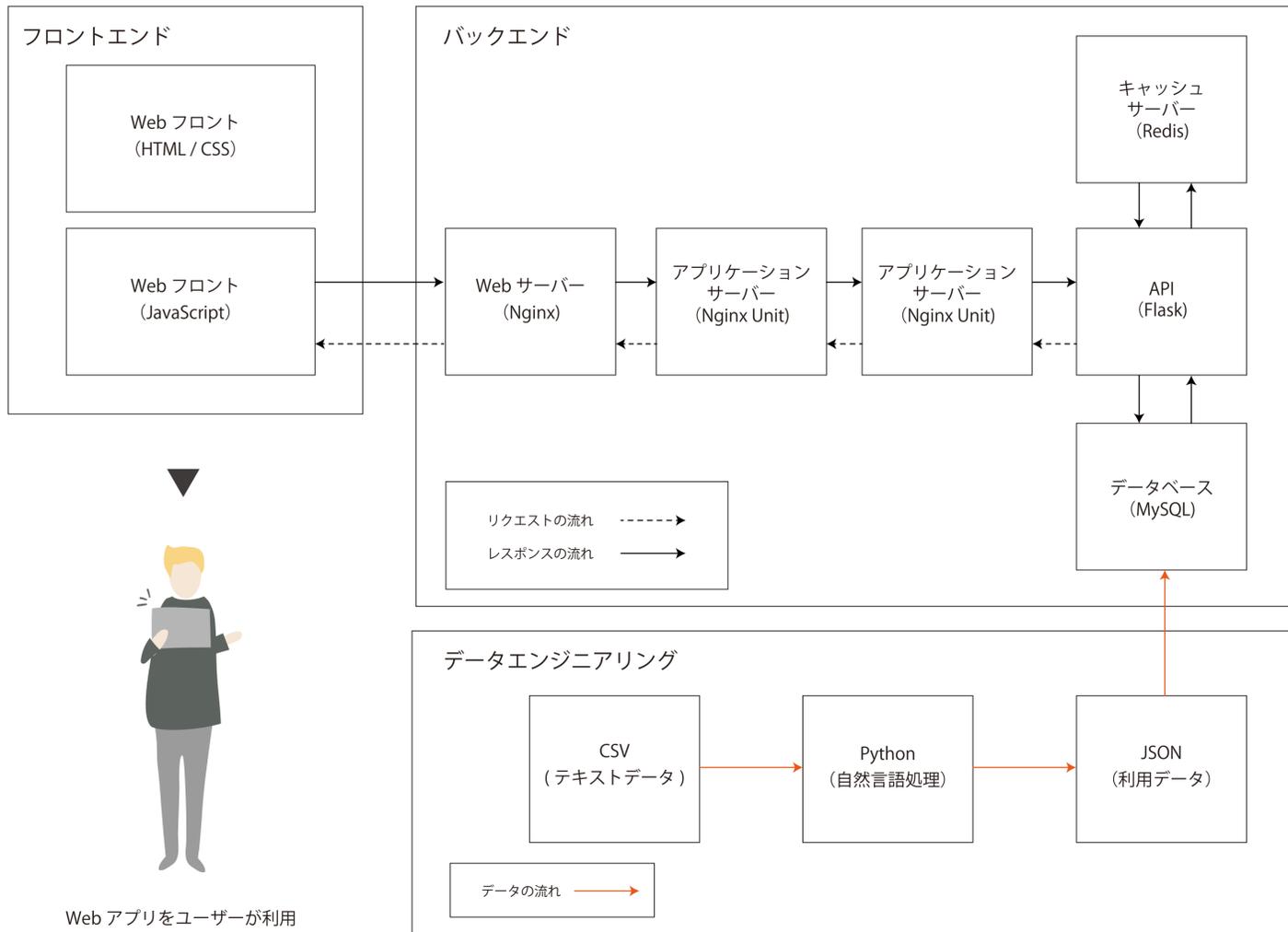


Web アプリのシステム構成図



ビッグデータの処理高速化を実現

データエンジニアリング班とバックエンド班が奮闘

アルゴリズムで計算量を最適化

データの書き出しをより早く

データエンジニアリング班では Python を用いて大規模なデータを分析する際の懸念点として、実行時間が長くなってしまったり、ノイズが多く含まれてしまったり、といったものがあります。そこで、自然言語処理を用いて、過去約 3 年分の新聞記事のテキストデータを整理・抽出することで、ノイズを減らしたり、適切な情報量に変換させてデータを扱いやすくする取り組みを行いました。具体的な処

高負荷に耐えるサーバーを実現

バックエンド班では、データエンジニアリング班が出力したデータをもとに API を作成しました。API 作成時の懸念点として、ユーザからリクエストからレスポンスまでの時間がユーザにストレスを与えるほどのものになることがありました。この問題の解決方法として、サーバ自体を良いものにする

というのももちろんありますが、「データ操作を DB 上で完結させる」「MySQL のエンジンをオンメモリ対応のものにする」といった手法を取りました。これらの手法によって、リクエストからレスポンスまでの時間を大幅に短縮することに成功し、ユーザビリティの向上に貢献しました。

漫画で追う時代の流れ

機械学習で潮流を紐解く試み

コンテンツ抽出班では、これまで見つけられていなかった新たな情報が新聞の中に存在しているのではないかと考え、大量にある紙面データを処理・活用することで画像的な観点からアプローチしました。紙面データにはテキストだけでなく、4コマ漫画や広告、レイアウトなど画像からしかわからない様々な情報が含まれています。考えており今後、実装予定です。



写真2:YOLOを用いた4コマ漫画検出結果



写真3:検出された画像の例

分かりやすい情報設計を追求

コンポーネントで整える

デザイン班では UI/UX データビジュアライゼーションにおけるデザインを担当しました。バックエンドからデータが渡される時の処理時間を考慮したローディング画面の設計、OOU を意識したアプリケーションの設計をアニメーションというアプローチ

```

1 CREATE TABLE article_data.word_count(
2   id INT AUTO_INCREMENT NOT NULL UNIQUE PRIMARY KEY,
3   published_date DATE NOT NULL,
4   word VARCHAR(128) NOT NULL,
5   word_count INT NOT NULL,
6   UNIQUE (word,published_date),
7 );
8
9 CREATE TABLE article_data.circle_parent(
10  id INT AUTO_INCREMENT NOT NULL UNIQUE PRIMARY KEY,
11  published_date DATE NOT NULL,
12  word VARCHAR(128) NOT NULL,
13  UNIQUE (published_date,word)
14 );
15
16 CREATE TABLE article_data.circle_child(
17  id INT AUTO_INCREMENT NOT NULL UNIQUE PRIMARY KEY,
18  parent_id INT NOT NULL,
19  word VARCHAR(128) NOT NULL,
20  UNIQUE (parent_id,word),
21  FOREIGN KEY (parent_id) REFERENCES article_data.circle_parent(id)
22 );
23
24 CREATE TABLE article_data.circle_grandchild(
25  id INT AUTO_INCREMENT NOT NULL UNIQUE PRIMARY KEY,
26  child_id INT NOT NULL,
27  word VARCHAR(128) NOT NULL,
28  UNIQUE (child_id,word),
29  FOREIGN KEY (child_id) REFERENCES article_data.circle_child(id)
30 );

```

写真1:DBのSQL構成

から実現しました。インタラクティブなシステムを実現するためにコンポーネントの概念を用いて情報を塊で配置しインタラクティブな見た目を担保しつつわかりやすい情報設計を目指しました。

HTML, CSS, JavaScript を用いて Web アプリケーションの実装を行いました。JavaScript では D3.js と jQuery を活用しインタラクティブなシステムの構築を実現しました。GitHub を用いてタスク管理を行いながら開発を進めました。タスク管理を行いながら各班と連携することで短い期間ながらもインタラクティブなシステムを効率的に開発を進めることができました。

コンポーネントを意識した設計

