, AUB) - Issam Fares Institute Seminar Room 2

This workshop is an introduction to programming for data analysis and visualization that requires no previous programming background. It will start by introducing the fundamental concepts of

programming – variables, expressions, selection structures, repetition structures. After that, our emphasis will be shifted toward writing programs to analyze data gathered from data sets stored in files of different format, such as comma-separated (csv) files or excel files. Among others, reading data into different data frames, combining and merging data, pivoting data, and handling missing values will be covered. Visualizing the analyzed data as different kinds of plots and charts will be also introduced. The language that will be used throughout this workshop will be Python because of its easy-to-learn syntax and powerful extensions. One of these extensions that will be exhaustively used is Python Data Analysis library or Pandas. Python can be downloaded for free [here](#).

(3) "Spatial Humanities and Digital Mapping" ([DJ Wrisley](#), AUB [@DJWrisley](#) and [Randa El-Khatib](#), AUB, [@Randa_DH](#)) - Issam Fares Institute Seminar Room 1

This workshop seeks to build participant appreciation for location-based humanities research, relevant kinds of spatial data and ways of visualizing them on different kinds of interfaces. Topics in this workshop include curating structured spatial data using [gazetteers](#) and other digital encyclopedic resources, faceted visualization of spatial information, linking maps and timelines, digital storytelling with maps, [geo-referencing](#) historical maps as well as the social and cloud-based world of mapping. Students will be introduced to [Google Maps](#)/[Earth](#), [MapWarper](#), [Palladio](#), [CartoDB](#), [TimeMapper,](#) [Animaps](#), [Odyssey](#) and [Neatline](#). All fields are welcome and no prior knowledge required. The workshop will be tailored to the specific research questions of the participants. Understanding of basic spreadsheets is desirable. As a group, we will look critically at the availability of map data and geospatial library holdings for the Arab region. Participants in this workshop are highly encouraged to attend the talk by data scientist Aurelia Moser from CartoDB on 23 February 1-3pm at AUB "Storytelling with Time, Space and Torque: Building Narrative with Thematic and Temporal Maps" [Event Poster](#)

(4) "Linguistic Annotation for Lesser-Studied Languages" ([Anaid Donabedian](#), [IRD](#)/AUB, Loubna Dimachki, Lebanese University, Veronique Traverso CNRS-France/AUB) - Center for Language Research and Teaching (CELRT) library, Fisk Hall 209

This workshop is an introduction to theory and practice of corpus annotation in lesser-studied languages with a language-oriented approach. It will explore different tools, compare their features, their convenience and ease-of-use for different kinds of corpora and different types of annotation for different purposes (linguistic analysis, content analysis, conversational analysis). It will cover (1) the oral corpus transcription charts, methods and tools, (2) the preparation of the set of properties for written or transcribed corpus annotation depending on the underlying research question (3) the identification of unities to be annotated, (4) most common annotation and glossing charts with a focus on linguistic approach, (5) the tools and methods of annotation depending on the language and the purpose, including semi-automatized annotation tools. The workshop will be include small presentations of research projects, using the methods discussed in the workshop. Participants will test the different techniques on a corpus of their choice. They will be asked to come with two Unicode-encoded files in a language of their choice (either lesser-studied or not), one of them around 2000 signs, and one between 50000 and 100000 signs. All fields are welcome where there is interest in language-specific annotation including linguistic features (language description, language variation, language use, content analysis) even as a part of a study in another field (anthropology, sociology, literature, history, psychology, etc.).

(5) "Creating a Digital Exhibit with Omeka and Neatline" ([Alex Gil](#), [@elotroalex](#), Columbia U) - IT/Academic Core Processes and Systems lab

Participants will download [Omeka](#), a server-based platform for creating digital exhibits used by many cultural institutions and libraries, along with its plug-in [Neatline](#) for creating digital narratives with

maps and timelines. Together they will mount a digital exhibit. No particular coding experience is necessary. Experience with metadata basics is desirable.

(6) **"From Analogue Documents to Electronic Texts: Introduction to TEI XML Editing in Multilingual Environments"** (Till Grallert, @tillgrallert, Frei Universität Berlin/Orient Institut Beirut) - Fisk Hall 2nd floor English lab

This five-day workshope will provide a hands-on introduction to digital editing practices and tools employing a set of XML-based technologies: The Extensible Markup Language (XML) itself (Specifications, Wikipedia), the XPath data model (Specifications, Wikipedia), the declarative language to transform XML (XSLT: Extensible Stylesheet Language Transformations, Specifications, Wikipedia), and finally a specific "dialect" of XML, the Text Encoding Initiative's TEI P5. The workshop will cover the process from the original creation of digital files using TEI XML to the publication of XML files using XSLT. To do so in a meaningful way, each participant will encode a single issue of one of Beirut's many newspapers covering the restoration of the Ottoman constitution in summer 1908. The American University of Beirut's library holds a vast number of newspapers, including, but not limited to, *Lisān al-Ḥāl*, *Thamarāt al-Funūn*, *al-Bashīr*, *al-Iqbāl*, and *al-Nashra al-ʿUsbūʿiyya*, from which the examples will be selected. Finally, we will problematise the applicability of the TEI to non-European languages written in non-Latin scripts, particularly Arabic and Ottoman. No previous programming skills or knowledge of mark-up languages is required, even though previous experiences are an advantage. The workshop will be adapted to the needs of registered participants. Participants for whom Arabic is not their main research interest, will be encouraged to bring their own 4 page text in which writing direction, non-Latin alphabet, non-Western units, dates or names, or any other feature of multilingualism is of particular importance.

(7) **"New Forms of Scholarly Production"** (Matt Hiebert, @HiebertMatthew, U Victoria) - Arts and Humanities Initiative, Bldg 37 - large room

This workshop seeks to introduce new platforms and methods for online production and publishing of scholarly editions, books, articles, and electronic literature. Participants will be taught presentational markup (HTML5 and CSS3) and will learn about digital humanities pertinent concepts such as metadata, open source software, game-design models, and social knowledge creation. During the week, participants will develop their own web-based project using one of production platforms to be covered: CommentPress (WordPress), Scalar, Twine, and WikiMedia. Participants are asked to bring their own laptop computer. All fields are welcome. No prior knowledge required.

(8) **"Digital Project Management"** (Lynne Siemens, @lynnelynne53, U Victoria) Arts and Humanities Initiative, Bldg 37 - small room

This offering will cover the basics of project management from project definition to project review upon completion. Topics such as budget setting and controls, risk management, critical path scheduling, software tools, creating and sustaining teams, and related Internet resources will also be discussed. Material will be covered through lectures, discussions, and case studies.