

演題のテキストマイニング分析から見た 医学情報サービス研究大会

西村 飛俊

愛知淑徳大学大学院文学研究科図書館情報学専攻

1. 研究目的

近年、テキストマイニングという技術が注目されるようになった。これは多くのテキストデータをあるアルゴリズムを用い特殊な形式に書き出すことで、新しい情報を得るということを目的とした技術である。自由記述した文章などに適用することが出来るが、これのある書誌の集合に適用することも可能である。そこで、医学情報サービス研究大会の発表演題の集合に対してテキストマイニングを行い、この大会の特徴の把握をすることを目的に実験を行った。

2. 研究方法

医学情報サービス研究大会の20年分の論題群を対象とし、テキストマイニングソフト“Polaris”を用いて解析した。“KeyGraph”はPolarisが出力するテキストマイニング結果を独自のアルゴリズムによって図示したものである。また、解析にあたって形態素解析ソフト“MeCab”を使用し、形態素解析を行っている。

論題群の解析は3年分ごとに行った。ただし19回～20回は2年分の解析になっている。

3. 結果と考察

下図が出力結果の一例である。3年ごとに出力することにより、論題の中で関心が集まった主題の経年変化を捉えることが出来た。特に新技術に関しては登場とともに論題に現れている場合もあり、CD-ROM媒体が徐々に関心をなくし、ネットワーク検索の時代へと移行していく様が見てとれた。また会の専門性が次第に整えられ、「患者サービス」と「EBM」(Evidence Based Medicine 根拠に基づいた医療)が中心的な話題になっていったこともわかった。

4. 結論

会に参加している人ならば経験的に感じ取れるだろう事柄を、テキストマイニングを用いて理論的にとらえることに成功した。ただし、Polarisから得られるキーワード群から他の連想をする人もいることは確実である。

今回の発表では、各回のマイニング結果を掲載する。大会の当時を思い出してもらったり、過去の大会のキーワードを参照して過去を知ってもらったりするのに使われたら幸いである。

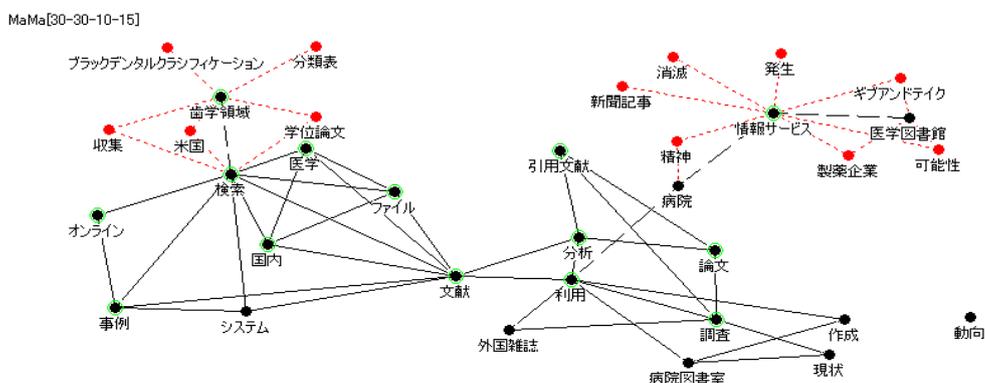


図 KeyGraph を用いたテキストマイニングの一例