

自然言語処理

第10回:

自然言語処理の応用1:

概略 (overview)

David Ramamonjisoa

連絡先、Webページなど

◆ メールアドレス

david@iwate-pu.ac.jp

◆ URLアドレス

<http://p-www.iwate-pu.ac.jp/~david/NLP>

評価方法

- ◆ 100点満点で評価
 - 各回20点×5回
- ◆ 各回の出席
 - 5点(届けあり欠席3点)
- ◆ 各回の課題(発表会)
 - 内容:10点満点で採点
 - 提出:5点

内容(暫定版)

- ◆ 第1回(12月10日): 自然言語処理の応用1
 - 概略(overview)
- ◆ 第2回(12月17日): 自然言語処理の応用2
 - テキスト処理(n-gram, collocation, ...)
- ◆ 第3回(12月24日): 自然言語処理の応用3
 - 言語モデル、辞書、Wordnet、分類
- ◆ 第4回(1月7日): 自然言語処理の応用4
 - 質問応答システム、テキストマイニング
- ◆ 第5回(1月21日): 自然言語処理の応用5
 - 機械翻訳、自然言語理解、

自然言語処理

この会話
正しい文法で
話さないと
難しいから



「こ」=目的詞?
「いなか」動詞
「さー」感嘆符?
???

主語は誰?

理解
できない!!



こ、いなかさー、まい
ちよー、やばくね、



自然言語理解の応用

- ◆ インターフェース、エージェントシステム
 - コンピュータと人間の対話
 - 機械の音声制御(パソコン、機械、車、飛行機など)
- ◆ 文法とスタイルチェック、文書作成支援システム
- ◆ かなから漢字変換、入力予測システム、機械翻訳
- ◆ 質問応答システム、談話や文書の理解
- ◆ 情報検索と抽出: 正解に検索する、単語の意味
- ◆ 文書や物語の自動生成、自動要約、自動批評
- ◆ 電気商品、ゲームなど

自然言語理解のプロセス

音声認識(データ:音声周波数)(信号処理)

入力:周波数スペクトログラム

出力:単語、文字

形態素解析、構文解析(データ:文法)

入力:単語、文字

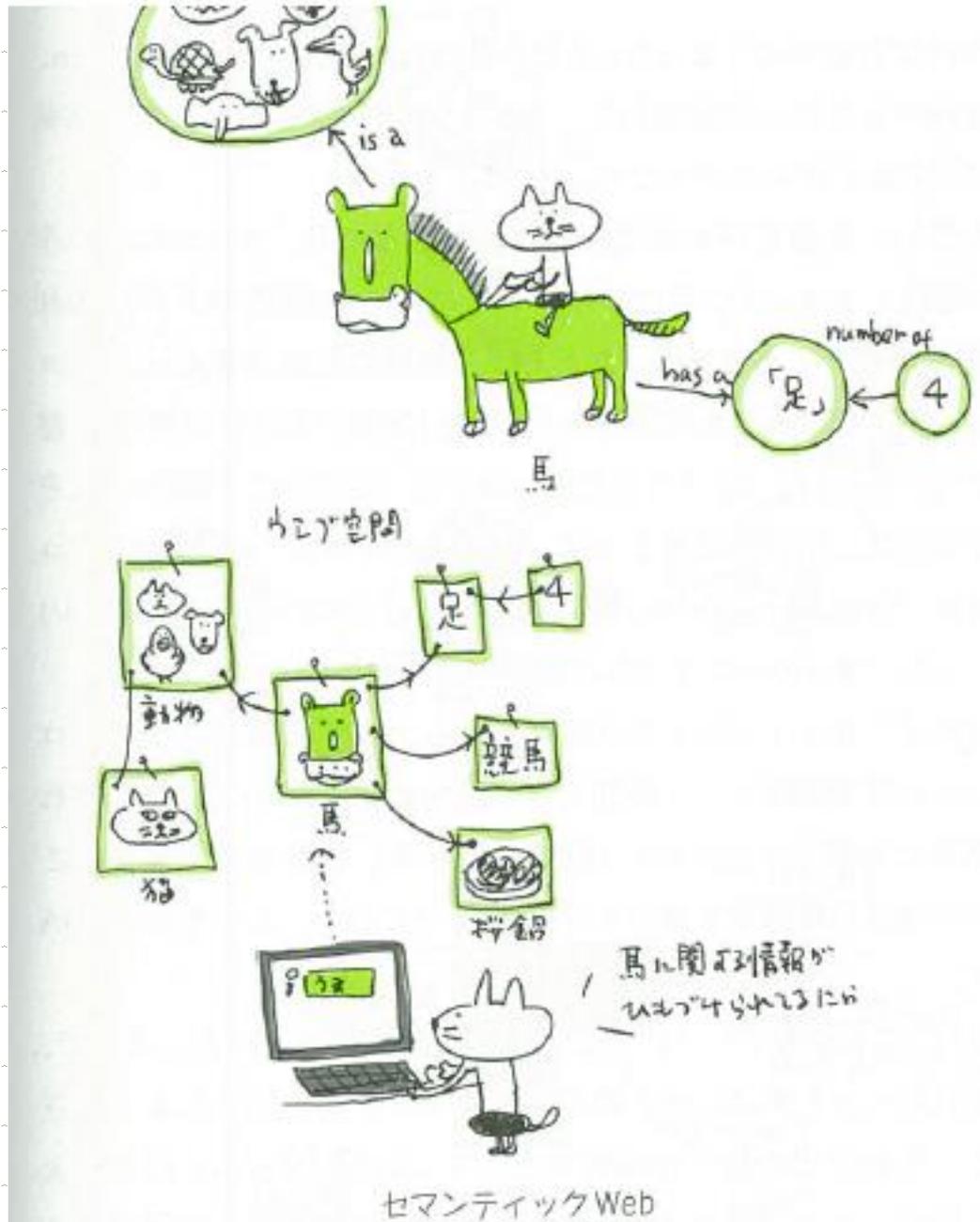
出力:文構造

意味解析、文脈解析(データ:辞書、辞典、オントロジー、文脈、など)

入力:句構造

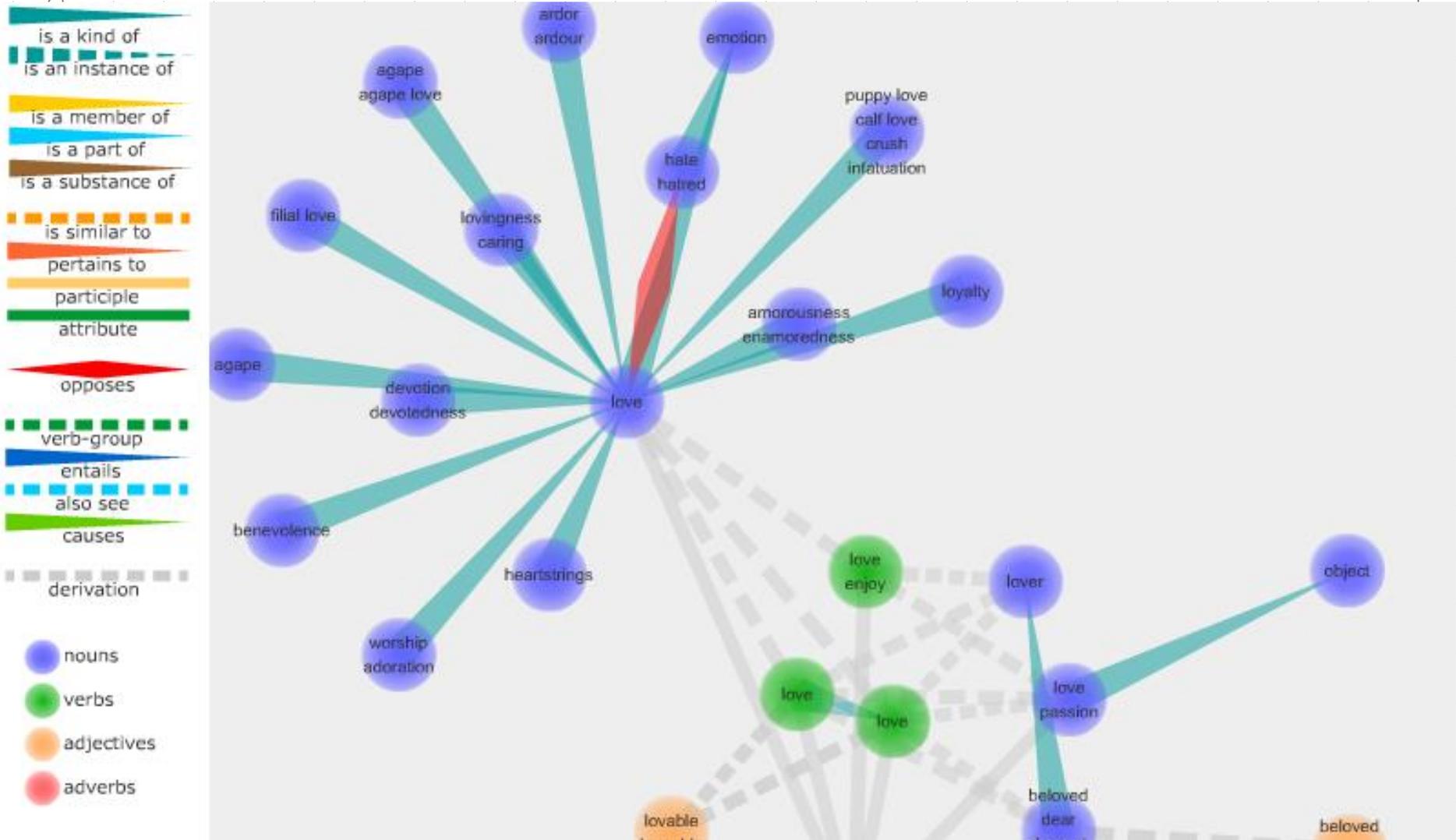
出力:意味素、知識表現

意味ネット



語の意味

◆ 例: love(愛する) - 反対: hate



質問応答システムとチャットボット

- ◆ 質問応答システムは、特定の種類の情報に対する質問をユーザから自然言語で受けつけ、その解答を返すようなコンピュータソフトウェアのこと。解答は文章か、質問に対する直接解答がほとんどである。検索エンジンを用いるものや百科辞典などを用いたり、さまざまなデータベースを基にして検索を行う。英語のサイトでは AskJeeves が有名。
- ◆ 基本的な処理の流れは、以下のようになる。
 - 質問文分析(質問文を単語や節など、検索エンジンのクエリに変換する)
 - 検索エンジン(ここではAND検索など通常のクエリを投げかけることが多い)
 - 絞込み(取得した文章から、解答部分を抽出する)
 - 結果表示(解答をスコア順に並び替え、表示する)

Gooラボの自然文検索

- ◆ アイルトンセナの優勝回数は？41回
- ◆ 標高が8500メートル以上の山は？ 2.エベレスト (8,844.43珠峰最新海拔高程)
- ◆ 東京都で最も長い川は？ 1.荒川 (関東)(173)
- ◆ ドラえもんの身長は？
 - 1. 129.3キロメートル, 2. 129.3センチメートル, ...
- ◆ 初めて人類が月に行ったのはいつですか？
- ◆ 日本の国技って何？
- ◆ 58代横綱は誰？

情報検索の入力文の支援システム

◆ Googleの例（校正支援）

自然言語	
自然言語処理	1,730,000 件
自然言語検索	7,190,000 件
自然言語処理学会	2,140,000 件
自然言語処理研究会	1,280,000 件
自然言語処理 意味解析	725,000 件
自然言語とは	4,620,000 件
自然言語処理 入門	587,000 件
自然言語理解	841,000 件
自然言語検索を使用する	5,050,000 件
自然言語処理 求人	522,000 件
	閉じる

機械翻訳システム

◆ Google 言語ツールの例

テキスト、ウェブページ、ドキュメントを翻訳

テキスト、またはウェブページの URL を入力するか、[ドキュメントをアップロード](#)

The boy went to school.

原文の言語:

英語

翻訳

翻訳する言語:

日本語

英語から日本語に翻訳



[ローマ字表記を表示](#)

少年は学校に通った。



[翻訳を改善する](#)

その他の応用

- ◆ テキストマイニング (Web, Twitter, Blogs, ...)
- ◆ テキストから情報抽出
- ◆ 音声自動認識: 対話システム
- ◆ 物語自動生成
- ◆ 障害者のインターフェース支援システム
- ◆ セマンティックウェブ
- ◆ 文書の自動要約システム
- ◆ オントロジーの研究
- ◆ 情報抽出

自然言語の理解

人工知能の完全版、完璧版



課題:文書の理解

「Webの発展とともに自然言語処理が注目されています。この本では、自然言語処理の一つの応用である“質問応答システム”を紹介します。

人間の質問に答える人工知能は、SF小説や映画に何十年も前から登場していましたが、現実的になってきたのはつい最近です。膨大な文書から適切な文章を探し出す全文検索技術や、人間の言葉を処理できる自然言語処理技術、高度な自然言語処理を低コストで実現機械学習技術などの進歩が必要でした。

自然言語処理とは、簡単に言ってしまおうと、私たちが普段話す言葉をコンピュータに理解させるための技術です。Web検索や機械翻訳、音声アシスタントなどが自然言語処理の応用である。」

その文書について、以下の内容を求める

- ◆ 文の数,文字数,単語の数,
- ◆ 名詞のリスト, 動詞のリスト,形容詞のリスト,固有名詞のリスト
- ◆ 形態素解析Chasenで抽出した語のリストと数
- ◆ 構文解析Cabochaの結果

- ◆ Google翻訳を用いて英語で訳した文
- ◆ 一つ英文の要約