

貸出履歴を利用した蔵書検索結果の
リランキングに関する研究

筑波大学

図書館情報メディア研究科
2014年3月

陳 睿

目次

第1章	はじめに	3
1.1	研究背景：図書館の蔵書検索システムの現状	3
1.2	研究目的	3
1.3	本論文の構成	4
第2章	先行研究及び関連研究	5
2.1	図書館における貸出履歴の活用	5
2.2	図書館の貸出履歴を利用した研究	5
2.2.1	貸出履歴を利用した図書推薦	5
2.2.2	貸出履歴を利用した検索システム	6
第3章	貸出履歴の分析	7
3.1	分析対象	7
3.2	分析方法	7
3.2.1	NDC分類について	7
3.2.2	分析内容	9
3.3	分析結果	9
3.3.1	学科ごとの利用傾向	9
3.3.2	時期による利用傾向の変化	26
第4章	貸出履歴を利用した蔵書検索結果のリランキング手法	29
4.1	リランキング手法概要	29
4.2	利用者のグループ化	30
4.3	利用者プロファイルの作成	30
4.4	検索結果のリランキング	32
4.4.1	書籍のNDC分類番号の抽出	32
4.4.2	関連度の計算	33
第5章	実験と評価	34
5.1	評価用正解データの作成	34
5.1.1	概要	34
5.1.2	実験協力者	34
5.1.3	実験に用いたデータ	34

5.1.4	データ収集用システムの構築	35
5.1.5	データ収集の手順	36
5.2	評価	36
5.2.1	比較手法	36
5.2.2	評価ツールと指標	37
5.2.3	評価用正解データ	37
第6章	結果と考察	38
6.1	平均精度	38
6.2	考察	38
6.3	今後の課題	41
第7章	結論	42
第8章	謝辞	43
	参考文献	44

第1章 はじめに

本章では、研究の背景と目的を述べる。

1.1 研究背景：図書館の蔵書検索システムの現状

現在、図書館で扱う情報の量は爆発的に増加し、利用者の情報要求も様々なものとなっている。しかし、図書館の一般的に利用されている蔵書検索システムは利用者の違いは考慮せずに、同じ検索語に対して同じ結果が提示されるという問題がある。利用者にとって必要な情報が検索結果リストの上位提示されていれば、利用者は短時間で情報を探ることができる。しかし、そうでない場合、利用者は自分がほしい情報を収集するために提示された膨大な検索結果リストの確認作業を行わなければならない。これを絞り込むためにANDやNOTによる絞り込み検索が利用されるが、このような検索は利用者に高い能力を要求する。検索能力が高い利用者なら、効果的な絞り込みが行えるであろうが、そうではない利用者にとっては、効率的な絞り込みは行えず、検索漏れが発生するか、やはり大量の確認作業を行うことになってしまう。

利用者はプロフィール（趣味、知識世界など）の違いにより、求めている情報は異なる。しかし、この情報要求の違いは検索語だけから読み取ることは困難である。例えば、「インターネット」という検索語で検索する場合を考えると、経済学専攻の学生であれば、インターネットに関する統計や経済についての情報を求めていることが多いかもしれない。それに対して法学専攻の学生の場合、インターネットに関する法律について情報を求めていることが多い傾向があると考えられる。このように、利用者のプロフィールが異なれば、同じ検索語であっても求めている検索結果も異なると考えるのが自然である。しかし、現在の蔵書検索システムは利用者のプロフィールを考慮せず、どの利用者に対しても全くに同じ検索結果が提示されている。

1.2 研究目的

そこで本研究では以上述べた問題点を解決するために、利用者の情報要求に合わせて蔵書検索結果を効果的にリランキングすることを目的とする。そのために本研究では図書館の貸出履歴を利用して利用者のプロフィールを作成する。利用

者のプロフィールを基にリランキングを行い、これより利用者の目的とする書籍を検索結果の上位に集中させる。利用者の情報収集の負担を軽減することで効果的な検索作業が可能になると考えられる。

1.3 本論文の構成

本論文では、第2章で関連研究を挙げ、その概要と課題について述べる。第3章で大学図書館の蔵書検索システムの利用者の利用特徴を把握するため、貸出履歴を分析した結果を述べる。第4章は貸出履歴を利用した蔵書検索結果をリランキングする手法を詳述し、第5章では提案手法を評価するための実験とその結果と考察を述べる。最後に第6章で結論を述べる。

第2章 先行研究及び関連研究

本章では図書館における貸出履歴の活用の現状及び貸出履歴を活用した図書推薦と検索支援を扱った研究を紹介し、研究の方向性を述べる。

2.1 図書館における貸出履歴の活用

現在、Amazon¹のような利用履歴を用いた推薦システムが数多く実用化されてきている。また、閲覧履歴を利用した Web 検索の個人化と検索支援に関する研究も盛んに行われている。しかしながら、図書館での貸出履歴の利用については、個人情報漏れる危険性があることや、利用者の思想調査に利用される恐れがあるなど、様々な問題の可能性が指摘されてきた経緯があり、貸出履歴の積極的な利用は行われてこなかった。

利用者の貸出履歴利用に対する意識に関しては、公共図書館における利用履歴の活用に関する意識調査をした佐浦ら [1] による研究がある。佐浦は図書館の利用者 400 人を対象に、図書館の利用頻度別の図書館の貸出記録に関する意識調査を行っている。その結果、利用者の 71.5% が過去に借りた本の履歴を確認できるサービスを使いたいと回答し、利用頻度を問わず、図書館がそのような利用者サービスを目的に貸出記録を保存することについても 7 割以上が「容認できる」どちらかといえば容認できると考えていることを明らかにしている。

2.2 図書館の貸出履歴を利用した研究

図書館貸出履歴を利用した研究としては推薦システムに関する研究と図書館検索システムの検索支援に関する研究がある。

2.2.1 貸出履歴を利用した図書推薦

貸出履歴から図書の属性を利用して図書推薦を行う研究に原田らの研究 [2][3] がある。原田らは図書館の貸出履歴を使用して実際に図書推薦実験を行った。貸出履歴としては、ある大学図書館で原田らは 200 冊以上 500 未満の図書を借りた学生

¹<http://amazon.com/>

の貸出履歴 272751 件を用いている。原田らは図書の貸出特徴（図書の分野と貸出日付）に基づいて重み付けを付与して図書推薦を行う。具体的には NDC 番号（日本十進分類法：Nippon Decimal Classification）を重み付けに用い、借りた年月によって貸出履歴を分割してそれぞれに推薦を行った。推薦実験の結果、図書の分野と貸出日付の両方を使った重み付け手法が高い評価を得ることが示された。

2.2.2 貸出履歴を利用した検索システム

利用者の属性を利用した研究として、當山 [4] の利用者プロフィールを考慮した連想検索 OPAC の構築という研究がある。當山の手法は、利用者の属性を考慮し、分野単位で利用者をグループ化する。その上で、利用者の所属するグループの資料に対する貸出件数に基づく重みを算出して検索結果の順位を修正する。実験の結果、利用者の専攻分野による重みを検索結果の表示順に反映するシステムの有効性が検証された。

図 2.1 に占めたように以上述べた原田らの研究 [2][3] は図書の NDC 分類、貸出日付を重視して図書の推薦を行ったが、利用者の属性が考慮されないため、どの利用者に対しても同じ結果が推薦される。一方、當山の研究 [8] は利用者の所属分野を重視して検索結果のリランキングを行うが、図書の属性（NDC 番号、貸出日付）が考慮されない、書籍一冊一冊をリランキングするため、計算量が膨大になり、新入館の書籍に漏れを発生する可能性が多い。いずれの研究も結果の有効性を主張しているが、貸出履歴を活用するための属性の適切な組み合わせはまだ十分に把握できていないと考えられる。本研究では、當山の主張する所属分野を考慮に入れた検索システムを実現するが、そのために、原田らの利用した貸出履歴に含まれる図書の NDC 番号を活用する。

	貸出履歴	利用者属性(分野)	図書の属性	
			分野(図書の NDC 区分)	貸出日付
原田ら	○	×	○	○
當山	○	○	×	×

図 2.1: 関連研究比較

第3章 貸出履歴の分析

本章では、学生利用者の属性と図書属性の関係を明らかにするため行った貸出履歴の分析について述べる。

大学に所属する学生が大学図書館を利用する目的は主に学習・研究と考えられるため、その所属分野と大学図書館に求める情報の傾向には関係性があると考えられる。この関係を明らかにするため、A大学の図書館の貸出履歴から学生の分野情報と貸出書籍のNDC番号を抽出し、各分野に所属している学生の傾向を分析した。

3.1 分析対象

本研究の分析対象は、A大学の図書館の2006年4月から2012年3月までの貸出履歴である。データの項目には、利用者番号、所属、資料番号、資料名、著者名、貸出区分、所在コード、請求記号、貸出種別（貸出の種類。“一般貸出”、“相互貸借貸出”など）がある。

本研究において、利用するデータの抽出を行った。利用者のうち、学部学生の一般貸出のデータを抽出し、今回の分析で対象とするデータと定めた。このデータのうち、請求記号が当てられない図書はNDC番号の付与が困難であるため、分析対象外とした。また、図書の利用のみを分析するため、視聴覚資料も分析対象外とした。

3.2 分析方法

3.2.1 NDC分類について

NDC(日本十進分類法)とは、図書館の蔵書を主題別に配架するために日本図書館協会が定めた図書分類法である。一般に流通している書籍のほとんどには国立国会図書館・公立図書館・MARC提供企業によってNDC番号が付与されている。また、国立国会図書館はデジタルアーカイブポータル「PORTA」というWebサービスにおいてISBNなどの書籍の識別番号をキーとしてNDCを含む書誌情報を検索できるAPIを提供しており、図書館の環境によらずにNDCを取得することも可

能である。そのため、NDCは書籍の内容を容易に把握する特徴データとして活用できる可能性がある。NDCの第2次区分までの分類番号と主題を表3.1に示す。

表 3.1: NDC(日本十進分類表)2次区分表

00	総記	10	哲学
01	図書館	11	哲学各論
02	図書雑誌学	12	東洋思想
03	百科事典	13	西洋哲学
04	一般論文 講演集	14	心理学
05	逐次刊行物	15	倫理学 道徳
06	団体	16	宗教
07	新聞	17	神道
08	叢書 全集 選集	18	仏教
09	貴重書 郷土資料	19	キリスト教
20	歴史	30	社会科学
21	日本史	31	政治
22	アジア史 東洋史	32	法律
23	ヨーロッパ史 西洋史	33	経済
24	アフリカ史	34	財政
25	北アメリカ史	35	統計
26	南アメリカ史	36	社会
27	オセアニア史 両極地方史	37	教育
28	伝記	38	風俗習慣 民俗学 民族学選集
29	地理 地誌 紀行	39	国防 軍事
40	自然科学	50	技術 工学
41	数学	51	建設工学 土木工事
42	物理化学	52	建築学
43	化学	53	機械工学 原子力工学
44	天文学 宇宙科学	54	電気工学 電子工学
45	地球学 地学	55	海洋工学 船舶工学 兵器
46	生物科学 一般生物学	56	金属工学 鋳山工学
47	植物学	57	化学工業
48	動物学	58	製造工業
49	医学 医療学	59	家政学 生活科学
60	産業	70	芸術 美術
61	農業	71	彫刻

62	園芸	72	絵画 書道
63	蚕糸業	73	版画
64	畜産業. 獣医学	74	写真 印刷
65	林業	75	工芸
66	水産業	76	音楽 舞踊
67	商業 新聞	77	演劇 映画
68	運輸 交通	78	スポーツ 体育
69	通信事業	79	諸芸 娯楽
80	言語	90	文学
81	日本語	91	日本文学
82	中国語 その他の東洋の諸言語	92	中国文学 その他の東洋文学
83	英語	93	英米文学
84	ドイツ語	94	ドイツ文学
85	フランス語	95	フランス文学
86	南スペイン語	96	スペイン文学
87	イタリア語 両極地方史	97	イタリア文学
88	ロシア語	98	ロシア ソビエト文学
89	その他の諸言語	99	その他の諸文学

3.2.2 分析内容

分析は分野別の貸出傾向と同じ分野の時期による利用者傾向の変化の2つを行った。貸出履歴データを学科ごとに貸出件数、図書のNDC分類別貸出件数を集計し、専攻分野と貸出書籍のNDC分類の関係と特徴を分析する。また、貸出状況と貸出時期の関係と特徴を明らかにするため、学科ごとに月ごとの貸出件数、図書のNDC分類別貸出件数について集計し、時期による利用傾向の変化を分析した。

3.3 分析結果

3.3.1 学科ごとの利用傾向

学科ごとの利用傾向を分析した結果、学生の所属する学科と貸出書籍の主題には密接な関係があることがわかった。さらに、これらは、あるNDC番号の書籍が圧倒的に多い学科と貸出書籍が多様なNDC番号を持つ学科の2つに分けられることが分かった。以下では、貸出書籍の各NDC番号の割合をグラフにしたものを示して、これらの傾向を説明する。

1. 貸出書籍のNDC番号が多様である学科：

図3.1と図3.2にa学科2011年度と2011年12月のa学科における貸出書籍

のNDC第1次区分の割合である。例えばNDC10とあるのはNDCの第1次区分に分類される書籍を表す。このa学科では、2類の書籍の貸出が全体の約3割を占めているが、0類、1類、3類などの貸出件数が多い。また図3.3と3.4のように、b学科の貸出書籍は0類が1/4を占めている他に、多くの分類にわたって、貸出されていることが分かった。このような学科は全部9学科であった（他に図3.5～図3.13）。

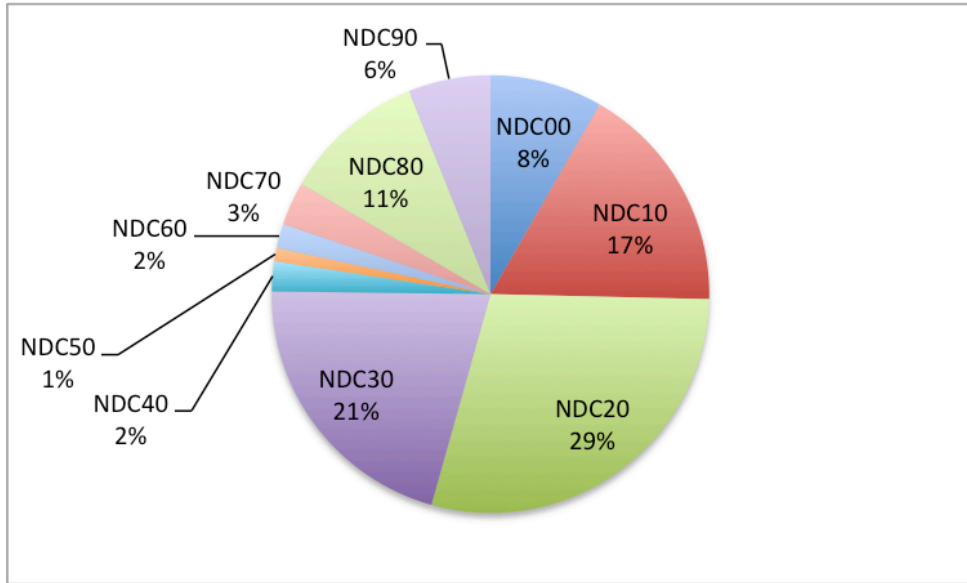


図 3.1: 貸出書籍の NDC1 次分類の割合, a 学科 2011 年度

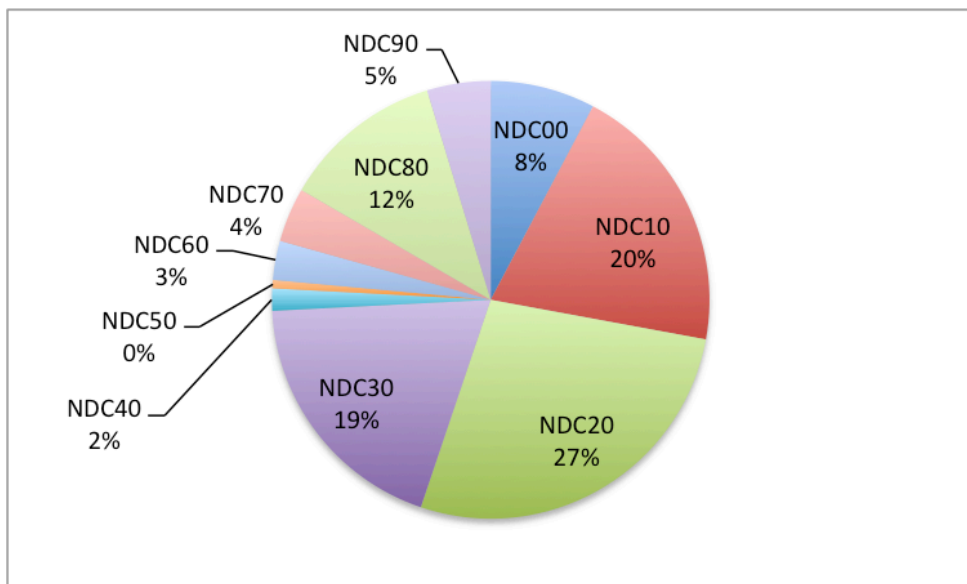


図 3.2: 貸出書籍の NDC1 次分類の割合, a 学科 2011 年 12 月

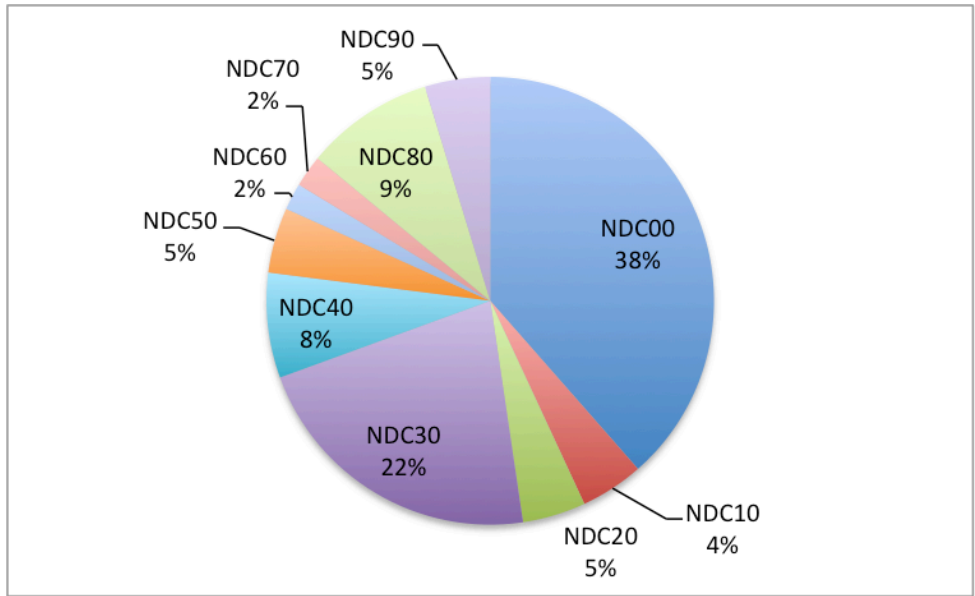


図 3.3: 貸出書籍の NDC1 次分類の割合. b 学科 2011 年度

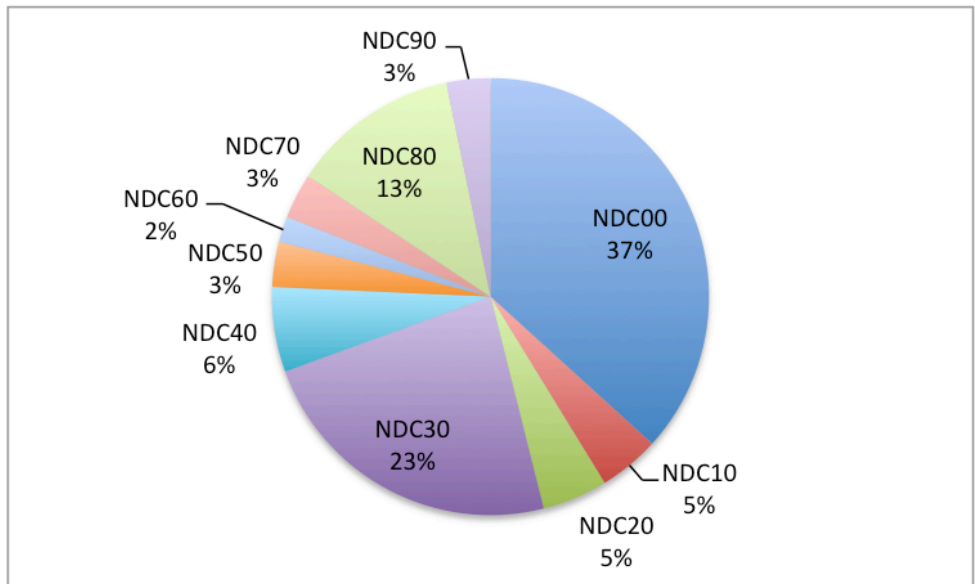


図 3.4: 貸出書籍の NDC1 次分類の割合. b 学科 2011 年 12 月

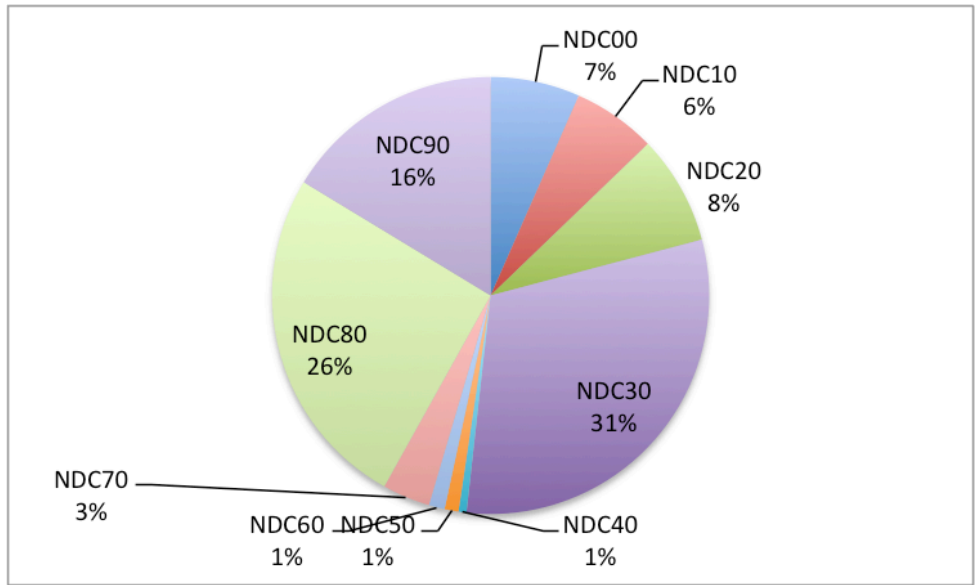


図 3.5: 貸出書籍の NDC1 次分類の割合, d 学科 2011 年 12 月

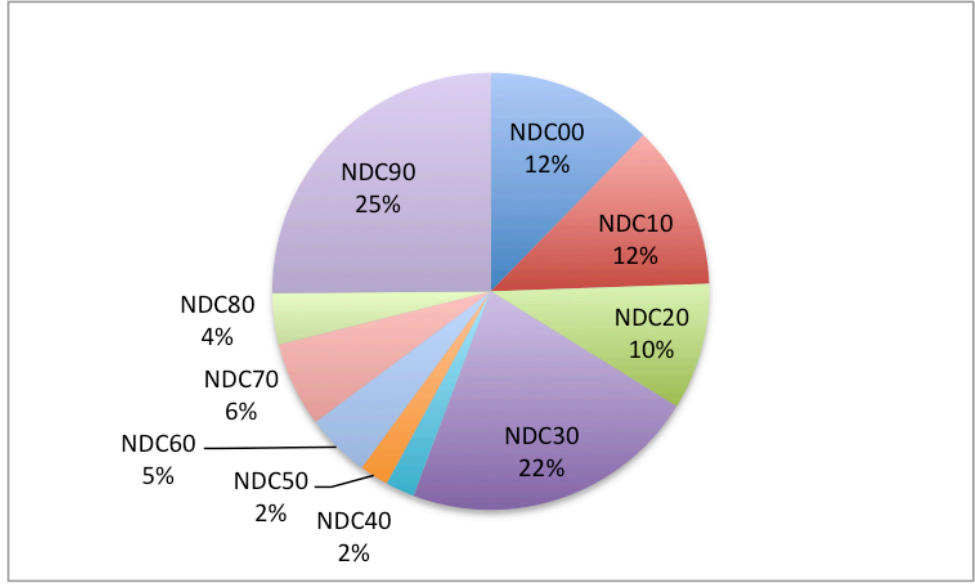


図 3.6: 貸出書籍の NDC1 次分類の割合, h 学科 2011 年 12 月

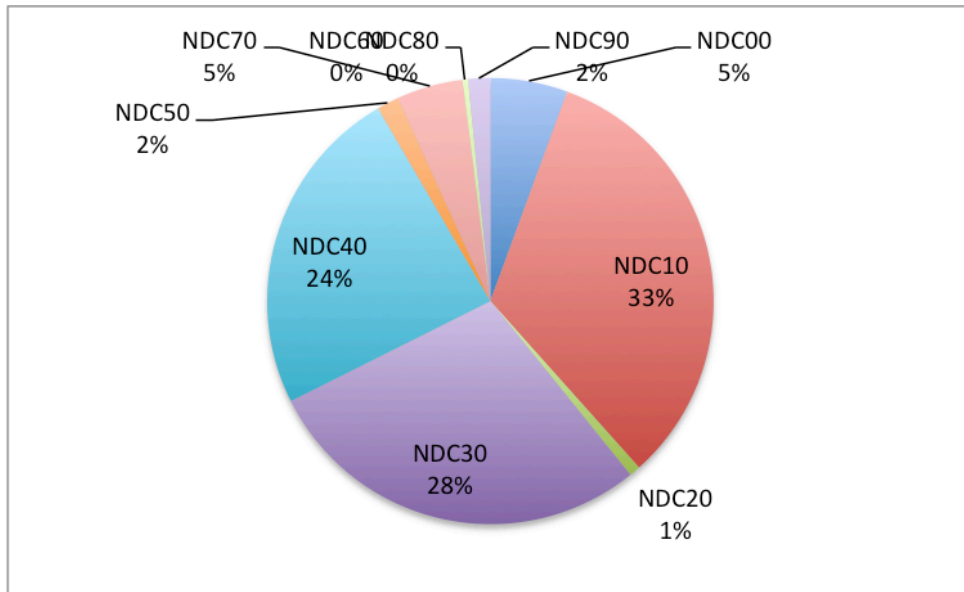


図 3.7: 貸出書籍の NDC1 次分類の割合, m 学科 2011 年 12 月

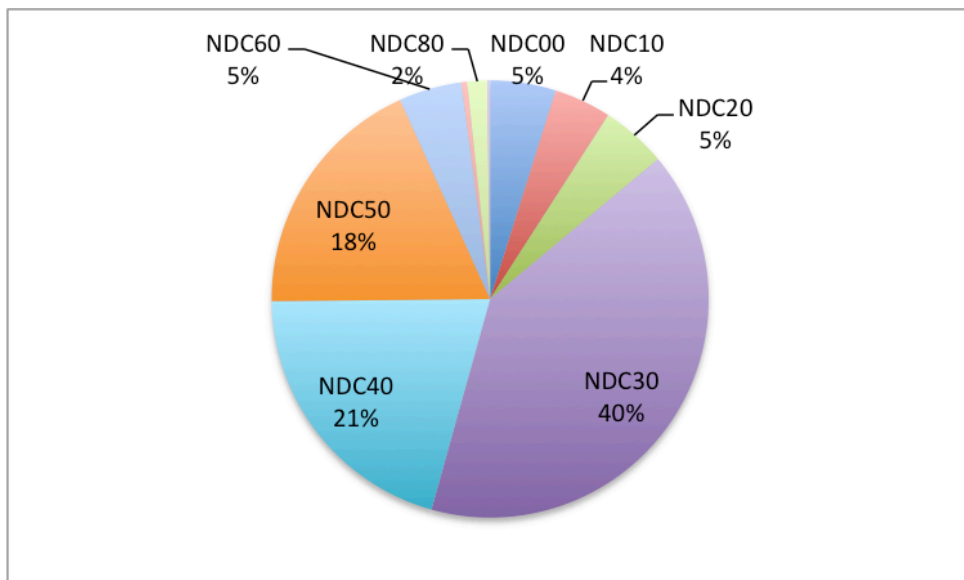


図 3.8: 貸出書籍の NDC1 次分類の割合, v 学科 2011 年 12 月

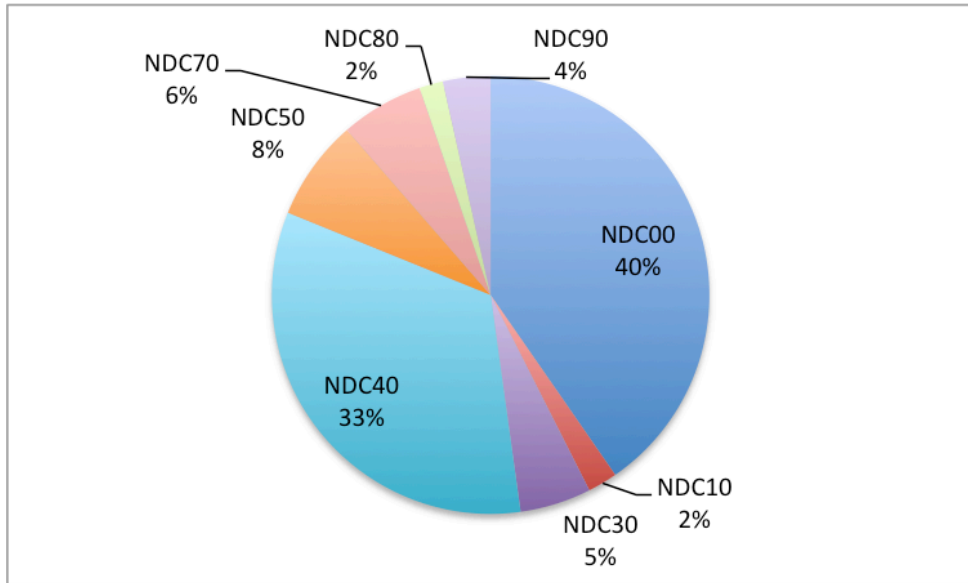


図 3.9: 貸出書籍の NDC1 次分類の割合, w 学科 2011 年 12 月

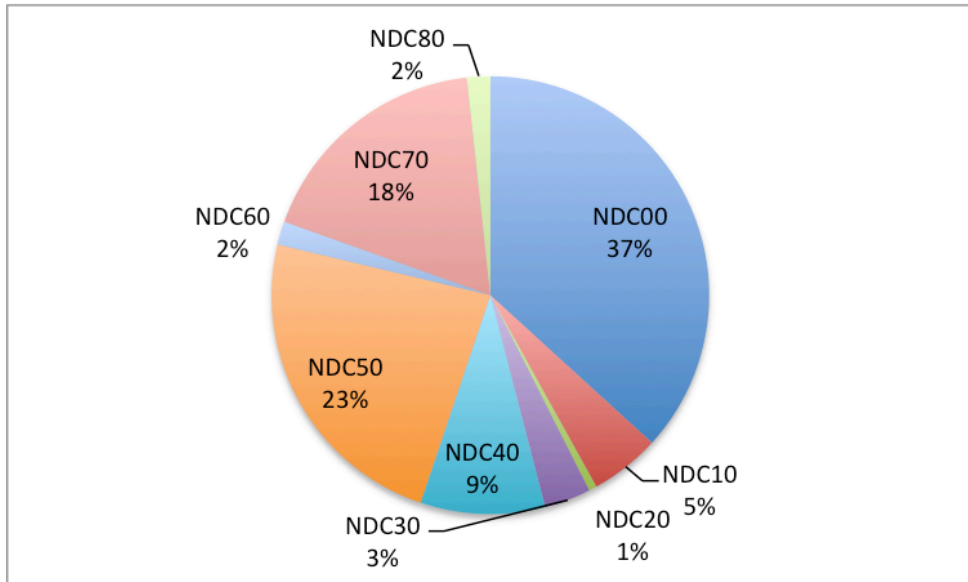


図 3.10: 貸出書籍の NDC1 次分類の割合, x 学科 2011 年 12 月

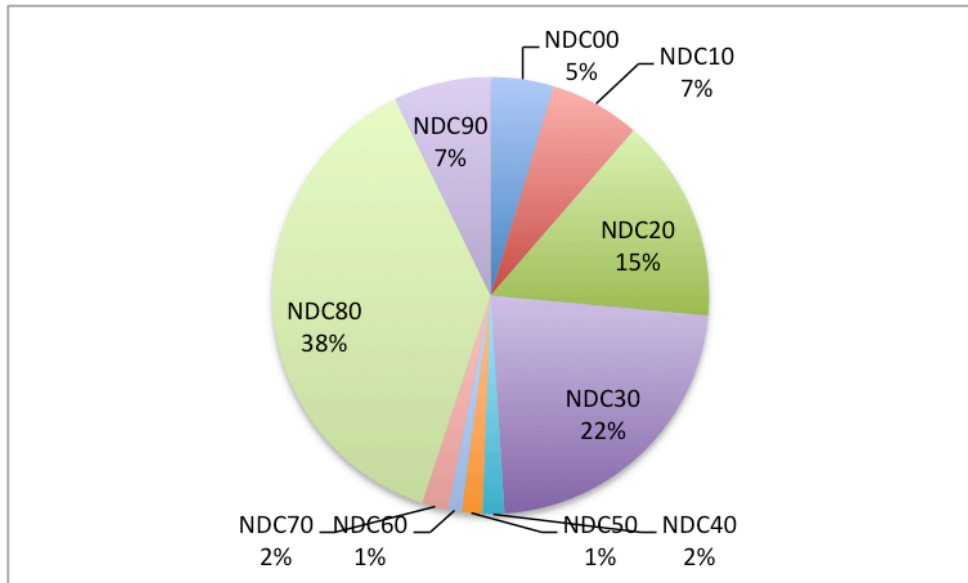


図 3.11: 貸出書籍の NDC1 次分類の割合, i 学科 2011 年 12 月

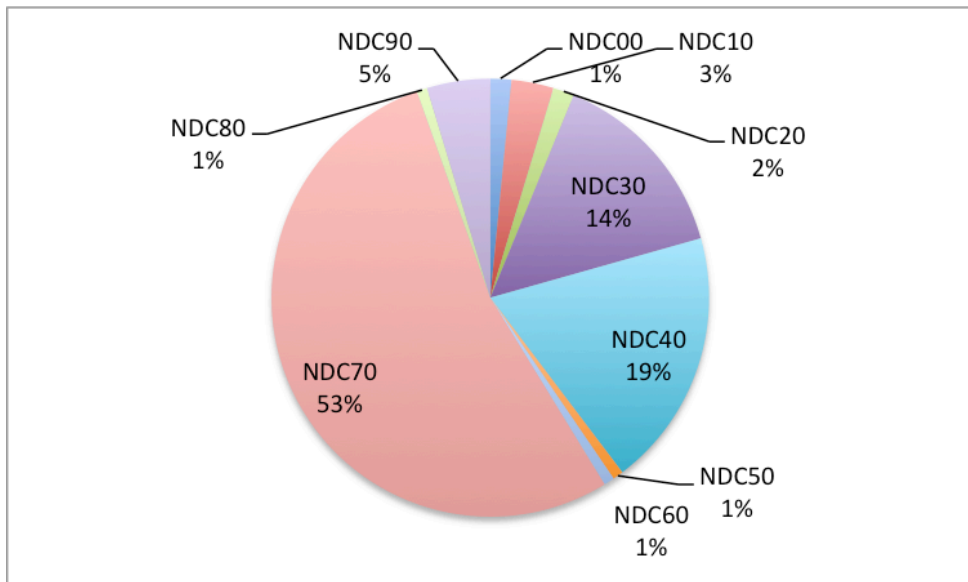


図 3.12: 貸出書籍の NDC1 次分類の割合, y 学科 2011 年 12 月

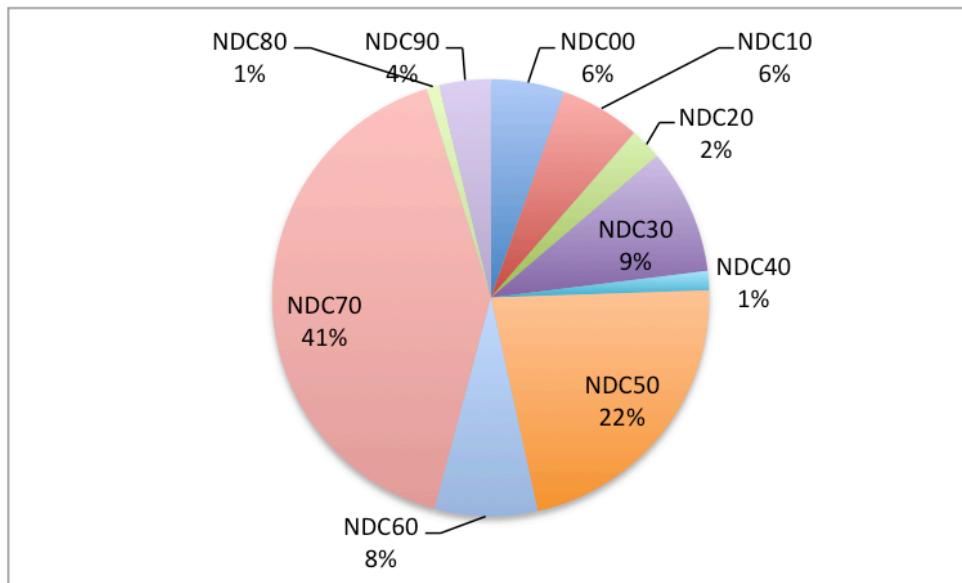


図 3.13: 貸出書籍の NDC1 次分類の割合, z 学科 2011 年 12 月

2. ある NDC 番号の貸出書籍が圧倒的に多い学科 :

図 3.14 と図 3.15 に c 学科の 2011 年度と 2011 年 12 月の貸出書籍の割合を示した。4 類の図書の貸出が全体の約 7 割を占めており、NDC4 類の貸出件数が圧倒的に多い。このような学科は全部 12 学科であった（他に図 3.16～図 3.27）。

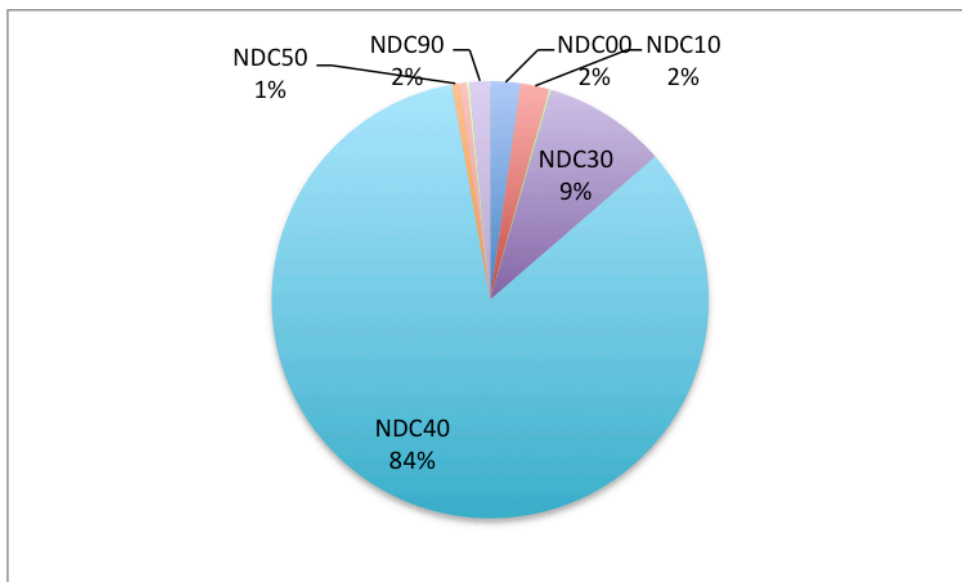


図 3.14: 貸出書籍の NDC1 次分類の割合, c 学科 2011 年度

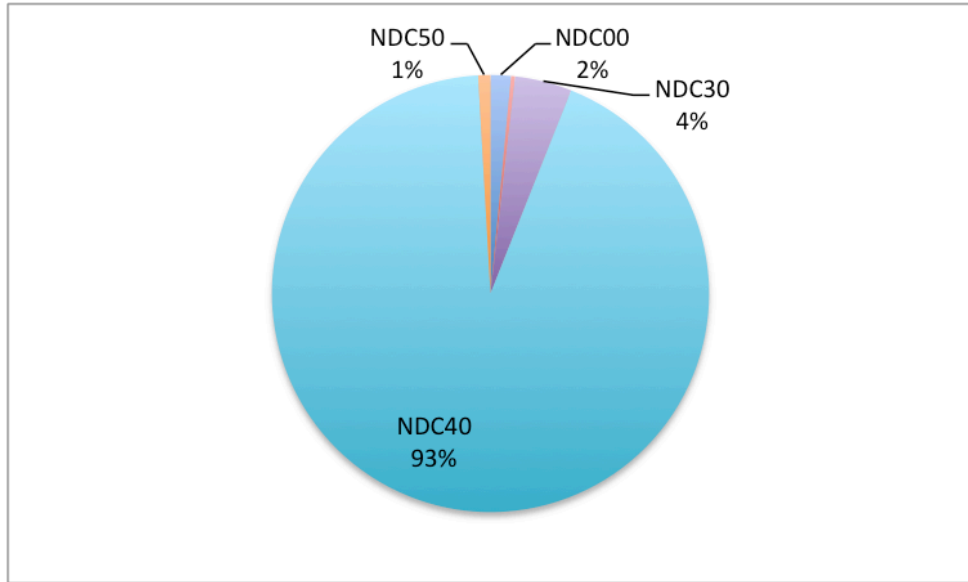


図 3.15: 貸出書籍の NDC1 次分類の割合. c 学科 2011 年 12 月

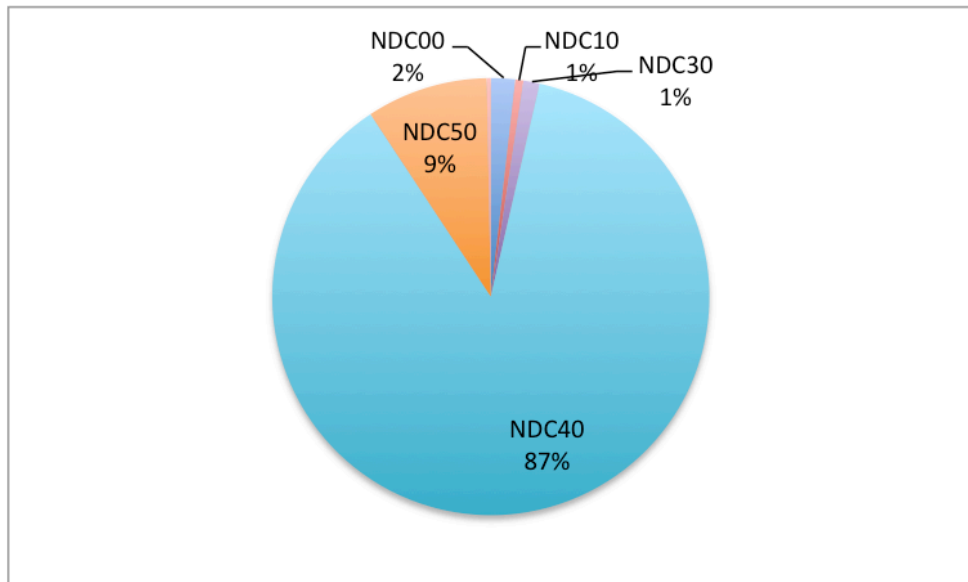


図 3.16: 貸出書籍の NDC1 次分類の割合. f 学科 2011 年 12 月

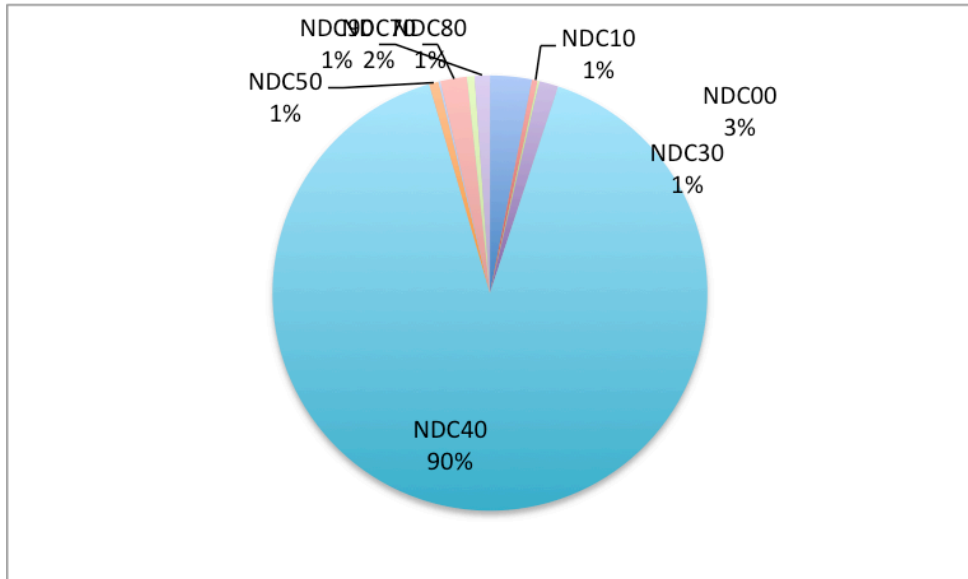


図 3.17: 貸出書籍の NDC1 次分類の割合. g 学科 2011 年 12 月

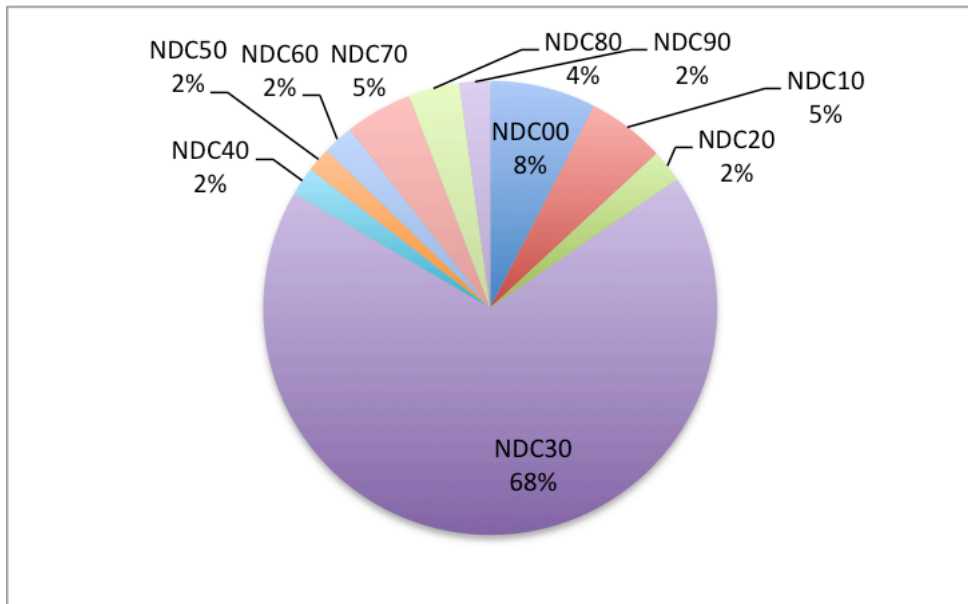


図 3.18: 貸出書籍の NDC1 次分類の割合. j 学科 2011 年 12 月

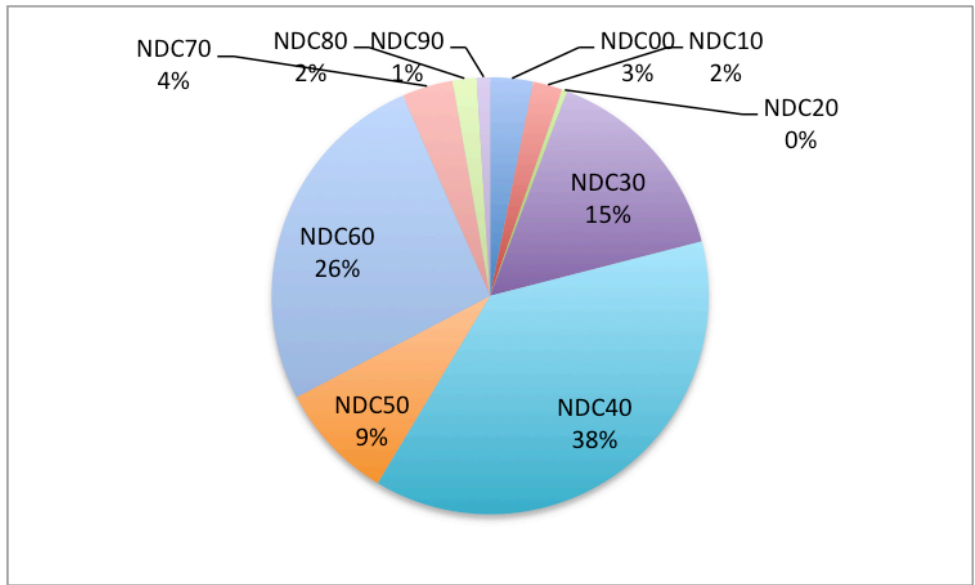


図 3.19: 貸出書籍の NDC1 次分類の割合. p 学科 2011 年 12 月

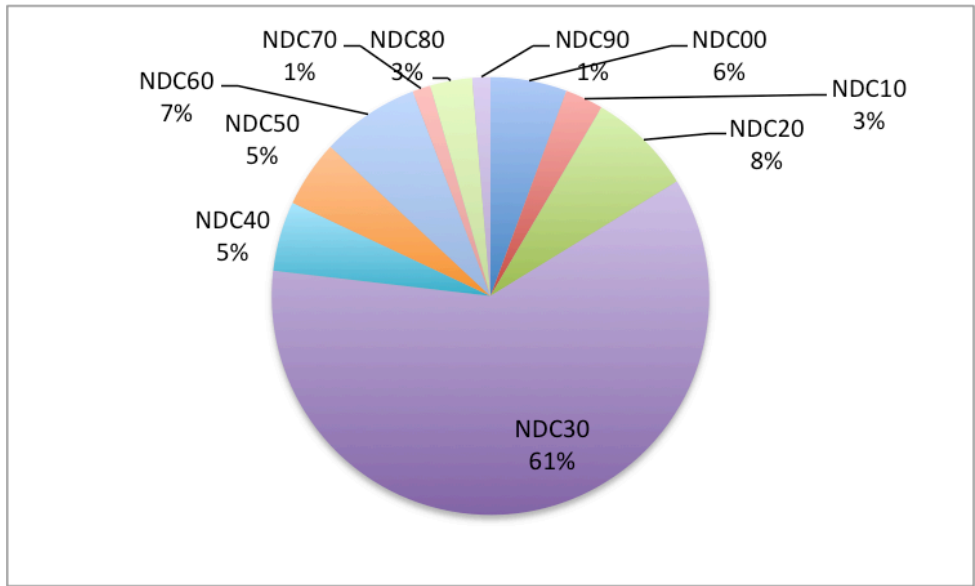


図 3.20: 貸出書籍の NDC1 次分類の割合.k 学科 2011 年 12 月

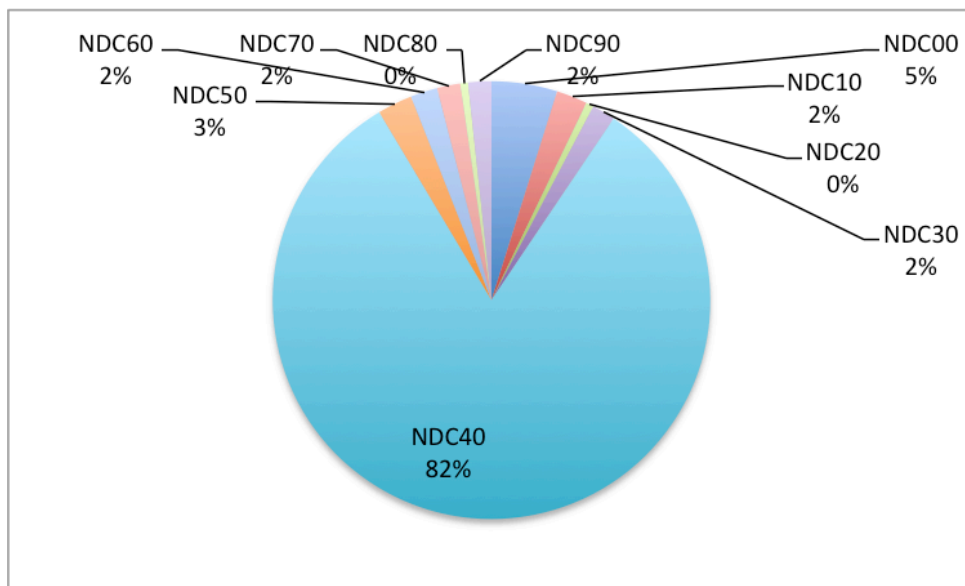


図 3.21: 貸出書籍の NDC1 次分類の割合, o 学科 2011 年 12 月

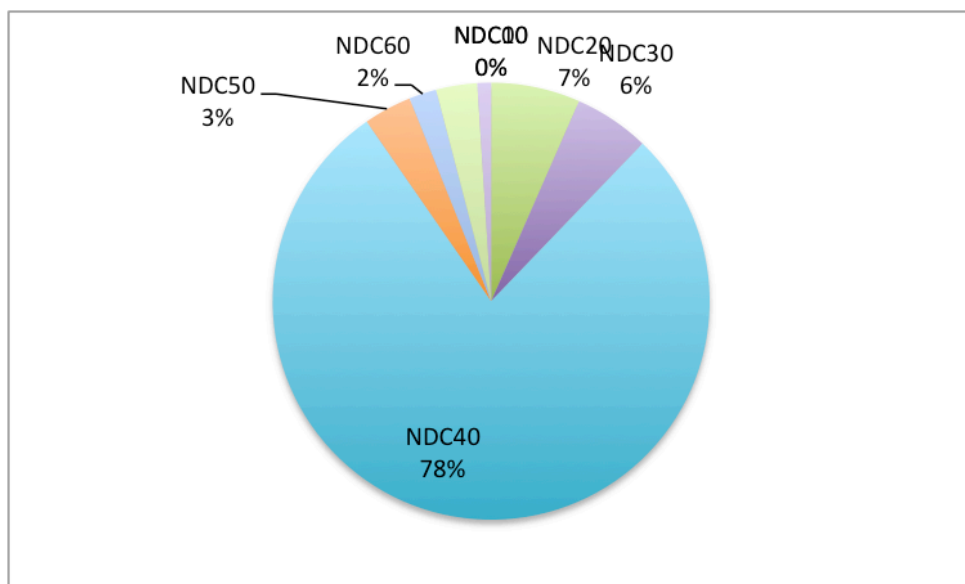


図 3.22: 貸出書籍の NDC1 次分類の割合, q 学科 2011 年 12 月

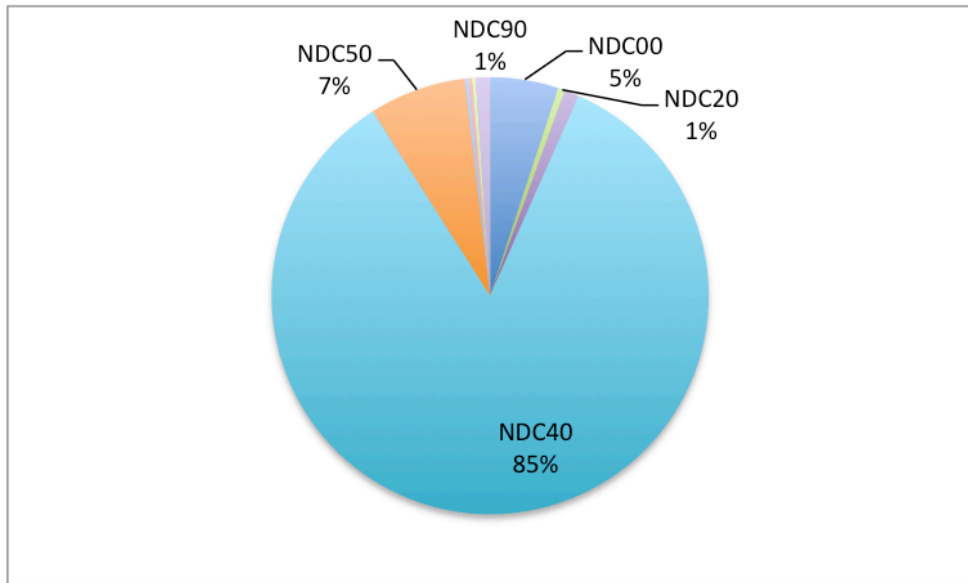


図 3.23: 貸出書籍の NDC1 次分類の割合, r 学科 2011 年 12 月

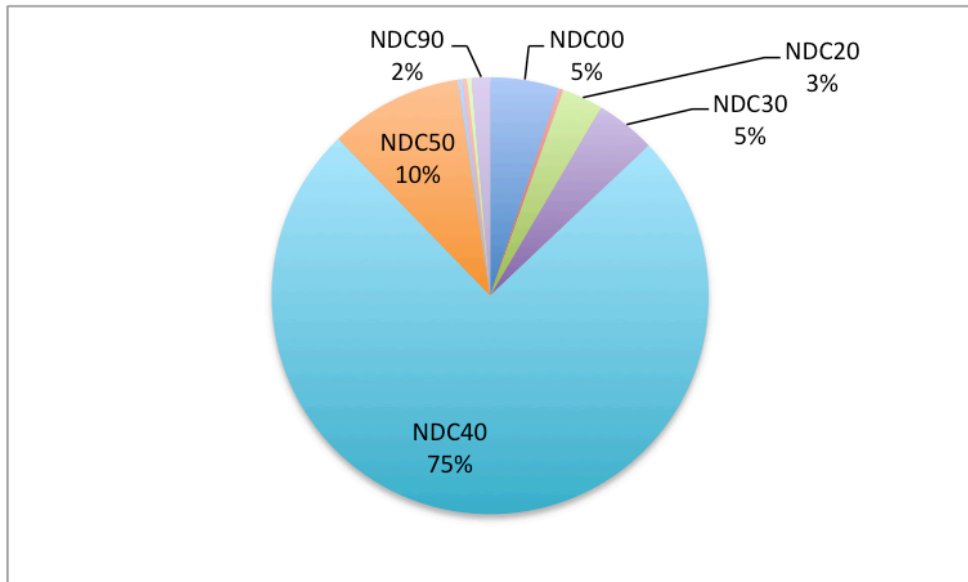


図 3.24: 貸出書籍の NDC1 次分類の割合, s 学科 2011 年 12 月

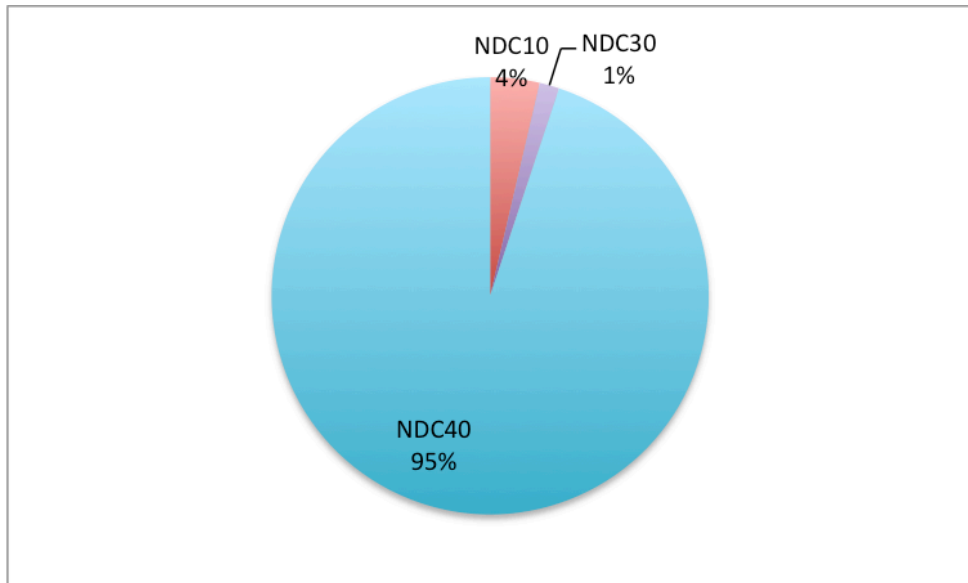


図 3.25: 貸出書籍の NDC1 次分類の割合, t 学科 2011 年 12 月

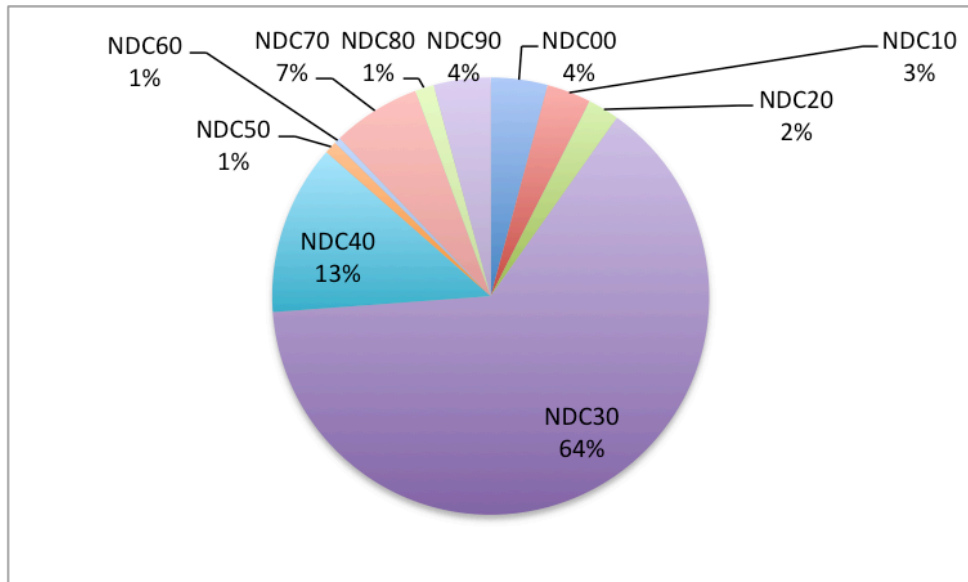


図 3.26: 貸出書籍の NDC1 次分類の割合, n 学科 2011 年 12 月

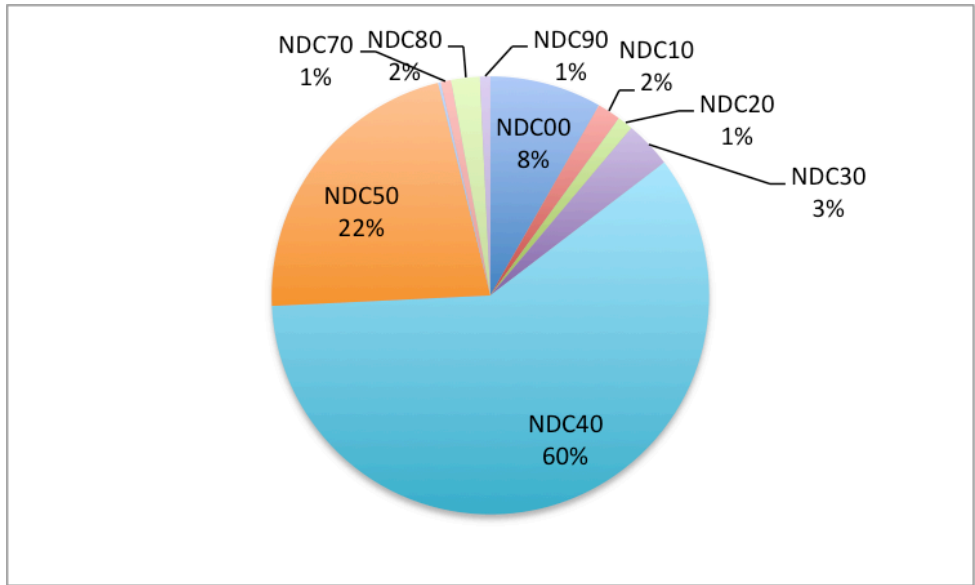


図 3.27: 貸出書籍の NDC1 次分類の割合, u 学科 2011 年 12 月

図 3.28 と図 3.30 のように、X 学部の e 学科と f 学科どちらも 4 類の図書の貸出件数が圧倒的に多く、全体の 8 割以上を占めている。しかし NDC の 2 次区分で分析すると、図 3.29 と図 3.31 のに示した結果になった。e 学科は貸出した 41 類（数学）の図書が 4 類図書の 9 割以上を占めている。f 学科は貸出した 43 類（化学）の図書が 4 類図書の 7 割を占めている。これは NDC の 1 次区分だけでなく 2 次区分まで見なければ、学科の傾向を正確に捉えることかできないことを示している。

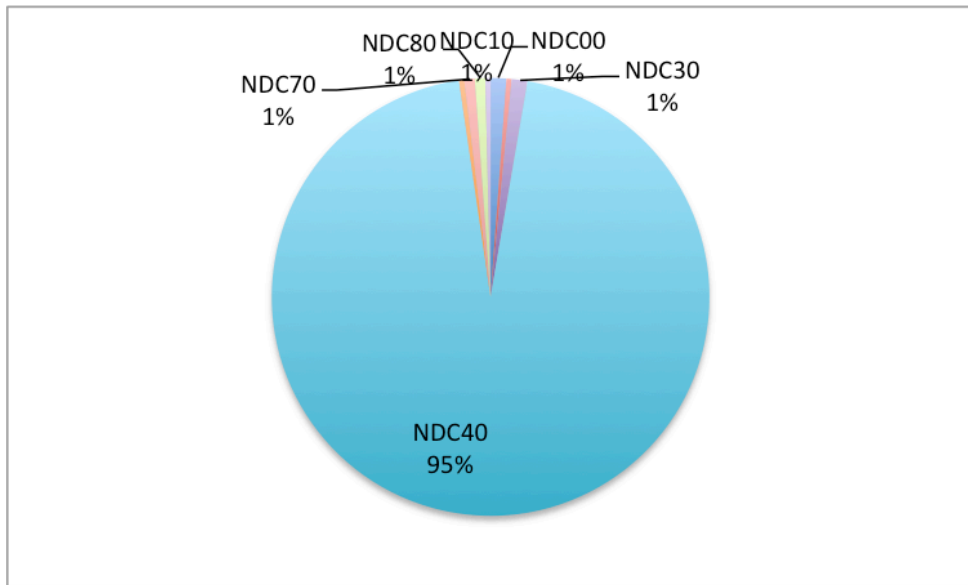


図 3.28: 貸出書籍の NDC1 次分類の割合, e 学科 2011 年 12 月

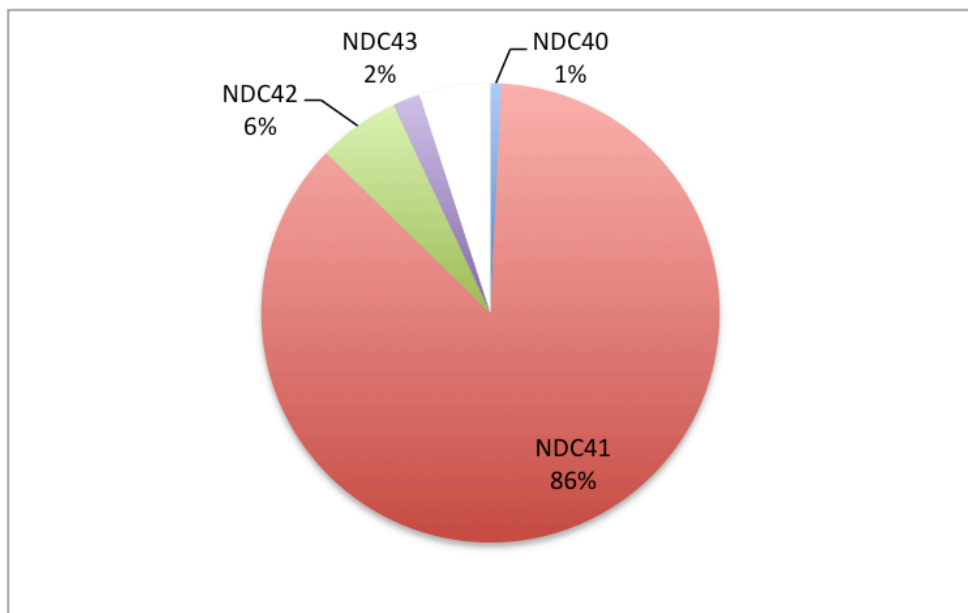


図 3.29: 貸出書籍の 4 類のみの NDC2 次分類の割合, e 学科 2011 年 12 月

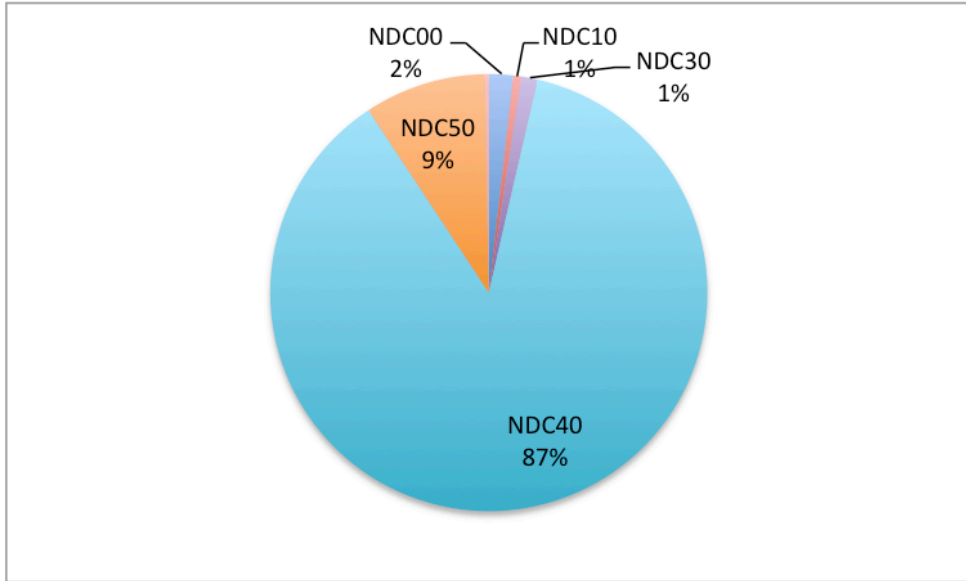


図 3.30: 貸出書籍の NDC1 次分類の割合, f 学科 2011 年 12 月

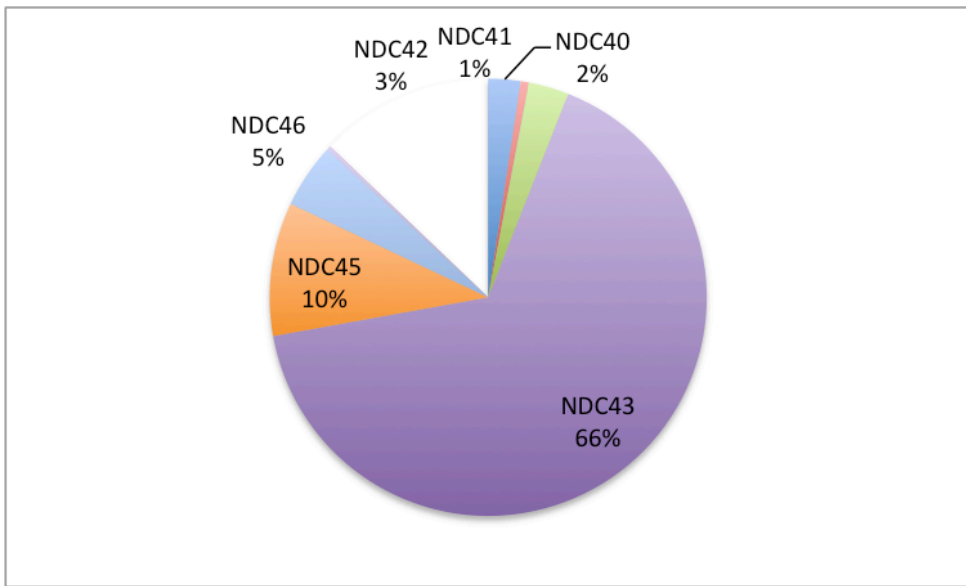


図 3.31: 貸出書籍の 4 類のみの NDC2 次分類の割合, f 学科 2011 年 12 月

3.3.2 時期による利用傾向の変化

図 3.32 次に時期による利用傾向の変化をみた。まず、貸出件数変動である。A 大学 2011 年度各月の貸出総件数の変化を示した。6 月と 11 月の貸出件数が多いが、休み時期（8 月、3 月）の貸出件数が少ない。

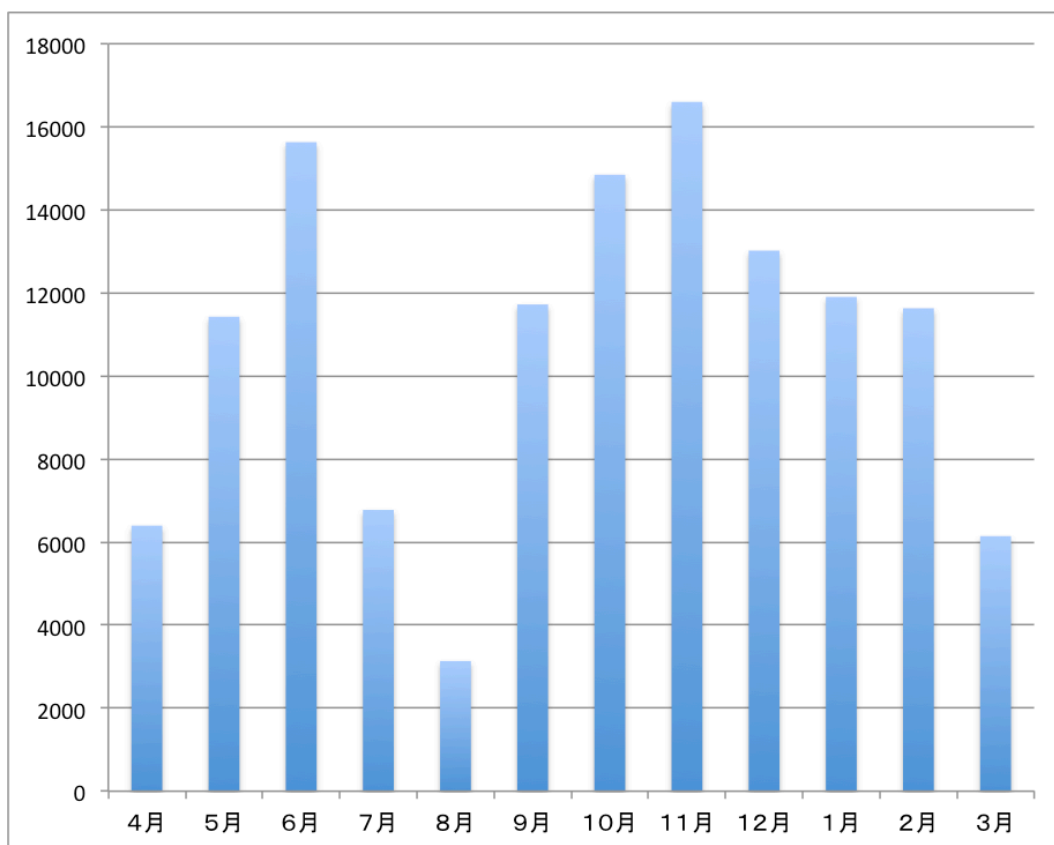


図 3.32: A 大学 2011 年度学科生各月の貸出件数

1. 時期により変化がほとんどない

図 3.33 に c 学科 2011 年度月ごとの利用傾向の変化を示した。図の通り、c 学科の貸出傾向はは月によりあまり変化がない。

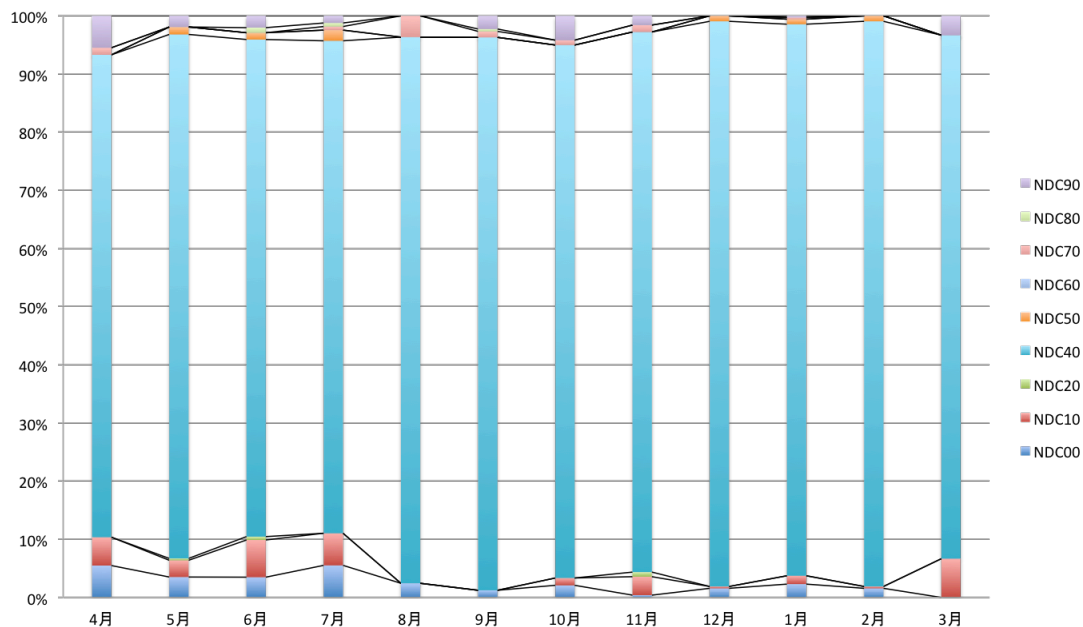


図 3.33: c 学科 2011 年度月ごとの利用傾向の変化

2. 時期により変化が大きい

図 3.34 に b 学科 2011 年度月ごとの利用傾向の変化を示した。図の通り、b 学科の貸出傾向はは月ごとにより変化が大きかった。

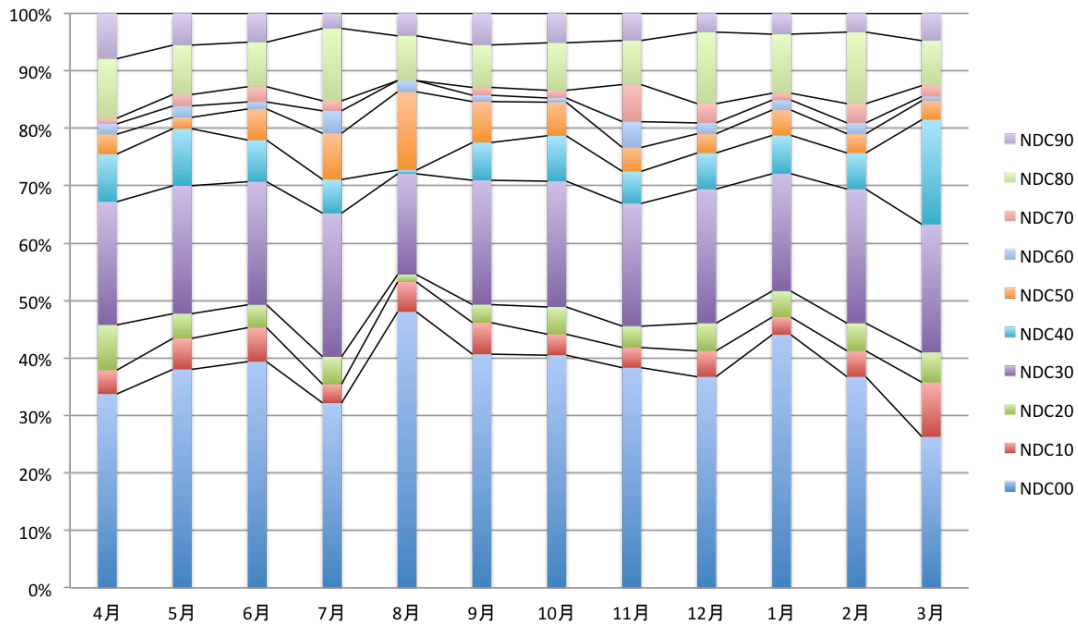


図 3.34: b 学科 2011 年度月ごとの利用傾向の変化

第4章 貸出履歴を利用した蔵書検索結果のリランキング手法

本章では第3章で分析した結果を利用した蔵書検索結果のリランキング手法について述べる。

4.1 リランキング手法概要

第3章分析の結果を利用し、次の手法を提案する。まず利用者を所属分野でグループ化する（利用者のグループ化）。次に、グループに所属する利用者の履歴を用いてグループのプロファイルを作成する（利用者プロファイルの作成）。最後に、作成したプロファイルに基づいて通常の検索結果の各資料に重み付けを行い、ランキングを修正する（検索結果のリランキング）。図4.1にリランキング手法の概略図を示した。

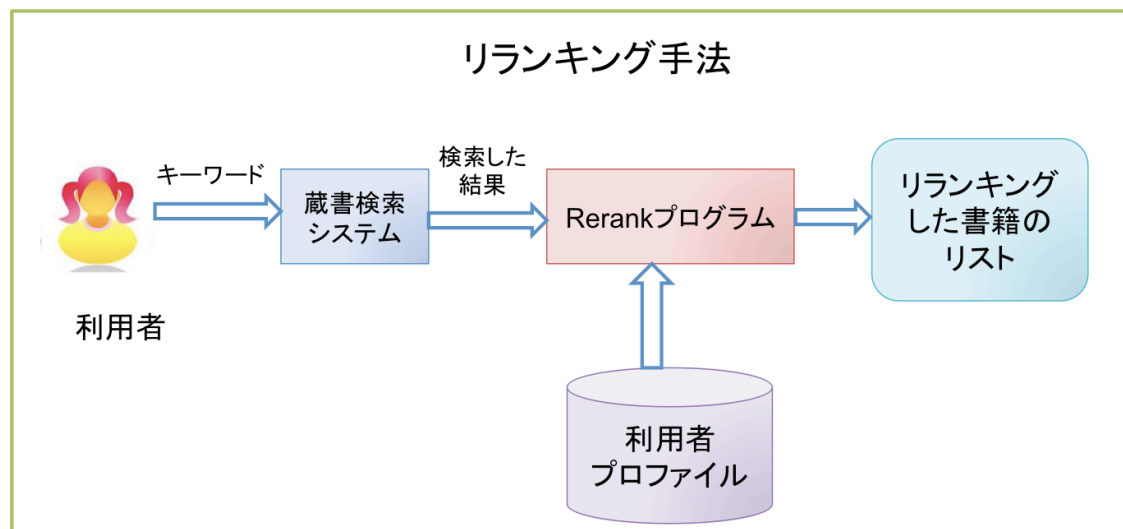


図 4.1: リランキング手法

4.2 利用者のグループ化

第3章の分析により学生の所属する分野と借り出す書籍のNDC区分には関連があることが明らかになった。そこで、利用者を所属分野毎にグループ分けをし、それぞれでプロファイルを作成することを考える。まず、図4.2に示した通り、利用者を所属分野としてグループに分ける。ここで、所属分野は各学生の学科とする。

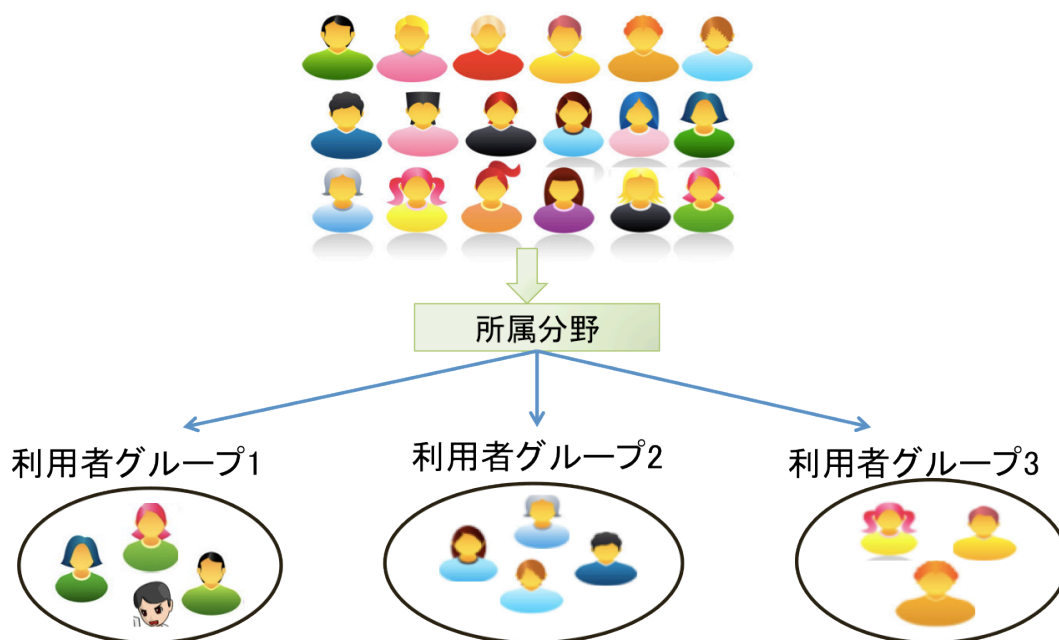


図 4.2: 利用者をグループ化する

4.3 利用者プロフィールの作成

グループ毎に貸出履歴を集計して時期ごとのプロフィールを作成する (図4.3)。利用者プロフィールはNDC分類の割合として作成する。表1はNDCの一次区分のみを考慮に入れた場合のプロファイルの例である。b学科は過去NDC分類「00」の書籍が37%で一番貸出された。c学科はNDC分類「40」の書籍が93%で一番貸出された。

表 4.1: 利用者プロフィールの例、2011 年 12 月分.

NDC	00	10	20	30	40	50	60	70	80	90
学科 b	37%	5%	5%	23%	6%	3%	2%	3%	13%	3%
学科 c	1%	0%	0%	4%	93%	1%	0%	0%	0%	0%

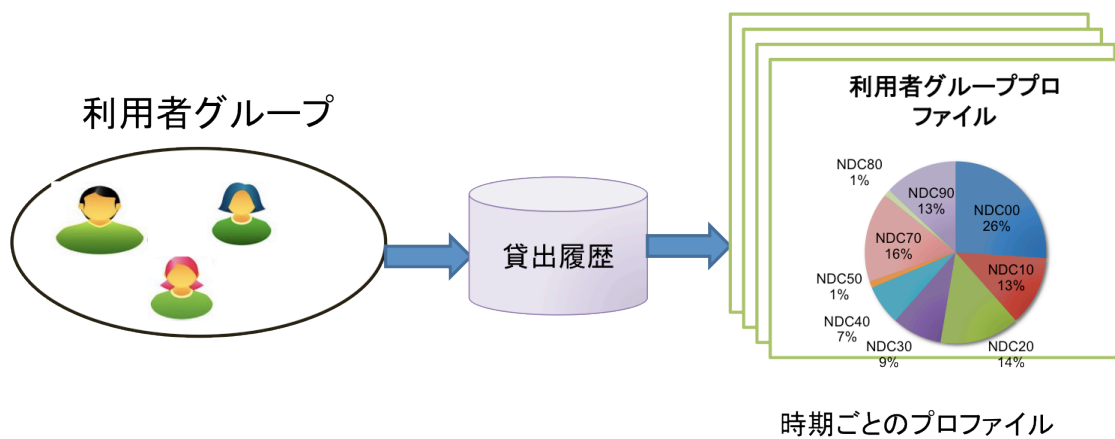


図 4.3: 利用者プロフィールの作成

4.4 検索結果のリランキング

検索結果をリランキングする方法を図4.4に示す。検索結果の書籍のNDC分類番号を抽出し、利用者グループの貸出履歴から作成した利用者グループのプロファイルと貸出時期を基に各書籍の関連度を求め、ランキングを修正する。

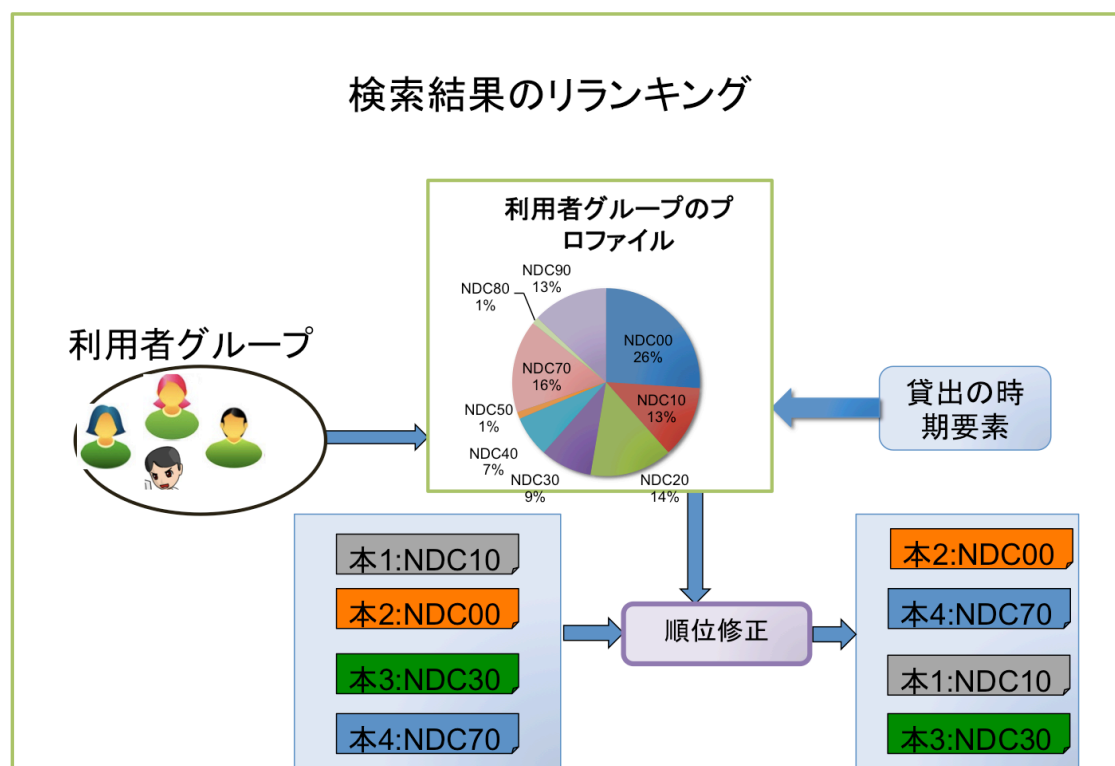


図 4.4: 検索結果のリランキング

4.4.1 書籍のNDC分類番号の抽出

検索結果のリストに付いている書籍の請求番号からNDC分類番号を一つずつ抽出する。請求記号とは、分類番号と著者記号（著者名などの頭文字アルファベット1文字）と巻数・年数表示などを合わせたものである。NDC分類番号は2次区分まで抽出する。例えば、書籍「プログラミング言語Ruby」の場合、007.64-F31という請求番号から00というNDC分類を抽出する。

4.4.2 関連度の計算

各書籍の関連度は蔵書検索システムが出力した順位に基づく「初期関連度」と利用者プロフィールから得られる各書籍のNDC番号の「貸出割合」の積で求める。

$$\text{関連度} = \text{初期関連度} * \text{貸出割合} \quad (4.1)$$

ここで、初期関連度は蔵書検索システムが出力したランキング結果の順位(rank)から、以下の数式で求める。

$$\text{初期関連度} = 100 \div \text{rank} \quad (4.2)$$

また、貸出割合は各書籍に付与されているNDC番号に対応する貸出割合を利用者プロフィールから抽出する。最後に関連度の大きい順に書籍を並べ替えることでリランキングを完成する。

第5章 実験と評価

本章では、提案手法の有効性を明らかにするための評価実験について述べる。

5.1 評価用正解データの作成

5.1.1 概要

本手法の有効性を評価するために、キーワードとそれに対する正解集合のデータを作成した。具体的な手順は、まず実験協力者を募り、キーワードを考えてもらった上で、こちらの用意した評価データ収集用の検索システムで検索してもらった。次に出力された結果に対する評価値を付与してもらった。

5.1.2 実験協力者

A 大学に在籍している学生 9 名に協力してもらった。これら実験協力者の内訳は b 学科に所属する 3 年生 5 名、c 学科に所属する 2 年生 2 名である。

5.1.3 実験に用いたデータ

本実験では b 学科と c 学科に所属している学生の A 大学図書館での貸出履歴を利用した。期間は 2011 年 4 月から 2012 年 3 月まで一年間である。なお、NDC 番号を判定するために、請求記号が当てられない書籍は分析対象外とした。また、書籍のみを分析するために、視聴覚資料も分析対象外とした。実験に用いた専攻分野と貸出件数を表 5.1 に示した。

表 5.1: 実験に用いた専攻分野と貸出件数

所属	貸出件数
b 学科	8564
c 学科	3397

5.1.4 データ収集用システムの構築

A大学の図書館の蔵書検索システムは、APIを提供している。本研究ではこのAPIを使って評価用正解データ収集システムを実装した。

システムのインターフェスを図5.1に示した。名前、学籍番号を入力し、所属を選択した上で、3つのキーワードを入力し検索ボタンを押す。ここではキーワード1に対する検索結果が表示される。提示される検索結果は1件毎に、書名、著者、出版年、表紙画像が、図5.2のように提示される。各書籍は5段階で評価ができるようになっている。すべての評価を行った後で、検索ボタンをクリックすることでキーワード2の検索結果が表示される。同様にして、キーワード2の結果、キーワード3の結果の評価を行うように誘導する。

*****書籍検索システム*****

-----基本情報-----

名前:

学籍番号:

所属を選択してください:

-----キーワード-----

キーワード1:

キーワード2:

キーワード3:

-----検索スタート-----

これから検索をスタートします。“検索”ボタンをクリックしてください、入力したキーワード1の検索結果が提示されます。

図 5.1: 実験用システムインターフェス

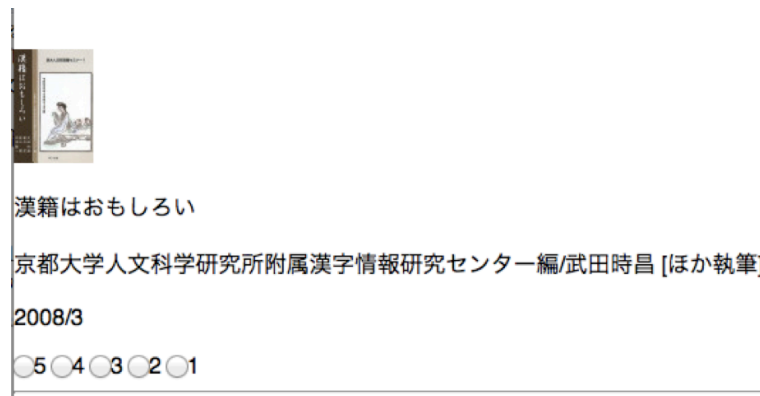


図 5.2: 提示された書籍項目

5.1.5 データ収集の手順

まず実験協力者はキーワードを利用して評価データ収集システムで検索を行ってもらった。実験協力者が2013年12月時点を想定して自分の学科の授業の予習復習やレポート、あるいは卒業研究に関する書籍を探すためのキーワードを想定してもらい、検索を行ってもらった。

次に検索した結果に対して興味の度合いを評価してもらった。図5.2に示した評価項目に出力された検索結果（最大100件）に対して、想定した「自分の学科の授業の予習復習やレポート、あるいは卒業研究に関する書籍」としての必要な度合いを5段階で付けてもらった。5は必要性が一番高い書籍とし、1は必要性が一番低い書籍とした。

5.2 評価

5.2.1 比較手法

次の3つの手法による検索結果を比較評価する。

1. OPAC-A：A大学図書館が現在使用している蔵書検索システム
2. RERANKAVR：本研究で提案したリランキングプログラム（2011年度一年間の貸出履歴データ）
3. RERANK12：本研究で提案したリランキングプログラム（2011年12月分一ヶ月分の貸出履歴データ）

5.2.2 評価ツールと指標

本研究では情報検索システム評価用テストコレクション構築プロジェクト¹で提供されているツールキット NTCIREVAL²を用いた。ここで出力される指標のうち平均精度を用いた。

5.2.3 評価用正解データ

作成した評価用正解データはキーワード 17 個である。

表 5.2: 評価用正解データ

所属	実験協力者	KeyID	キーワード	件数
b 学科	user1	KEY1	データマイニング	32
		KEY2	学校図書館	100
	user2	KEY3	Ruby	55
		KEY4	システム	100
	user3	KEY5	絵本	100
		KEY6	子供	100
		KEY7	読み聞かせ	28
	user4	KEY8	レファレンスサービス	37
		KEY9	情報	100
		KEY10	図書館	100
c 学科	user5	KEY11	看護過程	100
		KEY12	ヘンダーソン	60
		KEY13	呼吸器	100
		KEY14	肺炎	41
	user6	KEY15	看護記録	31
		KEY16	肺炎	41
		KEY17	貧血	60

¹<http://ntcir.nii.ac.jp/>

²<http://research.nii.ac.jp/ntcir/tools/ntcireval-ja.html>

第6章 結果と考察

本章では実験の結果と考察を述べる。

6.1 平均精度

表 6.1 及び図 6.1 に各キーワードによる検索結果の平均精度を表す。

全体の平均精度は、OPAC-A が 0.5198、RERANKAVR が 0.5510、RERANK12 が 0.5042 であり、RERANKAVR が最も精度が高く、RERANK12 にが最も精度が低かった。

また、学科ごとの平均を見ると、b 学科 (KEY1~KEY10) では RERANKAVR が 0.5944 で最も精度が高かった。RERANK12 が 0.5955 で精度が OPAC-A より高かったが、RERANKAVR より低かった。OPAC-A が 0.5621 で最も精度が低かった。それに対して、c 学科 (KEY11~KEY17) では RERANKAVR と RERANK12 の精度は 0.4889 で同じであった。

次に個別のキーワードごとの精度を見ると、b 学科の場合は本手法 (RERANKAVR、RERANK12) によつて OPAC-A より上がるものもあれば、下がるものもあったが、c 学科の場合、本手法よつて精度が下がるものはなかった。

6.2 考察

表 6.1 で示した通り、本研究の RERANKAVR は蔵書検索システム OPAC-A より精度が高くなった。この結果は b 学科、c 学科に分けて見た場合も同様であった。これはこの 2 学科に関しては本研究が提案する貸出履歴を利用した 1 年分のリランキングが成功したと言える。一方、12 月分の履歴を利用した RERANK12 は b 学科の場合は良い結果を得ることができなかったのに対して、c 学科では RERANKAVR と同等の結果であった。1 つの原因として、年間の利用傾向の変動があげられる。c 学科は図 3.33 が示すように、一年間を通して利用傾向がほとんど変化しない。従つて、RERANK12 と RERANKAVR の結果が変わらなかった。これに対して、b 学科の RERANK12 の結果が悪かった原因は図 3.34 が示すように月によって大きく利用傾向が変わることに加え、学期制の変更があると考えられる。実験協力者に想定してもらった 2012 年 12 月の授業と、利用履歴の 2011 年 12 月の授業の傾向が

違っている可能性がある。1ヶ月分を利用する場合には授業内容も考慮に入れる必要があるだろう。

表 6.1: キーワードごとの平均精度とその平均

keyID	Keyword	OPAC-A	RERANKAVR	RERANK12
KEY1	データマイニング	0.6706	0.5991	0.5991
KEY2	学校図書館	0.7177	0.7845	0.7845
KEY3	Ruby	0.6170	0.7524	0.7524
KEY4	システム	0.1891	0.1595	0.1481
KEY5	絵本	0.2613	0.4198	0.4198
KEY6	子供	0.3313	0.3901	0.3903
KEY7	読み聞かせ	0.6928	0.6928	0.6928
KEY8	レファレンスサービス	0.9730	0.9730	0.9730
KEY9	情報	0.9689	0.9621	0.9624
KEY10	図書館	0.1988	0.2104	0.2104
平均 [key1-key10]		0.5621	0.5944	0.5933
KEY11	看護過程	0.9635	0.9635	0.9635
KEY12	ヘンダーソン	0.4370	0.6192	0.6192
KEY13	呼吸器	0.2095	0.2095	0.2095
KEY14	肺炎	0.2434	0.2463	0.2463
KEY15	看護記録	0.2078	0.2078	0.2078
KEY16	肺炎	0.6516	0.6565	0.6565
KEY17	貧血	0.5029	0.5198	0.5198
平均 [key11-key17]		0.4594	0.4889	0.4889
平均 [key1-key17]		0.5198	0.5510	0.5503

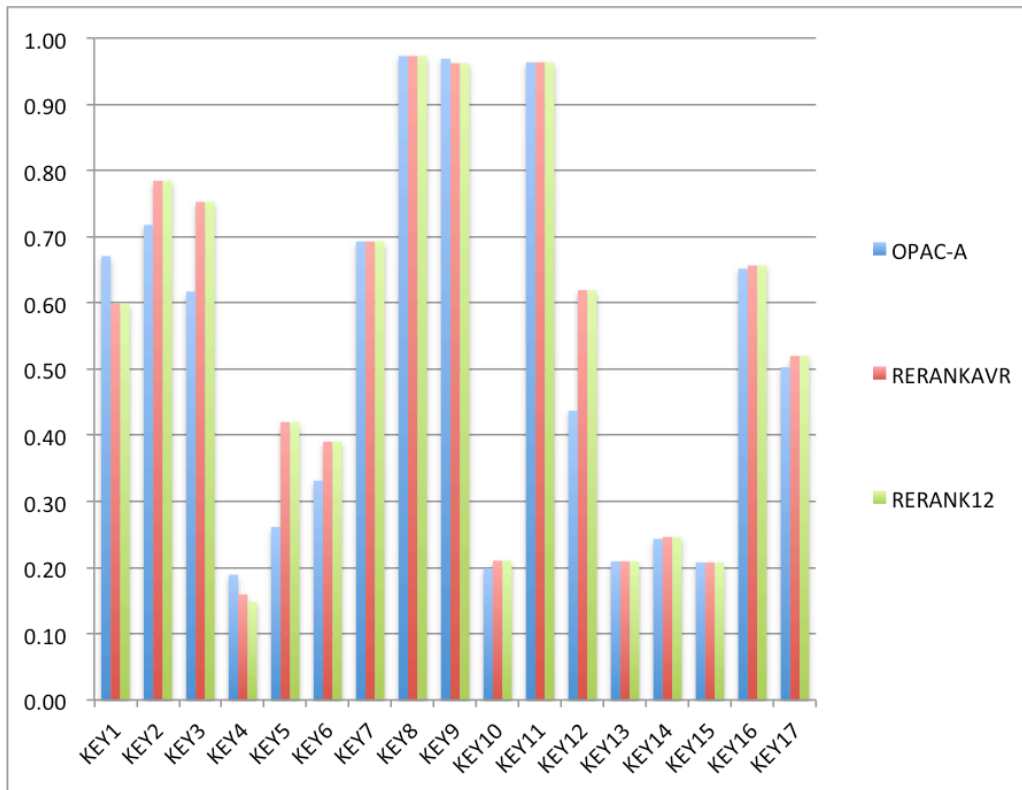


図 6.1: キーワードごとの平均精度とその平均

6.3 今後の課題

先ず他の学科に対する評価である。本研究は実験協力者が集まらなかったため、評価実験は2つの学科しか実施できなかった。本手法の有効性を高めていくため、他の学科での評価も必要である。

次に、より細かい個人の情報要求を満たすことのできるプロフィールの構成手法を検討する必要がある。本研究は分野によって利用者プロフィールを作成したが、実際に同じ分野でも、個人の情報要求の差がある、例えば、学年によって、情報要求が違う可能性がある。そのため、より個人化のプロフィールを作成すれば、利用者毎の利用傾向を正確的に把握できると考えられる。また、本研究は大学生を対象として行ったが、将来、大学院生の貸出履歴を分析し、同様に貸出書籍の主題と利用者の属性を明らかにして利用者プロフィールの構成する手法を検討する必要があると考える。

また、より細かい図書の属性を利用してプロフィールの構成手法を検討することも今後の課題となる。本研究は図書のNDC分類と貸出時期を図書の属性としたが、貸出傾向に影響する他の要素もあると考える。例えば、図書の在庫期間、蔵書数などの要素も分析し、図書のNDC分類と組み合わせて利用者プロフィールを構成すれば、より良い利用者プロフィールを作成できる。

最後に、図書の属性と利用者の属性を正確に把握し、より良い組み合わせ方法を検討することは今後の課題だと考えられる。近年、大学図書館においては、MyLibraryのような個人向けサービスの基盤が整いつつある。これらの認証機能と組み合わせ、さらに柔軟な個人の情報要求を満たすことのできる蔵書検索システムが構築できるのではないかと期待している。将来は図書館システムと学習システムの連携により個人に特化した検索の実現が可能になるかと考えられる。

第7章 結論

現在の図書館蔵書検索システムは同じ検索語に対して同じ結果が提示されるという問題がある。しかし、利用者によって情報要求が異なるため、現状の検索結果で全ての利用者が満足できるわけではない。そこで本研究では利用者の属性を利用して、検索結果のリランキングを行い、利用者の情報要求に合わせて蔵書検索結果を効果的にリランキングすることを目的とした。

本研究では利用者の属性抽出のために、図書館の貸出履歴に着目し、利用者ごとにプロフィールを作成することとした。利用者プロフィールの作成にあたって、A大学図書館の2006年4月から2012年3月の貸出履歴を分析し、利用者属性と貸出書籍の主題の関係を明らかにした。その結果を基に利用者をグループ化し、グループに所属する利用者の貸出履歴を用いてグループのプロフィールを作成した。作成したプロフィールに基づいて検索結果の各書籍に重み付けを行い、ランキングを修正する手法を提案した。提案するリランキング手法の有効性を検証するため評価実験を行った。被験者はA大学の2つの学科に所属する7名の学生である。被験者が選んだキーワードを利用してA大学の蔵書検索システムの出力する検索結果と本手法によりリランキングした結果とを比較した。比較の指標として平均精度を求めた。

実験の結果、本手法によるリランキング結果はA大学の蔵書検索システムより平均精度が高くなった。2つの学科限定ではあるものの貸出履歴に含まれる書籍の主題を利用することで、利用者の求める書籍の順位を上げることに成功したと言える。

今後の課題は図書属性と利用者属性を正確に把握し、より良い組み合わせ方法を検討することと蔵書検索システムの構築である。

第8章 謝辞

本研究を進めるに当たって、ご指導頂きました、松村敦先生、宇陀則彦先生、逸村裕先生に深く感謝申し上げます。また、研究生を含めて筑波大学に在籍している3年間、日本語も、研究もいつも熱心的に指導して頂いたチューターさんの皆様：石川里佳子さん、常川真央先輩、米島まどかさんに感謝致します。そして、実験を協力頂いだ被験者の皆様に感謝致します。

参考文献

- [1] 佐浦敬之, 辻慶太. 公共図書館における利用履歴の活用に関する意識調査, 第 57 回日本図書館情報学会研究大会発表要綱, 2009, p.1-4.
- [2] 原田隆史, 増田浩佑. 貸出記録を用いた図書推薦システムにおける重みづけの変更, デジタル図書館, 2010(38), p.54-66.
- [3] 原田隆史. 図書館の貸出履歴を用いた図書の推薦システム, デジタル図書館, No.36, 2009, p.22-31.
- [4] 當山仁健. 利用者のプロフィールを考慮した連想検索 OPAC の構築, 情報の科学と技術, 56(11), 2006-11-01, p.520-525.
- [5] 辻慶太, 滝沢伸也, 佐藤翔, 池内有為, 池内淳, 芳鐘冬樹, 逸村裕. 図書館の貸出履歴を用いた図書推薦システムの有効性検証, 図書館界 64(3), 176-189, 2012-09-01.
- [6] 小野永貴, 常川真央. Web 時代にあるべき未来の図書館サービスの胎動—貸出履歴の議論を超えた shizuku2.0 の実現へ, 情報管理 53(4), 2010, p.185-197.
- [7] 山口真也. 私立大学図書館における個人情報・プライバシー保護:貸出記録の管理方法と消去の必要性を中心に, 沖縄国際大学日本語日本文学研究, 2005, 10(1), p.1-24.
- [8] 辻慶太, 滝沢伸也, 佐藤翔. 図書館の貸出履歴と書誌情報を用いた図書推薦システムの有効性, 図書館界 65(4), 2013-11, p.253-267.