BA3 – Digital Humanities

**Time Allocation per Class (2 hours total, with 10-min break)**

* **First Half (~50–55 min)**: Lecture/instruction
  + Context, concepts, and “why this matters.”
  + Live demonstration
  + Discussion/Q&A to connect technical content to Korean Studies.
* **Break (10 min)**
* **Second Half (~50–55 min)**: Hands-on
  + Students replicate/demonstrate.
  + Short guided exercise (corpora provided + step-by-step goals).
  + Wrap-up discussion: “What did you see? What questions do you have? How could this connect to your thesis work?”

**🔹 Six-Week Course Plan (*Draft*)**

**Week 1: Introduction to DH, GitHub, and Data Management**

* **Lecture/Discussion**: What is DH? Why text-as-data matters in Korean Studies. How computational tools can complement close reading. Importance for thesis-writing (even if not directly used).
* **Hands-on**:
  + Set up GitHub accounts, create repositories.
  + Introduce FAIR principles (Findable, Accessible, Interoperable, Reusable).
  + Host syllabus, datasets, and repos on GitHub.
* **Orange Tutorials**:
  + *01: Welcome to Orange*
  + *02: Data Workflows*
  + *03: Widgets and Channels*  
    *(group these as “Orientation to Orange”)*

**Week 2: Preprocessing Text (with Aron v/d Pol)**

* **Lecture/Discussion**: Why preprocessing matters; tokenization, stopwords, cleaning, normalization. Show examples with Korean text (common stopwords, particles, etc.).
* **Hands-on**: Work with pre-prepared Korean corpora, clean them in Orange.
* **Orange Tutorials**:
  + *16: Text Preprocessing*
  + *19: How to Import Text Documents*  
    *(group as “Text Input and Cleaning”)*

**Week 3: Basic Text Analytics — Descriptive Patterns**

* **Lecture/Discussion**: Frequency, keywords, word clouds, what “counts” mean in DH. How to move from descriptive to interpretive claims.
* **Hands-on**: Run simple frequency/word cloud analyses, visualize token distributions.
  + Small-group work with pre-prepared corpora (e.g., ROK presidential speeches, DPRK speeches, progressive vs. conservative newspapers).
  + Run word frequency and word cloud analyses; identify and interpret top 10 terms.
  + Compare focal keywords (e.g., 민족 vs. 경제) using collocations or simple contrasts.
  + Apply clustering/projection (Tutorials 17 & 20) to explore whether texts group by theme or author.
  + Groups share brief observations: What did you see? What surprised you? How might this connect to a thesis question?
* **Orange Tutorials**:
  + *17: Text Clustering*
  + *20: Multivariate Projection – Freeviz*  
    *(group as “Exploring Text Patterns”)*

**Week 4: Classification & Prediction in Text**

* **Lecture/Discussion (≈50 min): What classification does, how supervised learning works, and how these methods relate to thesis writing (e.g., categorization of texts, labeling sources).**
* **Hands-On (≈50 min):**
  + **Apply sentiment classification to sample corpora.**
  + **Evaluate results with model scoring.**
  + **Discuss strengths and limitations of applying machine learning to Korean historical and literary data.**
* **Orange Tutorials:**
  + ***18: Text Classification***
  + ***06: Making Predictions***
  + ***07: Model Evaluation and Scoring*  
    *(grouped as “Text Classification & Evaluation”)***

**Week 5: Clustering, Similarity & Topic Modeling**

* **Lecture/Discussion (≈50 min): Unsupervised approaches for discovering patterns without labels. How clustering reveals structure in corpora, and how topic modeling identifies recurring *themes* as distributions of words. Emphasis on interpretive limits: clusters and topics are heuristic tools, not ground truth.**
* **Hands-On (≈50 min):**
  + **Apply hierarchical and k-means clustering to Korean corpora (e.g., political speeches, newspaper articles).**
  + **Compare clusters across corpora.**
  + **Run Topic Modeling in Orange (LDA via Text Mining add-on) to extract themes; interpret top words per topic.**
  + **Discuss: What kinds of research questions are better suited to clustering vs. topic modeling?**
* **Orange Tutorials:**
  + ***05: Hierarchical Clustering***
  + ***11: k-Means* and *12: k-Means Explained***
  + ***13: Silhouette***
  + ***(Supplement with Topic Modeling widget demo — no official short video, but simple workflow)*  
    *(grouped as “Clustering & Topic Modeling Methods”)***

**Week 6: Student Mini-Project & Wrap-Up**

* **Lecture/Discussion**: How to design a simple text-as-data project: from research question → corpus → preprocessing → analysis → visualization → interpretation.
* **Hands-on**:
  + Students pick one of the corpora provided.
  + Run one descriptive + one analytical method (e.g., word cloud + clustering, or classification + evaluation).
  + Present short write-up in GitHub repo.
* **Wrap-Up**: Revisit thesis-writing. Show how this skillset can strengthen arguments, even if not directly used.