HW0: Getting Used to GradeScope

CSCI 662: Fall 2025

out: August 25, 2025
due: August 29, 2025, 11:59:59 AoE

The purpose of this assignment is to get you familiar submitting your code to GradeScope, and to help us make sure our autograders are working. This assignment should take under an hour – it is not intended to be difficult.

Code To Write

- import `spaCy` and load the `en_core_web_md` model.
- write a function `tokenize_string()` that takes a string as input and returns a list of tokens
- write a function `get_pos_tags()` that takes a sentence as a string and returns a list of POS tag lists, where each list contains the simple POS tag, the detailed POS tag, and the syntactic dependency.
- write a function `is_in_vocab()` that takes a token as a string and returns `True` if the token is in the `en_core_web_md` model's vocabulary and `False` if the token is out of vocabulary (OOV).
- write a function `sentence_lengths()` that takes a file path as an argument (you can assume a plaintext file), reads the document, and returns a list of the number of tokens in each sentence of the document.
- write appropriate docstrings for all 4 functions and a very brief `README.md`

Coding Requirements

- This assignment will use `spaCy` (`https://spacy.io`). Use spaCy to implement the functions described above.
- Use spaCy's `en_core_web_md` model. The autograder will download this model for you, using `python -m spacy download en_core_web_md`. You should then be able to load it using `spacy.load()`
- All your functions should be in a python file named `solution.py`. The autograder will not run if your submission is not named correctly.
- All homework submissions must include a `requirements.txt`. This should include all dependencies for your code. Do not include unnecessary/unused dependencies. We will install your dependencies using `pip install -r requirements.txt`.
- All homework submissions should include a `README.md` file telling us how to use your code. For HW0, this will be trivial - one sentence or less is fine.
- Make sure the GradeScope auto-scoring script runs and gives reasonable results. There are no hidden test cases - WYSIWYG!

Coding Recommendations

- Your code will be run on an Ubuntu 22.04 system using Python 3.10.
- To ensure similar behavior on your system and ours, we strongly recommend including package versions in your `requirements.txt`.
- The autograder will not work and you may lose points if your files are not named correctly. Check autograder output to verify you have submitted all 4 required files (`solution.py, requirements.txt, README.md, hw0_report.pdf`

Your Report

For other homeworks, you will be expected to write a report detailing what you did and why you did it. For HW0, we are primarily interested in making sure you know how to write and compile a LaTeX document. Your submitted PDF file should be named `hw0_report.pdf`.

For this class, all your written materials should use ACL style files: https://github.com/acl-org/acl-style-files. On line 5, change `review` to `preprint`.

There are many ways to write and compile LaTeX; I generally use Overleaf (www.overleaf.com) for minimal headaches, but I have colleagues who abhor Overleaf and greatly prefer to compile on their own machines. Do what works for you. You can manage citations in LaTeX however you like; we recommend using `natbib`, which is already integrated with the ACL style files.

For HW0, your "report" should include:

- a few sentences about why you enrolled in this class and what you hope to learn
- a citation (in ACL style) to a paper you enjoyed recently. This is primarily to make sure you don't have headaches with LaTeX citations and bibliographies later.
- Did you have any problems submitting to GradeScope? describe what happened so we can debug. (screenshots appreciated but not mandatory)

There is no minimum length for your report. It shouldn't be more than a page, and could be a lot less.

Grading

For HW0, grading will be mostly participation based - if you submit the right files with the right info, you'll get full credit. Grading will be roughly broken down as follows:

- 50% – Did you submit the required files? is your code correct? Did you implement what was asked for, and did you do it correctly? (This part is autograded)
- 50% – did you submit a PDF document, typeset in LaTeX using ACL style, with the required info?

Rules

- This assignment does not count towards your final grade, and late submissions will be allowed without using late days. However, you **must** submit HW0 before you can submit any other assignments in this class.
- This is an individual assignment. You may not work in teams or collaborate with other students. You must be the sole author of 100% of the code and writing you turn in.
- You may not look for solutions on the web, or use code you find online or anywhere else.
- Generative language, code, and vision models (e.g. ChatGPT, Llama 2, Midjourney, Github Copilot, etc.; if you are unsure, ask and don't assume!!) can be used with the following caveats:

  - You must declare your use of the tools in your submitted artifact. If you don't declare the tool usage but you did use these tools, we will consider that as plagiarism.
  - For code and image generation, you must indicate the prompt used and the output generated.
  - For text generation you must provide either a link to the chat session you used to help write the content or an equivalent readout of the inputs you provided and outputs received from the system. You will lose credit if "the AI" is doing the work rather than you.

- Failure to follow the above rules is considered a violation of academic integrity and is grounds for failure of the assignment or, in serious cases, failure of the course.

- We use plagiarism detection software to identify similarities between student assignments, and between student assignments and known solutions on the web. Any attempt to fool plagiarism detection, for example the modification of code to reduce its similarity to the source, will result in an automatic failing grade for the course.

- If you have questions about what is and isn't allowed, post them to Slack!