CMSC 673 Graduate Assessment, Fall 2023

CMSC 673 — Introduction to Natural Language Processing

Item	Summary	Points
Checkpoint 1	Monday October 9th	10
Checkpoint 2	Monday November 6th	25
Checkpoint 3	Friday November 17th	65
Final turn-in	Friday December 8th	100

All due dates are 11:59 PM Baltimore time.

For the core graduate assessment (worth 40% of the CMSC 673 grade), choose **ONE** of the following assessments, (GA1)–(GA5). Four of them will require you to implement NLP algorithms. One of them is to write a 4 page (conference-style) literature review paper on a particular topic. While each person is free to choose which of these five options they will complete, I encourage each person to strongly consider how each may be able to further their career/degree goals and choose the assessment that maximizes progress toward their goal. For example, someone writing a disseration (doctoral or masters) may wish to do the literature review paper, since it could help form the basis of the related work chapter of a dissertation. On the other hand, someone may wish to really dive deep into an aspect of NLP implementation.

The possible assessments are divided into an "Implementation Track," and a "Paper Track". Both tracks consist of advanced topics that we *may* very lightly cover, or mention, in the course. However, we will not spend significant time on any of them.

Implementation Track

(GA1) Implement a linear chain conditional random field and evaluate it on a sequence tagging task, such as named entity recognition, part-of-speech tagging, syntactic chunking, or a task of your choice.

(GA2) Implement the inside-outside and CKY algorithms for probabilistic context free grammars, and evaluate it on semi-supervised syntactic parsing.

(GA3) Using the UD corpora, implement and evaluate a neural arc-standard dependency.

(GA4) Re-implement the "Character-Aware Neural Language Models" work of Kim et al. (2016, AAAI).

Paper Track

(GA5) Write a literature review paper in which you select one of a set of topics, and (i) identify, (ii) analyze, and (iii) synthesize modern approaches for the topic you choose.

Allowed and Disallowed Resources Regardless of the track you choose, the only reference materials you may peruse, read, skim, look at, use, or otherwise consult are

(Y1) published textbooks (online versions or drafts are acceptable);

(Y2) peer-reviewed papers;

(Y3) arXiv preprints;

(Y4) slides from tutorials or presentations for papers given at conferences;

(Y5) course materials from this class (such as slides, lecture discussion, and Discord notes and discussion);

(Y6) the official documentation of community-standard libraries, like numpy, scipy, sklearn, Pytorch, and huggingface;

You may receive sanctioned UMBC writing assistance, such as through the GSA Writing Advisor.[1]

You are **not** allowed to peruse, read, skim, look at, use, or otherwise consult any of the following:

(N1) blog posts;

(N2) any sort of article, writing, manuscript, or slides not covered by the "Y" list (unless specifically approved in advance, in writing);

(N3) code or pseudo-code available online from existing implementations;

(N4) other online courses and their materials.

The use of any of the "N" items will be considered a serious academic integrity violation and will result, *at a minimum*, in an automatic 0 on this entire assessment (i.e., a 0 will be recorded for 40% of your entire course grade).

The rest of this document is divided into two main sections: one describing the "Implementation Track" tasks, milestones, and deliverables; and the other describing the "Paper Track" tasks, milestones, and deliverables. All options have four components: three checkpoints and the final turn-in. All checkpoints share the same due dates. While what is required for the milestones depends on whether you choose the implementation track or paper track, the first checkpoint/milestone is the same for all options.

The following table summarizes what is due when:

Item	**If you select from the Implementation Track:**	**If you select from the Paper Track**	**Due**
Checkpoint 1	Selection of one of (GA1)–(GA4), and initial list of resources	Selection of topic for (GA5), and initial list of papers	Monday October 9th
Checkpoint 2	Initial report discussing progress, hurdles, and other challenges; and git repo containing at least three non-trivial/starter commits	Complete first draft and paper source	Monday November 6th
Checkpoint 3	Feedback/suggestions on another student's initial report	Review	Friday November 17th
Final turn-in	Completed code, full git repository, completed writeup, and document summarizing how the feedback was used	Completed final version, paper source, and summary-of-changes document	Friday December 8th

[1] `https://gsa.umbc.edu/writing-advisor/`