

メタデータ共有システムによる「知識」の構造化と構築の支援

上松 大輝（総合研究大学院大学・国立情報学研究所）、棕本 輔（鶴見大学・二松学舎大学）

本発表では、主にデジタルアーカイブをめぐる人間の「知識」構築を支援すべく構想・開発中のメタデータ共有システムについて紹介する。また、それをを用いた実験・実践を通して考察を進めている、人間の「知識」の構造的な側面をデジタル情報として扱うための意味論的な定義についての検討を行う。

近年、ChatGPT に代表される大規模言語モデル（LLM）を用いた生成系人工知能（Generative AI）など AI の進化が目覚ましい。LLM は膨大な量のテキストデータからパターンを抽出し、文脈を考慮して確率的に文章を生成、自然言語処理の実行が可能である。一方、人間が文章をテキストデータとして「生成」する場合には、個々の内在的なメッセージやイメージのようなものを、様々な意図に基づき文字言語としてコード化する。ここでコード化の前提となる意図には、さらに前提としての様々な知識が存在する。そうした知識には「単語の意味理解」「文章の構造理解」「文脈理解」「専門分野の知識」があると考えられる。「単語の意味理解」と「文章の構造理解」には、基本語彙や文法といった言語習得者にとって共通の知識が含まれている。また、「文脈理解」は、コンテキストの理解や解釈に必要な知識、どの分野の語彙を使用すべきか判断するための知識と考えられる。「専門分野の知識」には、ドメイン内でのみ用いられる単語や解釈、グループや個人が持つ知識が挙げられるが、LLM の学習対象データには含まれていない場合がある。Wikipedia や Wikidata 等の百科事典や、インターネット上に公開される情報は、集合知としてより多くの人理解可能な、一般化された知識へと収束するためである。

それに対して、本発表におけるメタデータ共有システムでは、よりミクロな地域コミュニティやテーマ別のアーカイブのようなドメイン固有の知識の構造化と、それを通じた新たな知識構築の支援を目指している。メタデータ共有システムは、「メタデータ”を”（効率的に）共有するシステム」ではなく、「メタデータ”によって”（資料の存在や資料をめぐるコミュニケーションについて）共有するシステム」を目指しており（水島・棕本・上松，2020）、「戦争関連資料」というテーマへの適用・応用も現在進められている（棕本・上松，2021）。そこでは、意味論的なリンクに基づく情報の提示と、メタデータ付与の簡易・簡便化によって、さらなるリンクの形成が促され、新たな資料やメタデータの発見を支援していく。

しかしさらなる課題として、実世界における様々な事象についての「知識」やそれに基づく我々の意図は、すべからず個々の自由な「解釈」と切り離せない。それを踏まえた上で「知識」の構造的な側面をデジタル情報として扱う際には、意味論的な定義をどのように考えるべきなのか、本発表では記号学・記号論における議論を参照しながら検討する。

参考文献

- 水島久光・棕本輔・上松大輝「コミュニティアーカイブ連携のためのメタデータスキーマについて」（デジタルアーカイブ学会・第 5 回研究大会/デジタルアーカイブ学会誌 4 巻 s1 号, 2020) https://doi.org/10.24506/jsda.4.s1_s27
- 棕本輔・上松大輝「戦争関連資料をつなぐメタデータ共有システムの構想」（『ライブラリー・リソース・ガイド（LRG）』第 36 号・特集「戦争の記憶と記録」, 2021.8） <https://arg-corp.jp/2021/08/13/lrg-24/>