Vol.2017-CH-114 No.9
2017/5/13

デジタルアーカイブの利活用を促進する情報検索技術の研究を通して感じた課題

未代誠仁[†1]

概要：本稿では，デジタルアーカイブの利活用促進を目的とした情報検索技術の研究を行ってきた筆者が，その過程で感じたいくつかの課題を取り上げ，それらに対する所見を述べる．筆者の研究を含めて，人文科学とコンピュータ研究会における様々な研究はフロンティア領域として扱われている．そのことは筆者にとって一種の自負心につながっているが，その一方で「なぜ何年経ってもフロンティアなのか」という疑問の対象にもなってきた．筆者は勤務先となる大学において，総合科学系という研究組織で研究業績の評価を受け，リベラルアーツ学群という教育組織で教鞭をとっている．これらは既存の学部の役割を，学部とは異なる区切りによって担う組織群である．このような組織群に身を置く筆者の視点から研究領域とフロンティアについて考えたとき，それらは人材評価と教育カリキュラムによって区切られた領域と境界になっているのではないか，という考えを否定することができなくなった．本稿は，様々な既存組織が構成する領域・境界を否定するためのものではない．それらを研究上の課題として位置づけ，今後の研究活動において何ができるのかを議論するための一つの起点となることを目指すものである．

キーワード：人文科学とコンピュータの研究に関する課題

1. はじめに

デジタルアーカイブにおいて「いかなる情報をいかなる形で利用するのか」という利用時の課題は，「いかなる情報をいかなる形で保存するのか」という設計・実装時の課題と強く連動する．したがって，デジタルアーカイブを作成・提供する研究者が利用者としての視点も持ち合わせることは重要である．特に，膨大な情報を適切に利用するための情報検索技術に対する配慮は不可欠である．

デジタルアーカイブの情報検索を考える上で，メタデータの定義は重要である．DBMS が標準的にサポートする書式のメタデータであれば，その定義自体が情報検索技術となる．また，複数のデジタルアーカイブにおいて共通の定義を用いれば，利用者に対してシームレスな利用手段を提供できる．柔軟性を許容しながら定義に統一性を与える Dublin Core は，前述の理想の具現例といえる．

DBMS から独立してメタデータを定義し，それを扱う情報検索技術を構築することも可能である．筆者は，デジタルアーカイブの検索技術が「デジタルアーカイブの作り手の軌跡を記録する媒体」であり得ると考え，パターン認識技術を応用した古文書デジタルアーカイブの検索に関する研究に従事してきた．この研究は，渡辺らが構成した人文科学，工学などの研究者によるグループの活動[1]を母体としている．当時存在した標準的なメタデータ／DBMS の機能といったものとは別の視点で古文書と情報技術を結ぶという上記グループの志向が，筆者の研究活動およびその成果に繋がる軌跡を生み出していることは事実である．

本稿では，デジタルアーカイブの利活用促進を目的とした情報検索技術の研究を行ってきた筆者が，その過程で感じたいくつかの課題を取り上げ，それらに対する現時点での所見を述べる．なお，課題の選択と考察について，視野が筆者の活動範囲に留まり，DBMS からある程度独立した情報検索の話題に偏ることを予め断っておく．また，筆者が勤務先となる大学においてリベラルアーツ学群，および総合科学系という組織に所属し，学部とは異なる視点から研究領域とその境界を見ている点についても予めご了解いただきたい．

2. いくつかの課題

2.1 人文科学と工学における研究期間の違い

本稿で述べる情報検索技術は，人文科学と工学の境界にある研究課題である．多くの場合，研究の起点には人文科学のコンテンツ／研究者と工学の技術／研究者との出会いが存在する．その出会いが強い結合に育つためには，両分野の研究期間の違いに対する配慮が必要である．

人文科学分野のコンテンツを扱うデジタルアーカイブの構築には長い時間が必要である．また，一度公開されれば長期に渡って安定的に提供されることが望ましい．これを担う人文科学の研究者に対しては，長期間に渡って安定したメタデータ定義の上で研究活動を行うことが求められる．情報検索の機能についてもこの期間に合わせた安定的な運用が必要である．

このことは，新しい手法・手段を次々に生み出す短い期間での研究活動が強く求められる工学系の研究者には課題となる．特に，大学・大学院の研究室では学生・院生の在籍期間が単位となることが多い．このため，情報検索機能の長期運用に必要な各種メンテナンス，運用基盤となるサーバの更新に伴う再実装などの活動に関わることへの困難が存在する．

研究期間の違いを意識し，両分野の研究者にとって共益

†1 桜美林大学
J.F.Oberlin University

の関係を模索することは，多くの研究者が協力して情報検索技術を生み出すための重要な礎となるはずである．

2.2 運用コストと利用者のニーズ

情報検索技術の提供に追加コストが必要となる場合は，その捻出計画についても議論しておく必要がある．

人文科学分野におけるデジタルアーカイブの価値は，利用者の多寡だけで決まるものではない．しかし，利用者の増減が評価と運用に影響することは事実である．長期の安定した運用に適したコストを実現するために，工学的な工夫が必要である．また，万が一にもコストの捻出が困難となった場合にデジタルアーカイブの価値が大幅に損なわれることのない設計も重要である．

同時に，稼働中のデジタルアーカイブが新たな研究予算の獲得に繋がるような研究活動の継続性も重要となる．このことは，工学分野の短期間での研究活動との相性を改善することにも貢献する．

2.3 情報公開に関する業務と評価システムの課題

多くのデジタルアーカイブが Web 経由で公開されているが, DBMS や Web サーバの標準的な機能以外で情報検索技術をする場合には，公開担当機関における各種の情報公開ポリシーを遵守する体制にも配慮が必要である．

ポリシーの順守は，利便性向上とセキュリティ確保の両面で重要である．これらは流動的課題であり継続的な対応が不可欠であるが，研究機関がデジタルアーカイブに関する直接的な研究課題として評価しない場合，研究者として業務に関わる際の制限事項となる．

特に，各機関の情報公開に関わる部署が実施する定常的なメンテナンスで対応できない場合は，予算・人材などの確保を通して体制を確立する必要が生じる．当該人材として研究者を充てる場合は，業績評価システムおよびポストについての十分な配慮が求められる．

2.4 継続的な情報の追加と整理に関する課題

人文科学分野の研究成果が継続的に生み出される中で，デジタルアーカイブの拡充もまた継続的な課題となる．扱う課題によっては，その期間は数十年を優に超えることを考慮しなければならない．DBMS の機能に対する情報検索技術の依存度を問わず，継続的な情報の追加を支援するための情報技術，および作業の担い手となる人材の確保は不可欠である．

3. 所見

以上に取り上げた課題について，その多くは人材への評価に関するものではないかと思っている．

本稿で取り上げたような課題に対しては，筆者が述べるよりも前にいくつかの研究機関が効果的な解を導き出し，実践していると思っている．また，人材に関していえば，研究が人材を選ぶということ，逆に人材が研究を選ぶかどうかということも考慮しなければならない．ただし，それらを踏まえた上で，(1) 境界領域に生まれた研究が境界を残したまま進行する限り，隣接する研究分野のニーズを共に満たすというミッションが生じること，(2) それが個々の研究者にとって容易ではないこと, そして, (3) 人材に対する評価基準が個々の研究内容とは無関係な境界を構成すれば研究者の活動は難しくなること，について筆者の問題意識を汲み取っていただけると幸いである．

筆者は，勤務先において総合科学系という研究組織，およびリベラルアーツ学群という教育組織に所属している．これらは，学部という枠組みに照らせば複数を跨ぐような位置にあり，また，研究組織と教育組織の境界は別々のものとして設定されている．各教員は既存研究領域の境界に対して比較的自由な活動を実施し，その成果を教育内容に反映することが可能である．このような組織群と学部との詳細な比較を行うことは他に譲るが，少なくとも筆者が研究業績に関する評価を受け, 教育活動を行うにあたっては，フロンティア領域にいるという意識が少なからず薄らいでいるように感じている．

ただし，既存の学部や研究領域の枠組みで実施されている人材評価および教育活動の成果については肯定されるべきものである．フロンティア領域に位置づけられる研究に関わる各自が，そのような枠組みの中における研究活動のあり方と研究成果の表現方法を見出し，実現し，そこに生身の研究者が常識的な範疇で存在し得ることを示し続けるために，筆者なりにできることを考えていきたいと思っている．

4. おわりに

本稿では，デジタルアーカイブの利活用促進を目的とした情報検索技術の研究を行う中で筆者が感じた課題と所見について述べた．人材評価に対しては，査読付きシンポジウムであるじんもんこんに加えて，論文誌特集号の発刊などを通した取り組みが既に実現されてきた．このような取り組みが今後も発展し，より多くの研究者が人文科学とコンピュータに関する研究に関わることができるようになることを願っている．

参考文献

[1] 渡辺晃宏（代表）他, 科学研究費補助金 基盤(S)「推論機能を有する木簡など出土文字資料の文字自動認識システムの開発」, 2003-2007 年度.