

MOTIVATION

Background

- 国際化に伴う外国語学習者の増加
- 外国語学習における辞書の重要性
- 言語学習の系統的研究の不振

Problem

- 母国語と表記の異なる言語の学習は困難
- マイナーな言語ほど良質な辞書が少ない
- 同様に、効率的な言語学習環境の不足

GOAL

- 文字が読めなくても検索できる初学者向けのUI
- 上級者の使用にも堪える良質な辞書データの整理
- Web辞書の利点を活かした外国語学習支援の実装

辞書を引くことの難しい言語 ～ペルシア語編～

どこからどこまでが文字?

辞書引き例

میبینمتان (私はあなたを見る)

↓

دیدن (見る:不定形)

辞書にこの単語がない!?

こんな文字読めません!

ペルシア語は、アルファベットが単独、語頭、語中、語尾のどこに位置するかによりそれぞれ形が異なり、変形する。また動詞の活用は多岐にわたり、さらに、本来スペースで区切る二語以上の語が例外的に結合する場合がある。(左下図参照)

Non nativeにとって難読性が高い！
辞書を引くために語学力が必要！

初学者にはハードルが高い

ペルシア語について

ペルシア語はイランイスラム共和国とその周辺の国々で使用される言語であり、その話者人口は約5,000万人である。アラビア文字を使用し、右から左に書字する。ちなみに英・仏語と同じ印欧語族であり、広い同種族言語を持つ。

不定形 使用例文 短縮形

دیدن: "من شما را می بینم." = "میبینمتان."

「見る」 短縮時に消失 → 私はあなたを見る ←

動詞と名詞の活用例およびスペースの省略

文字連結時に生じるアルファベット変形例

形態	単独	語頭	語中	語尾
ペルシア語フォント	ه	هـ	هـ	هـ

※すべて同一文字

Our Solution

システム概要

プロジェクト趣旨

日本人向け多言語学習コンテンツとして開発

リーディングチュウ太→日本語学習者向け

Rikai.com→英語から日本語の単一方

辞書引きの難しい言語へのアプローチ(現状)

- ◇スクリーンキーボード
- 文字の読めない言語に対応
- ◇活用解析
- 文法による語形の変化に対応

文字が読めなくても大丈夫！

活用規則を解析し見出し語を表示！

“دیدن”

“میبینمتان”

活用され変形された単語を入力しても...

ユーザー (初学者)

メインプログラム

検索

辞書データ

活用解析

単語の活用形や文章の短縮形など、見出し語に無い文字列が入力された場合は、形態素解析を行い文字列を原形にまで分解し、再度検索をかけることで変形した単語から見出し語の辞書引きを行う。

検索語 人称接尾辞 人称語尾 語根 接頭辞

میبینمتان = دیدن + متان

見出し語	発音記号	品詞	語根	接頭辞
دانشستان	dānestān	verb	دان	1=知る, 2=思う, 3=...
دیدار	didār	Noun	دید	1=訪問, 2=面会, 3=...
دیدن	didān	verb	دید	1=見る, 2=会う, 3=...
دیر	dīr	adjective	—	1=遅い

元の文字列を形態素に分解、語根で再検索。

スクリーンキーボード

○日本語 ○فارسی ○品詞 ○動詞検索

●前方一致 ●部分一致

検索結果

سلام: こんにちわ

سلامانی: 平安

سلامانی: 床屋

سلامانی: 趣味

文字が表示・入力されるたびに候補を絞り込む

アルファベットが書かれたボタンにポインタを重ねると適切に変化した形が表示される

keyboard

چ	ج	ث	ت	پ	ب	ا	آ
س	ژ	ز	ر	ذ	د	خ	ح
ف	غ	ع	ظ	ط	ض	ص	ش
ه	و	ن	م	ل	گ	ک	ق
							ی

学習支援機能&単語データ

音声再生と単語暗記テスト機能を搭載する。収録する単語に頻度を関連づけ、その数値を基に選別した単語を、訳を隠した状態でレベル別に表示し、利用者の自主学習の効率化に寄与する。

目標収録語数80,000語のうち、主要2,500語に関してはWebや古典文学等の幅広いジャンルから厳選した例文とネイティブの発音データを付与し、その単語が実際にどのように使用されているかを文字と音で紹介する。

Future work

- ★辞書データの作成
- 目標80,000語の収録
- Native speakerの発音採録
- ★学習支援機能の実装
- 音声再生&単語テスト
- 文書一括辞書引き機能