

図書館評価のための概念枠組みの開発とその方法

隋 亥華 (慶應義塾大学大学院文学研究科) zuyikk2010@a8.keio.jp

1. 研究の目的

従来の図書館評価の研究において、いくつかの概念枠組み(フレームワーク)の研究が行われてきたが、図書館評価研究に求められる様々な概念や用語を網羅し、さらには実際の図書館評価で検証されたものは多くない。そこで、本研究はそうした網羅的・包括的な概念枠組みの開発を目的とする。その後、実際の図書館評価報告書をもとに、その妥当性を検証する予定である。

2. 候補用語の抽出

形態素解析において、専門用語を自動抽出するツールは数多くある。中でも、マンチェスター大学によって開発された TerMine¹⁾ は特にユーザーインターフェースが簡潔かつ便利なので、本研究でもこれを使用する。更に、それは、C-value という独特のアルゴリズム²⁾ を利用して開発された専門用語自動抽出ツールである。C-value とは、一般的な用語の出現頻度を一つずつ数えて合計する手法ではなく、候補用語のユニット性と用語性を求め、より正確に計算できる手法である。それは、入れ子(nested terms)構造を持つ長い文字列の抽出に優れているとされている。

$$C\text{-value}(a) = \begin{cases} \log_2 |a| \cdot f(a) & a \text{ is not nested,} \\ \log_2 |a| (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b)) & \text{otherwise} \end{cases}$$

a= 候補文字列の長さ

f(a)= 全文書中に出現した頻度

Ta= a に含まれる候補文字列の集合

P(Ta)= a に含まれる候補文字列の数

本研究では、図書館評価に関する雑誌論文のタイトルとキーワードと抄録を取り上げてそれらのテキストを抽出した。具体的には、「web of science (WOS)」により、下記の検索式をもとに、分野 (information science library science)、言語(English)、データベース=SCI-EXPANDED、SSCI、A&HCI、タイムスパン=全範囲と絞り込んで、612 雑誌論文を得た。(検索日 2013. 6. 8)

```
<タイトル=(library evaluat*) OR  
タイトル=(library measur*) OR  
タイトル=(library assess*)>
```

そのうち、ダウンロードでき、テキストに変換できるのは 317 論文であった。これらの雑誌論文について題目、キーワード、抄録をテキストファイルで WOS から自動抽出した。また、それらの全文の PDF ファイルをもダウンロードし、テキストに変換した。文字化けなしで、テキスト化できたのは 283 論文であった。

3. TerMine によって分析した結果

TerMine を用いて、上記の雑誌論文の題目、キーワード、抄録から成るテキスト集合に対して専門用語を自動分析した。その結果、専門用語の候補文字列として 2, 311 文字列を得た。同時に、その専門用語候補に用語としての重要度を反映するスコア付けを自動的に行った。より高い C-value 値を得た候補用語は閾値が 3.5 以上の 124 複合名詞である。そのうち、図書館評価の概念枠組みを開発するうえで、関連性の低い名詞

(this article、short formなど)、機構や組織名 (elsevier inc、elsevier science inc、bbc scotland library service など) の複合名詞を除き、整理した結果、残ったのは65複合名詞であった(表1)。

表1 TerMineによる候補用語のC-value (一部)

No	Term	C-value
1	public library	57.799999
2	web site	48.388085
3	service quality	45.479699
4	digital library	43.827587
5	library service	37.803921
...
65	Measurement system	3.5

4. 雑誌論文における専門用語の分布

上記で得た65複合名詞(専門用語)間において、どのような関係を持つかということ を明らかにするため、専門用語間の共起関係を分析した。本研究において、専門用語の共起というのは、任意の2つの専門用語が同一論文において共出現する場合を言う。前提としては、題目、抄録などから抽出した専門用語を283論文全体において、共出現する状況を明らかにする必要がある。まず、出現結果を妨げないように、テキスト中の参考文献と掲載されている雑誌名などに関する事項を手作業で削除した。次に、図書館評価に関する専門用語とそれらの関係を網羅的に把握する必要があるため、同一の意味を指し、表現が違う専門用語を追加し、分析した。

さらに、C-value値によって得た65専門用語をそれぞれの論文全文において、専門

用語ごとに探索し、分析した結果を表2に示す。表2では、それぞれの用語が同一の論文において出現する場合「1」と記入し、出現しない場合は「0」と記入した。

表2 雑誌論文における専門用語の出現状況(一部)

Article \ Term	A1	A2	A3	A4	A5
public library	0	0	0	0	0
web site	0	0	0	1	0
service quality	0	0	0	1	0
digital library	1	0	1	0	0
library service	0	0	1	1	0

5. 階層クラスター分析

クラスター分析(cluster analysis)は、似通った個体あるいは変数のグループ化を行うための分析手法である。65専門用語を雑誌論文における分布データ、即ち、用語間の共起度により、SPSSの階層クラスター分析により分類した。分類の基準は用語間の距離とし、その距離を示す方法がいくつかある。本研究では、連結しようとする小グループ間でそれぞれの個体間の組み合わせで距離を調べて、その平均距離が小さいような連結を優先するという平均連結法を用いた³⁾。分析結果を図1に示す。図1により、区切り線以内、分類されたケースは18クラスターあり、以下のように解釈することができるだろう。

・クラスター1: ケース5から30にかけては、一つのグループになる。そのすべては

ある特定の図書館の一般的な構造(人、サービス、資源、システム、情報技術、図書館環境)と図書館評価に関する一般的な手順(サービスの質を測定するために、パフォーマンス測定を用いてサービスを評価する)の一環だと考えられる。

- ・クラスター2: ケース 6 から 45 にかけてが、図書館コレクションに関するものである。

- ・クラスター3: ケース 37 は、図書館の効率性というものである。

- ・クラスター4: ケース 8 から 35 にかけては、図書館リテラシーに関するものである。

- ・クラスター5: ケース 31 と 55 は、図書館評価方法に関するものである。

- ・クラスター6: ケース 57 から 63 にかけては、戦略計画に関するものである。

- ・クラスター7: ケース 51 から 7 にかけては、図書館のアウトカム評価に関するものである。

- ・クラスター8: ケース 29 は、図書館空間というものである。

- ・クラスター9: ケース 64 は、因子分析というものである。

- ・クラスター10: ケース 11 から 25 にかけては、医学図書館、国家図書館のような図書館の種類というグループとみなすことができる。

- ・クラスター11: ケース 65 は、図書館評価(測定)システムというものである。

- ・クラスター12: ケース 1 と 14 は、公共図書館ということを示している。

- ・クラスター13: ケース 12 と 59 は、大学図書館を評価の実例として考える際の教育効果のことを示している。

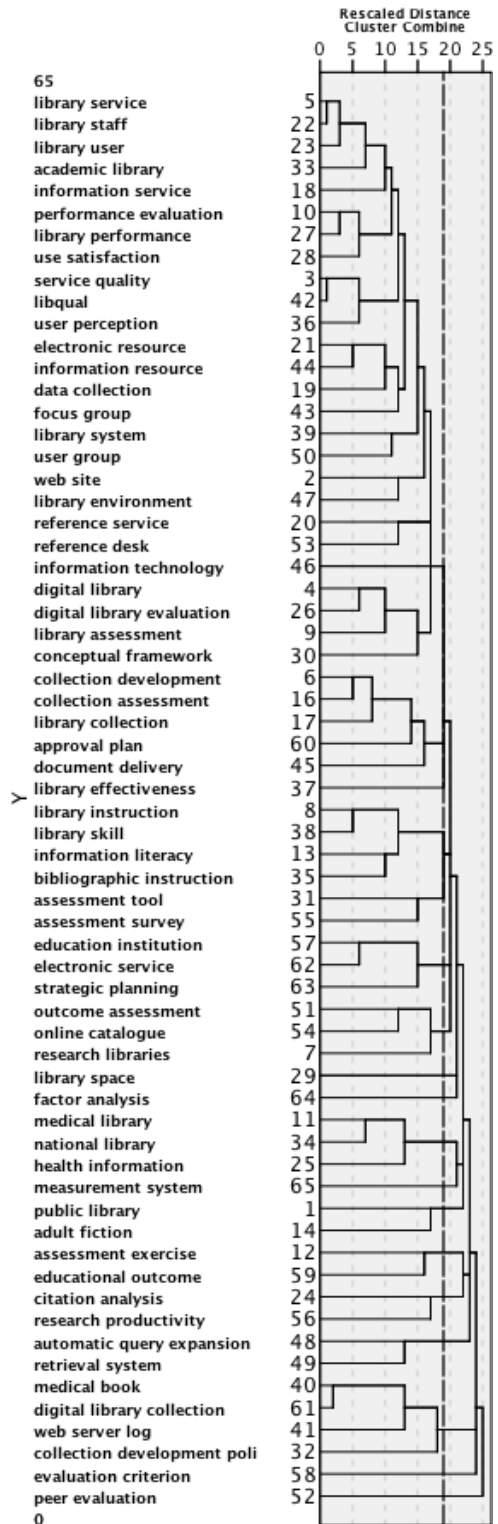


図 1 デンドログラム

- ・クラスター14: ケース 24 と 56 は、大学図書館を評価の実例として考える際の研究効果のことを示している。
- ・クラスター15: ケース 48 と 49 は、図書館の検索システムということを示している。
- ・クラスター16: ケース 40 から 32 にかけては、デジタルライブラリーのコレクションに関するものである。
- ・クラスター17: ケース 58 は評価の規準である。
- ・クラスター18: ケース 52 は図書館員による勤務査定の領域である。

6. 図書館評価のマクロな概念枠組み

デンドログラムの分析を踏まえ、それらを反映させた図書館評価に関する概念枠組み(図2)を構築した。この図の上の部分は大学図書館を例として図書館評価の一般的な構造を示し、同時に、図書館評価がそ

の図書館業務のすべての中に位置づけられることも明らかにした。図の下部は本研究で得られた重要な図書館評価に関する専門用語とそれらの相互関係に基づいて構成した評価システムである。これは、マクロの視点から図書館評価の概念枠組みを表すものと考えられる。

注・引用文献

- 1). <http://www.nactem.ac.uk/software/termine/>, (参照 2013-6-25)
- 2). Frantzi, K., Ananiadou, S. and Mima, H. Automatic recognition of multi-word terms. International Journal of Digital Libraries 2000.3(2), p. 117-132.
- 3). 村瀬洋一, 高田洋 等. SPSS による多変量解析. オーム社, 2007, 349p.

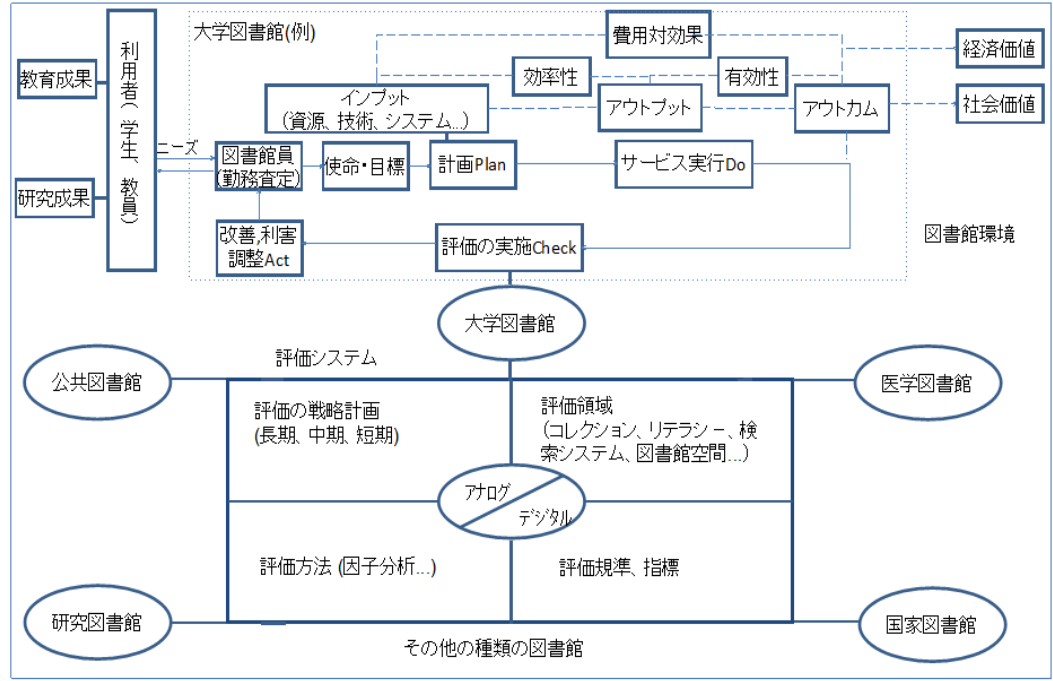


図2 図書館評価のマクロな概念枠組み