LT4234: LINGUISTIC COMPUTING

Effective Term
Semester B 2022/23

Part I Course Overview

Course Title
Linguistic Computing

Subject Code
LT - Linguistics and Translation
Course Number
4234

Academic Unit
Linguistics and Translation (LT)

College/School
College of Liberal Arts and Social Sciences (CH)

Course Duration
One Semester

Credit Units
3

Level
B1, B2, B3, B4 - Bachelor's Degree

Medium of Instruction
English

Medium of Assessment
English

Prerequisites
LT3232 Computing Programming for Language Studies

Precursors
Nil

Equivalent Courses
CTL3234 Computational Linguistics II, CTL4234 Linguistic Computing

Exclusive Courses
Nil

Part II Course Details

Abstract

This course aims to teach students essential programming techniques for linguistic computing and language technology tasks, focusing on the underlying programming concepts and principles, using a general-purpose programming language.

Course Intended Learning Outcomes (CILOs)

	CILOs	Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Analyze computer programming tasks chosen from typical linguistic computing and language technology applications.		x	x	x
2	Draw up specifications for computer programs for the selected tasks.		x	x	x
3	Apply essential programming concepts, including basic data structure and algorithms to design and implement computer programs for the selected tasks competently and creatively.			x	x

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: Accomplishments
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

Teaching and Learning Activities (TLAs)

	TLAs	Brief Description	CILO No.	Hours/week (if applicable)
1		Learning and implementing algorithms for tasks selected from off-line and online linguistic data processing, sentence splitting, tokenization and word segmentation, lemmatization, parts-of-speech tagging, and preliminary sentence parsing.	1, 2, 3	

2		Lectures on the above selected tasks to explain and illustrate the basic issues involved and necessary programming techniques for a practical solution for each of them.	1, 2, 3	3 hours
3		Readings of lecture notes and selected chapters from textbooks.	1, 2, 3	
4		Homework assignments to help students to solve their programming problems and perform demos of their programs for the above tasks.	1, 2, 3	

Assessment Tasks / Activities (ATs)

	ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Assignments to design and implement (the kernel part of) the programs for the above tasks;	1, 2, 3	50	
2	Demos of running programs in tutorials;	1, 2, 3		
3	Quizzes (optional)	1, 2, 3		

Continuous Assessment (%)

50

Examination (%)

50

Examination Duration (Hours)

2

Assessment Rubrics (AR)

Assessment Task

1. Assignments

Criterion

Knowledge, attitude, ability, creativity, accomplishment and performance in completing and/or presenting demons and/or assignment

Excellent (A+, A, A-)

Excellent knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Excellent, creative application of computing and programming knowledge to basic tasks of linguistic computing. Very active participation and high marks/ performance.

Good (B+, B, B-)

Good knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Good application of computing and programming knowledge to basic tasks of linguistic computing. Active participation and good marks/ performance.

Fair (C+, C, C-)

Adequate knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Fair application of computing and programming knowledge to basic tasks of linguistic computing. Adequate participation and fair marks/ performance.

Marginal (D)

Basic familiarity with the subject matter. Marginal ability to apply basic computing and programming knowledge to basic tasks of linguistic computing. Marginal participation and marginal marks/ performance.

Failure (F)

Poor familiarity with the subject matter. Poor ability or fail to apply computing and programming knowledge to basic tasks of linguistic computing. Poor participation and poor marks/ performance.

Assessment Task

2. Demos of running programs

Criterion

Knowledge, attitude, ability, creativity, accomplishment and performance in completing and/or presenting demons and/or assignment

Excellent (A+, A, A-)

Excellent knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Excellent, creative application of computing and programming knowledge to basic tasks of linguistic computing. Very active participation and high marks/ performance.

Good (B+, B, B-)

Good knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Good application of computing and programming knowledge to basic tasks of linguistic computing. Active participation and good marks/ performance.

Fair (C+, C, C-)

Adequate knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Fair application of computing and programming knowledge to basic tasks of linguistic computing. Adequate participation and fair marks/ performance.

Marginal (D)

Basic familiarity with the subject matter. Marginal ability to apply basic computing and programming knowledge to basic tasks of linguistic computing. Marginal participation and marginal marks/ performance.

Failure (F)

Poor familiarity with the subject matter. Poor ability or fail to apply computing and programming knowledge to basic tasks of linguistic computing. Poor participation and poor marks/ performance.

Assessment Task

3. Quizzes

Criterion

Marks

Excellent (A+, A, A-)

Excellent knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Excellent, creative application of computing and programming knowledge to basic tasks of linguistic computing. Very active participation and high marks/ performance.

Good (B+, B, B-)

Good knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Good application of computing and programming knowledge to basic tasks of linguistic computing. Active participation and good marks/ performance.

Fair (C+, C, C-)

Adequate knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Fair application of computing and programming knowledge to basic tasks of linguistic computing. Adequate participation and fair marks/ performance.

Marginal (D)

Basic familiarity with the subject matter. Marginal ability to apply basic computing and programming knowledge to basic tasks of linguistic computing. Marginal participation and marginal marks/ performance.

Failure (F)

Poor familiarity with the subject matter. Poor ability or fail to apply computing and programming knowledge to basic tasks of linguistic computing. Poor participation and poor marks/ performance.

Assessment Task

4. Examination

Criterion

Marks

Excellent (A+, A, A-)

Excellent knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Excellent, creative application of computing and programming knowledge to basic tasks of linguistic computing. Very active participation and high marks/ performance.

Good (B+, B, B-)

Good knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Good application of computing and programming knowledge to basic tasks of linguistic computing. Active participation and good marks/ performance.

Fair (C+, C, C-)

Adequate knowledge of major issues, concepts, principles, techniques and skills in linguistic computing. Fair application of computing and programming knowledge to basic tasks of linguistic computing. Adequate participation and fair marks/ performance.

Marginal (D)

Basic familiarity with the subject matter. Marginal ability to apply basic computing and programming knowledge to basic tasks of linguistic computing. Marginal participation and marginal marks/ performance.

Failure (F)

Poor familiarity with the subject matter. Poor ability or fail to apply computing and programming knowledge to basic tasks of linguistic computing. Poor participation and poor marks/ performance.

Part III Other Information

Keyword Syllabus

Text data processing.
Basic Web text processing: Web crawling and text extraction from Web page.
Sentence splitting: splitting of text into sentences.
Tokenization and word segmentation: Token vs. word; identification of tokens/words in English and Chinese texts.
Lemmatization: conversion of variant word forms to their base forms (lemmas).
Part-of-speech tagging: assignment of part-of-speech tag to words.
Preliminary sentence parsing: context-free grammar and chart parsing.

Reading List

Compulsory Readings

	Title
1	Lecture notes/slides for the course
2	Selected papers/chapters on topics of linguistic computing
3	Selected tutorials on key tasks of programming and implementation for linguistic computing
4	Selected topics of Java programming from the Java Tutorials Online provided by Oracle at https://docs.oracle.com/javase/tutorial/
5	Online API (Application Programming Interface) specification for selected Java classes needed in the programing for this course

Additional Readings

	Title
1	Relevant chapters in the recommended reading list or form online materials
2	Advanced and/or related topics of programming and implementation
3	Advanced and/or related topics of Java programming from the Java Tutorials Online provided by Oracle at https://docs.oracle.com/javase/tutorial/