Strathmore
UNIVERSITY

SCHOOL OF COMPUTING AND ENGINEERING SCIENCES

BACHELOR OF SCIENCE IN INFORMATICS AND COMPUTER SCIENCE

ICS 3204: COMPUTER GRAPHICS

CAT 1 - 20 Marks

TIMELINES

Class Activities	Dates
Group Discussions and Solutions	**25th - 29th Sept, 2023**
Group Presentations	**2nd - 6th Sept, 2023**

Data File:

MASSIVE Dataset by AMAZON

Instructions:

Revisit your **Codelabs** to solve these questions. You have to upload everything on your **GitHub** accounts. Remember we have to still show our contribution graphs for full scores. This task will be explained in class so you have to make sure that you attend the session.

QUESTION 1 - Python3 Development Environment **10 Marks**

1. Set up a new Python3 Development environment for this assessment. Install all the dependencies that you think will be relevant.
    a. Build a Python3 project with the structure of projects in PyCharm then import the **MASSIVE** Dataset mentioned on the Data File above.
    b. In this dataset, the pivot language is English, given that all the **`ids`** of the languages are matching, generate a **`en-xx.xlxs`** file for all the languages. In this question use the **`id`**,

`utt` and the `annot_utt`. **Do not** use **Recursive** algorithms in this solution as they have a time complexity of $O(n^2)$, which is bad for memory.

   i. Have a look at **Flags** to help you run this on your `generator.sh` files

---

QUESTION 2 - Working with Files **10 Marks**

2. For English (en), Swahili (sw) and German (de), generate separate `jsonl` files with test, train and dev respectively.
3. Generate one large json file showing all the translations from `en` to `xx` with `id` and `utt` for all the **train** sets.
   a. Pretty print your `json` file structure.
4. Upload all the files to your Google Drive Backup Folder.
   a. Upload all the changes to GitHub
   b. Write a clean `readme.md` file

Presentation: Your Slides should show the following information

1. Introductions
   a. Group Members with admission numbers
2. Solution
   a. What approach did you take?
   b. How did the team collaborate on the project?
3. Code Demo
   a. Walk us through the codebase
   b. Walk us through your GitHub
      i. Show insights, readme, and any other relevant information.