SCHOOL OF COMPUTER SCIENCE

MSC STUDENT RESEARCH PROGRESS REPORT FORM

Note: The activities reported here cover the period starting from the beginning of your Masters studies, or since the last time you wrote a M.Sc. progress report.

Name: Michael Haaf

Supervisor(s): Morgan Sonderegger

Year in the Program: 2

Dates of Applicable Time Period: September 2018 to January 2022

1. **Courses:**

    a. Courses Taken and Grades:

Code	Title	Grade	Credits	Semester
COMP 598	Topics in Computer Science 1	A	3	Fall 2021
COMP 547	Computer Networks	B+	4	Fall 2021
COMP 547	Cryptography & Data Security	A-	4	Fall 2020
COMP 596	Topics in Computer Science 3	A	3	Fall 2020
COMP 766	Advanced Topics Application 1	B+	4	Winter 2019
COMP 551	Applied Machine Learning	A	4	Winter 2019
COMP 598	Topics in Computer Science 1	A-	3	Fall 2018
COMP 550	Natural Language Processing	A	3	Fall 2018

Topic course subjects:

Topic code	Semester	Subject
COMP 598	Fall 2021	Introduction to Data Science
COMP 596	Fall 2020	Fron Natural Language to Data Science
COMP 766	Winter 2019	Sociocultural and Epistemic Understandings of Computer Science
COMP 598	Fall 2018	Mathematical Foundations of Machine Learning

    b. Courses Audited: N/A
    c. Courses TA'd:

Code	Title	Semester
COMP 424	Artificial Intelligence	Winter 2022
COMP 550	Natural Language Processing	Winter 2019
COMP 189	Computers and Society	Fall 2018
COMP 251	Algorithms and Data Structures	Fall 2018

   d. Courses Taught: N/A

2. **Publications (indicate if they are refereed):**

   a. Journals: N/A
   b. Conferences:

   Haaf, M. & Fortier-Dubois, E. (2019). Local stylometric features for authorship attribution in French fiction. CSDH/SCHN Digital Humanities Conference 2019, Vancouver, CA. Poster presented by Michael Haaf.

   c. Other: N/A

3. **Software or System Built:** N/A

4. **Talks Given at McGill or Elsewhere, Workshops Attended, Research Visits:** N/A

5. **Prizes and Awards:** N/A

6. **Committee Service (if relevant to research):** N/A

7. **Summary of Research Progress (Note: Items 7 and 8 together should not exceed one page):**

   - Research Assistantship onboarding:
     - account created and development software set up on Compute Canada, Roquefort, and Mimi servers.
     - diagnostic alignment tasks (English and Cantonese corpuses) complete.
     - transcript to testgrid processing script implemented.
     - syllable datatype with constructor for list of phonemes unit/function tested and implemeted.
     - serializer/deserializer pattern unit/function tested and implemented to generalize data processing across languages and data sources.
   - Capstone project onboarding:
     - development software (MFA and polyglotDB) installed and run.
     - initial research into package development (conda, architecture, dependencies) complete.

8. **Research Plan for the Next 4 Months:**

- Research Assistantship tasks:
  - work with Connie to implement robust raw corpus -> MFA alignment pipeline generalized across languages.
  - codebase refactoring tasks (implement generalizable patterns, unit testing)
- Capstone project tasks:
  - Short term goals (defined and completed on a monthly basis):
    * publish polyglotDB usage tutorials (5-10 over the course of 4 months) demonstrating variety of useful applications for public non-expert consumption
    * publish polyglotDB installation tutorials, investigate common issues experienced by non-expert users
  - Long term goals (complete by May 1 2022):
    * repackage polyglotDB for automated install to work across common architectures (proposal: create condaforge package for polyglotDB)
  - Strech goals (if all other goals are complete):
    * implement known "breakfast experiments": demonstrate polyglotDB utility in novel "quick" research problems
    * create new "breakfast experiements": conduct novel research using polyglotDB

Overall Research Progress:

Satisfactory _____, **NOT Satisfactory** _____

Comments of the Supervisor(s):

- TBD

Student's signature: _____

Supervisor's signature: _____

Date: January 31, 2022