

## **Scenario 1**

**Text** -> "activation of **vy9vδ2** t cells by non-peptidic antigens induces the inhibition of subgenomic hcv replication"

**Desired Tokens** -> [activation, of, **v, 9v, 2**, t, cells, by, non, peptidic, antigens, induces, the, inhibition, of, subgenomic, hcv, replication]

**Not Desired** -> [activation, of, **v9v2**, t, cells, by, non, peptidic, antigens, induces, the, inhibition, of, subgenomic, hcv, replication]

Treat any character other than A-Z, a-z, 0-9 as special characters, and handle them as you would handle any punctuation: replace with " " and split.

## **Scenario 2**

**Do not** create a txt file with .json extension, and dump the outputs. Use the **json.dump** function provided by the json package

## **Scenario 3**

Be mindful of duplicates :)

## **Scenario 4**

**Do not** output the postings list for all the dictionary terms. You are required to output the postings list **only for the query terms**.

## **Scenario 5 (Very Important !!!)**

Check the names of the keys of the output json. They **must match** with the provided sample json.

For example:

1. **"postingsList"** and **not** "postingsLists" or "postings List"
2. **"daatAnd"** and **not** "daat AND" or "DaatAnd"
3. **"results"** and **not** "result" or "Results"
4. **"num\_docs"** and not "num\_doc" or "num docs"
5. **"num\_comparisons"** and not "num comparison" or "comparisons"

## **Scenario 6**

Make sure you are merging postings lists in an optimized order. Many of you lost points due to num\_comparisons outside the acceptable range of +/- 5%.

## **Scenario 7**

**Do not** generate intermediate files

### **Scenario 8**

Do not try to read corpus and query using r+

### **Scenario 9**

Corpus or query file location hardcoded in the code

### **Scenario 10**

Packages used in the program not imported

### **Scenario 11**

KeyError, IndentationError, NameError, pandas.errors.ParserError, list AttributeError, IndexError: list index out of range, TypeError: 'int' object is not iterable.

### **Scenario 12**

The username in the file name submitted does not match with the folder name. For example, few of you have named your script as <UB Person number>\_project2.py, instead of <UB person name>.