Theme/Title Bachelor Thesis AGREEMENT

Student's first name and last name: Daniel Avram Pop

Specialization: Computer Science in English

Class academic years: 2018-2021

First name and last name of the scientific advisor: Adrian Ioan Sterca

Thesis domain: Natual Language Processing

Thesis title: *A Comparison of Some Data-driven Extractive Summarisation Strategies for Generating Abstracts of Theological Journals' Articles*

Bachelor thesis topic description:

The topic of the Bachelor thesis will be described, minimum 2-3 paragraphs in natural language.

Living in a time where quintillions of bytes of data are created daily, a way to get the most important part of it has become a must. Automatic summarisation is the natural language processing task that concerns itself with extracting the most relevant information from a given document (or a collection of such) and putting it into a human-understandable format. One of the main strategies in achieving that goal is weighting the importance of each sentence in the whole document and join together the ones that are the most fitting in what would be a (supposedly) coherent résumé commonly known as "extractive summarisation".

The aim of my research is to apply two such methods to scientific texts, a subset of all the texts available online that has (at least) one unique feature: such texts almost everytime are paired with an author-written abstract, which is vital in evaluating the accuracy of the machine-generated summary. Narrowing the topic further, I would focus on New Testament Exegesis articles which have one more special quality: their text combine English with Ancient Greek and specialised technichal terms in the field of hermeneutics (mostly in Latin and German).

With this endeavour, I hope to be able to single out that features of a (very) technical text that one has to bear attention to when training an automatic summarisation algorithm and also, try to deduce whether heavy NLP tasks (like word lemmatization) worth their computational effort when taking into consideration the gain they bring to the accuracy of the final result.

Date, **Student signature,**

3.3.2021

Scientific coordinator signature,