

自然言語処理論

1. 概論

1

授業の進め方

- 講義(火,木)
- オフィスアワー(火)
 - 演習、補講
- 講義に関する連絡
 - 授業中
 - ネットニュース frontier.lecture
 - ホームページ
 - ◆ <http://www2.jaist.ac.jp/~kshirai/local/i223/>

2

授業の進め方

- 講義日程
 - ホームページ参照
- 教科書
 - 自然言語処理, 長尾真編
- 配付資料
 - /home/lecture/is/i223/06-1-2/
 - <回数>.pdf, <回数>l.ps.gz, <回数>s.ps.gz
- 評価
 - 中間試験, 期末試験 (50%ずつ)

3

講義の内容

- 概論
- オートマトンと言語
- 文法(4回)
- 形態素解析(2回)
- 辞書と概念階層
- 言語の統計モデル
- テキスト処理(3回)
- 応用

4

言語

- 形式言語(formal language)
 - 記号列の部分集合
 - ◆ ex. 0と1から構成される記号列のうち、0と1を同数含むものの集合
- 人間が設計する言語
 - プログラミング言語
 - エスペラント語
- 自然言語(natural language)
 - 人間が日常生活で自然に使う言語
 - 日本語, 英語, 中国語…

5

自然言語処理とは?

- natural language processing (NLP)
 - 計算機で自然言語を「処理」すること
 - 人工知能の研究の一分野

6

自然言語処理の歴史

- 本格的な研究は1960年代
 - 計算機による記号処理
 - 機械翻訳
 - ◆ 直接翻訳方式
- 研究対象の広がり
 - 対話システム
 - 情報検索
 - 自動要約

7

自然言語処理

- 自然言語を理解する(解析する)
 - 文や文章の内容を理解する
 - 浅いレベルから深いレベルまで
 - ◆ キーワードの検出
 - ◆ 文の構造, 意味まで理解
- 自然言語を生成する
 - 翻訳結果の生成
 - 応答生成
 - 要約生成

8

自然言語を理解する

- 形態素解析
 - 単語区切りを決める
 - 品詞を決める
- 構文解析
 - 文の構文的な構造を決める
- 意味解析
 - 文の意味解釈を決める
- 文脈解析
 - 文脈や状況から文の解釈を決める

9

形態素解析

- morphological analysis
- 単語の境界を決める
 - くるま/で/まつ
 - くる/まで/まつ
- 単語の品詞を決める

I	water	the	plants	in	the	garden
代名詞	動詞	冠詞	動詞	前置詞	冠詞	名詞
	名詞		名詞			

- 活用の処理をする
 - 複数形, 動詞の活用変化

10

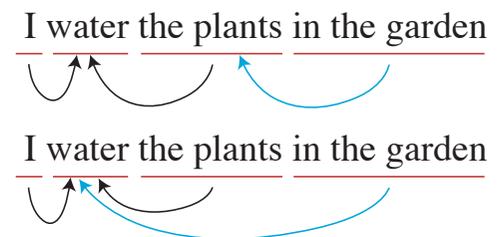
形態素解析に必要な知識

- 辞書
 - 単語と品詞の対応関係
 - ◆ もも: 名詞
 - ◆ water: 名詞, 動詞
 - 活用型など
- 語尾変化の規則
 - 英語: 複数形, 三単現, 過去形
- 品詞間の接続関係
 - ○ 代名詞 動詞
 - × 代名詞 名詞

11

構文解析

- syntactic analysis
- 文の構文的な構造を決める
- 単語の修飾・被修飾関係を決める



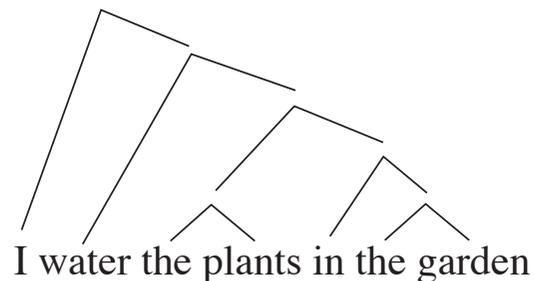
- I ate a dinner in the restaurant の場合は下の解釈のみ可能

12

構文解析に必要な知識

● 文法

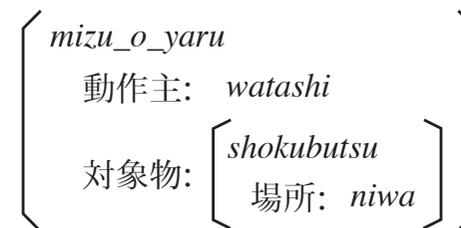
- 文の構文的なまとまり(句,節)を記述
 - ◆ 冠詞と名詞で名詞句を作る
 - ◆ 前置詞と名詞句で前置詞句を作る
- 構文解析結果は木構造で表現される



13

意味解析

- semantic analysis
- 広義の意味解析
 - 文の意味を決めること
 - 計算機が保持できる形式に変換する
 - ◆ I water the plants in the garden



14

意味解析

● 狭義の意味解析

- 単語間の修飾関係の種類を決める
 - ◆ ex. 主語と動詞の関係
 - ◆ 動作主: John played tennis
 - ◆ 対象物: The stick broke
- 単語の意味を決める
 - ◆ plantは「植物」か「工場」か

15

意味解析に必要な知識

- 格フレーム
 - 個々の動詞について、構文的関係(格)と意味的關係の対応を記述
 - ◆ water₁
主格→動作主, 目的格→対象物
 - ◆ break₁
主格→対象物
- 連想関係
 - waterはplant(植物)とは意味的に関係があるが, plant(工場)とはない

16

文脈解析

- discourse analysis
- 前後の文脈を見て、始めて文の解釈が一意に決まる場合がある
 - I water the plants in the garden
 - ◆ Iとは誰か?
- 単独の文ではなく、文章(文の集合)を解析の対象とする

17

文脈解析の例

- 照応解析(anaphora resolution)
 - 代名詞の先行詞を特定する
 - 省略されている要素を特定する
 - ◆ 太郎は映画を見た
(太郎は)(映画を)面白いと思った
- 談話構造解析(discourse structure analysis)
 - 文章の構造を理解する
 - ◆ 序論, 本論, 結論

18

自然言語解析の流れ

- 形態素解析
- 構文解析
- 意味解析
- 文脈解析

- 個々の処理の出力が後続する処理の入力となる

19

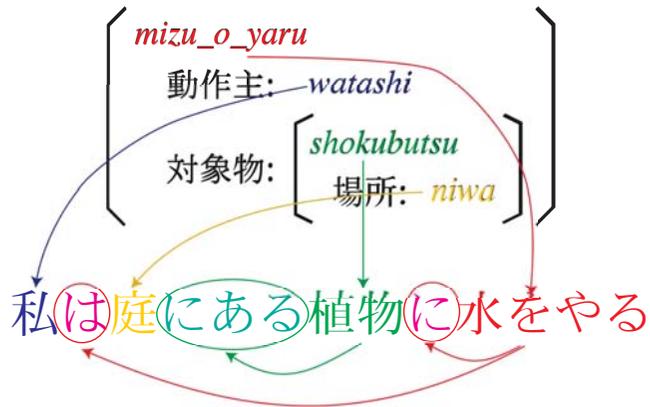
自然言語生成

- natural language generation
- 定型文のみ生成
 - 電話, ATMなどの自動応答
- テンプレートによる生成
 - ただいまから<時>時<分>分ちょうどをお知らせします
 - 発音の整合性を考慮する必要がある

20

自然言語生成

● 意味解析結果からの生成



21

応用

● 自然言語処理の様々な要素技術の組み合わせ

- 機械翻訳
- 対話システム
- 自動要約
 - ◆ 自動抄録

22

応用

● 情報化社会の発展

- 電子化テキストの爆発的な増加

● 大量のテキストを扱う技術

- 情報検索

- ◆ 文書集合の中から、自分が欲しい文書だけを取り出す

- 情報抽出

- ◆ 文書の中から自分が欲しい情報を抽出する

- ◆ 新製品発表の記事
 - 製品名, 発売日時, 価格

23

講義の内容

● 概論

● オートマトンと言語

● 文法(4回)

↙ 解析

● 形態素解析(2回)

↙

● 辞書と概念階層

← 解析に用いる知識

● 言語の統計モデル

● テキスト処理(3回)

↘ 応用

● 応用

↘

※ 解析、応用が中心

24