

Thai Question Answering Program from Thai Wikipedia

Speech and Text Understanding Research Team (STU), NECTEC

What is a question answering system

- Question Answering System (QAS) is an information retrieval system that automatically generates an concise answer of a question posed by human in natural language



Objectives

- To build a first Thai QA dataset
- To encourage the QA corpus to be available as the standard corpus for research and development of QA algorithms
- To develop an algorithm to answer a question from Thai Wikipedia
 - Retrieving small snippet of text contained an answer
 - Finding an exact answer



Dataset

- Question-answer pairs created by linguists on a set of Wikipedia articles
 - An answer is a word, segment of text, or span appearing on a part of the corresponding reading passage
 - A question is a simple (factoid) questions [what, where, who, when, which, how many]
- Dataset is approximately 10,000 question-answer pairs
 - Simple dataset is 100 question-answer pairs
 - Development dataset is 4,000 question-answer pairs
 - Validation dataset is 1,000 question-answer pairs
 - Evaluation dataset is 5,000 question-answer pairs

Evaluation

- We use exact match (EM) and F1 metrics, computed on common substring of word level between the predicted answer and the gold answer

Gold standard#1

ឆ្នាំ

Gold standard#2

កីឡា	ឆ្នាំ
------	-------

Gold standard#3

ប្រភេទ	កីឡា	ឆ្នាំ
--------	------	-------

Prediction#1

ឆ្នាំ

1) EM = 1, Precision = 1/1 = 1, Recall = 1/1 = 1, F1 = (2*1*1)/(1+1) = 1

2) EM = 0, Precision = 1/1 = 1, Recall = 1/2 = 0.5, F1 = (2*1*0.5)/(1+0.5) = 0.66

3) EM = 0, Precision = 1/1 = 1, Recall = 1/3 = 0.33, F1 = (2*1*0.33)/(1+0.33) = 0.49

Prediction#2

ឆ្នាំ	ឆ្នាំ
-------	-------

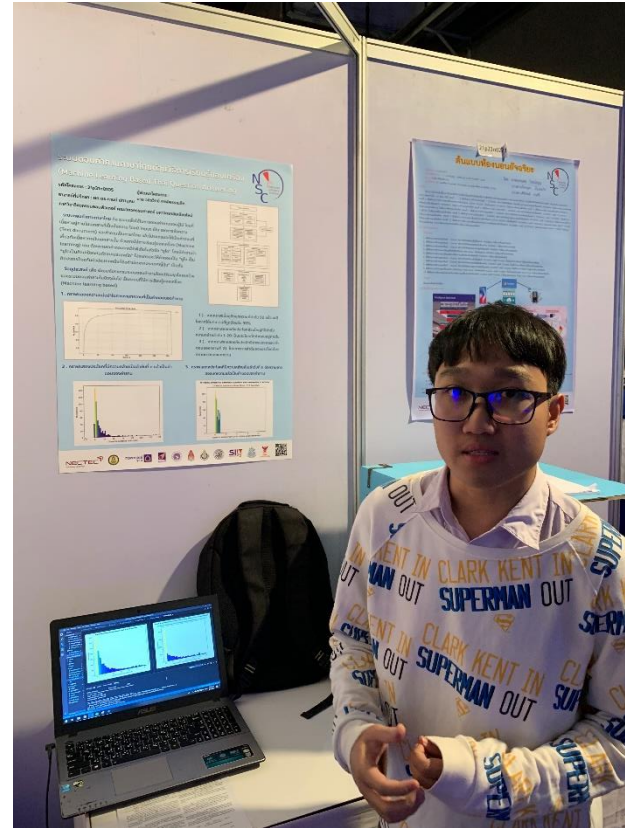
1) EM = 0, Precision = 1/2 = 0.5, Recall = 1/1 = 1, F1 = (2*0.5*1)/(0.5+1) = 0.66

2) EM = 0, Precision = 1/2 = 0.5, Recall = 1/2 = 0.5, F1 = (2*0.5*0.5)/(0.5+0.5) = 0.50

3) EM = 0, Precision = 1/2 = 0.5, Recall = 1/3 = 0.33, F1 = (2*0.5*0.33)/(0.5+0.33) = 0.39

Final round

5 teams through to the final round of NSC 2019



Evaluation Result (Second round)

Project ID	Model	Document Retriever (Accuracy)	Document Reader	
			EM	F1
21P31N0105	BiLSTM	36.4	5.0	7.39
21P31N0225	CNN + Rule-based	32.12	23.56	31.11
21P31W0001	BiLSTM	72.10	26.6	37.43
31P31I0095	BiLSTM + Full Aware Attention	46.50	8.20	13.26
21P31C0457	Pattern knowledge + Ontology	N/A	N/A	N/A

Evaluation Result (Final round)

Project ID	Model	Document Retriever (Accuracy)	Document Reader	
			EM	F1
21P31W0001	BiLSTM	83.50	34.80	45.96
21P31N0225	CNN + Rule-based	63.51	10.54	12.86
21P31N0105	BiLSTM	38.0	4	6.06
31P31I0095	BiLSTM + Full Aware Attention	53.38	9.62	15.38
21P31C0457	Pattern knowledge + Ontology	N/A	N/A	N/A