

北陸先端科学技術大学院大学 研究室レポート

JAISTAR

Japan Advanced Institute of Science and Technology, Announcement of Researchers

Information
Science

情報科学研究科
白井 研究室

自然言語処理

—コンピュータを用いて人間の活動を支援する—

科学技術のフロンティアを拓く



北陸先端科学技術大学院大学

北陸先端科学技術大学院大学
情報科学研究科 白井 研究室 ● ● ●

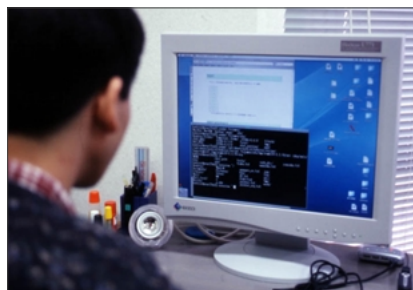
自然言語処理

—コンピュータを用いて人間の活動を支援する—

コンセプトは、コンピュータを用いて
人間の知的な活動を支援すること。

斬新なアイデアで自然言語処理の研究に取り組む若手研究者。

School of
Information Science
Shirai
Laboratory



自然言語処理とは

プログラミング言語などの人工言語に対し、日本語や英語など通常の人間同士のコミュニケーションに用いられる言語が自然言語である。自然言語をコンピュータに処理させるための技術を自然言語処理という。人工知能研究が始まって以来、世界中でもっとも盛んに研究が行われている分野のひとつである。英語を日本語に翻訳するような機械翻訳、人間と会話できるコンピュータなど、多様な切り口で自然言語処理の研究が行われているが、白井清昭教授が目指すのは「コンピュータを用いて人間の活動を支援する」こと。応用志向が強いテーマに取り組んでいる。

特に力を注いでいるのが、自然言語処理技術を応用した人間の知識獲得の支援である。私たちは日常生活の中でさまざまな方法で新しい知識を収集している。たとえば知らない単語があれば辞書を引いて意味を調べるだろう。インターネットを用いて情報を検索する機会も多い。ところがこのように新しい知識を獲得しようとする際、人間は思わぬ苦勞を強いられることがある。こうした問題は自然言語処理の技術を応用することで解決されるという。

日本語学習者向け 読解支援システム

外国語を学ぶ際に誰もが感じるのは、文中の知らない単語を辞書で引いた場合に、その単語がもつ複数の意味の定義文をすべて読み、どれが適当なのかを判断するという作業の煩雑さである。白井助教授は東工大との共同研究で、日本語学習者向けの読解支援システムを構築した。このシステムの特徴は、単語の意味にスコアをつけてその順序で表示することである。たとえば解析したい日本語文「電話をかけながらメモを取る」を入力すると、名詞、動詞といった単語ごとの情報が得られる。ここで動詞“取る”をクリックすると単語の概念が示される。“取る”には複数の意味があるが、ここでは“記録する”という適切な意味が一番に表示される。

どうやって正しい単語の意味を選択するのか。白井助教授は知識源として大量の例文集(コーパス)を用い、単語がどの意味でどんな文脈に現われやすいかといった傾向を自動的に学習させている。もうひとつの知識源は辞書そのものである。たとえば国語辞



日本語読解システム「あすなろ」
<http://hinoki.ryu.titech.ac.jp/>



典に掲載されている単語の用例との類似度を測ることにより、正しい意味を選択できる。このシステムはWEBで公開されていて誰でも使うことができる。日本語学習のみならず、他言語の学習も可能なシステムへの発展が望まれる。

インターネットを用いた知識獲得を支援

ネット上で情報を検索したが、知りたい情報になかなかとど着かなかったという経験は誰にでもあるだろう。白井助教授はWEBアクセス支援という切り口で、自然言語を使った質問応答システムの研究をスタートさせている。

ユニークな点はコンピュータがユーザーとのインタラクティブなやりとりを通して適切な回答を返すという点である。たとえばユーザーが「ワールドカップの優勝国はどこですか?」という質問を入れた場合、コンピュータが「ワールドカップの種目は何ですか?」と聞き返してくる。

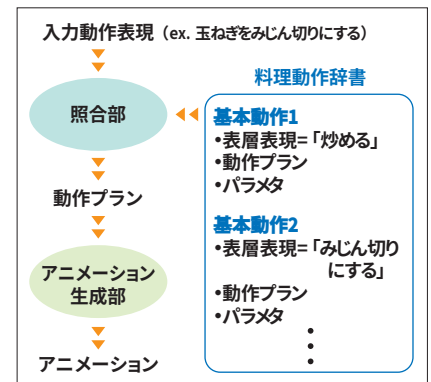
技術的には自然言語の曖昧性が課題となる。この問題をクリアするためには、たとえばワールドカップということばを含む文章をWEB上の記事などから探し出しその前後の文章を自動的に解析させることが必要となる。これによりコンピュータはワールドカップがさまざまな意味を持つことを学習し、その意味を限定するような問い返しをさせれば適切なやりとりができる。

料理のレシピ理解を支援

レシピの理解を支援するシステムというユニークな研究開発にも着手している。料理レシピからアニメーションを生成するという自然言語処理の研究はよく行われているが、そのほとんどはレシピを解析して一連の流れを見せるもので、限られたレシピしか扱えない。これに対し白井助教授が目指すのは、さまざまな料理の“動作”をアニメーションに変換することである。

自然言語処理の技術が活かされるのは、動作を表す表現の多様性への対応である。たとえば“みじん切り”“みじんにする”は同じ動作を指す。コンピュータに表現のバリエーション、パターンを吸収させることがポイントとなる。

純粹な興味から自然言語処理という分野に足を踏み入れたという白井助教授。その興味は今も尽きない。東京工業大学の助手を経て2001年にJAISTに着任したばかりの若手研究者。その斬新な発想に注目が集まっている。



アニメーション生成までの流れ



Research Interests

北陸先端科学技術大学院大学 情報科学研究科

助教授 白井清昭

Associate Professor Kiyooki Shirai

白井研究室 情報科学棟III TEL:0761-51-1216 FAX:0761-51-1149
http://www.jaist.ac.jp/nlp/lab/index-jp.html
E-mail kshirai@jaist.ac.jp

主な研究テーマ

1) 日本語学習者のための読解支援システム

日本語学習者は、日本語のテキストを読むとき、意味のわからない単語があれば辞書を引く。このとき、1つの単語は複数(場合によっては数十個)の意味を持つことがあるが、その全ての意味の定義文をいちいち読むのは大変である。そこで、単語がテキスト中でどの意味で使われているかを判断し、その意味の定義文だけを提示すれば、日本語学習者への利便性が増す。我々は、テキスト中の単語に対して辞書を引き、そのテキスト内で使われている意味に近い順序で辞書の定義文を提示することにより、日本語学習者の読解支援を助けるシステムの開発に取り組んでいる。具体的には、大量の例文集(コーパス)を用い、単語のどの意味がどのような文脈に現われやすいかといった傾向を自動的に学習する。

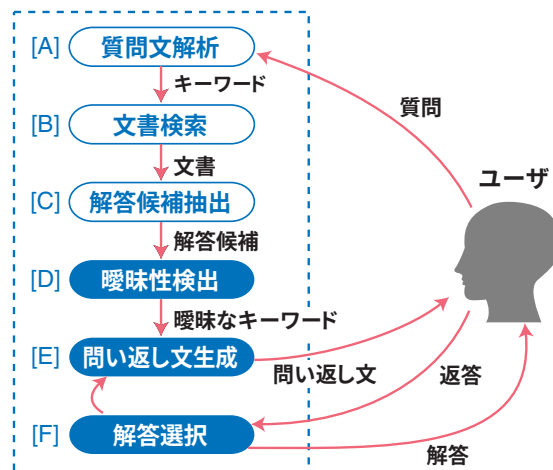


読解支援システムのスナップショット(東工大との共同研究)

2) 情報探索の支援

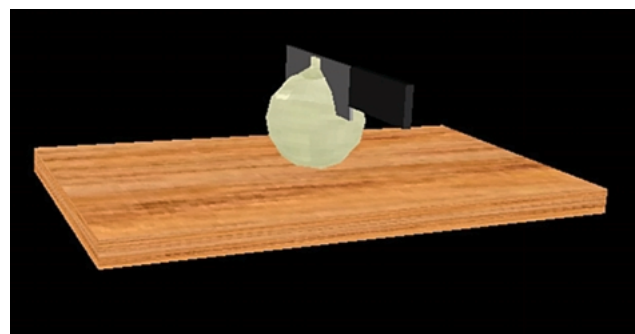
ユーザとの対話に基づく質問応答システム

情報化社会の進展に伴い、WWWや新聞記事といった大量のテキストデータベースが利用可能になり、これらから必要な情報を効率的に探索したいという要求も大きくなりつつある。このような要求に対し、我々は、質問文に対する解答をテキストデータベースから検索し、ユーザとの対話を通じて適切な解答を選択する質問応答システムの構築を目指している。例えば、「ワールドカップの優勝国はどこですか」といった質問が入力されたとする。ところが、ワールドカップにはサッカー、ラグビーといった様々な種類があり、それに応じて解答が異なる。そこで、システムからユーザに対して「どのスポーツのワールドカップですか」といった問い返しを行い、ユーザの返答によって適切な解答を選択し、提示する。



3) アニメーション生成による料理レシピの読解支援

料理レシピには、「かつらむき」「板ずりする」など、様々な動作表現が使われている。しかし、料理になじみのない人にとっては、調理動作自体がよくわからない場合がある。我々は、このような動作表現に対応するアニメーションを自動的に生成することにより、人間の料理レシピの理解を支援する研究に取り組んでいる。料理レシピをアニメーションに変換するためには様々な要素技術を必要とするが、我々は動作表現の多様性をどのように取り扱うかという問題に特に積極的に取り組んでいる。例えば、「みじん切り」は「みじんに切る」「細かく刻む」など、様々な表現で表わされる。このような動作表現の同一性を認識し、多様な動作表現に対して適切なアニメーションを生成することを目指している。



生成したアニメーションの例(たまねぎのみじん切り)

Faculty Profile



北陸先端科学技術大学院大学 情報科学研究科

助教授 白井清昭

Associate Professor Kiyooki Shirai

<学位>

東京工業大学工学士(1993)、東京工業大学工学修士(1995)、
東京工業大学工学博士(1998)

<職歴>

東京工業大学大学院情報理工学研究科助手(1998)
北陸先端科学技術大学院大学情報科学研究科助教授 (2001-)

白井研究室 情報科学棟III TEL:0761-51-1216 FAX:0761-51-1149
http://www.jaist.ac.jp/nlp/lab/index-jp.html E-mail kshirai@jaist.ac.jp

専門

自然言語処理, 機械学習

主な研究課題

自然言語処理用知識の自動獲得

高度な自然言語処理を行うためには、辞書やシソーラスなど、様々なタイプの自然言語処理用知識が必要となる。このような知識をコーパスや機械可読辞書などから自動的に獲得する。また、既存の自然言語処理用知識に対して、規模を拡大したり新たな情報を付与することにより、これらをより豊かにすることも試みる。

- (1) 機械可読辞書からの上位・下位関係の獲得とその応用
- (2) コーパス・辞書を利用した既存シソーラスの拡張
- (3) コーパスからの文法獲得

自然言語処理技術を用いたWWWアクセス支援

World Wide Webは様々な情報を持っている反面、人間が知りたい情報にたどりつくまでに多大な労力を要することも多い。そこで、我々は自然言語処理技術を応用して人間のWWWへのアクセスを支援することを目指している。

- (1) ポータルサイトの自動生成
- (2) ウェブを対象としたイベント追跡システム
- (3) 対話的質問応答システム

統計的自然言語解析

複数の解析結果の候補の中から正しい解を選択する曖昧性解消問題は、自然言語解析における重要な問題である。大量のテキストから学習された様々なタイプの統計情報を用いて、この問題を解決する手法を探る。

- (1) 構文解析
- (2) 語義曖昧性解消

主な著書・論文・講演発表

- Learning a Robust Word Sense Disambiguation Model using Hypernyms in Definition Sentences, Kiyooki Shirai, Tsunekazu Yagi, 20th International Conference on Computational Linguistics, 917-923, 2004.8.23
- SENSEVAL-2 日本語辞書タスク, 白井清昭, 自然言語処理, Vol. 10, No. 3, pp. 3-24, 2003/04
- Word Sense Disambiguation using Heterogeneous Language Resources, Kiyooki Shirai, Takayuki Tamagaki, First International Joint Conference on Natural Language Processing, 614-619, 2004/3/22
- Construction of a Word Sense Tagged Corpus for SENSEVAL-2 Japanese Dictionary Task, Kiyooki Shirai, Proceedings of the Third International Conference on Language Resources and Evaluation, pp. 605-608, 2002/05
- Decision Lists for Determining Adjective Dependency in Japanese, Taiichi Hashimoto, Kosuke Nishidate, Kiyooki Shirai, Takenobu Tokunaga, Hozumi Tanaka, Machine Translation Summit VIII, 151-156, 2001/9/18
- Integration of heterogeneous language resources: A monolingual dictionary and a thesaurus, Takenobu Tokunaga, Yasuhiro Syotu, Hozumi Tanaka, Kiyooki Shirai, 6th Natural Language Processing Pacific Rim Symposium, 135-142, 2001/11/27
- 自然言語解析のためのMSLRパーザ・ツールキット, 白井清昭, 植木正裕, 橋本泰一, 徳永健伸, 田中 穂積, 自然言語処理, Vol. 7, No. 5, pp. 93-112, 2000/11
- An Empirical Evaluation on Statistical Parsing of Japanese Sentences using Lexical Association Statistics, Kiyooki Shirai, Kentaro Inui, Takenobu Tokunaga, Hozumi Tanaka, the 3rd Conference on Empirical Methods in Natural Language Processing, pp. 80-87, 1998/
- Automatic Extraction of Japanese Grammar from a Bracketed Corpus

Kiyooki Shirai, Takenobu Tokunaga, Hozumi Tanaka, Natural Language Processing Pacific Rim Symposium, pp. 211-216, 1995/12

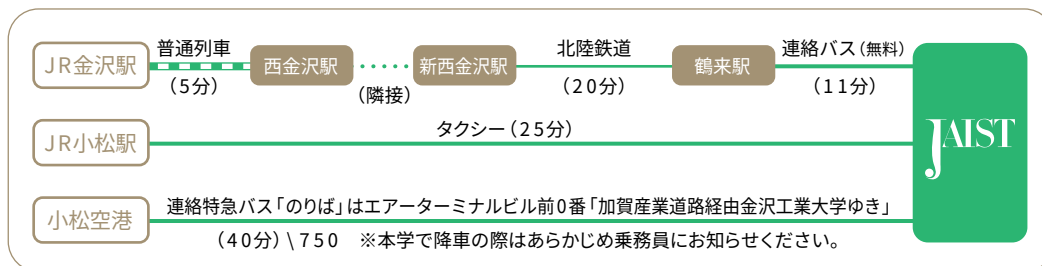
- 統計的構文解析における構文的統計情報と語彙的統計情報の統合について, 白井清昭, 乾健太郎, 徳永健伸, 田中穂積, 自然言語処理, Vol. 5, No. 3, pp. 85-106, 1998/07
- 括弧付きコーパスからの日本語確率文脈自由文法の自動抽出, 白井清昭, 徳永健伸, 田中穂積, 自然言語処理, Vol. 4, No. 1, pp. 125-146, 1997/01
- SENSEVAL-2 Japanese Dictionary Task, Kiyooki Shirai, SENSEVAL-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems, 2001/7/5
- 意味辞書を利用するための形態素変換規則の自動獲得, 森田勝, 白井清昭, 第9回言語処理学会年次大会, 横浜国立大学, 2003/3/18
- 読解支援システムのための語義曖昧性解消に関する研究, 玉垣隆幸, 白井清昭, 第9回言語処理学会年次大会, 横浜国立大学, 2003/3/18
- 辞書定義文を用いた低頻度語のための語義曖昧性解消モデルの学習, 白井清昭, 八木恒和, 情報処理学会自然言語処理研究会 SIGNL-158-20, 通信総合研究所, 2003/11/6
- ポータルサイト自動作成の試み, 白井清昭, 菅井俊介, 平野健児, 星正人, 第10回言語処理学会年次大会, 東京工業大学, 2004/3/16
- コーパスと辞書定義文中の上位概念を用いた頑健な語義曖昧性解消, 白井清昭, 八木恒和, 第10回言語処理学会年次大会, 東京工業大学, 2004/3/16
- アニメーション生成のための料理動作辞書の構築, 白井清昭, 大川 寛志, 情報処理学会自然言語処理研究会, 広島市立大学, 2004/11/4
- 対話型質問応答システムにおける問い返し文生成に関する基礎研究, 白井清昭, 徳江英範, 情報処理学会自然言語処理研究会, NHK放送技術研究所, 2005/1/11

国際貢献・国内貢献等

- 言語処理学会, 会員, 1998 -
- 情報処理学会, 会員, 1998 -
- 人工知能学会, 会員, 1998 -
- The Association for Computational Linguistics, 会員, 2003 - (社) 電子情報技術産業協会, ヒューマンインターフェース技術関係委員会委員, (2002-)
- 言語処理学会 査読委員 (1999-)
- 情報処理学会「自然言語処理研究会」連絡委員 (1999-2001)
- 情報処理学会「自然言語処理研究会」幹事 (2002-)
- 言語処理学会 評議委員 (2000-2003)
- 情報処理学会 学会誌編集委員 書評・ニュース分野担当 (2000-2003)
- 言語処理学会 第7回年次大会 プログラム委員 (2001)
- 言語処理学会 第9回年次大会 プログラム委員 (2003)
- 言語処理学会 第10回年次大会 プログラム委員 (2004)

共同研究の希望テーマ

- WWWを対象とした自然言語処理
- 説明文書からのアニメーション生成



JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

国立大学法人 **北陸先端科学技術大学院大学**

〒923-1292 石川県能美市旭台1-1 TEL:0761-51-1111 (代表)
<http://www.jaist.ac.jp/index-jp.html>

産学連携に関するお問い合わせ

先端科学技術研究調査センター

TEL 0761-51-1070 FAX 0761-51-1944

E-mail ricenter@jaist.ac.jp HPアドレス <http://www.jaist.ac.jp/ricenter/index.html>

学術協力課 連携推進室 産学連携係

TEL 0761-51-1906 FAX 0761-51-1919 E-mail renkei@jaist.ac.jp

白井研究室

TEL:0761-51-1216 FAX:0761-51-1149

<http://www.jaist.ac.jp/nlp/lab/index-jp.html> E-mail kshirai@jaist.ac.jp