

病理情報のデータベース化

－会議録を活用した病理学的情報の集約－

中西陽子¹⁾，佐藤 永²⁾，原 一貴²⁾

¹⁾日本大学医学部病態病理学系腫瘍病理学分野

²⁾筑波大学情報学群知識情報・図書館情報学類

【はじめに】医学分野の中で病理学は，臨床医学と基礎生命科学の接点にあり，ヒトの病気について，臓器から細胞レベルまでの形態の変化や差異に重点を置いて研究する形態学を基盤とした学問分野である。これまでに蓄積された成果に基づいた知見により病理診断が行われている。しかし近年，分子生物学の飛躍的な進歩により癌の治療薬の開発が目覚ましいことから，治療効果が期待できる患者を選定するための病理学的検査分野の早急な情報整備が必要となってきた。そこで本研究では，疾患ごとに異なり，加速度的に更新される診断，治療に関連する情報を速やかに集約するため，会議録を利用したファクトデータベースの構築と，個別化医療への支援を目指した二次利用の有用性について検討することを目的とした。

【方法】2000年から2015年に開催された日本病理学会総会会議録約22,400件より，発表数が多い乳癌研究ならびに肺癌研究に関する会議録計624件を対象とし，会議録本文より抽出した単語を対象データとした。対象会議録はScanSnapSV600（富士通）を用いてテキスト化した。単語の抽出と単語間の関係解析はSocial Insight（ユーザーローカル）を用いて行った。

【結果】日本病理学会総会における肺癌研究，乳癌研究の発表数はそれぞれ10年で1.2倍，1.6倍，15年で2.0倍，2.6倍と増加傾向であり，抽出された単語数は，肺癌計119,969語，乳癌計59,242語，抄録1件あたりの単語数は肺癌423語，乳癌363語であった。ワードクラウドによる単語の出現頻度解析では，各年度のtopicsが癌の種類によって異なる状況が浮き出された。また，出現頻度の高い単語は，解析年度毎の病理診断や癌治療に必須の分子や状態を示していた。次に，共起ネットワークにより，出現頻度の高い単語間の関係を解析した結果から，各年度での主流の概念が示されたとともに，周辺に出現する新たな概念が，癌種別，年度別に見出された。

【考察】本研究では，時事的な速報性に富みながらも一過性となっている会議録情報を電子化して蓄積し，これまでの画像情報ではなく，文字情報の解析を試みた。この結果，年度毎の疾患の病理診断や治療を取り巻く重要事項が特徴的に捉えられ，継続的に更新可能となることが期待された。今後さらに，抽出された分子や薬剤および疾患名などについての各種情報の関連付けをより具体的に行い，実用性の向上を図る必要があると考える。

【謝辞】本研究は，平成24年度～平成26年度科学研究費補助金（課題番号26540171）により実施した。