MediReport

FA19 CMPE-272-01 Group 14 Project

Why do we need MediReport?

- **Ever been confused and perplexed by the complicated medical jargon you've seen on your medical reports?**

- **Have you spent ages trying to Google and figure out what your medical results mean?**

- **Have NO FEAR, cause we at MediReport, have THE CURE!**

What MediReport does.

- **Our application enables end-users who do not have medical backgrounds, to interpret the meanings of various medical/clinical terms that are present in their medical reports.**

- **The application provides the users with personalized results about their medical conditions, which can be translated into multiple languages.**

Vision behind MediReport

- Presents health care system relies on the assumption that patients can understand complex written information presented to them in their reports.

- If people do not understand health information, they are less likely to take necessary actions for their health or make appropriate health decisions.

- With MediReport, we aim to **bring down the heavy reliance of people on their health care practitioners** for common health problems leading to **reduced medical bills** and, to **reduce the time involved in taking appropriate health decisions**.

Flow of control
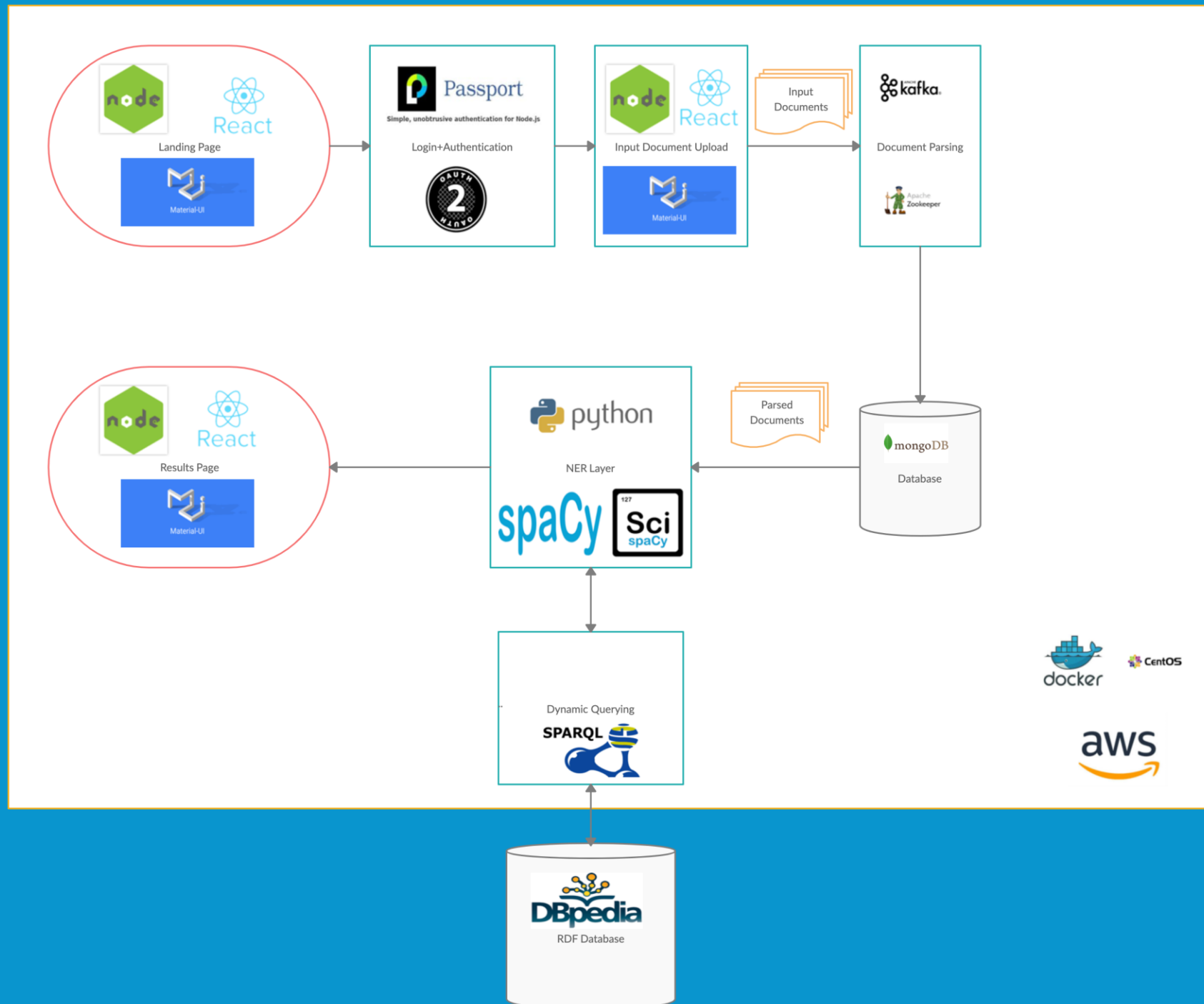
Technology Stack
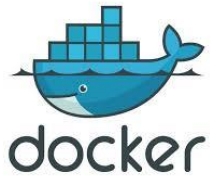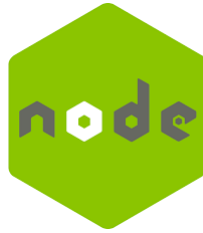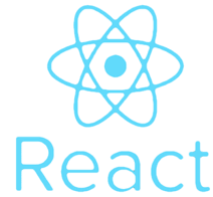
Frontend

- **React** has been used to create a **Single Page Web application**.

- **Passport.js**, an authentication middleware for Node.js, has been to generate JSON Web Tokens (**JWT**) for user authentication.

Backend – Node.js

- **The backend is built using Node.js with Express framework**

- **node.bcrypt.js library has been used to encrypt passwords and stored in the Mongo DB.**

- **The backend communicates with Kafka to achieve scalability by queuing API requests.**

Backend – Python

- **The Python backend uses Klien server for REST API handling and, uses spaCy and ScispaCy's models for processing biomedical, scientific or clinical text to perform NER.**

- **The named entities are searched in DBpedia using dynamically generated SPARQL and the result is returned to the UI.**

Personalization – Multiple languages supported

- English
- Arabic/ٱلْعَرَبِيَّة
- Spanish/Español
- French/Le français
- German/Deutsch
- Chinese/官话
- Italian/Italiano
- Japanese/日本語
- Dutch/Nederlands
- Polish/Polski
- Portuguese/Português
- Russian/русский язык

Clinical term

Term -
cholesterol

Explanation -
الكُولِسترول هو مادة دهنية شمعية أساسية في تكوين أغشية الخلايا في جميع أنسجة الكائنات الحية. بالإضافة إلى ذلك يلعب الكولسترول دورا أساسياً في الاستقلاب الحيوي (التمثيل الغذائي). من الكوليسترول نوعان أحدهما طيب ويجب أن تكون نسبته في الدم أعلى (HDL) مفيد والأخر ضار للصحة. النوع المفيد وهو بروتين دهني مرتفع الكثافة أو من 40 مليجرام/ديسيلتر. والنوع الضار يجب أن تكون نسبته في الدم أقل من 100 مليجرام/ديسيلتر، وهذا يسمى (LDL). بروتين دهني منخفض الكثافة أو

Potential diagnosis -
cholesterol imbalance can lead to the following conditions -

فرط كوليسترول الدم

Two other personalization

- Used data from **Institute for Health Metrics and Evaluation (IHME)**, an independent population health research centre at University of Washington, Medicine to provide age and gender-based disease recommendation

Write With Transformer

Get a modern neural network to auto-complete your thoughts.

- Used GPT-2, a large transformer-based language model trained on a dataset of 8 million web pages, to provide auto-complete functionality.

MediReport's personalized recommendation   ✕

Hi v@v.com! For a male in the age group 15-49, below are the most common diseases. **<NLG>** We want you to be healthy. **</NLG>** Please close this box to see the food items and complications associated with your medical test.

Cardiovascular diseases
Self-harm and interpersonal violence
Transport injuries
Neoplasms
Unintentional injuries
HIV/AIDS and sexually transmitted infections
Respiratory infections and tuberculosis
Digestive diseases
Diabetes and kidney diseases
Substance use disorders
Enteric infections
Chronic respiratory diseases
Other infectious diseases

CLICK TO PERSONALISE MORE!

Cloud deployment

- **The application is divided into multiple micro services:**
  - **Frontend**
  - **Backend**
  - **Kafka Backend**
  - **NLP Python backend**
  - **Kafka server**
- **Each of these micro services has been containerized using Docker and is running on AWS EC2 instances.**
- **There are multiple clones of the docker instance running the application, which are prefaced with a load balancer to provide redundancy in case of excessive traffic loads.**

Security vulnerabilities testing

- The OWASP Zap (Open Web Application Security Project Zed Attack Proxy) security tool has been used to check for security vulnerabilities of the web application.

- It allows web application security testers to perform fuzzing, scripting, *spidering,* and *proxying* in order to attack web apps.

# ZAP Scan report

ZAP Scanning Report

Summary of Alerts

Risk Level	Number of Alerts
High	0
Medium	2
Low	5
Informational	0

Alert Detail

Medium (Medium)	X-Frame-Options Header Not Set
Description	X-Frame-Options header is not included in the HTTP response to protect against 'ClickJacking' attacks.
URL	http://localhost:3000/sockjs-node/537/iyiayyc2/htmlfile?c=_jp.akmcfzf
Method	GET
Parameter	X-Frame-Options
URL	http://localhost:3000/sockjs-node/855/xcizabsz/htmlfile?c=_jp.aif1ix2
Method	GET
Parameter	X-Frame-Options
URL	http://localhost:3000/sockjs-node/791/c3505eia/htmlfile?c=_jp.absib3b
Method	GET
Parameter	X-Frame-Options
URL	http://localhost:3000/home
Method	GET
Parameter	X-Frame-Options
URL	http://localhost:3000/chart
Method	GET
Parameter	X-Frame-Options
URL	http://localhost:3000/create
Method	GET

Project submitted by Group 14

| Team member	Student ID
Vini Aswal	014554371
Sakshi Mahendru	014528618
Anjana Menon Cherubala	014428154
Animesh Swain	014509248
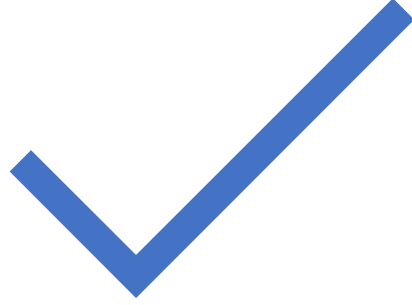
Transformer, based solely on attention mechanisms
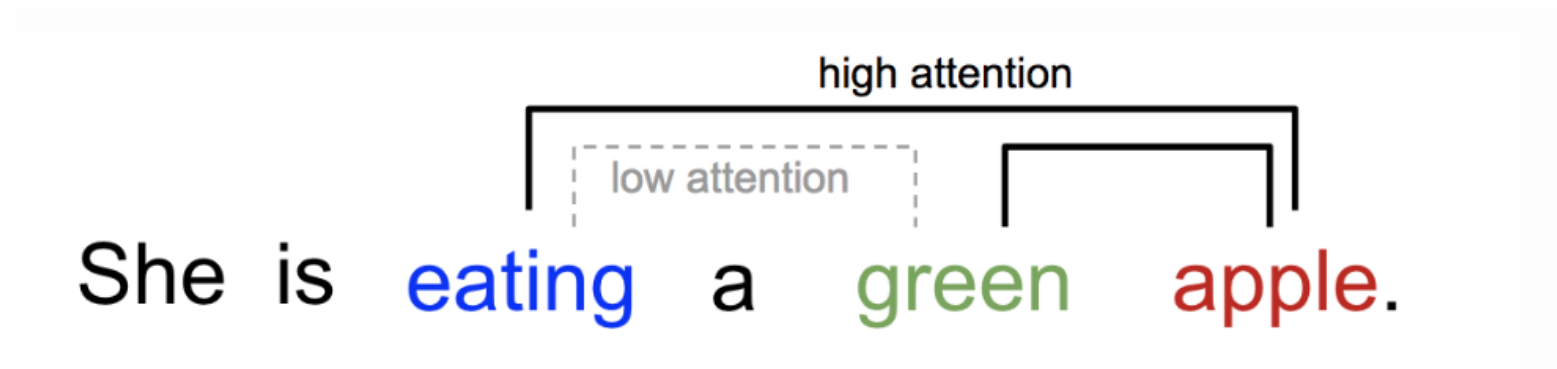
What's Wrong with Seq2Seq Model?

- An **encoder** processes the input sequence and compresses the information into a context vector (also known as sentence embedding or "thought" vector) of a *fixed length*. This representation is expected to be a good summary of the meaning of the *whole* source sequence.
- A **decoder** is initialized with the context vector to emit the transformed output.
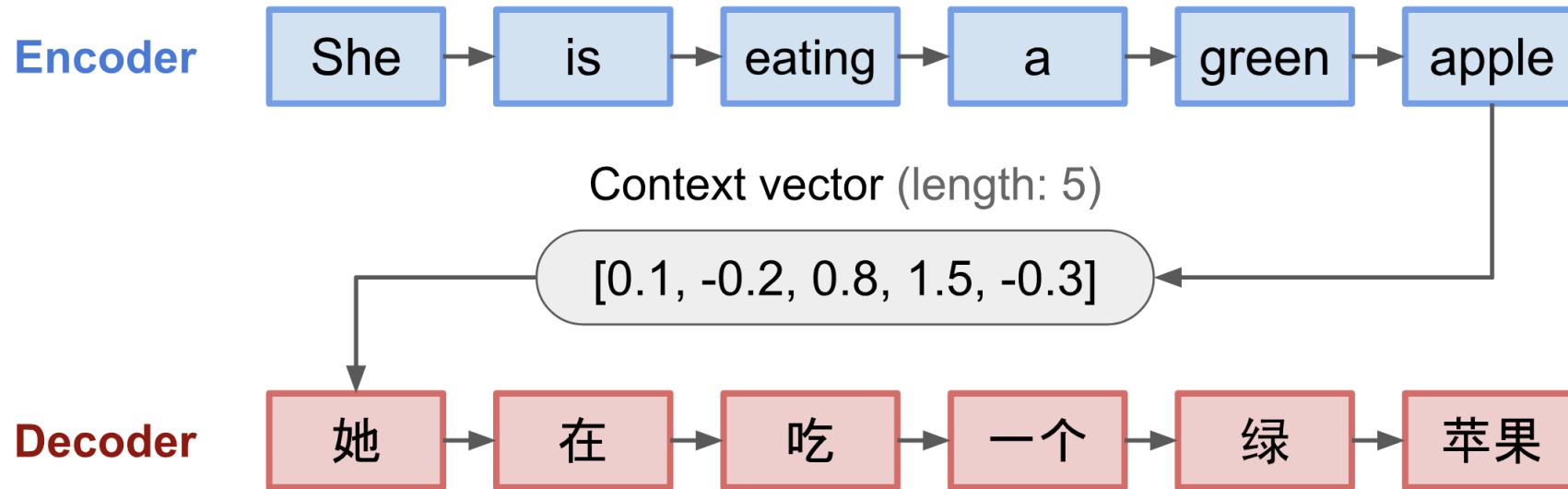
A critical and apparent disadvantage of this fixed-length context vector design is incapability of remembering long sentences.

Attention! How it works?

Rather than building a single context vector out of the encoder's last hidden state, the secret sauce invented by attention is to create shortcuts between the context vector and the entire source input.

(Target) y_{t-1} y_t

S_{t-1} S_t

Context vec

Global alignment weights

$a_{t,1}$ $a_{t,T}$

$a_{t,2}$ $a_{t,3}$

$\overrightarrow{h_1}$ $\overrightarrow{h_2}$ $\overrightarrow{h_3}$ $\overrightarrow{h_T}$

$\overleftarrow{h_1}$ $\overleftarrow{h_2}$ $\overleftarrow{h_3}$ $\overleftarrow{h_T}$

X_1 X_2 X_3 X_n (Source)

Decoder: RNN with input from previous state + dynamic context vector.

Attention layer: parameterized by a simple feed-forward network

Additive Attention

Encoder: bidirectional RNN