COMP492 SENIOR DESIGN PROJECT II

PROPOSAL FORM

DEPARTMENT OF COMPUTER ENGINEERING

Project Name

Turkish Question Generation Model

Project Summary (Abstract)

Our goal in this project is to develop a Turkish Question Generation System using available online sources such as Wikipedia. Question Generation Systems are capable of generate various logical questions from the given text input. QG Systems are prevalent in several computer applications such as chatbots, automated grading systems etc. Development of such a system is particularly important since there are not any examples of online tools and available datasets in Turkish. However, QG research has already reached a notable level in other languages such as English and there are available datasets for this task. For instance, SQuAD by Stanford University is one of the well- known datasets used for this task in English.

In this project, we are aiming to create the first Turkish Question Generation Dataset of moderate size. This step is required for developing learning-based solutions that we're going to apply in this project. We plan to apply different machine learning based techniques for questions of generating varying forms (Who-What-When etc.) and assess their performances on the dataset we'll gather.

Furthermore, in the next semester, we are going to explore the applicability of deep learning models on this topic. Outcome of this research will be integrated into a larger project which focuses on Automated Question Answering system for the online lectures in the Turkish Language. (This is a joint project with EE Dept. under the supervision of Ebru Arisoy Saraclar)

Keywords

Automatic speech recognition, Natural Language Processing, Deep Learning, Sentiment Analysis, Education Technology, MOOC (Massive Open Online Courses).

Hardware and Software Requirements

- GPU Optimized Server. Will be used for developing deep learning models, storing large amounts of data and as a collaborative environment.

- UNIX Based Environment(s). Given the design and integration of de-facto programming languages, frameworks, tools with Unix based operating systems, along with the flexibility and low-level tools it has to offer, we concluded that UNIX based operating systems will be most suitable for the development process.

Project Tasks, Time Plan and Deliverables

Task	Start & Due Dates	Deliverable	Evaluation Criteria	Objective
Project Proposal	09/10/2020 23/10/2020	Proposal with clear goals.	Readable, clear	Expressing our intent for the research project and our tasks for this semester.
Development Environment Preparation	19/10/2020 10/11/2020	A working environment	Installment of necessary tools, programming languages etc.	Embracing the concepts of the operating system, programming language, frameworks and best practices to be used during the development stages
Literature Review and data collection	19/10/2020 30/11/2020	A collection of related papers and a training dataset for QG.	Moderate size dataset and collection of recent publications	Learning the foundations of the learning-based methods, developing insight about the concepts. Learning state-of-the art in this research area. Collecting the first Turkish QG dataset.
Question Generation System Prototype	01/11/2020 01/01/2021	Prototype System	A working prototype with limited capabilities	Generating questions of different forms from a given text.
Final Report	01/01/2021 08/01/2021	Final Report	Complete, readable, formatted	Covering the work done in detail.

Project Team and Authority Information

Proposal Date	08/03/2020
Academic Term of Project Delivery	2020-2021, Spring semester
Project Team Members	Alp Gokcek, #041701014, Computer Engineering (Major)
	Erdal Sidal Dogan, #041701076, Computer Engineering (Major)
Advisor(s)	Seniz Demir