# ポートフォリオ

[Mairu Matsuyama]





### クリエイターをお探しの企業様へ

この度はお忙しい中、選考にお時間をいただき誠にありがとうございます。 津山高専の電子・情報システム工学を専攻しておりました松山摩偉児と申します。 本スライドでは、私の経歴や実績について述べさせて頂いております。 なので、御社とクリエイターとして契約した場合にどんな経験が活かせるのか、 また、どんな貢献ができるのかを判断する材料にしていただければ幸いです。



#### **PROFILE**

## 1 プロフィール



電子・情報システム工学専攻

松山 摩偉児

## 経歴

2014年 3月 津山市立 久米中学校 卒業 2014年 4月 津山高専 情報工学科 入学 2019年 3月 津山高専 情報工学科 卒業 2019年 4月 津山高専 専攻科 入学 2021年 3月 津山高専 専攻科 卒業





#### **PROFILE**

## 1 プロフィール



電子・情報システム工学専攻 松山 摩偉児



### ・サイトリンク

GitHub: https://github.com/maturu

Twitter: <a href="https://twitter.com/maturu410">https://twitter.com/maturu410</a>

• 資格

基本情報技術者試験 合格

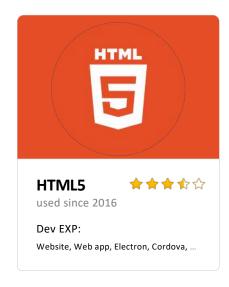
・ 受賞歴等アイチャレ2017 優秀賞 受賞

情報処理学会 第81回全国大会 出場

令和元年電気学会論文発表賞 受賞

#### **SKILL**

## 2 メインスキル – フロントエンド



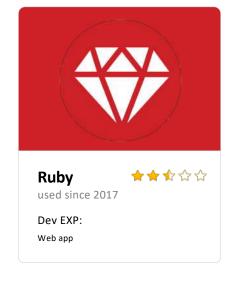


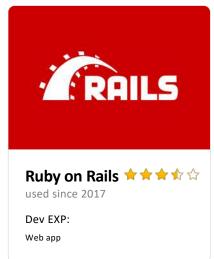


- HTML5
   経験年数 5年
   Webサイト制作、Webサービス開発、
   Electronによるアプリ開発などに利用
- CSS3 経験年数 5年 Bootstrap, Semantic UI など
- JavaScript経験年数 7年jQuery

#### **SKILL**

## 2 メインスキル – バックエンド







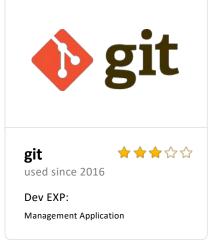
- Ruby 経験年数 4年 Railsで利用
- Ruby on Rails
   経験年数 4年
   インターンと公式チュートリアルを経て
   Webサービス開発のバックエンドに利用
- MySQL経験年数 4年WebサービスのDBとして利用

#### **SKILL**

## 2 スキル - その他







- Python
   経験年数 5年
   機械学習やスクレイピングに利用
   Chainer, Pytorch, OpenCVなど
- dockerRailsの環境構築
- gitプロジェクト管理、個人アプリの管理GitHub, Bitbucketなど
- その他 Node.js, PHP, C/C#, …

#### **WORKS**

## 3 プロジェクト - Webサイト制作

### 株式会社 吉信工務店

父が務めている株式会社 吉信工務店様から,ホームページ作成のご依頼を受け、HTML, CSS.

JavaScriptを使用し、Webサイトを制作しました。スマートフォンに対応するため、レスポンシブデザインで設計し、当時、SVGが流行していたため、パスアニメーションを加えて取り入れました。

2019年8月頃、先方と話し合いを行い、モックアップや試作を始め、2週間ほどで制作。

### 使用言語など

HTML5, CSS3, JavaScript, jQuery, Vegas.js

CSSフレームワークを使わず、基本的に直にコーディングしています。フレームワークに縛られずに自分のやりたいように書くのが好きです。しかし、とにかく完成させることだけを目標に一人で作ったので、コードの美しさや可読性などが疎かになってしまいました。

http://yoshinobu-1968.jp/index.html





#### **WORKS**

## 3 プロジェクト - Webサービス事業

### ケトサーチ KetoSearch

近年、日本の飲食業における栄養成分表示の普及率が低いこと、その原因として、飲食業の経営が厳しく栄養成分表示を導入するのに手間がかかり、店にとってメリットが無いことが問題になっています。

この問題を解決するため、食品の栄養成分を簡単に計算し、投稿・検索・閲覧・評価できるWebサービスを個人開発しました。エンドユーザーは、スマートフォンが使える幅広い層を想定しています。

開発にあたり、飲食店の方への取材、清潔感のある テーマ色、グラフィカルな成分表示、QRコードの生成 など様々な工夫を行いました。

2020年3月にリリース。開発期間は3ヶ月ほど。現在はサービス改善をしつつ、事業展開しています。

### 使用言語など

HTML5, CSS3, JavaScript, jQuery, Ruby, Ruby on Rails, Slim, Sass, MySQL, Heroku, Docker https://www.ketosearch.org/

Q



#### **STUDY**

# 4 研究: レポートのWebサイトによる剽窃した部分の自動検出システムの開発

令和元年度(第70回)電気・情報関連学会中国支部連合大会 ・剽窃したと レポートの 会和元任度(第70回)雷尔、情報関連学会由国支部連合大会 )比較表は, の Web サ レポートのWebサイトによる剽窃した部分の自動検出システムの開発 の詳しい情 松山 塵偉児\* 尾 冠澤\* (津山工業高等専門学校) Development of an automatic detection system for plagiarized parts from website on the report Mairu Matsuyama, Fang Guan-Shen (National Institute of Technology in Tsuyama College) 大学や専門学校などで出題される課題に対して, Web サ イトから剽窃をおこない不正なレポートを提出する学生が 増加している。このような行為は学生への正確な成績評価 をより困難にさせ、著作権侵害の可能性もある。 剽窃行為は2つの種類に分けられる:(1)学生間に発生す るレポート同士による剽窃と、(2) 学生個人のレポートと Web サイトによる剽窃である。近年、インターネットと情 報技術の発展に伴い、後者による剽窃問題が深刻化した。 レポートが剽窃したものかを確認する作業は膨大な時間と 労力を要するため、教師への大きな負担となる。 現稲度が多く、他の文書では出現稲度が少ない単語を重要 これを解決するため、我々は Web サイトをもとに剽窃し な単語として評価する。式(1)は TF-IDF 重みの計算式で たレポートの自動検出手法を提案し開発した[2]. この手法 ある.  $n_{t,d}$  は文書 d における単語 t の出現回数,  $\sum_{i=1}^{n} n_{i,d}$ は、学生から提出されたレポートと出題課題をもとに自動 は文書 d における全ての単語の出現同数の和、N は総文書 で Web 検索を行い、検索結果の内容と比較することで剽窃 数、df(t) は単語 t が出現する文書数を表す. されたかどうかを判断するものである。しかし、これは手 法のみであり実際の教師が使用するには利便性・眼瞭性・ 宇田性が欠落している. したがって本研究は、我々が提案した剽窃検出手法[2]を この値の上位3つの単語を特徴語として抽出する。 もとにレポートの剽窃部分や剽窃率をわかりやすく確認で に剽窃した きるインターフェースを開発する. 剽窃検査には、Smith-Waterman アルゴリズムを用いる |物部分や割 (図 2). 一つの Web サイトの文字列を  $\langle x_1, x_2, \cdots, x_n \rangle$ , 検 2 剽窃検出システム スを開発し 査したいレポートの文字列を $\langle y_1, y_2, \cdots, y_m \rangle$ とする。まず, 2.1 システム概要 結していな 表  $V(V \in R^{(n+1)\times(m+1)})$  を作成し、0 行と0 列を値0 で初期 我々が提案した検出手法[2]の概要図を図1に示す。ま 実用性の向 化する.次に、セルスコアを以下の数式に従って表の空白の ず、教師がレポート課題の中から適当な単語を選択し、入 と期待する. セルの左上から右下まで計算する.ただし、セルの値を $v_{i,j}$ 力する、次に、提案システムが各レポートに対して含まれ ていないが、 スコアパラメータを (gap, match, mismatch) = (-1, 1, -1)る単語の TF-IDF 重みを計算する。これを元に上位3つの 単語をレポートの特徴語として抽出する。課題から選択さ れた単語と特徴語を検索語として Web 検索をおこない、検 此二:Smith 案結果の上位 k 件を剽窃元候補の Web サイトとし取得す  $v_{i-1,j} + gap$ , :含むコード る、次に、取得した各 Web サイト k件とレポートの文章を ).2, pp.981-比較し剽窃部分を検出する。比較には2つの文字配列から 類似する部分配列を検出する Smith-Waterman アルゴリズ ,たレポート ムを用いる[1]、最後に、剽窃部分がレポートに占める割合

> セルスコアの最大値を開始セル $V_{i',j'}$ とし、左上の方向 に沿って値が0のセル $V_{i'',j''}$ に辿り着くまでトレースパッ

> クをおこなう. トレースパックをおこなった経路と対応す

る入力文字列の部分配列は類似部分として識別し標記す

る、V から i'.....i'' 列をV から除く、得られた部分行列

を剽窃率として算出し、この値が設定された関値を超えた 場合、そのレポートを剽窃されたものとする。

特徴顛抽出では、TF-IDF 重みによりレポートに含まれ

る単語の重要度を評価する。この手法では、ある文書で出

22 特徵顯抽出

### 概要!

近年, インターネットの急速な発展により, 多くの大学, 専門学校では課題としてWebサイトから盗用した不正なレポートを教師に提出する学生が増加しています。

このような行為は、学生への成績を正確に評価することを困難にし、著作権侵害の問題を引い起こす可能性があります。

また,このようなレポートは手動で検出する必要があるため,教師への負担が大きくなります。

この問題を解決するため、論文では、Webサイトからの学生レポートの盗用を自動的に検出するシステムを提案しています。

### 実績

- ① 情報処理学会 第81回全国大会出場
- ② 第70回 電気・情報関連学会中国支部大会
  - → 令和元年電気学会論文発表賞 受賞

リンク: Paper submitted to The 81st National Convention of IPSJ

# 4 研究: ディープラーニングを用いたSQLの自動生成に関する研究

₹ 33.3%.

よの実験結 ま正解クエ り論理精度

|と一致し | |平均値

のいずれも

とがわか

関係を表す

こ影響が出

の平均値

寝が 0.5%

の精度が向

けして補助

これによ

WHERE

伝搬し、こ

集約関数の

ネットワー

L生成全体

ルを提案し

らの手法と

体のほうが

HERE の多

hort-term

#### ディープラーニングを用いた SQL の自動生成に関する研究

松 山 摩 偉 児 津山高専 電子・情報システム工学専攻

#### 1. はじめに

関係データベース(RDB)は、表形式のデータ集合を互いに関連付け、その関係を利用したデータベース(DB)のことであり、大量の情報が保管できる。しかし、RDB を利用するにはプログラム初心者には習得が困難な。SQL を理解する必要がある。さらに近年、システムエンジニアが不足しており、何度も同じような SQL を記述しなければならないことで、エンジニアの負担も大きくなっている。

これを解決する研究として、Zhong らの Seq2SQL<sup>1)</sup> と呼ばれる手法がある。彼らの研究では、大きく二つの貢献をしている。

一つは、自然音語の問合せ文、SQL クエリ、および Wikipedia の 2424l 個の HTML テーブルから抽出し た SQL テーブルの 80654 個の注釈付きコーバスをま とめた WikiSQL と呼ばれる DB を提供したこと。

もう一つは、自然言語を対応する SQL クエリに翻訳するための DNN である SeqZSQL を提案したこと である。しかし、これは SQL の「MAAX"、"MIN"、 "COUNT"、"SUM"、"AVG"]のような集約関数の生成 に対する再現率が低く、正しく予測されない可能性 がある。

本研究は、SeqSSQL の集約開数の生成に注目し、 自然言語の問合せ文と集約開数の依存関係を表す補 助情報を加えることで再現率の向上と SQL 生成全 体の精度の向上を図る。また、Zhong らの手法と概楽 手法を用いた実験を実施し、結果の比較分析を行う。 ただし、データセットは彼らが提供した WikiSQL を 用いる。

#### 2. Seq2SQL

Seq2SQL の概要図を図1に示す. Seq2SQL では自然言語の問合せ文とテーブルのカラム名を入力することで SQL 文が生成される.



図 1 Seq2SQL モデルの概要図

Seq2SQL は三つの機能により構成される: (1) 集約 関数を生成する部分, (2) SELECT のカラムを選択する 部分、(3) WHERE の条件文を生成する部分である。こ のモデルはまず、SQL クェリの集約開敷を生成する。 ただし、集約開敷を必要としたい場合に対応する NULLの操作も追加する。次に、SELECTのカラムに対 応する入力テーブルの列を指定する。最後にポインタ を使用して、SQL クェリの条件文を生成する。つまり、 Sog2SQL は3つの学習ネットワークを有している。



図2 集約関数の生成ネットワーク

本研究は集約関数の生成に注目するため、図 2 に集 約関数の生成ネットワークを示す、図2の入力xは全て のカラム名、SQL トークン、問合せ文の連結で定義さ れる:

 $lpha_t^{inp} = W^{inp} h_t^{onc}$  (2) 計算された $lpha_t^{inp}$ を Softmax 関数によって正規化し  $eta^{inp}$ を計算する。正規化した注目度 $eta^{inp}$ によって重み 付けされたエンコーダーの出力 $h^{onc}$ の合計を $K^{agg}$ とす。t

$$K^{agg} = \sum_{t} \beta_{t}^{inp} h_{t}^{enc}$$
(3)

Kagg は注目度が高い特徴量が保持され、注目度が低い特徴量は破棄される. つまり、集約関数に関する特

### 概要!

関係データベースは、テーブルと呼ばれる単純なデータ形式を利用したデータベースのことで、大量の情報を保管できます。

しかし、テーブルのデータにアクセスするには、習得が難しいデータベース操作言語であるSQLを理解する必要があります。

この問題を解決するために、先行研究では、ディープラーニングを用いて自然言語の問合せ文からSQLクエリを生成する Seg2SQLという新たな手法を提案していました。

ただ、残念なことに、この手法ではSQLの集約関数の生成精度が 低いという欠点を持っていました。

そこで本論文では、自然言語の問合せ文と集約関数の依存関係 を調査し、Seq2SQLの集約関数生成ネットワークにその依存関係 を表す補助情報を加える新たな学習モデルを提案しました。 **INTERNSHIP** 

## 5 インターン

- ・株式会社 オプティム
  - 2017年3月13日から約一週間
  - HTML/CSS/TypeScrpitを用いた機械学習ラベリングツールの開発
- ・株式会社 リブセンス
  - ・2017年8月14日から8月25日まで
  - Ruby on Railsを用いた社内Webツールの開発改善
- 山梨大学 西崎博光研究室
  - ・2019年8月19日から8月23日まで
  - Chainerを用いたDLの学習とオリジナルアプリの開発

#### **TASK**

## 6 今後の課題

- React, Vue.jsによるフロントエンド開発
- docker, gitの知識の深化
- ・コードの可読性向上
- UI/UX向上のためのWebデザイン知識獲得