

# メディアアーカイブ分野 [知-13]

(学術情報メディアセンター 大規模テキストアーカイブ研究分野)

教授 森 信介 forest@i.kyoto-u.ac.jp  
http://plata.ar.media.kyoto-u.ac.jp/



古来、人類の知は文書に記録されてきました。本分野では、これを理解し新たな知を記述できるコンピューターの創造を目指しています。そのための基盤として、自然言語の理解とそのために必要となる基礎的な自然言語処理の研究を行っています。加えて、データ分析や未来予測などのコンピューターの思考、あるいは映像や音声などの他のメディアを言葉で説明する自然言語生成について研究を行っています。

## 言語理解・生成

**Procedural text**

- 両手鍋で油を熱する。セロリと青ねぎとニンニクを加え、1分ほど炒める。
- パイヨンと水とマカロニと胡麻を加えて、パスタが柔らかくなるまで煮る。
- 刻んだセージをまぶす。

**Flowgraph [Mori+, LREC14]**

**Cooking robot [Bollini+, ISER13]**

**Smart kitchen [Hashimoto+, IPMU08]**

Cut garlic  
Then add it

**Intelligent search [Yamakata+, Socinfo13]**

自然言語理解は人工知能の究極の目標の1つです。レシピを代表とする手順書をターゲットとして自然言語理解に取り組んでいます。また、(調理)映像などからの手順書(レシピ)の自動生成にも取り組んでいます。

## コンピュータの思考の解説

- 盤面の画像認識
- 深層学習によるシンボルグラウンディング
- 言語表現の自動生成

- 東京大学他と協力

カメラ

ゲームを題材にコンピューターに自身の思考を言葉で説明させる方法の研究をしています。機械学習がよりブラックボックス化する中で、ホットな研究テーマです。証券市場の予測とその解説にも対象を広げています。

## ビッグデータからの言語知識の獲得

- 仮名漢字変換ログ
- テレビ放送

**世界的な統計的仮名漢字変換**

**あいびーえす さいほう**

音声から読みを学習!!

ウェブテキスト、テレビ放送、言語処理ツールの利用ログなどから言語知識を獲得し、言語処理に活用することに取り組んでいます。ツイッターなどの言語解析では最高水準のツールを持っています。

## 言語処理ツールの開発・公開

- 形態素解析 (日本語・中国語)
  - KyTea  
<http://www.phontron.com/kytea/>
- 係り受け解析
  - EDA  
<http://plata.ar.media.kyoto-u.ac.jp/tool/EDA/>
- 固有表現認識
  - PWNER

**4. どんどん研究・開発して公開しましょう!!**

形態素解析(単語分割・品詞推定)や係り受け解析などの基礎的な言語処理について研究し、論文発表に留まらずツールの公開も行っています。他の研究テーマで共通に用いられる重要な研究課題です。