Week 3 Milestones and Progress Report

Timothy Laurent

Project description:
This project seeks to bring modern NLP analysis to a document reader interface to help facilitate reader comprehension. This interface could assist potentially assist individuals with learning disabilities, cognitive deficits such as ataxia, and readers of lower proficiency to understand more of what they read. The project as envisioned will focus on open information extraction and coreference resolution to provide triples of information to the user.

Progress
This week focussed on combining the output of AllenNLP's coreference model with the open information extraction OpenIE model. To do this, the entities from the coreference model were combined with the arguments in the open information extraction model. Since the OpenIE model operates on the sentence level, its output had to be translated from the BIO tagging format it uses to indexed spans adjusted for the sentence offset. Several techniques were attempted for merging overlapping entities, in the end entities that shared a starting offset were merged into a canonical entity otherwise a new entity was created. Now that there is a normalized representation of the entities and their relations, next steps will be to visualize that in a react web view.

Research
To test the efficacy of this merged model approach, two questions from the Standford Question and Answer Dataset SQUAD were selected and analyzed with the merged model, see appendix 1 for the questions. A string representation of the model output was created to allow for validation of the effectiveness of the method while developing the web UI. An example of model output is below, the first column is the strings of all predicted entity coreferences spans, then the subjects entity index, the specific subject mention, the verb,specific, object mention, Object entity mentions, and modifier. It is immediately obvious that there are errors in the coreference resolution model, but the basic structure of the data looks helpful to understanding. Perhaps the coreference issue could be helped by some distant supervision by matching queries to wikipedia or another knowledgebase.

A version of this extracted information will be given to survey participants to determine if it can facilitate comprehension.

Subject	S index	s mention	Verb	o mention	Object	Modifier

themselves Many species	0.0	Many species	utilize	mechanisms that appear to be precursors of these aspects of vertebrate immunity	mechanisms that appear to be precursors of these aspects of vertebrate immunity mechanisms	nan
the first vertebrates Prokaryotes RNA CRISPR sequences them to block virus replication through a form of RNA interference phage they CRISPR them	1.0	Prokaryotes	possess	acquired immunity	acquired immunity acquired	nan
	1.0	Prokaryotes	possess acquired	immunity	immunity	nan
	1.0	CRISPR sequences	to retain	fragments of the genomes of phage	fragments of the genomes of phage	nan
	1.0	they	have come	into contact with in the past	into contact with in the past	nan
	1.0	them	to block	virus replication	virus replication	nan
It	2.0	It	is	likely that a multicomponent , adaptive immune system arose with the first vertebrates	likely that a multicomponent , adaptive immune system arose with the first vertebrates	nan

a multicomponent , adaptive immune system	5.0	a multicomponent , adaptive immune system	arose	with the first vertebrates	with the first vertebrates	nan
invertebrates	8.0	invertebrates	do	not generate lymphocytes or an antibody - based humoral response	not generate lymphocytes or an antibody - based humoral response	nan
	8.0	invertebrates	do not generate	lymphocytes or an antibody - based humoral response	lymphocytes or an antibody - based humoral response	nan
an antibody	15.0	an antibody	based	humoral response	humoral response	nan
mechanisms that appear to be precursors of these aspects of vertebrate immunity mechanisms	19.0	mechanisms	appear	to be precursors of these aspects of vertebrate immunity	to to be precursors of these aspects of vertebrate immunity	nan
	19.0	mechanisms	to be	precursors of these aspects of vertebrate immunity	precursors of these aspects of vertebrate immunity	nan
Immune systems	27.0	Immune systems	appear	even in the structurally most simple forms of life	even in the structurally most simple forms of life	nan
bacteria	30.0	bacteria	using	a unique defense mechanism , called the restriction modification system	a unique defense mechanism , called the restriction modification system	nan

					to protect
	30.0	bacteria	called	the restriction modification system	the restriction modification system	to protect themselves from viral pathogens
the restriction modification system	35.0	the restriction modification system	to protect	themselves	themselves Many species	from viral pathogens
viral pathogens	41.0	viral pathogens	called	bacteriophages	bacteriophages	nan
a system	51.0	a system	uses	CRISPR sequences	the first vertebrates Prokaryotes RNA CRISPR sequences them to block virus replication through a form of RNA interference phage they CRISPR them	to retain fragments of the genomes of phage
the past	66.0	the past	allows	them to block virus replication through a form of RNA interference	the first vertebrates Prokaryotes RNA CRISPR sequences them to block virus replication through a form of RNA interference phage they CRISPR them	nan
Offensive elements of the immune systems	73.0	Offensive elements of the immune systems	are also	present in unicellular eukaryotes	present in unicellular eukaryotes	nan

studies of their roles in defense	77.0	studies of their roles in defense	are	few	few	nan

Next steps
- A React viewer that shows the model output will be created hopefully allowing hoverover highlighting of the main text when mousing near the model output and vise versa.
- Create final view of the model output for the comprehension survey and deliver it.

Appendix 1 SQUAD QUESTIONS

Question 1.

It is likely that a multicomponent, adaptive immune system arose with the first vertebrates, as invertebrates do not generate lymphocytes or an antibody-based humoral response. Many species, however, utilize mechanisms that appear to be precursors of these aspects of vertebrate immunity. Immune systems appear even in the structurally most simple forms of life, with bacteria using a unique defense mechanism, called the restriction modification system to protect themselves from viral pathogens, called bacteriophages. Prokaryotes also possess acquired immunity, through a system that uses CRISPR sequences to retain fragments of the genomes of phage that they have come into contact with in the past, which allows them to block virus replication through a form of RNA interference. Offensive elements of the immune systems are also present in unicellular eukaryotes, but studies of their roles in defense are few.

Invertebrates do not generate what type of cells that are a part of the vertebrate adaptive immune system?
- *Ground Truth Answers:* lymphocytes lymphocytes lymphocytes

What is the main defense mechanism of bacteria known as?
- *Ground Truth Answers:* the restriction modification system restriction modification system restriction modification system

The restriction modification system is used by bacteria for protection from what pathogens?
- *Ground Truth Answers:* bacteriophages viral bacteriophages

What is the system by which prokaryotes retain phage gene fragments that they have previously come in contact with?
- *Ground Truth Answers:* CRISPR CRISPR sequences CRISPR

Where is it unlikely that the first multicomponent, adaptive immune system arose?

- *Ground Truth Answers:* `<No Answer>`

What arose in the latest vertebrates?

- *Ground Truth Answers:* `<No Answer>`

What mechanisms do many species not utilize?

- *Ground Truth Answers:* `<No Answer>`

What kind of immunity do prokaryotes not have?

- *Ground Truth Answers:* `<No Answer>`

What has often been studied in unicellular eukaryotes?

- *Ground Truth Answers:* `<No Answer>`

Question 2.

When a T-cell encounters a foreign pathogen, it extends a vitamin D receptor. This is essentially a signaling device that allows the T-cell to bind to the active form of vitamin D, the steroid hormone calcitriol. T-cells have a symbiotic relationship with vitamin D. Not only does the T-cell extend a vitamin D receptor, in essence asking to bind to the steroid hormone version of vitamin D, calcitriol, but the T-cell expresses the gene CYP27B1, which is the gene responsible for converting the pre-hormone version of vitamin D, calcidiol into the steroid hormone version, calcitriol. Only after binding to calcitriol can T-cells perform their intended function. Other immune system cells that are known to express CYP27B1 and thus activate vitamin D calcidiol, are dendritic cells, keratinocytes and macrophages.

What does a T cell extend when it encounters a foreign pathogen?

- *Ground Truth Answers:* a vitamin D receptorextends a vitamin D receptorvitamin D receptor

What is the active form of vitamin D known as?

- *Ground Truth Answers:* calcitriol calcitriol steroid hormone calcitriol

What is the nature of the relationship between T-cells and vitamin D?

- *Ground Truth Answers:* symbiotic relationship symbiotic symbiotic relationship

What gene is responsible for converting calcidiol into calcitriol?

- *Ground Truth Answers:* gene CYP27B1 CYP27B1 gene CYP27B1

Other than T cells, what other immune cells express CYP27B1?

- *Ground Truth Answers:* dendritic cells, keratinocytes and macrophages dendritic cells, keratinocytes and macrophages dendritic cells

What does a pathogen do when it encounters a T-cell?

- *Ground Truth Answers:* <No Answer>

What cell binds to vitamin C?

- *Ground Truth Answers:* <No Answer>

What vitamin do T-cells have a parasitic relationship with?

- *Ground Truth Answers:* <No Answer>

What gene is expressed by the B-cell?

- *Ground Truth Answers:* <No Answer>

What gene converts calcitriol into calcidiol?

- *Ground Truth Answers:* <No Answer>

Additional Questions:

On an 1-5 scale from strongly disagree to strongly agree, rate your agreement with these statements.

I feel confident about my answers.

I enjoyed this quiz.