IDK Wintersemester 2021/22 Basisseminar II:

Einführung in die Digital Humanities zum Erwerb von Kenntnissen über digitale Konzepte, Methoden und Techniken zur Aufbereitung, Analyse und Auswertung digitaler Daten. Freitag 10.00 Uhr bis 12.00 Uhr, Raum B 011A. Ansprechpartner: Dr. Christian Riepl (riepl@lmu.de).

A. Hinführung

1. **Begrüßung, Vorstellung, Einführung und Überblick (29.10.2021, Riepl)**
    a. Begrüßung
    b. Kurzvorstellung der IT-Gruppe Geisteswissenschaften
    c. Kurzvorstellung der Promotionsvorhaben
    d. Einführung: Digitalisierung - Data Literacy - Digital Humanities - Digital Philology
    e. Überblick zur Veranstaltung
    f. Überblick und Erläuterungen zum DHVLab

2. **Grundlagen: Jenseits von Windows und Office (5.11.2021, Riepl)**
    a. Zeichenkodierung: Wie kommt die Schrift in den Rechner? - ASCII, Unicode, Betacode, Transliteration, Transkription
    b. Textverarbeitungsprogramme (Word) und Texteditoren (gvim, notepad++): Graphische und logische Strukturierung und Auszeichnung
    c. Datenstrukturen: Dateinamen, Dateiformate, Verzeichnisse
    d. Das Konzept der „Regulären Ausdrücke (regular expressions)“: Zeichenmuster, Regeln, Aktionen - Automatisierung
    e. Einfache Werkzeuge zur Datentransformation: UNIX-Shell und UNIX-Tools

B. Grundlagen der Datenverarbeitung und Modellierung

3. **Einführung Programmierung in Python Teil 1 (12.11.2021, Frank)**
    a. Grundlagen
    b. Installation und Editoren
    c. Datentypen, Variablen, Listen, grundlegende Operationen
    d. Einfache Funktionen und Methoden (z.B. Stringmethoden, Eingabe, Ausgabe etc.)

4. **Einführung Programmierung in Python Teil 2 (19.11.2021, Frank)**
    a. Konditionale (und Logik), Schleifen
    b. Dateiverarbeitung (lesen, schreiben)
    c. Reguläre Ausdrücke
    d. Eigene Funktionen, Modularisierung, Fehlerbehandlung

5. **Einführung XML Teil 1 (26.11.2021, Frank)**
    a. Grundlagen strukturierte Daten
    b. Anwendungsszenarien (Literaturwissenschaften)
    c. Wohlgeformte XML-Dokumente
    d. Dokumentengrammatik (DTD)

6. **Einführung XML Teil 2 (3.12.2021, Frank)**
    a. XPath-Abfragen
    b. Einblicke in XSLT
    c. XML-Verarbeitung in Python (mit BS4)
    d. TEI P5-Standard für die Kodierung von Textdokumenten

7. **Einführung SQL und Neo4J (10.12.2021, Frank)**
    a. Grundlagen relationale Datenbanken
    b. MySQL und die SQL-Abfragesprache (in Auswahl: CREATE, SELECT)
    c. Grundlagen von Graph-Datenbanken (inklusive Neo4J-Installation)
    d. Neo4J und die Cypher-Abfragesprache (in Auswahl: CREATE, MATCH)

C. Analysemethoden, Annotation und Datenvisualisierung

8. **Digitalisierung und manuelle Annotationsmethoden (17.12.2021, Englmeier/Schön)**
    a. Manuskriptdigitalisierung / OCR (Transkribus, OCR4All)
    b. Tools für manuelle Annotation und Transkription (Squirrel)

9. **Wiederholung und Fragen (7.1.2022)**

10. **Methoden der Korpuslinguistik (14.1.2022, Gacia/Wisiorek)**
    a. theoretische Grundlagen (Repräsentativität, balanced usw.)
    b. Ressourcen: Standardkorpora und lexikalische Ressourcen (Stopwort-Listen, Wortlisten, WordNet)
    c. Korpusauswertung
        - Korpora einlesen / Text Processing (Wortlisten)
        - Frequenzlisten (auch Wordcloud)
        - Konkordanzlisten / KWIC
        - Kollokationen, ngrams
        - Verteilungsplots
        - Lexical Diversity
    d. Beispiel: Anwendung auf Beispielkorpus

11. Automatische Annotationsmethoden (21.1.2022, Wisiorek)

a. Lemmatisierung
b. POS-Tagging & Morphologische Features
c. Syntaktische Annotation (Dependency Parsing, IOB-Tagging)
d. NER (Named Entity Recognition)
e. Sentimentanalyse
f. Beispiel: Extraktion von NER-Daten aus Korpus

12. Methoden der Datenanalyse (28.1.2022, Wisiorek)

a. Dataframes
b. Datenauswertung (deskriptive Statistik)
c. (explorative Methoden: Clustering mit scikit-learn; oder in 4 bei Stilometrie)
d. Datenvisualisierung
e. Netzwerkanalyse
f. Beispiel: Analyse der NER-Daten (u.a. Personennetzwerk im Korpus)

13. Methoden der Textanalyse (4.2.2022, Wisiorek)

a. annotierte Korpora einlesen (Tupellisten; Formate: json (Twitter), XML-Korpora)
b. Keyword-Extraction / Topic-Modelling
c. Kollokationsnetzwerke
d. Stilometrie (Clustering mit Features der quantitativen Stilistik)
e. Beispiel: Analyse der Korpus-Text-Daten (u.a. Genres in Brown-Corpus)

D. Forschungsdaten und Nachhaltigkeit

14. Forschungsdatenmanagement (11.2.2022, Bayer)

a. Standards und Normdaten
b. FAIR-Prinzipien
c. Open Access (Creative Commons Lizenzen)
d. Repositorien
e. NFDI