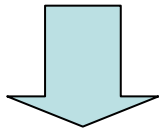


分散支援型人工知能(仮)

栗原一貴

分散支援型人工知能(仮)とは

- 機械だけでは解決できない問題を, 細かく分割して多数の**人間の知識**により解決するシステム. cf. Folding@home 多数の計算機力で解決
 - 参加する一人一人には, 全体で何を行っているのかわからないほど問題が分割されている.
 - 機械には難しいが, 一般人なら誰でも持っている能力を少しずつ借りる. cf. 遠隔医療システム ←少数のプロの能力
- 

【ただし私は自然言語処理は専門ではありません】

自然言語処理の流れと 各段階における妥当な問題形式

アナログ資料

画像取り込み

OCR

デジタル資料

形態素解析

構文解析

意味解析

文脈解析

・この文字列には何と書かれていますか？

・この文(句, 節)を品詞分解しなさい。

・この文(句, 節)の構造として正しいものを選びなさい。

・この文(句, 節)におけるこの単語の意味を選びなさい。

・この指示語の内容を答えなさい。

・この文章Aは次の文章Bと関係がありますか？

本システムが完成した場合の 仮想運用デモ

出来たらよいと思っていることリスト

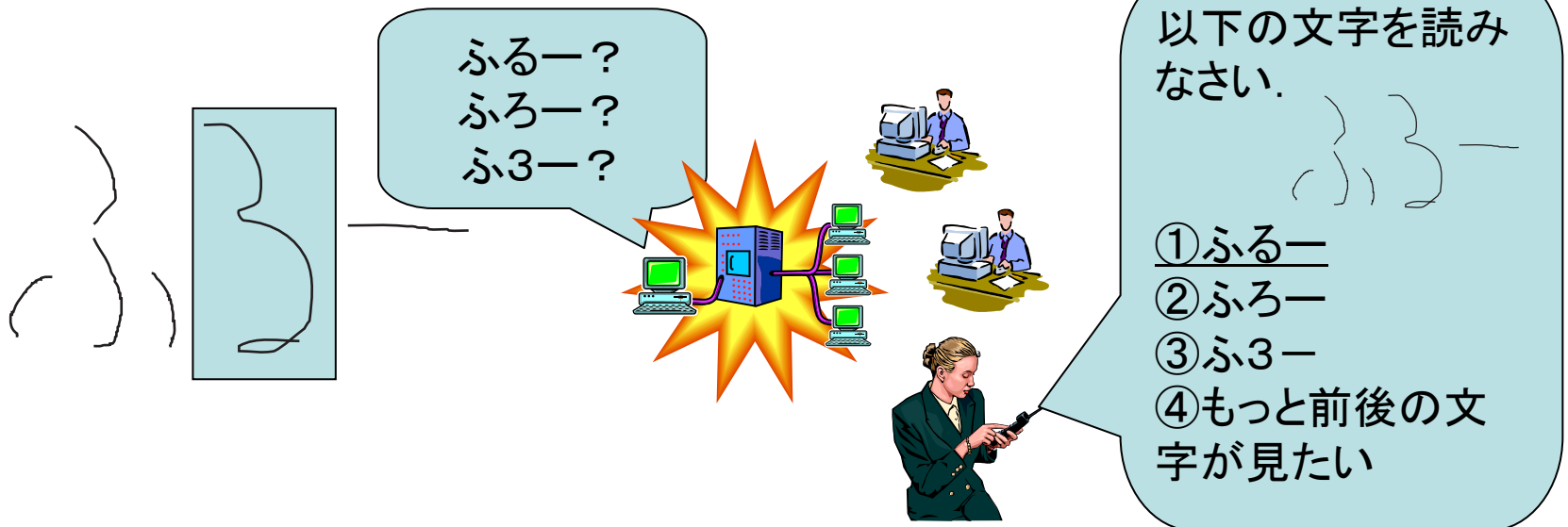
- 現代語解析済みコーパス作り
- 文脈マッチング型検索
- 対話システム
- (翻訳システム)

現代語解析済みコーパス作り

- 現在，自然言語処理のための解析済みコーパスは人手により作られている。（憶測）
- 「現代語」のコーパスなら，実際に大勢の現代人に解析してもらった方がいい．公平で，生きた情報だから．
- 自然言語研究のためのコーパスを自然言語処理＋人手で作る（！？）
- 将来的には，人手に頼る部分を減らして行きたいので，システム自身の教育のためにも．

1. 生コーパスを作る

- 現代語のアナログ資料(画像)を用意する
- OCRにかけ、文字として読めそうな部分を切り出し、幾通りかの認識文字を求める
- 曖昧なところはサポーターにたずねる



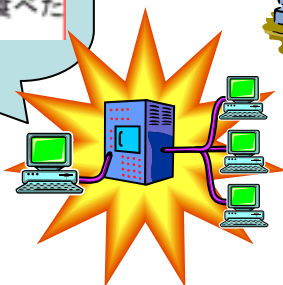
2. 解析

- 形態素解析, 構文解析, 意味解析, 文脈解析を自動で行い, 候補を求める
- 曖昧なところはサポーターにたずねる

太郎は私と
カレーを食べた。

太郎
私 は カレー を 食べた

太郎は 私
カレー を 食べた

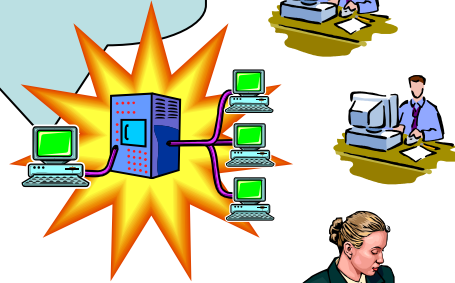


文の構造として適切なものを選びなさい。
「太郎は私とカレーを食べた。」

- ① 太郎
私 は カレー を 食べた
- ② 太郎は 私
カレー を 食べた
- ③ もっと前後の文が見たい

それはどう考えてもおかしい。

滑稽だ？
奇妙だ？
不合理だ？



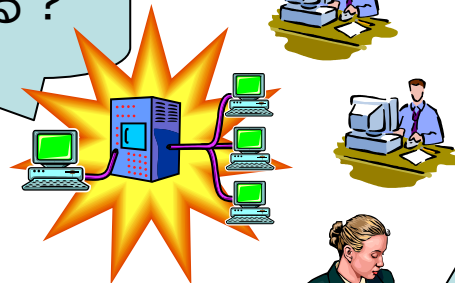
あまりに短すぎて文脈がわからない場合もある。

次の「おかしい」の意味で適切なものを選びなさい。
「それはどう考えてもおかしい。」

- ①滑稽だ
- ②奇妙だ
- ③不合理だ
- ④もっと前後の文が見たい

およいでからはしる

泳いでから走る？
泳いで蚊等は知る？



ただの漢字変換の問題か。

単語の区切り方として適切なものを選びなさい。
「およいでからはしる」

- ①泳いでから走る
- ②泳いで蚊等は知る
- ③もっと前後の文が見たい

文脈マッチング型検索

- ある文章Aが，ある文章Bに関係しているかどうか調べる.
- 国語の作文問題の添削，苦情メール・アンケート処理，長い文章が「要は賛成なのか，反対なのか」を判定する 等.
- 文脈解析の問題

例：国語の作文試験の添削支援

問. 「中学生のボランティア活動への参加の強制について①賛成②反対のどちらかの立場をとって論じなさい。」

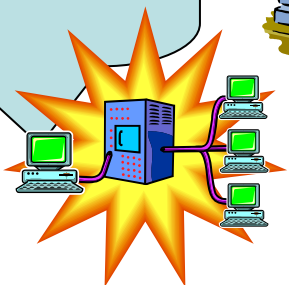
(実際に栃木県の中学生向け模擬試験に出題された問題より)

- 文章内容があらかじめ指定された賛成, 反対と一致しているか.
- 「効率」という観点で論じていると減点.
- 誤字, 脱字
- 起承転結の段落構成か.

(採点基準の指針の例)

作文

この作文（もしくはその一部）は、賛成意見？ 反対意見？

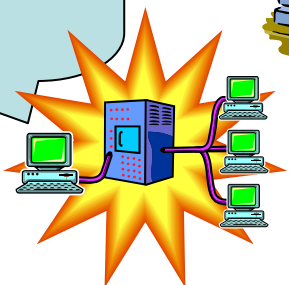


以下の文章内容は、「～」という意見に賛成ですか。反対ですか。それとも曖昧ですか。

- ① 賛成
- ② 反対
- ③ 曖昧
- ④ もっと前後の文が見たい

作文

「効率がよいから全員強制の方がよい」という論旨かな？



以下の文章内容は、「効率の面で全員強制の方がよい」という論旨ですか。

- ① Yes
- ② No
- ③ 曖昧
- ④ もっと前後の文が見たい

どのくらいの長さの文章を一人に読ませるのか。

対話システム

- リアルタイム性が要求される
- 厳密に答えられなくても、個性の範疇になる
- 無根拠にいい加減なことを答えるロボットは信用されないが、本システムでは、発言の根拠を背後の人間がもっている。

好きな
食べ物
は？

人間



知っている食べ物リ
ストはデータベース
にあるけど、「好き」
の対象はわからない
よ～.

「好き」って
何？
牛丼が好きで
す.

バ



僕はその中では
牛丼が好きかな.



僕はその中では
牛丼が好きかな.



僕はその中では
うどんが好きかな.

サポーター(人間)

翻訳システム（可能性の示唆のみ）

- 翻訳ページにある「テキスト翻訳」は逐語訳的で性能が悪い
- しかし訳し先の言語を母国語とする人には、大体何が言いたいかが理解できる場合がある
- →ちゃんとした文章に整形してもらおう
- 結局2ヶ国語話せる人が必要になるような気がする。

やはり問題となるプライバシー

- 個人情報をお細切れにして散布するやうなもの.
- 人名を偽名に変換する処理は可能か？

難問：

松井選手，メツじゃマツイ。（AERA地下鉄中刷りより）

【松井でないと成立しないし，そもそも形態素解析が難しい。】