

インタビューテキストの事例分析による ライブラリアンに求められる能力調査

金川久美子、鄒楽晗、戸崎祐輔、中藤哲也、廣川佐千男（九州大学）

1. はじめに

図書館は世界の貴重な歴史資料を保管する場としての機能を持つ、伝統のある組織である。インターネットの普遍化と資料のデジタル化により図書館のサービス領域は拡大しており、ライブラリアンが担う業務内容は多岐にわたる。多様な利用者のニーズに応えるためにライブラリアンに求められる能力も高度化・専門化されてきているが、人員や予算などの経営資源が縮小する中で、その実態は明らかになっていない[1]。本稿では、香港大学の Dickson Chiu らによる世界各地の図書館のライブラリアンに対するインタビュー記録の提供を受け、現在のライブラリアンにどのような能力が求められるかを分析した。具体的には、インタビュー記録から能力に関する記述を含んだ文を手で識別し、機械学習による分類と関連語の可視化を用いることにより、ライブラリアンの能力を表す特徴語を抽出した。

2. インタビューテキスト

2-1. 東アジア関連資料館、およびライブラリアンへのインタビュー

本稿で分析対象としたのは、Cho らの著書[2]に掲載された第3章までのインタビュー記録である。表1に示すとおり、インタビューを受けたのは、東アジアに関する資料を保管している図書館や資料館の専門的ライブラリアンであり、単純なアンケート形式ではなく、対話的に深く尋ねたインタビューとなっている。一般の人あるいは学生が利用する教育機関の図書館や公的図書館とは位置づけが違うが、分析データとして36件ものインタビュー記録は貴重である。したがって、我々はライブラリアンの必要な能力について偏りなく収集している適切なデータだと判断した。本稿は、その最初の3つの章だけだがテキストマイニングによる分析の有効性を実証する試みとして取り組んだ。使用したデータの分量を表2に示す。

表1 分析対象データ

章	章題名	対象者	所属
1章	Inside the Vatican Appstolic Library	研究者、ライブラリアン、兼キュレーター	バチカン図書館東アジア資料館
2章	A Library of the Venice and the Giorgio Cini Foundation's Far East Centre for Comparative Studies of Cultures and Spiritualities	ライブラリアン	イタリア、ベニス、ジョルジオ・チーニ財団
3章	From the Shang to the Qing at the British Library	主任キュレーター	英国、ロンドン、大英図書館

表 2 分析対象データの分量

章	質問数	文の数	ページ数
1 章	35	428	15
2 章	16	72	7
3 章	24	220	12

2 - 2. インタビュー形式と特徴

これらのテキストは半構造化によるインタビューの質問文と回答文から構成されており、回答文を本調査の分析対象とした。対面式、Skype、電子メールによる会話的な質問を用いて、12 か月わたって対象者 36 人にインタビューが行われている。回答者はすべて任意としている。インタビューは、自由回答式質問法により「なぜ」「何」「どうやって」などの疑問符ではじまる質問を行い、回答者は質問項目について回答者自身の言葉で自由に表現するようになっている。

スカイプと e-mail によるフォローアップの質問も実施し、インタビュー項目では最終学歴、現在の職務内容、技術的能力、専門的能力などについて尋ねている。これはインタビュー調査の一貫性を維持する目的で、質問者はすべての回答者に類似の質問項目で質問している。

3. 機械学習による分析

テキストマイニングは文章を単語に分割し、それらの出現頻度や共起関係を分析することで有益な情報を得ることができる。我々はライブラリアンに必要な能力を表す単語の特徴を分析するため、インタビューテキストにテキストマイニングを適用した。

3 - 1. 質問文、回答文、必要能力の文の区別

分析対象としたインタビューの PDF ファイルからインタビューのテキストデータのみを手作業で識別し、質問文と回答文をそれぞれ文単位分けた。対象データの総単語数は 1618 語である。特に、回答文はライブラリアンの能力についての文を識別し、426 個の回答文のうち 57 の文が抽出できた。本稿では、機械学習と単語関連図を適用することで、ライブラリアンに必要な能力を回答文から抽出することを試みた。

3 - 2. 機械学習による能力記述文の識別と特徴語

機械学習 SVM は特定の条件を満たす文書群を機械的に判別し、特徴語の抽出と、各特徴語の重要度を評価できる [3]。表 3 は回答文におけるライブラリアンに必要な能力を表す特徴語を機械学習で識別した時の識別性能を示している。著者らが人手で選定した「ライブラリアンに必要な能力を含んだ回答文」を正例とし機械学習を適用した。その結果、1618 語の単語のうち、180 個(能力に関する特徴語 90 個と能力に関連しない特徴語 90 個)の特徴語で最適識別率は 94% となった。なお表 3 の 1 行目は全単語を使った場合の惜別性能を示す。これらの結果から、インタビューのテキストデータに機械学習を適用することにより、ライブラリアンの能力を 180 語と言う少数の単語でとらえられることがわかった。

表 3 機械学習による識別性能

N	precision	recall	F-measure	accuracy
0	0.2099	0.3337	0.2334	0.8120
70	0.7850	0.4790	0.5944	0.9311
80	0.7976	0.5610	0.6365	0.9404
90	0.7783	0.5752	0.6469	0.9401
100	0.8450	0.5200	0.6274	0.9389
200	0.3870	0.4032	0.3336	0.8607
300	0.2270	0.3337	0.2402	0.8155
400	0.1932	0.3194	0.2179	0.8090

4 キーワード関連図からみたライブラリアンの能力

図1は、ライブラリアンに必要な能力を含む回答文を、テキストマイニングを適用し、単語の共起関係をマップとして可視化したものである[4, 5, 6]。まず頻度の低い単語から始め、それよりも頻度の高い単語で、最も関連の強い単語一つ選び枝を延ばすことを繰り返し、ツリーを構成している。これにより、ライブラリアンに必要な能力の特徴語がわかるだけでなく、回答者が業務をする上で重要なキーワードを複数の単語群から解釈することができる。

以下からは、出現頻度の高いキーワードをA, B, C, Dの4つの単語群に分けて考察する。

4 - 1 A, B, C グループの特徴

Aグループは language から knowledge と important が共起している。回答者は言語について言及している。Bグループは systems から inputs と、 experience から subject と addition が共起している。Cグループは skills から needed と new が共起している。回答者は技術について言及している。Dグループは member から5つの単語が共起している。他のグループと比べる共起する単語数は最も多い。これらの単語群からAグループは言語に関わる回答、Cグループは技術に関わる回答を行ったと推定できる。

Choらは[2]、インタビュー対象者に、特定の業務に関する「技術」について尋ねただけではなく、「個人の経験」や「業務を行った理由や考え」についての質問を同時に行ったと述べている。これは図1のA, B, Cのグループの共起関係と一致している。具体的には、Aグループの language-important-good の回答文は「We can look at an item from a language, historical or archaeological perspective, so I would say that knowing language, history and archaeology and having research skills are all important abilities for being a good curator.」とあり、著者らが選定したライブラリアンに必要な能力を含んだ単語の共起関係と符合している。

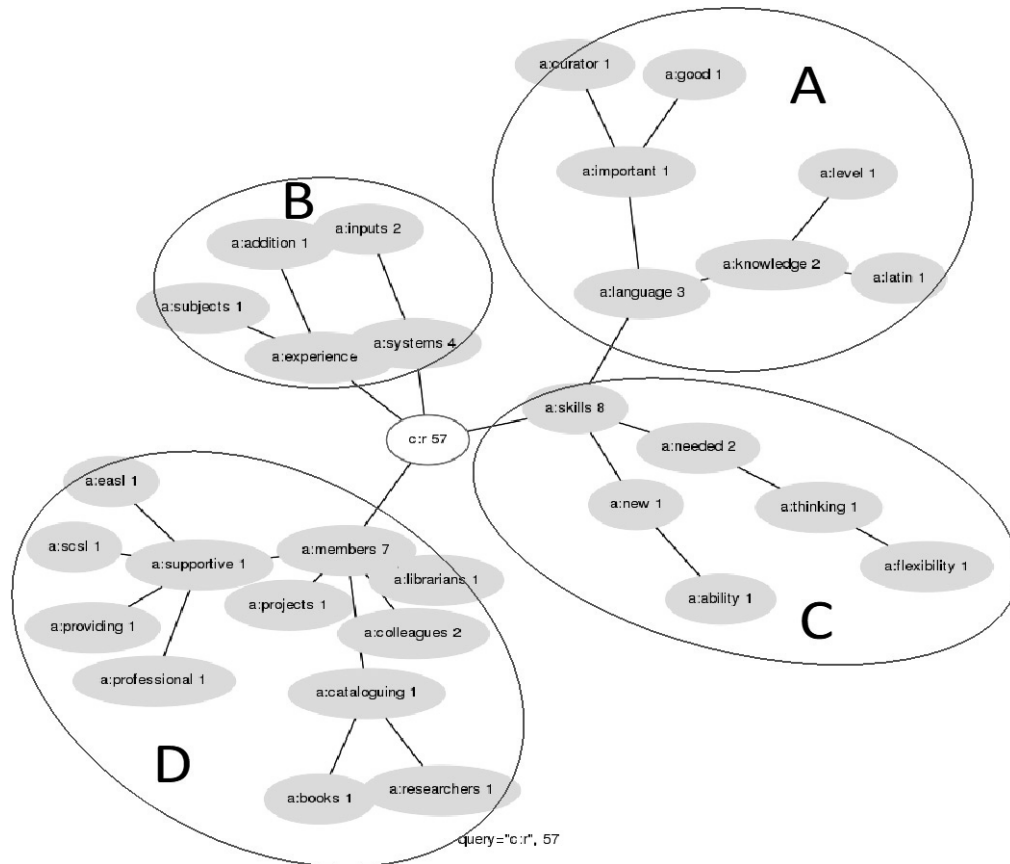


図 1 単語の共起関係マップ

4-2 Dグループの特徴

Dグループは members-cataloging-books (仲間-目録-本) や members-supportive-providing (仲間-協力的な-提供する) など、仲間との連携した業務について言及していると推定できる。Dグループには、Choらのインタビューで設計された「技術」「経験」「考え」に関連する単語は見当たらない。しかし、ライブラリアンの能力には、仲間との共同作業が重要であることが推察された。

5. 関連研究

インタビューは、直接観察できない感情、思想、意図、過去の行動などを見つけ出す目的のために用いられる。インタビューで得られたデータをテキストマイニングによって分析する手法は社会学や看護学において多く行われている [5, 6, 7, 8, 9]。

寺下 [10] は、インタビューによるデータ収集の利点は、調査者と回答者の顔が見えることによる、高い回答率と回答の質の高さにあると指摘している。

6. まとめと今後の課題

本調査結果により、2つのことが明らかになった。1点目は機械学習 SVM によってライブラリアンに必要な能力が、分析対象のインタビューテキストデータ 1618 語のうち、180 語によって 90% 識別できたことである。2点目はライブラリアンに必要な能力を表す特徴

語の関連図から、「技術と経験」、「技術と理由」、「技術と考え」に関する単語が共起していたことである。これはインタビュー実施者である Cho らが、仕事や業務に関わる「技術」について尋ねるだけでなく、関連する「経験」やその仕事が必要となった「理由」や「考え」も合わせて尋ねているというインタビューの設計者の意図と符合する。

今後の課題は、これらの分析手法を他の 33 名のライブラリアンインタビューテキストにも適用することである。また、本調査と一般的にいわれているライブラリアンの能力比較を行う予定である。本稿では回答文のみを分析対象としたが、質問文の中に含まれる単語とライブラリアンに必要な能力を表す回答文との単語の対応を調査する予定である。

【参考文献】

- [1] 長谷川豊祐, “変化への適応：大学図書館における業務分析と業務管理手法,” 情報の科学と技術 = J. Inf. Sci. Technol. Assoc., vol. 61, no. 8, pp. 311-316, 2011.
- [2] A. Cho, P. Lo, and D. K. W. Chiu, *Inside the world's major East Asian collections: one belt, one road, and beyond*. Chandos Pub., 2017.
- [3] Y. Adachi, N. Onimura, T. Yamashita, and S. Hirokawa, “Standard Measure and SVM Measure for Feature Selection and Their Performance Effect for Text Classification,” in *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services*, 2016, pp. 262-266.
- [4] S. Hirokawa, B. Flanagan, T. Suzuki, C. Yin, *Learning Winespeak from Mind Map of Wine Blogs*, Proc. HIMI 2014, Part II, LNCS 8522, pp. 383-393, 2014
- [5] 加藤千佳、城丸瑞恵、いとうたけひこ, “テキストマイニングを用いた病棟看護師の実習指導に対する語りの分析,” 昭和大学保健医療学雑誌, no. 8, pp. 23-33, 2011.
- [6] 渡辺裕一, “行政職員から見た限界集落の現状と課題：行政職員へのインタビューに対するテキストマイニング分析から,” 武蔵野大学人間科学研究年報 = *Annu. Bull. Musashino Univ. Inst. Hum. Sci.*, no. 4, pp. 69-80, 2015.
- [7] S. L. Hunt and C. J. Bakker, “A qualitative analysis of the information science needs of public health researchers in an academic setting,” vol. 106, no. April, 2018.
- [8] 内藤みよ子, “レファレンス・インタビュー,” 看護と情報：日本看護図書館協会会誌, vol. 9, pp. 75-80, 2002.
- [9] 木村朗, “乳がん・前立腺がん経験者のインタビューテキストデータから集団機械学習ランダムフォレストによる検診行動の推定の試み：DIPEX-Japan のテキストデータ二次分析,” 群馬パース大学紀要 = *Bull. Paz Coll.*, no. 18, pp. 19-25, 2014.

- [10] 寺下貴美, “質的研究方法論: 質的データを科学的に分析するために,” 日本放射線技術學會雑誌, vol. 67, no. 4, pp. 413-417, 2011.