

2020年度セミナー・特別研究

担当教員 馬 青

これからの連絡方法

- すべてチームズの「馬研チーム」の「特別研究生チャンネル」にて行う
 - メンバー同士の情報交換も
 - 適切にチェックしないことはセミナーの評価に影響を与える
 - 緊急の場合はメールで連絡する場合もある
 - 迷惑メール解除設定
 - 教員: qma@math.ryukoku.ac.jp
- なお、メールで連絡するとき、必ず名前を名乗ってください。
- 大学メール (t*****@mail.ryukoku.ac.jp) を必ず使えるように設定しておくこと
 - 詳細メール、大学・学部・学科からの大事なメールは大学メールに送付するので使えるように
 - 携帯に転送設定をする

注意事項

- セミナーは授業。必ず出席すること。セミナーの欠席は「どうしても」の理由がなければ認めない。無断欠席すると、大きく減点される。セミナー(100点での評価)が不合格だけでも卒業できない
- 特別研究は、それなり分量の研究結果がないと、卒業できない。主担当教員がOKを出しても他の先生の審査に通らない。
前例多数あり
- 後期授業開始後、週4日・1日6時間以上実験室にて研究を行うことを義務付ける(出席チェックを取る。自宅での研究は原則不可)。この条件をクリアできなければ指導を行わない
- 就職活動は大事だが、セミナー・特別研究も大事。片一方だけではうまくいかない

マイクロソフトチームズ

- PCのデスクトップアプリがあり、使うように
- スマホのアプリもあり、使うように
- 馬研のすべてのチャンネルの通知をオンにするように
- オンラインセミナーはこのツールを利用する

通年のスケジュール

- 4/9～7/16 木10:00-13:00 セミナー
 - 輪講の形で自然言語処理の基礎を学ぶ
 - 計算機の使い方、必要なソフトのインストールなど
 - 研究テーマ探しのための討論・相談
- 7/16(木): **第1回の研究発表**(1人10分。テーマ、理由・動機、計画、準備状況等)
- 夏休み: 研究テーマを決定し、本格的に研究をスタートさせ、一定の成果を獲得
- 9/18(金)4,5講時: **第2回の研究発表**(1人10分。進捗状況について報告)
- 9/29から、**火4,5講時**
 - 科学技術英語 & 研究進捗状況報告
 - 輪講の形で自然言語処理関係の英論文を読む
- 11/24: 全実験を終了し、**第3回の研究発表**(卒論着手がOKかどうかのチェック)
- 11/24: 卒論の書き方の指導・質問応答
- A:12/18(金)まで、B:1/6(水)までに**卒論Ver.1**を提出
- Aは12/21(月)に、Bは1/8(金)までに指導する。Ver. 2の提出は指導時に指示する。
- 2021/1/15(金) **最終版卒論**を教務委員へ提出(すぐ返却、実験室待機)
(審査教員に提出まで(1/22(金))、最終修正可能。ただし、発表の準備もしなければならない)
- 2021/2/3,4(水木)の**特別研究審査会**に向け、発表の準備を行う
その間、3回以上の発表練習を行う
卒論提出・特別研究審査会までのスケジュールなどの詳細は以下を参照

<http://www.math.ryukoku.ac.jp/~www-qma/info/database/index.html#500>

研究テーマについて

キーワードは「言葉」+「プログラミング」

- 「言葉」と「プログラミング」というキーワードさえ入っていればどんな研究テーマでもOK
- ただし、それだけでは面白くないかも。**+α**が付いていればOK。この**+α**とはたとえばこれがあると便利（助かる、面白い、楽しい・・・）であったり、何かの見えない事実や結果が研究によってわかるようになったり、ということである

これまでの卒論を勉強すること

- これまでの卒論を読むことにより、馬研ではどのような研究を行ってきたか、または、言語処理の研究ってどのようなものかがわかる
- 興味のあるテーマは再度深く研究していくのもよい
- そこからヒントを得て新しいテーマを見つけるのもよい

これまでの特別研究テーマ

ゲーム

- コトバテトリス
- コトバ五目並べ
- マジカルバナナ連想ゲーム
- ポケモンシリトリ
- 語彙力測定・文字パズルゲーム
- クロスワードパズルの作成支援
- 人工無脳
- 祭り・神社・ゆるキャラのクイズゲーム
- なぞかけの自動生成

- 単語の多義性解消
- 専門用語の自動抽出
- 略語の辞書構築
- 日本語表記ゆれの解消
- 同義語の抽出

研究

応用

- Webデータからの「流行語」自動発見
- 計算機による俳句理解の試み
- 「渋谷弁」(ギャル語)の翻訳
- 図書・音楽・映画検索システムの作成
- Wikipediaから武将関連情報の抽出
- 特定トピックのBlogデータの抽出と意見分類
- コピペレポート発見システム
- 英語文書閲読支援
- 青空文庫蔵書検索システム
- スマートフォンの道案内アプリ
- 個人化レシピ提案システム

使用言語

- C
- Java
- Perl
- Ruby
- Python
- HSP

研究データベース

- 本資料は

<http://www.math.ryukoku.ac.jp/~www-qma/info/index.html>

にある

- 研究を遂行していく上で、参考になる資料やデータなどは

<http://www.math.ryukoku.ac.jp/~www-qma/info/database/index.html>

にある。なにかを捜すときはまずこのページから捜してください

計算機環境について

- 「研究データベース」を用いながら説明
- 実験室に行って確認
- Chasen, MeCab, CaboChaありなしの確認

セミナーについて

輪講に使う資料

- 大学院「自然言語処理特論」講義に使っている資料
- これをベースに、上記資料に挙げている参考文献などを調べて、内容を $+\alpha$ にして輪講を行う
- 輪講の形式は資料をみんなに配って行ってもよく、パワーポイントを使ってもよい。両方を使い分けしてもよい

輪講の割当て

- 形態素解析(p1~p7) + 構文解析(CYK法) + 文節分割: 1名
- 構文解析(チャート法) + 確率型構文解析 + 依存文法解析: 2名
- 情報検索1 (p68~p75) + α : 2名
- 情報検索2 (p76~p90): 2名
- 機械学習手法:
 - 決定リスト、Naïve Bayes: 2名
 - 三層パーセプトロン(p52~p58): 2名
 - 自己組織化神経回路網モデルSOM
 - サポートベクターマシンSVM (p59~p67、ただしp63,64のソフトマージン除く):
 - Deep Learning: SdA, DBN, LSTM/RNN, CNN...

自然言語処理の紹介

自然言語処理とは

- 人工言語
 - プログラミング言語のように、人間によって設計された言語
- 自然言語
 - 日本語のような自然発生的に生まれた言語
- 自然言語処理
 - 自然言語をコンピュータに処理(解析・理解・生成)させる一連の技術。計算言語学とも言う

主な応用

- 仮名漢字変換処理(日本語入力)
- 情報検索・情報抽出・文章要約・質問応答(Google など)
- 機械翻訳
- 音声認識・音声合成(TTS: Text To Speech, テキストの読み上げ)
- 言語生成(ロボットの会話など)
- スペルチェック(英語)・日本語変換チェック(変換? 返還?)・構文チェック・・・これらはすでにWordなどに実装されている

アプローチ

- 人間は言語処理が得意なので、人間を手本にするのが一番。しかし、
- 言語学や心理学などの諸分野における長年の研究にもかかわらず、人間における言語処理メカニズムは解明されていないのが現状。したがって、
- 人間の言語処理メカニズムと無関係に、コンピュータに適する自然言語の解析方法の確立と、既存の研究成果に基づいた効率的なアプリケーションの開発を目的としている
- 用いる方法：規則・統計・機械学習
- 必要な基本技術：形態素解析・構文解析・意味解析・文脈解析

単語レベルでの研究

くるまでまつ



くる | まで | まつ

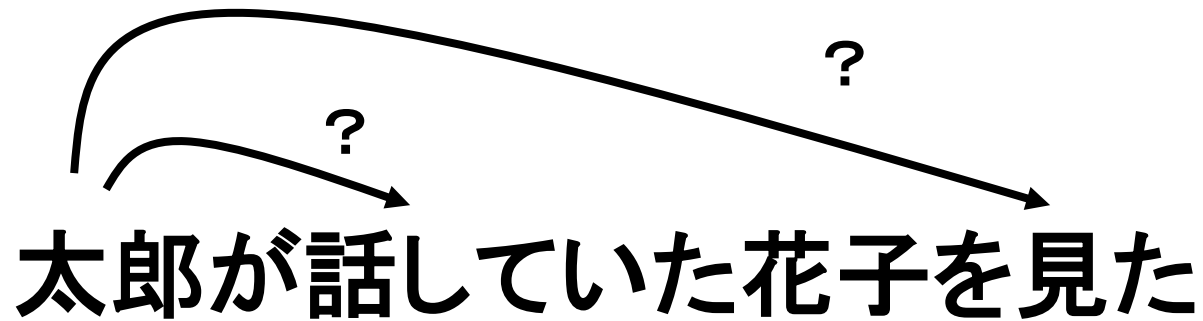
くるま | で | まつ



来る | まで | 待つ

車 | で | 待つ

文レベルでの研究



意味レベルでの研究

彼に期待をかけた → 希望・期待を持つ

絵をかけた → 物をぶら下げる