Sri Lanka Institute of Information Technology

PROJECT REGISTRATION FORM

(This form should be completed and submitted on or before 3.00 PM, Friday 3rd March, 2017)

The purpose of this form is to allow final year students of the B.Sc. (Hon) degree program to enlist in the final year project group. Enlisting in a project entails specifying the project title and the details of four members in the group, the internal supervisor (compulsory), external supervisor (may be from the industry) and indicating a brief description of the project. The description of the project entered on this form will not be considered as the formal project proposal. It should however indicate the scope of the project and provide the main potential outcome.

PROJECT TITLE	Adaptive Artificial Intelligent Q&A Platform
RESEARCH GROUP	Machine Learning
PROJECT NUMBER	(will be assigned by the lecture in charge)

PROJECT GROUP MEMBER DETAILS: (Please start with group leader's details)

	STUDENT NAME	STUDENT NO.	CONTACT NO.	EMAIL ADDRESS
1	Akram M.R	IT14109386	0774605347	rifhan.akram1@gmail.com
2	Deleepa Perera	IT14103100	0779544211	deleepa.p@gmail.com
3	Saad M.S.M	IT14109072	0779303127	saad.sahibjan@gmail.com
4	Singhabahu C.P	IT14126802	0771770077	mails.chamara@gmail.com

SUPERVISOR

Mr. Yashas Mallawarachchi		
Name	Signature	Date

CO-SUPERVISOR (will be assigned by the Supervisor, if necessary)

Mr.Anupiya Nugaliyadda		
Name	Signature	Date

EXTERNAL SUPERVISOR (if any, may be from the industry)

Name	Affiliation	Contact Address	Contact Numbers	Signature/Date

ACCEPTANCE BY CDAP MEMBER

Name	Signature	Date

PROJECT DETAILS

Brief Description of your Research Problem:

The rate of information growth in the world has increased exponentially over the past decade. With the increase in availability of digital input and output devices people are creating and consuming more digital information than ever before. In any given field there is a vast amount of information and knowledge that can be put to use to solve problems. For example, when there is a requirement for some knowledge a human being will take a long period of time to process all the available knowledge to him/her and to come up with the most optimal solution. Assuming that a given human being is rational, that person will be able to make the most optimal decision if they have all the required information. But due to various practical issues this is not possible. A computer that is connected to a vast network of other computers over a fast network will be able to process much more information than a human is capable of. If this type of tool is made available to people via an easy to use interface (such as over a website or mobile application) it will make people more efficient and enable them to make more rational decisions.

However due to the large amount of knowledge it is hard for a single algorithm to be able to understand the content. Deep learning has the capability of learning this information and provide only the required information on demand.

Another problem is that although the correct data may exist in the corpus the information must be formatted into answers after understanding the question correctly. Providing the most relevant output or answer for the understood content is a major problem faced at present because the information growth rate is extremely high.

Description of the Solution:

The system that we have proposed is an Adaptive Artificial Intelligent Q&A Platform. The idea behind the project is to build a platform that is able to intelligently learn from a given corpus and be able to answer logical questions. Questions and answers were chosen as the best way to interface with the data and it also gives a straightforward response to the questioner. The system must be able to understand the context of the question being asked and must be able to find the relevant information in the corpus that is available. This requires the use of Deep Learning and NLP and NLU techniques.

One important area is question understanding. Since the user will be asking questions from the system, it must be able to identify what the user is asking. We have to use the Stanford Parser, In order to understand the context of the question we will have to carry out semantic analysis as well. Semantic analysis will help us understand the meaning of certain phrases in the question

sentence. We will also have to identify the subject of the question. This will involve using techniques such as POS (parts of speech) tagging and phrase analysis. One approach that has been suggested for semantic analysis is to use a lexicon and word nets. However more efficient approaches have been suggested by using neural networks (such as recurrent neural nets) and distributional sentence models for text analysis and to perform tasks such as sentiment analysis and paraphrase detection.
The optimal neural network will be chosen after experimenting with a set of candidate neural network techniques.
For the purpose of this research we will be narrowing down the scope of the domain the Q&A platform will address to and attempt to produce logical answers for questions asked in a narrow spectrum that is decided based on the availability of a corpus.

Main expected outcomes of the project:

The main outcome of the project will be to cater patients in a specific domain in the medical field such as (e.g. – Medical Emergency, Cancers) (subjected to the availability of a corpus).
Note - We have formally requested NHS for their Emergency Care Dataset and currently waiting for their reply.

What will be the inputs and expected?
Input: Question related to the medical domain
Output: Answer for the question.

The output will not be direct diagnosis rather suggestions and further information for the patient that will enable him/her to further understand the condition. A sample use case would be, a patient suffering from some emergency situation can simply ask a question from the system.

E.g. -
Q: I am having a swollen wrist, what can be the causes?
A: Possible Causes for your swollen wrist could be,
Muscle strain, Tissue Damage or a Fractured Wrist.

WORKLOAD ALLOCATION (Please provide a brief description about the workload allocation)

MEMBER 1	*Akram M. R.*

Mapping of an understood question to an answer will be done using a deep neural network. In order to understand map the question to an answer, the system must be designed using a specific subset of activation functions. From our research so far we have identified that Recurrent Neural nets are the best optimized for analyzing text. Since our corpus will be composed of mostly natural text and some structured data in text form we will attempt to create a suitable network. According to our research, the Recurrent Net is more efficient than normal Feedforward Neural nets however this is subject to change with further research. Either way the most optimum network will be selected.

MEMBER 2	*Deleepa Perera*

Training the deep neural network will be another important task that must be carried out for the system to work properly. A common technique is to use backward propagation to balance the weights used in the tensors of the neural net. An issue that we will have to tackle with backward propagation is gradient descent. Another way we found to do this is to use a Restricted Boltzmann Machine. This will allow us to feed in an unstructured set of inputs and allow the neural net to structure the inputs by itself. Then the fine tuning can be done by backward propagation.

When using the ontology approach what would normally happen is that the semantics of the question would be mapped to certain hotspots in the ontology graph. Since we are still researching the deep neural network approach, we are not sure how this would work.

MEMBER 3	*Saad M. S. M.*

The main objective of this component is to identify the semantics of a question in order to process an accurate answer. Commonly there are two possible approaches to build a semantic analyzer.

1. Rule Based
2. Machine Learning

The Machine Learning approach will be under supervised learning and the algorithm is highly depended on an annotated corpus. To achieve semantic analysis, we will need to identify parts of speech through the Stanford parser and perform syntactic analysis further steps need to take in order to achieve and identify semantics of a question.

MEMBER 4	*Singhabahu C.P.*

Once the data that is relevant for the answer sentence has been identified, it is important to match the semantics of the answer sentence to the question. When a user has asked a question such as:

Q: When was Sri Lanka Railways founded?

If the answer sentence that was located looks like the following:

A: The Ceylon Government Railway was founded in 1858 and was subsequently re-named to Sri Lanka Railways

We can see that the information we want is there, but it is not direct and there is some extra information as well. So this component will have to construct the correct answer and construct an answer that has the highest semantic similarity to the question asked. Sometime during this process the system might have to ask the user more questions to identify the semantics of the question correctly as well.

DECLARATION

"We declare that the project would involve material prepared by the Group members and that it would not fully or partially incorporate any material prepared by other persons for a fee or free of charge or that it would include material previously submitted by a candidate for a Degree or Diploma in any other University or Institute of Higher Learning and that, to the best of our knowledge and belief, it would not incorporate any material previously published or written by another person in relation to another project except with prior written approval from the supervisor and/or the coordinator of such project and that such unauthorized reproductions will construe offences punishable under the SLIIT Regulations.

We are aware, that if we are found guilty for the above mentioned offences or any project related plagiarism, the SLIIT has right to suspend the project at any time and or to suspend us from the examination and or from the Institution for minimum period of one year".

	STUDENT NAME	STUDENT NO.	SIGNATURE
1	Akram M.R.	IT14109386	
2	Deleepa Perera	IT14103100	
3	Saad M.S.M	IT14109072	
4	Singhabahu C.P	IT14126802	