



HSS302: Big Data and Language (Edu 3.0 course)

🕒 Class Time

MTWTh 13:00 – 16:00

📍 Location

To be announced

📖 Credit

3

👤 Instructor

Professor Seonmin PARK ([seonminpark\(at\)kaist.ac.kr](mailto:seonminpark@kaist.ac.kr))

Tel. 042-350-4679, Office @1212, N4

📚 Required Materials

A laptop if possible

Course Summary

Course Description: In this course, students will understand characteristics of language through big data. Students will learn how to collect and analyze big data, and find linguistic features from the data. A number of approaches to the linguistic analysis of written and spoken texts will be discussed.

Goals & Objectives:

- To understand linguistic features in texts based on big data
- To apply scientific knowledge to linguistic analysis

Course Evaluation

Student Evaluation:

20% Participation & Attendance / 5% Proposal / 25% Mid-term Exam / 25% Presentation / 25% Final paper

Attendance Policy: 3 unexcused absences may result in failure of this course / documentation for excused absences must be provided to professor by student. (Over 15 minutes late = 1 absence)

	Contents		Contents
1	Introducing this course	9	Big data and vocabulary: ANTCONC
2	Characteristics of data: spoken texts	10	Big data and syntax: TagAnt
3	Characteristics of data: written texts	11	Group presentations
4	Register and big data	12	Group presentations
5	Big data analysis tools: BNC	13	Big data for different languages
6	Big data analysis tools: COCA	14	Limitations of big data analyses
7	Big data analysis tools: MICASE	15	Future direction of big data
8	Mid-term Examination	16	Final Examination

Note: This is a tentative schedule for this course. The instructor reserves the right to make changes to this syllabus based on the needs/interests of the students and/or any changes in circumstances.

Grade Scale:

Grade	100- points	4.3- points
A+	97- 100	4.3
A0	94-96	4.0
A-	90-93	3.7
B+	86-89	3.3
B0	83-85	3.0
B-	78-82	2.7
C+	75-77	2.3
C0	72-74	2.0
C-	67-71	1.7
D+	64-66	1.3
D0	61-63	1.0
D-	53-60	0.7
F	0-52	

Late Policy:

In all cases, extensions must be applied for via email before the due date and should be supported with appropriate documentation (medical certificate etc). No extensions will be granted because you have not adequately organized your time. Presentations will be expected to be done at the beginning of the next class.

Academic Honesty:

All work submitted in this class is expected to be the original work of the student whose name appears on it and the work was prepared originally for this course. Academic dishonesty includes cheating, plagiarism, unauthorized collaboration, falsifying academic records, and any act designed to avoid participating honestly in the learning process. If there has been any academic dishonesty in preparing for the assignment, you will receive a zero for that particular assignment.

Attendance:

You are expected to attend every class of the semester. Your attendance and class participation grade will decrease every time you are late. If you arrive late four times, it will be considered as one absence. If you arrive more than 15 minutes late for class, you are considered absent. Each unexcused absence will result in a 2% deduction from your final grade. 3 or more unexcused absences will result in an automatic F in the course. In all cases, excused absences must be requested via email at least one day before the class and should be supported with appropriate documentation. Students sleeping, texting, or playing games etc. in class may also be regarded as absent.