

「デジタル技術によるアーカイブズの開き方 ～ “Japan Search” が目指す世界～」

高野 明彦

国立情報学研究所／東京大学大学院 教授



【略歴】

1980年東京大学数学科卒。東京大学博士（理学）。2001年より現職。1980～2000年、日立製作所の基礎研究所、中央研究所等に勤務。2002年より東京大学大学院情報理工学系研究科コンピュータ科学専攻教授併任。専門は関数プログラミング、連想情報学、デジタルアーカイブ。研究成果を活用して、「WebcatPlus」「新書マップ」「想・IMAGINE」「Book Town じんぼう」「TIMEMAP」を構築・運用している。また、共同研究を通じて「文化遺産オンライン」「NHK放送文化アーカイブ」「日本アニメーション映画クラシックス」「渋沢敬三アーカイブ」などの構築も手がけている。著書に『検索の新地平』（角川）、『311 情報学』（岩波書店）、『検索エンジンは脳の夢を見る』（講談社）などがある。2017年より内閣府知財戦略本部・デジタルアーカイブジャパン実務者検討委員会座長。同委員会の監修の下、分野横断統合ポータル “Japan Search”（試験版）が2019年2月に公開された。

<https://scholar.google.com/citations?hl=en&user=v59zRTYAAAAJ>

<https://ci.nii.ac.jp/nrid/9000240034046>

<https://www.linkedin.com/in/akihiko-takano-8222204/>

【要旨】

爆発的に拡大する電子情報空間から求める情報を探し出すのに、現在はもっぱらキーワード検索が使われている。しかし、数個のキーワードを組み合わせるだけでは、検索の意図を表現しきれないことがよくある。自分がどんな情報を探しているかを、もっと的確に表現するための方法が絶対必要である。この目的のため、私たちは「連想検索」という新しい検索方法を提案してきた。連想検索では、いくつかの文書の集まりを選んで、それをクエリにできる。検索結果としては、クエリと近い内容の文書群が近さの順に並べられて返される。クエリとして選択する文書を変化させることにより、真の関連性フィードバックが容易に実現できる。文書と文書や単語と単語の類似性を計るための統計的計量を定義することにより、「連想」に数学的にも計算的にも健全な理論的基礎づけを与えることができるのである。

アーカイブズ、図書館、美術館や博物館、その他の文化的記憶に関わる施設は、それぞれ独立

に情報を蓄えてきたが、それらは互いに繋がっておらず、まるで別々のサイロのように孤立している。私たちの提唱する連想検索は、統一的なメタデータを必要としないので、このように孤立した知識のサイロの間に橋をかけて繋ぐことに特に役立つ。広い意味での文化的情報や記憶、知識などを集めてきて、それらの間に繋がりを作るための情報技術の研究を進めてきた。

これまでに連想検索に基礎を置く公開サービスを多数立ち上げて維持してきた。例えば、図書検索の Webcat Plus や新書マップ、文化財情報の文化遺産オンラインなどである。これらのサービスを、IMAGINE 技術で束ねることにより、それぞれ独立に作られた情報源を連想検索を使って、その場で相互に関連づけることができる。IMAGINE が提供するこの「連想の場」では、ユーザは自分の望む情報を表す文脈を対話的に作りあげることができる。その文脈探しにより、ユーザは自分の興味を敏感に反映した自分だけの文脈に出会えるのである。

私たちが提案する PONGEE という読書環境は IMAGINE インタフェースの一種と考えることができる。書籍の本文を表示すると、それがその場で分析されて、Wikipedia や国語辞典などと自動的に結びつけられて、結果がサイドノートとしてページの傍に表示される。単純な辞書引きによる関連づけだけでなく、表示中の本文をクエリとして他の情報源を連想検索した結果を出すこともできる。PONGEE を使ってアーカイブされた資料を閲覧することの意義について論じたい。

最後に、内閣府知財戦略本部の元で進めているデジタルアーカイブ ジャパンというプロジェクトについて紹介する。2019年2月には、”Japan Search” という新サービスを公開した。これは、日本全国から多様なデジタル情報のメタデータを収集して、一元的に提供するためのサービスで、国会図書館や国立公文書館の情報も集められている。今後 Japan Search に期待される役割についても論じる。