automatic quiz generation for educational knowledge testing

# Initial document

## Swansea University

### Auguste Tomaseviciute

2020 October

Table of Contents

[Initial document 1](#_Toc54277947)

[Swansea University 1](#_Toc54277948)

[Auguste Tomaseviciute 1](#_Toc54277949)

[Abstract 2](#_Toc54277950)

[What’s the motivation behind the project? 2](#_Toc54277951)

[Aims 3](#_Toc54277952)

[Objectives 4](#_Toc54277953)

[Deliverables 5](#_Toc54277954)

[What related work has been already done? 6](#_Toc54277955)

[Work methodology 7](#_Toc54277956)

[Chosen Waterfall methodology 7](#_Toc54277957)

[Chosen Agile SCRUM model 7](#_Toc54277958)

[Advantages of SCRUM 7](#_Toc54277959)

[Disadvantages of Scrum 7](#_Toc54277960)

[Alternative methodologies 7](#_Toc54277961)

[Requirement analysis 9](#_Toc54277962)

[Timeline 9](#_Toc54277963)

[What are the potential risks? 9](#_Toc54277964)

#### Abstract

#### What’s the motivation behind the project?

Since 19th century there has been rise in digital technology. While the life of an everyday human has drastically changed by digitalization, the ways of education and educational assessments has been the same since the 220 AD [Han dynasty standardized testing example]. Naturally, the means of delivering education has seen some changes – organizational and delivery means has moved online but the most of the worlds’ secondary schools still rely on the same structure. The structure where a teacher explains topic and later all students take some form of the same test to assess their knowledge about it. In the era where advertisement is curated to one’s taste, there is expectation for one-fits-all school model to educate future generation.

The road of personalizing school system is long one but the one of the crucial parts that could pave the way to is personalizing assessments. One the most popular means of assessment in secondary schools is multiple choice quiz (MCQ) [ref]. MCQ compromises of either text with a blank open space that is filled in by one of the answers or simple question with a suitable and multiple unsuitable answer.

Studies show [ref] that frequent testing that helps with long-term knowledge retrieval. But with limited assessment options from the educational sources (such as textbooks) and students’ tendency to share knowledge before and after tests, creating a suitable number of tests for frequent enough testing would prove to be impossible. Automatic quiz generation would help teachers by eliminating manual process of creating assessments and would allow focus of teaching part.

This paper will aim to evaluate methods of creating automatic multiple-choice quiz generator, understand the tools, and suggest on the best platform to deliver the quizzes.

#### Aims

The overall aim is to provide schools with tools to deliver personalized learning. Where students can follow the general outline for their studies but on their own pace. Where teachers act as mentors and educational content is provided not by the teacher in the typical classroom environment but rather students with the guidance are available to learn from everything that is available around them. In this scenario students are able to learn from real life examples, such as if one is learning about a combustion engine, he or she can attend a day at mechanics, where the information can be explained using real life examples. Or if the student is obliged to learn about derivatives, he or she can use videos (such as Khan Academy) to learn that. Teachers here would act guides, pointing students in the right direction and in scenarios where content is not widely available, creating it and making it accessible to everyone.

Such schools would be guiding students and allowing them to learn and explore and be curious about the surroundings. Iterative testing and mentoring meetings would not only allow to follow students’ progress but also distinguish one’s talents and allow them to excel on their own pace.

#### Objectives

Given the time, resource and political constraint the overall aim is outside the scope for this project. The objective of this project is to focus on only one of the challenges in this goal – assessment. More precisely, using technology to free teachers’ time and automating quiz creation for classroom use.

* Research and evaluate methods of creating automated quiz generator

*The main objective is to research what it takes to create automated quiz generator. Have solid guidelines and understanding what steps must be taken to implement one and what methods should be used.*

Apart the main objective of the paper, these are the personal and professional objectives that are in place for this project:

* Gain additional knowledge in programming in *Python* language.

*As per**IEEE Spectrum’s sixth annual interactive ranking of the top programming languages, Python is the most popular programming language. With its vast amount of specialized Machine Learning, it makes a great choice for this project.*

* Learn about Machine Learning and applying it Natural Language domain.

*As this project relies heavily on Natural Language processing and steps such as transforming statement sentences into questions involve Machine Learning, gaining extensive knowledge in both fields in a key step.*

* Understand work methodologies that can be used to do research.

*This objective was already achieved as per submission of the initial document.*

* Become proficient in utilizing productivity platform *ClickUp* for research workload management and following chosen work methodology.

*This multi-platform software allows to schedule tasks, visualize them in Grantt chart that will be beneficial for completing this project.*

#### Deliverables

The achievable deliverables for this project are to explore different ways of creating an automated quiz generator system. Along the way, exploring the what has been invented up to date and figuring out the most accurate way of creating one.

1. Find a suitable dataset for testing.

*As previously mentioned, the aim of the project is evaluating different methods of creating automated quiz generator. Finding a suitable data for testing is a crucial step in evaluating and comparing different combinations of such system.*

1. Evaluate different methods of ranking sentences in a given text.

*As not all of the sentences are informative enough to be used in quizzes. One of the key steps of this project is to be able to select the most important sentence in a text.*

1. Evaluate methods of ranking keywords in text to pick suitable answer.

*In addition to ranking sentences, keyword selection is also very important. Choosing the most essential word in a text will allow for the correct focus of the questions.*

1. Evaluate methods of transforming statement sentences into quiz questions.

*This step of the project will aim to evaluate different methods of quiz questions such as transforming sentence into ‘fill in the blank’, ‘true or false’ or full questions.*

1. Evaluate methods of picking incorrect but similar answers.

*The last step of the project is to evaluate different methods of generating similar domain but incorrect answers.*

#### What related work has been already done?

Automatic quiz generator is not a new concept. There are some commercially available products already. ‘Questo AI’ was a mobile application that was aiming to help students with revision. They were backed by IBM and Google Cloud which proves there is demand for such systems. Unfortunately, they stopped developing their app in 2019.

Another commercially available product is Quillionz which is created by Harbinger AI group. There software is great, it provides multiple functions such as automatic note taking, quiz generation. It has free – limited use of the application and premium options.

Upon online research, Deep Quiz was discovered which looks like an application hosted on AWS that upon given input create True or False questions. While the goal of the application is similar to this paper, the applications fails to recognize different sentences and does not provide with sufficient questions.

#### Work methodology

###### Chosen Waterfall methodology

###### Chosen Agile SCRUM model

For the purpose of this developing the project, Agile SCRUM methodology was chosen. Originally this method is for small teams but during this project an adapted version will be used.

The work is organized in short bursts called Sprints and in longer Program Increment PI that dictate the overall goal for the sprints.

Normally before each of the Program Increment there is a PI planning session. During this session the main goal for the PI is declared and dependencies between all of the teams are marked. Due to only one person working on this project, there will be no PI Planning and the goals will be assigned before the initial document release.

During this project sprints will be one weeklong with one main goal set for it. Each of these goals is divided into stories (tasks) that should not take more than a day to complete.

As listed down in a SCRUM guide these four events will be used:

* Sprint Planning (picking a goal from the backlog)
* Daily Scrum (review of the previous day, and sorting out tasks for the day)
* Sprint Review (review what was completed, what tasks need to be moved to another sprint)
* Sprint Retrospective (understand what when well, what needs to be improved)

This framework was picked for one main reasons - it is adaptive. As the scope of this project is hard to accurately to account for, iterative deployment and adapting the technique of working allows to continuously reconsider the scope and adjust the work around it.

###### Advantages of SCRUM

###### Disadvantages of Scrum

###### Alternative methodologies

Waterfall

#### Requirement analysis

#### Timeline

During the initial research is has been identified that process of creating Automated Quiz Generator AQG can be split into four main components:

1. Implementing sentence ranking algorithms.
2. Adding keyword ranking methods.
3. Turning statement sentences into question-like.
4. Generating additional similar domain answers.

Taking into consideration these four main research points. The Gantt-chart below visualizes these stages on a timeline, including any other project related dates.

As mentioned above the work week will be divided in 2 weeks sprints. Given that each this project aims to evaluate at least three different algorithms for each of these steps,

#### What are the potential risks?

As with any potential project there are numerous risks that could happen in the duration of this project. To help prevent and mitigate potential risk a Risk Value is calculated:

***Risk Value = Probability of Event \* Impact of the event***

Probability of the event is assigned a value between 0 and 1. 0 being ***none***and 1 being ***high probability****.*

Impact of the event is assigned a value in range (0 – 10). 0 being ***none*** and 10 being ***high impact****.*

**(Risk Value 0.5 \* 4 = 2.0)**

**Hardware failure -** one the most likely risks as all of the work is carried out on a single personal laptop. While the device is fairly new and it should not experience failure due old age, there is potential for this event to happened in case of accidental mishandling. An impact of such event is high, but with the prevention method listed below it can be significantly lowered. The impact would slow down work as another PC would have to be used.

**Prevention Method –** All of the work will be backed up after every single workday using GitHub. Here implementation and document files will be placed in a private repository.

**Mitigation plan –** in occurrence of such event, technical equipment could be borrowed from a family member or work could continue using equipment provided by university.

**(Risk Value 0.7 \* 8 = 5.6)**

**Unexpected illness or forced time-off -** this includes both physical and mental illness, death or illness of a close family member, life situation complications due to Covid-19. Due to the outgoing situation in the world the likelihood of such event is higher than normally. Impact of time off is quite high work would stop completely.

**Prevention method –** there is no prevention method for such event.

**Mitigation plan –** There will be scheduled week off after 5 sprints that can be used to catch up on any unexpected delays. In additional holiday time is also scheduled that can be used to catch up on.

**(Risk Value 0.4 \* 3 = 1.2)**

**Underestimation of workload –** due to the fact that topics surrounding Machine Learning and Natural Language processing are new to the author, there is a high risk of underestimating how long will it take to implement or evaluate one algorithm. Impact of such event is moderate as part of the project goal is to explore possibilities and failing at evaluating one is a tolerable.

**Prevention method –** Sprint planning allows to gain understanding about the scope of the task.

**Mitigation plan –** Extra time is added after each PI to account for such occurrence.

**(Risk Value 0.1 \* 3 = 1.2)**

**Development of algorithms that are not suitable for the purpose –** due to the complexity of the project is it possible that some of the algorithms that are evaluated and implemented will not be suitable for the project.

**Prevention method –** initial research over summer.

**Mitigation plan –** Extra time is added after each PI to account for such occurrence.