形態素解析を用いたトレンド情報の抽出と可視化システムの構築

日大生産工 (院) ○寺内貴洋 日大生産工 吉田典正
日大生産工 齋藤敏雄

1. はじめに

世界中に存在する様々な問題に対し、日々その解決策を見出す研究が行われ、その成果をまとめた論文が次々に発表されている。これらの研究は、時代背景の移り変わりに影響を受けるものであり、言い換えれば、過去の研究にはその時代のトレンドとなる“知識”や“技術”、“問題定義”などが多く含まれていることになる。

しかし、1つの分野に焦点を当てた場合でも、非常に多くの論文が存在しており、それだけ情報量も膨大であるといえる。研究者達は長い研究の中で感覚的にこれらの情報と推移を把握していると考えられるが、全てを把握することは困難である。

本研究では、上記の“知識”や“技術”、“問題定義”などとして読み取れるワードを『トレンド情報』と呼称し、過去数年の特定分野の論文を集積し、それらのトレンド情報を示すワードを検出し、論文の相関関係を可視化するためのシステムの開発を行った。その狙いは、ユーザーの研究分野における、過去のトレンド情報の推移の把握、及び今後のトレンドの予測・推定を行うことで、現在取り組んでいる研究の位置付けを明らかにし、さらに今後の課題の設定を支援することにある。

2.トレンド情報の抽出

2.1 対象情報

本研究では、特定分野として経営工学に関わる研究に焦点を当て、その中から「経営情

報学会」の論文を数年分収集した。近年は情報のデジタル化技術の進展とインターネットの爆発的な発展に伴い、論文もまたネットワーク上でデータベース化され、研究者であれば容易に閲覧できる環境が整っている。そのため、本研究では、論文や図書・雑誌などの学術情報を検索できるデータベース・サービス『CiNii』¹⁾を用いて収集を行なった。

2.2 抽出方法

本研究では、トレンド情報の検出手法として形態素解析エンジン『MeCab』²⁾を使用する。形態素解析とは、文を単語に区切り、品詞を同定する処理である。例として、経営情報学会の論文から「多様性の時代を生き抜くラーニング・リーダーシップの提唱」というタイトルを形態素解析にかけた結果を図1に示す。各品詞が解析され、その情報が表示される。また事前に専門用語を辞書として記憶させることもでき、辞書の参照により意味のある語句で区切られ、より精度を上げて解析が行われる。これにより、論文のトレンド情報と捉えられる語句を選別して抽出する。

本研究では、単純に「名詞」と「一般」の

```
多様性の時代を生き抜くラーニング・リーダーシップの提唱
多様 名詞,形容動詞語幹,***,多様,タヨウ,タヨ
性 名詞,接尾,一般,**,性,セイ,セイ
の 助詞,連体化,***,の,ノ,ノ
時代 名詞,一般,***,時代,ジタイ,ジタイ
を 助詞,格助詞,一般,**,を,ヲ,ヲ
生き抜く 動詞,自立,**,五段・カ行イ音便,基本形,生き抜く,イキヌク,イキヌク

ラーニング・リーダーシップ 名詞,一般,***,
の 助詞,連体化,***,の,ノ,ノ
提唱 名詞,サ変接続,***,提唱,テイショウ,テイショー
EOS
```

図1 MeCabの解析

Trend information extraction using morphological analysis
and building its visualization system

Takahiro TERAUCHI, Norimasa YOSHIDA, Toshio SAITO