

# ことばを計算機で処理する技術と その応用システムの開発

奥村 学 研究室

- 専門分野：自然言語処理、テキストマイニング、Webテキスト処理
- Home Page： <http://www.lr.pi.fitech.ac.jp/>



## 研究目的

奥村研究室では、ことばを計算機で処理する技術に関する研究と、その技術を用いた応用システムの開発を行なう。ことばの理解というテーマでは、これまで研究が難しいとされている、意味、文脈理解を中心に行なっていきたい。それと同時に、世の中で役に立ちそうなシステムの開発も行なっていきたい。具体的には、テキストの自動要約、テキストをわかり易く「言い替える」、テキストを他のメディア（音声、アニメーション）へ変換して提示するなど、テキストを「わかりやすく」提示する技術、WWW上の情報に対するテキストマイニング、障害者のコミュニケーション支援のためのシステムを開発していきたい。

## 研究テーマ

### 1. 人間の言語理解のモデルを目指して（頑健な自然言語の意味、文脈解析）

これまでの研究成果により、自然言語の解析技術のうち、形態素、構文解析がある程度の精度で、分野を限らないテキストに対して実行できるようになってきている。だが、その先の解析技術である、意味、文脈解析はその水準に達しているとは言えない。そこで、テキスト中の単語の語義の決定、照応/省略の解消、テキストの構造の決定などの意味/文脈解析の問題について個別に研究し、頑健な意味、文脈解析技術の構築を目指す。

### 2. 機械学習、統計的手法に基づいた自然言語処理

言語処理の手法としては、現在WWW上などに大量の言語データ（コーパス）が蓄積されるようになってきたことから、それらのコーパスを情報源として、統計的手法あるいは機械学習手法を利用することで、言語処理用知識を得て、その知識を利用する立場を採って、言語処理手法を開発している。たとえば、今後高度な言語処理を行なうためには、我々人間が持っている常識的知識を計算機も利用して言語処理を行なう必要がある。この常識的知識をコーパスから自動獲得する研究を行なっている。

### 3. WWW上のテキストデータのテキストマイニング

インターネットの普及にともない、Web上で多くの人々情報発信するようになってきている。このため、発信された人々の情報を収集し、それを利用したいという要求が高まっている。特に、Web上のテキストから社会の動向、意見を分析する手法が現在非常に注目を集めている。そこで、Weblog (blog) を収集、監視、マイニングするシステム blogWatcher を開発するとともに、Web上のテキストから意見を収集、分析するシステムの開発を行なっている（次ページの図は、blogWatcher の出力イメージ）。

### 4. テキスト情報の提示技術

近年電子化されたテキストが世の中に満ち溢れ、情報洪水という言葉が使われるようになってからかなり

The screenshot displays a search engine interface for 'ipod' on the bW3 platform. The main content area shows search results with titles, dates, and snippets. The sidebar on the right includes filters for 'PSP', 'WALKMAN', and 'ウォークマン', a 'もしかして?' section with 'ipod nano (236件)', and a 'バースト for ipod' section with a bar chart showing search volume over time. Below the chart is a '評判情報 for ipod' section with a table showing the ratio of positive and negative reviews, and a gender breakdown table.

ポジティブ・ネガティブな評判の割合	
ポジティブ	68%
ネガティブ	31%

  

男性・女性による評判の割合	
女性	40%
男性	59%

の歳月を経ている。このような大量のテキストから必要なテキストを検索する技術と同時に、(検索された)テキスト集合から人間が情報を効率的に入手できるよう、「わかりやすく」テキスト集合からの情報を提示する技術が求められている。テキストの要約、マルチメディア(音声、アニメーションなど)を用いたテキストの提示など、「わかりやすい」テキスト情報の提示技術を探求する。

## ● 教員からのメッセージ

2000年4月に精密工学研究所に着任して、早いもので13年が経ちました。研究室は、助教1名、博士課程の学生さん4名、修士課程の学生さん10名という、にぎやかな状態が定常になりつつあります。「ことば」の処理に興味がある、意欲のある学生さんが来て下さることを期待しています。各自がやりたいこと(夢)を目指して、楽しく一緒に研究しましょう。

## ● 参考文献・図書

1. 自然言語処理の基礎, 奥村 学, コロナ社, 2010.
2. 奥村 学, 白井清昭, 代表性のあるコーパスを利用した日本語意味解析, 人工知能学会誌, Vol.24, No.5, 2009.
3. 奥村 学, プログマイニング技術の最新動向, 電子情報通信学会誌, Vol.91, No.12, pp.1054-1059, 2008.