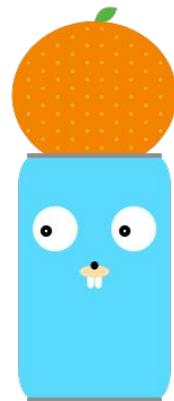


# DajareGoを支える技術

ダジャレによる豊かなコミュニケーションの裏側

@未来祭 LT会 (2022/10/08)



# 自己紹介

- 公立はこだて未来大学 システム情報科学部 学部1年生
- 思い立ったらなんでも作ります  
プログラミング、電子工作、3Dモデル、自律型サッカーロボット など作品多数
- 鉄道旅行、クイズなどが趣味
- Twitter: [@jugesuke](https://twitter.com/jugesuke)
- GitHub : [jugesuke](https://github.com/jugesuke)
- Website: <https://jugesuke.pages.dev/>



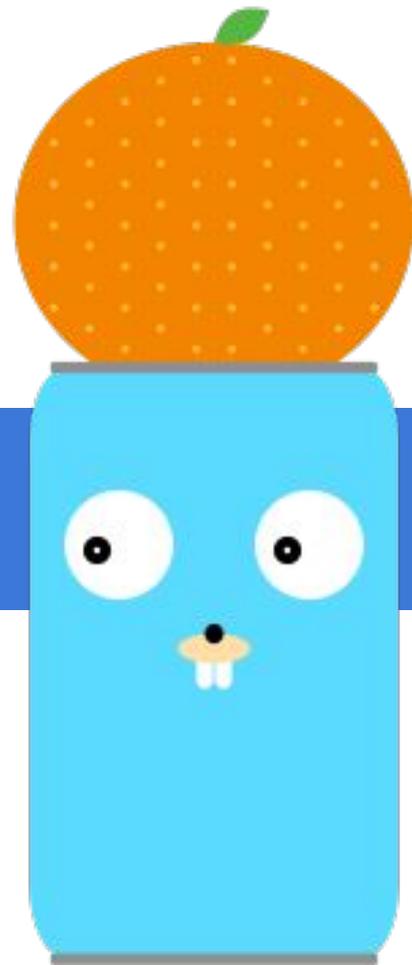
# 本日の内容

思いつきで開発を始めたら...

いつのまにか自然言語処理の研究が始まってしまった話

# DajareGo

由来: ダジャレでいこう!/ダジャレ×Go言語



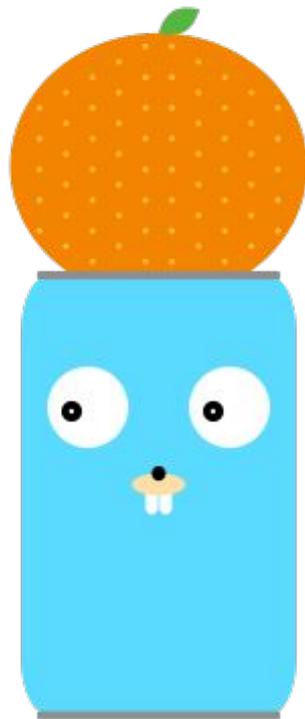
# DajareGoの概要

## Discord上でダジャレを検出するBot

言語: GO

### 使用ライブラリ:

- Kagome (形態素解析器)
- NEologd (形態素解析辞書)
- DiscordGo (DiscordBot作成ライブラリ)
- zap (ロギングライブラリ)



本作品は[技術展2022](#)コミュニケーション部門に出展しました(プレゼン一般公開は終了)

# きっかけ

Discordでの雑談で、**ダジャレ**を検出できたら面白いよねという話になったので、勢いで実装した

**ダジャレって何？**

# Google Scholar

ダジャレ 定義



- すべての言語    英語 と 日本語のページを検索

巨人の肩の上に立つ

# ダジャレを定義する

(シャレとは、) **同音または類音で意味の違う言葉を利用して言語表現の指示機能を複雑にする修辭的言語操作**

日本語の場合、ある意味(M1)と**言葉の音声(F)**がたまたま別の意味(M2)を持つ言葉と同じものが非常に多い。(中略)この二つのF(言葉の音声)が同じであったり、よく似てしまった時**シャレが成立する**。(F1 = F2 または F1 F2)  
そして、F1 と F2 が似ているほど M1 と M2 が隔たっているほど**シャレの効果は大きい**。

ダジャレのダは「駄句」「駄犬」「駄菓子」の「駄」。(中略)加工の仕方が下手なのは品質が劣るという意味の「駄」**ジャレで**、サービス精神が欠けている。

織田 正吉, 野村 雅昭, Wヤング, 長島 平洋, シャレ・ダジャレ学事始(ことはじめ)(第 19回研究会), 笑い学研究, 1999, 6 巻, p. 55-67 より引用

# ダジャレとは？

読みが同じまたは似ているがそれぞれ意味の違う言葉が含まれるフレーズ

アルミ缶の上にあるミカン  
アルミカン/ノ/ウエ/ニ/アル/ミカン

布団が吹っ飛んだ  
フトン/ガ/フットン/ダ

人民の人民による人民のための政治  
ジンミン/ノ/ジンミン/ニ/ヨル/ジンミン/ノ/タメ/ノ/セイジ

# どう処理する？

- 単語の読みをどう取得する？
- どうやって文章を単語に分解する？
- “言葉の音声の類似性”をどう判定する？

▶ ダジャレ判定を実現するにはこれらの課題を乗り越える必要がある

# 自然言語処理を知る

形態素解析をすれば、文章を単語に分解して読みを取得できるらしい！

DajareGoでは、Kagomeという形態素解析器を採用した

すもも	名詞, 一般, *, *, *, *, すもも, スモモ, スモモ
も	助詞, 係助詞, *, *, *, *, も, モ, モ
もも	名詞, 一般, *, *, *, *, もも, モモ, モモ
も	助詞, 係助詞, *, *, *, *, も, モ, モ
もも	名詞, 一般, *, *, *, *, もも, モモ, モモ
の	助詞, 連体化, *, *, *, *, の, ノ, ノ
うち	名詞, 非自立, 副詞可能, *, *, *, うち, ウチ, ウチ

# アルゴリズムを知る

編集距離・始端フリーDPマッチングを用いれば、あいまい検索が出来るらしい！

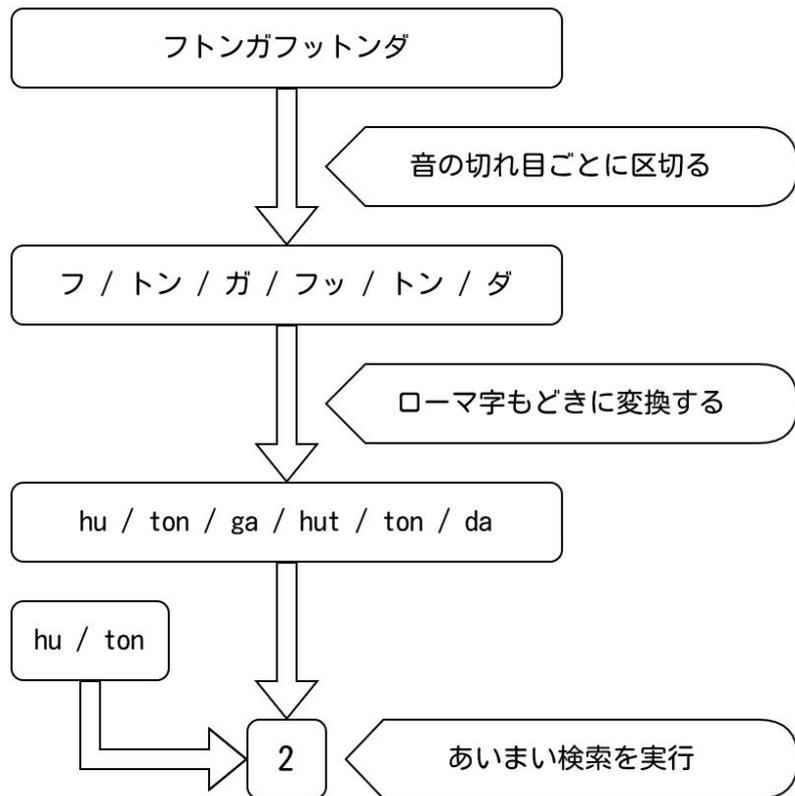
▶ これをベースにアルゴリズムを作れば良さそう

$$D_{i,j} = \min\{D_{i-1,j-1} + \delta(s_i, q_j), D_{i-1,j} + 1, D_{i,j-1} + 1\}$$

	-	ふ	っ	と	ん
-	0	1	2	3	4
ふ	1	0	1	2	3
と	2	1	1	1	2
ん	3	2	2	2	1

	-	ふ	と	ん	が	ふ	っ	と	ん	だ
-	0	0	0	0	0	0	0	0	0	0
ふ	1	0	1	1	1	0	1	1	1	1
と	2	1	0	1	2	1	1	1	2	2
ん	3	2	1	0	1	2	2	2	1	2

# あいまい検索を実装



## 編集コスト設定

置換：音節同士の編集距離

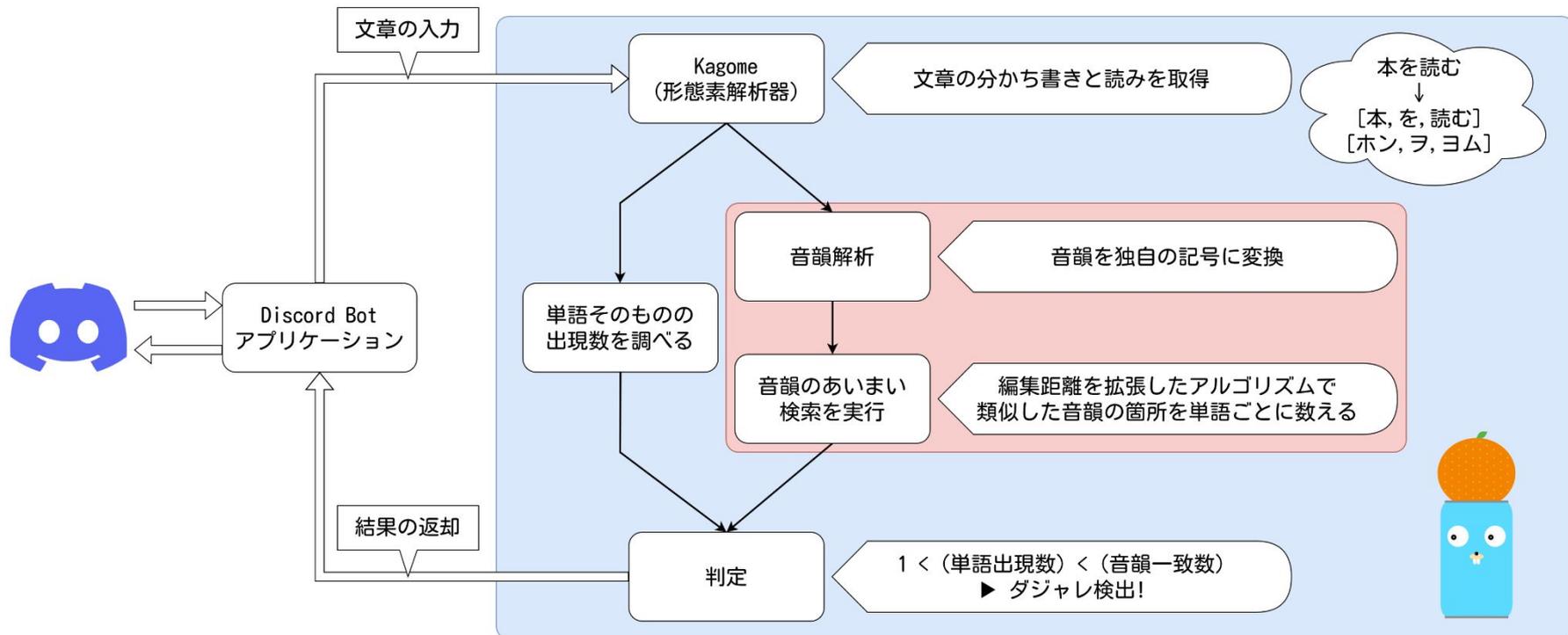
削除：+2

挿入：+2

合計コストが1以下なら類似と判定

	-	hu	ton	ga	hut	ton	da
-	0	0	0	0	0	0	0
hu	1	0	2	2	1	2	2
ton	2	2	0	2	3	1	3

# ダジャレ検出処理の概要



# 検出精度

データセット種別	正解率	正解率 (長文無視)
ダジャレ (n=1532)	70.1%	69.2%
非ダジャレ (n=5060)	91.0%	98.5%

小数点第二位を四捨五入している

ダジャレデータセットは以下より収集

<https://dajare.jp/ranking/>

<https://dajareshuu.web.fc2.com/>

非ダジャレデータセットは以下より収集

夏目漱石 『こころ』(青空文庫)

- **誤検出率の低減**を重視したチューニングに
  - 通常の会話を邪魔しないため
- “**猫が寝転んだ**” のような2文字一致のパターンは検出できない
  - これを検出しようとする**と非ダジャレの正解率が著しく下がる**
- **知識の必要なダジャレ**は現状検出できない
  - 英語を利用したもの、他の単語をもじったもの等

# 実際に使ってみた

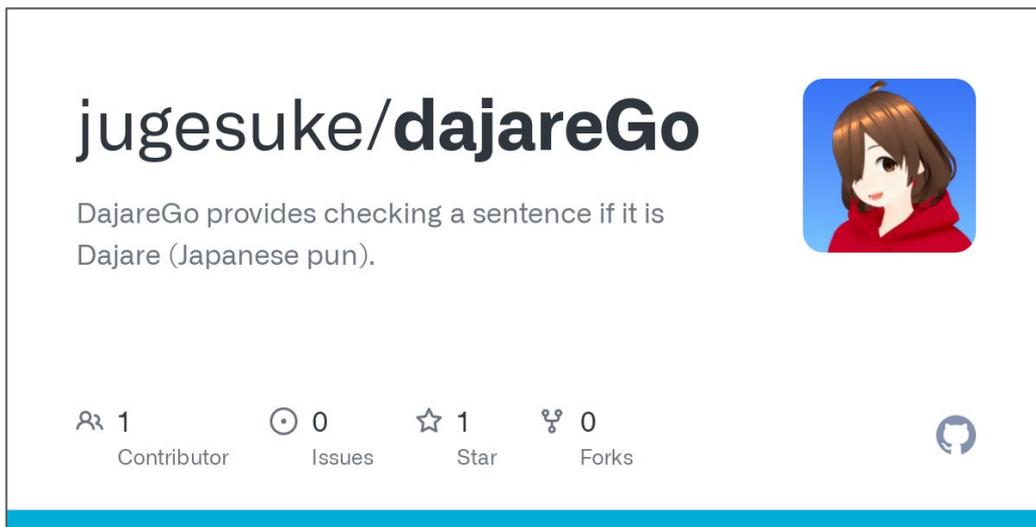
- DajareGoで遊ぶ投稿
  - 予期せず発したダジャレ
  - まさかの誤検出
- ▶ 新たな会話のきっかけとなった



**ダジャレで会話が活性化！**

# あなたの作品にも

ダジャレ検出ライブラリをGitHubにて公開中  
あなたなら、これを使ってどんな作品が作れそうですか？



The screenshot shows the GitHub repository page for 'jugesuke/dajareGo'. The repository name is displayed in large, bold black text. Below it, a short description reads: 'DajareGo provides checking a sentence if it is Dajare (Japanese pun)'. To the right of the text is a square profile picture of a woman with brown hair wearing a red hooded garment. At the bottom of the repository card, there are four statistics: '1 Contributor', '0 Issues', '1 Star', and '0 Forks', each with a corresponding icon. A GitHub logo is visible in the bottom right corner of the repository card.

<https://github.com/jugesuke/dajareGo>