

XML 文書としての電子カルテ検索システム

関 隆宏[†] 安元 裕司^{††} 和多 太樹^{††} 中島 直樹[‡] 廣川 佐千男^{‡‡}[†]九州大学大学評価情報室 ^{††}九州大学大学院システム情報科学府 [‡]九州大学病院医療情報部 ^{‡‡}九州大学
情報基盤センター

1 はじめに

政府は e-Japan 戦略の一環として「平成 18 年には電子カルテ普及率を 60 %に」という努力目標を掲げ医療の IT 化を進めている。病院内、病院間、あるいは医師、看護師、患者間のコミュニケーションの向上が進むことが期待される。様々な項目から構成されるカルテは、半構造化データの典型であり、その検索と分析手法の開発は、国民に対する医療の高度化のためのインフラと位置付けられる。今後蓄積され続ける電子カルテを、情報共有のためだけでなく、科学的根拠に基づく医療 (EBM Evidence based Medicine) の基礎データとして利用し、データに基づく科学的発見を支援する検索・分析システムが必要とされる。医師は、患者の症状と検査結果のデータに応じて、原因と対応を推定する。場合によっては、仮説として複数の原因を考え、それぞれに対する再検査項目を考慮する。このような高度の判断をするのが医師の最も重要な職務であるが、それは各々の医師の経験と知識に依存する。それらの知識は従来、書籍や論文として蓄積され長期間の教育により伝達されてきた。本研究では、このような知識だけでなく、蓄積されるカルテ・データを、医師の分析に役立てるシステムを作ることを目指す。

本研究では、医療カルテを対象として、それぞれの項目を検索の観点として捉える多面的検索方式を提案する。各文書のそれぞれの項目を独立した文書として捉え、個別インデックスを構成することにより、項目毎のクラスタリングを実現する。利用者は、複数の項目から任意の二項目を選択することにより、2つの観点をういたクラスタリング結果を見ることができる。さらに、各項目のクラスタから抽出した特徴的キーワード

を2つの軸に表示することにより、単純な可視化ではなく、可視化画面においてそれぞれのセルを二種類のキーワードで解釈することができる。このキーワードは、検索結果全体の概観を与えるだけでなく、検索絞込みにも利用でき、対話的な検索の効率が向上できる。具体的なシステムとして、著者の中島が所属する九大附属病院外科の341件の退院カルテを対象として検索システムを実現した。これにより、クラスタリング方式、キーワード抽出法、分類結果の分析を情報検索と医療という二つの観点からの分析が可能となった。

2 関連研究

医療カルテの電子化についてはデータベースや検索についての既存の技術を応用し、様々な企業によりビジネスレベルで進んでいる。研究動向の中心は、電子カルテの標準化と現場での入力インターフェース向上 [1, 2] やネットワーク化における安全性向上 [3, 4] の2点にある。このような標準化が達成されると、複数の機関によるデータ共有が可能となり、本稿で提案する分析システムの活用が期待できる。

本稿で利用する多面的検索方式は、著者の廣川等が九州大学教員データベースの開発において着想を得たものであり [5] これを電子カルテに適用するものである。カルテの解析については、科学技術振興機構 (JST) による「安全性に係わる社会問題の解決のための知識体系の構築」プロジェクトがあり、医療安全研究グループ (代表永井良三教授) による、数値的データについての統計解析機能をもつ電子カルテシステムの開発もある。しかし、本研究のアプローチは、数値的データだけでなく、カルテの項目に含まれる文書情報を半構造化文書として捉えるものである。

インデックス付と XML 検索の統合に関する研究としては、Raghav らの構造索引と転置リストとを統合して索引を作成する方法 [6]、絹谷らの XML 文書から利用者の問い合わせに最適な部分文書を、文書構造と文書内容の両者を利用して取り出す方法 [7]、森らのタイトルや見出し、強調語に着目して特徴量ベクトルを改善し、XML 文書検索システムの精度を向上させる方法 [8] などがある。しかしこれらはいずれも XML の個別の部分構造に着目する多面的検索ではない。本研究では XML 文書の各項目毎にインデックスを作成

Multipronged Retrieval System for Care Card as XML Documents

[†] Takahiro SEKI (tsekiuoc@mbox.nc.kyushu-u.ac.jp)

^{††} Yuji YASUMOTO (y-yasu@i.kyushu-u.ac.jp)

^{††} Taiki WADA (t-wada@i.kyushu-u.ac.jp)

[‡] Naoki NAKASHIMA (nnaoki@info.med.kyushu-u.ac.jp)

^{‡‡} Sachio HIROKAWA (hirokawa@cc.kyushu-u.ac.jp)

Office for Information of University Evaluation, Kyushu University ([†])

Graduate School of Information Science and Electrical Engineering, Kyushu University (^{††})

Department of Medical Informatics, Kyushu University Hospital ([‡])

Computing and Communications Center, Kyushu University (^{‡‡})

し、多面的な検索を行うシステムを扱う。

3 システム概要

本研究では、汎用連想計算エンジン (GETA)[†]を利用している。検索の対象としたい各項目に対し、WAMと呼ばれる GETA に固有のインデックスを作成する。ユーザーはウェブ上のフォームから検索したい語を入力し、どの項目で検索したいのかを選択する。ここで、選択された項目に入力された語を含む文書のリストが返され、このリストをユーザーが任意に選ぶ2つの項目で、それぞれ指定した個数にクラスタリングを行い、片方のクラスタで*i*番、もう一方のクラスタで*j*番に属するような文書が (*i*, *j*) 要素となるようなマトリックスを作成する。また、検索結果の絞込みを容易にするため、各クラスタごとに特徴語を抽出する。ユーザーはあるセルに含まれる文書が多い場合、そのセルに含まれる文書で再度マトリックスを作成するズームングを行うことにより、検索結果の絞込みをする。

今回対象とした XML 文書は「主訴」、「現病歴」、「既往歴」、「家族歴」、「入院時現症」、「入院時検査所見」、「入院後経過」、「退院時処方」、「特記事項連絡事項」、「生活歴」、「考察」、「退院後の方針」がフラットに並ぶ構造を持つ。このそれぞれの項目に対して形態素解析を行い、インデックスを作成した。

図1は文書中に「ポリープ」を含む文書を「既往歴」、「考察」で3×3に分割した例である。

既往歴	虫垂炎, 摘出, 平成, ポリープ, 椎, 多発, 緑内障, 頸, 筋炎	3	1	3	7
	切除, 大腸, ベ, ク, トミー, ポリ, 手術, 施行, 術後	4	4	1	9
	肺炎, 事項, 特記, basedow, disease, 才代, 早期, じんましん, 花粉	0	3	1	4
	総数	7	8	5	20
		臍, 腫瘍, 分泌, インスリン, 粘膜, 病変, 嚢胞, 早期, 症状, 診断		療法, 肝, 予後, 胆, 例, 治癒, 手術, 治療, 移植, 死亡	
		考察		総数	

図 1: 検索結果のマトリックス表示

4 まとめと今後の課題

本研究では XML 形式の電子カルテの各項目ごとにインデックスを作成し、ユーザーが選ぶ任意の観点で検索を行い、任意に選ぶ2つの観点で結果をマトリッ

クス表示することにより多面的な検索、分析を行うシステムを作成した。本システムの定性的および定量的な評価を行うことが今後の課題である。また、クラスタリング方法の検討などをする必要がある。

参考文献

- [1] 中島直樹, 田中直美, 関口直孝, 小林邦久, 井口登, 与志, 名和田新, 野瀬善明, 電子カルテ定型語彙集の有用性についての検討—入力 of 簡便化と診療の標準化, 医療情報学, Vol.22 (suppl. 1), p. 535, 2002
- [2] 田中直美, 中島直樹, 高野香子, 宮崎美穂, 井口登与志, 名和田新, 野瀬善明, 電子カルテ診療における患者が理解しやすいヒューマンインターフェイスの開発, 医療情報学, Vol 22 (suppl. 1) : p.362, 2002
- [3] 清本 晋作, 田中 俊昭, 三宅 優, 三田村 好矩, 医療情報流通プラットフォームにおけるセキュリティ機構の開発, 電子情報通信学会論文誌, Vol.J88-D1, No.2pp.378-390, 2005
- [4] 中島直樹, 坂本憲広, 三村和郎, 梅田文夫, 井口登与志, 名和田新, 公開鍵基盤を用いた広域分散型糖尿病電子カルテ開発事業について, 糖尿病ケア IT 革命, 別冊 80-85, 2002, 2002.
- [5] 廣川佐千男, 関隆宏, 安元裕司, 山田泰寛, 教員データに対する多面的検索システム, 情報処理学会研究報告 2005-DBS-137(II) pp.665-672, 2005
人工知能学会研究会資料 SIG-KBS-A501, pp.1-4, 2005
- [6] R.Kaushik, R.Krishnamurthy, J.F.Naughton, R.Ramakrishnan. On the integration of structure indexed and inverted lists, *SIGMOD 2004*, (2004).
- [7] 絹谷弘子, 波多野賢治, 吉川正俊, 植村俊亮. XML 文書の文書構造と内容を用いた部分文書の抽出手法, トランザクション「データベース」 Vol.43, No.SIG02, (2002).
- [8] 森康弘, 吉川正俊, 波多野賢治. 特徴ベクトルの要素間の影響を考慮した XML 文書検索手法の提案, DEWS2004, (2004).

[†] <http://geta.ex.nii.ac.jp/>