COMP 4446 / 5046 Assignment 2

Due: Monday, May 15th (ie., the start of week 12 of semester)

In this assignment, you will learn about a key part of NLP: data annotation. This is often the most critical part of work on a project. If you do not create accurate datasets for training and evaluation then it doesn't matter how good your model is, you will not be able to build an effective system.

The assignment has a series of stages. Note that after stage 1 you need to wait for us to send you a file before you can do stage 2 (and the rest of the assignment). We will respond within 3 business days.

0 - Forming Groups of 1-3 students

You may work on your own or in a group of 2 or 3 students. You need to do two things in this stage:

1. Form groups on Canvas. See instructions here.
2. Once you have your group, please write it down in this spreadsheet.

   Please record both the members in your group and the group ID you have in Canvas.

1 - Initial Annotation

In this section of the assignment, you will annotate data (~2,500 tokens of text) and develop annotation guidelines that describe your annotation process. This annotation should be done by you, not by an AI model.

1. Download the data from here.
2. Find the file that matches your group number (from the spreadsheet in stage 0).
3. Read the initial annotation guide (Google Doc or Word Doc). Note, the guide has been updated to include that nested named entities should be annotated. The guide has an example.
4. Each student in your group should independently, without discussion, annotate the file and keep notes on (a) examples of each category, and (b) explanations of what you chose to do in unusual cases, along with examples.

   Note: you may discuss technical decisions about what tool to use for annotation, how to set it up, debugging running it, etc.
5. Store your annotations as a ".txt" file with one line per annotation, in this format:

   ((line start, token start), (line end, token end)) - label

   For example, if a Person entity spanned the first two tokens in the third line, you would have:

```
                        ((2, 0), (2, 1)) - PER
```

For items that span a single token you may save it in one of two ways (either is acceptable):

```
                        ((0, 5), (0, 5)) - PER
                        (0, 5) - PER
```

Note:
- The numbering starts from 0
- Tokens are specified by splitting the file on whitespace
- Blank lines count when determining the line number

6. Meet as a group and create a new version of the annotation guide that adds:
   - New examples of each category that come from your data.
   - Discussion of unusual cases, with the decisions each of you made and what your group has decided is the best approach in future.

To do the annotation, you may use any tool you like. We recommend SLATE, which follows the file format described above. Some other free options are docanno, and INCEpTION. If you use a tool that has an auto-annotation mode or semi-automatic annotation mode (e.g. brat and prodigy have such modes), please do not use it in this assignment. All annotations should be done by you.

Submit - https://canvas.sydney.edu.au/courses/48399/assignments/446897

(a) PDF of your annotation guide
(b) Text files containing annotations, one text file for each person in your group

2 - Adjudication and Refinement

In this section, you will adjudicate disagreements in the annotations of your file. If your group has N members then you will be comparing N+! annotation files (the extra one is the one we provide to you). This adjudication should be done by you, not by an AI model.

1. Download the annotations we provide and find the file for your group.
2. Go through the annotations as a group and resolve every case where the annotations do not match. After doing this you should have a single file that is the agreed annotations.
3. At the same time, add and remove examples and explanations from your annotation guide so that it explains your decisions.
   a. For content you want to remove from the annotation guide, draw a line through the text (ie., a strikethrough).
   b. For content you want to add, include it in blue text so it is clearly different from the original text (which should be in black).

Note - you should not add or remove entire label types here. Always use the 6 types we specified in the initial annotation guide. You are only changing the guide to clarify how to do annotation for cases that might be ambiguous or tricky.

Submit - https://canvas.sydney.edu.au/courses/48399/assignments/452047

(a) PDF of the revised annotation guide with strikethroughs and blue text
(b) Text file of the final annotations

3 - Improved Annotation

Now, you will annotate another piece of text, using your revised guidelines. This annotation should be done by you, not by an AI model.

1. Download the data from here.
2. Find the file that matches your group
3. Independently, without discussion, each student in your group should annotate the file.

Submit - https://canvas.sydney.edu.au/courses/48399/assignments/452048

(a) Text files of your annotations, one text file for each person

4 - Evaluation Metrics

In this section, you will implement a metric to see how consistent your annotations are.

1. Implement F-Score (see lecture 8 or this wikipedia article). Note, to be considered a match, an annotation must have the same span and the same label.
2. Calculate F-Score for each pair of annotations in stage (1) of the assignment, including the annotations we provide. If you are working on your own this means calculating the F-Score between your annotations and the ones we provided. If you are in a group of 2 you will calculate three values (person A - person B), (person A - provided), (person B - provided). If you are in a group of 3 you will calculate six values.

   If you are working in a group with 2+ people then also calculate the average of the values.
3. Repeat the previous step using the data from stage (3) of the assignment. If you are working on your own, you should compare with the output in this file.

Submit - https://canvas.sydney.edu.au/courses/48399/assignments/452051

An ipynb file containing (a) your code for calculating the metrics and (b) the results of your calculations in (2) and (3).

5 - Model Evaluation

In this section, you will measure the accuracy on your data of three widely-used models.

1. Run [Flair](#), [SpaCy](#), and [Stanza](#) on your data. Note, you should use their 18-class NER models.

   The 18 classes those models produce include 5 of the ones we consider here. You should post-process the output of the models to remove cases where they use a label we are not using (e.g. TIME).
2. Evaluate on your data from stage 2 (ie., the adjudicated data).

   Note: You will need to map their output to our format. Sometimes tokenisation will not match up exactly. That's okay - it will impact scores, but you will still be able to compare the three.

Submit - https://canvas.sydney.edu.au/courses/48399/assignments/452052

(a) Text files containing the output of the three models on your adjudicated data (stage 2), in the format specified in stage 1
(b) A text file containing:

    Flair - SCORE
    SpaCy - SCORE
    Stanza - SCORE

Where 'SCORE' is replaced by the F-Score for comparing the model's output to your adjudicated annotations.

6 - [Bonus] Competition

This is an optional section where you train models for this task. We will provide all of the data students have submitted in stages 1, 2, and 3, which you can use for training. You will be tested on a separate dataset annotated by the tutors.

More details of the competition will be released later. It will also have a deadline 1 week after the main assignment deadline (May 22nd). There will be NO EXTENSIONS for the competition.

The competition can either be completed in the same group as for the rest of the assignment, or on your own.

If you receive bonus marks and they take your overall mark for the assignment over 20 (ie. 100%) then the bonus can count to your overall non-exam course mark.

Mark Allocation

The table below shows the value of each section, broken down across the items you submit.

Section	Value	Breakdown

1 - Initial Annotation	5	2 - Annotations 2 - Annotation guide examples 1 - Annotation guide explanations
2 - Adjudication and Refinement	5	2 - Adjudicated annotations 3 - Annotation guide updates
3 - Improved Annotation	2	2 - Annotations
4 - Evaluation Metrics	4	2 - Code 2 - Results for the two calculations using the code
5 - Model Evaluation	4	1 - Flair output 1 - SpaCy output 1 - Stanza output 1 - Scores for the three models

Note: Your annotations MUST match the file format we specified in stage 1. If they do not, you will score 0 for them.

Bonus points in the competition are awarded as follows:

- Top 25% of entrants, +1 point
- Top 10% of entrants, +2 points
- Top 2 entrants, +3 points