

# 新潟国際情報大学卒業論文データベースの概要と 論文表題の形態素解析による卒業論文の構成要素に関する考察

*Overview of the Graduation Thesis Database and the Study on the Functional Elements of  
Graduation Thesis using Morphological Analysis of Thesis Title*

高 木 義 和

## 要約

新潟国際情報大学卒業論文データベースは抄録と原文の登録システムと、抄録の検索システムから成る。登録システムは、著者である学生自らが、作成した卒業論文の抄録と原文の登録を Web 上で行う。卒業論文の検索システムは登録した抄録データを Web 経由で検索し結果を表示する。卒業論文検索システムは抄録の検索を行うと共に原文の参照機能を含んでいる。卒業論文データベースは OS を含め基本的に全てフリーのソフトを使用して構成した。登録されたデータの概要を把握するため、これまでに登録された表題の日本語形態素解析を行った。解析は形態素解析システム茶釜で切り出した用語を Perl で作成したプログラムにより処理を行った。その結果、使用頻度の高い、一般名詞とサ変接続名詞を組み合わせてみると典型的な卒業論文の表題と、卒業論文の構成要素を推定することができた。卒業論文の課題の設定や構成を考える上で、出現頻度の高い一般名詞一覧と、サ変接続名詞の一覧が参考になると考えられた。また、具体的な卒論作成や指導においては原文情報が重要になると考えられた。

キーワード：データベース、卒業論文、形態素解析、サ変接続名詞、登録、抄録

## 1. はじめに

新潟国際情報大学は 1994 年 4 月に開学し、最初の卒業生を 4 年後の 1998 年 3 月に送り出した。新潟国際情報大学は情報文化学部 1 学部で、情報文化学部は情報文化学科と情報システム学科の 2 学科から構成されている。両学科とも開学当初から卒業論文は必須科目となっており、開学以来これまで卒業生全員が卒業論文を作成し卒業している。地域に貢献できる人材の育成をめざして学部教育を実践する本学にとって卒業論文は、量的、質的に一定の品質を保持し、本学で継続的に生産できる知的創作物である。そこで卒業論文を情報として整理、蓄積し、必要な時にいつでも取り出せるデータベースシステムを構築することにより、将来本学の資産として活用できると考えた。

他大学における卒業論文の扱いは、研究室単位で Web 上に公開している例がもっと多いが、情報管理という観点からみると研究室が無くなると卒業論文に関する情報も消滅することから継続性に問題がある。次に多いのが学部・学科単位で題目を整理している例であるが<sup>2, 3, 4)</sup>、発表会などのため卒業年度における利用のために作成されていると思われ、各年度の情報が独立して

\* TAKAGI, Yoshikazu [情報システム学科]

整理されている。卒業論文のデータベースあるいは登録検索システムについての報告はあるが<sup>5,6)</sup>、その稼働についての報告は見つけることができなかった。小樽商科大学附属図書館では1952年以降の卒業論文、修士・博士論文のタイトル一覧をWeb上に公開しており検索機能も備えている<sup>7)</sup>。国際基督教大学図書館<sup>12)</sup>も同様のサービスを提供している。原文を保存している図書館の場合は学内の図書館で閲覧可能である。

本学では1998年3月に1期の卒業生を送り出すにあたり、卒業論文のデータベース化を教授会で提案し了承された。抄録と原文をデータベース化の対象として考えたが、当時はPCを使用していない教員も存在し、手書きで卒業論文が作成される場合も存在したことから、抄録を必須としてデータベースを作成することにした。卒業論文の原文登録は指導教員あるいは本人の判断に委ねた。データは抄録、原文ともデジタルデータでの扱いを前提とした。

卒業論文データベースは抄録と原文の登録システムと、抄録の検索システムから成る。登録システムは、著者である学生自らが、作成した卒業論文の抄録と原文をWeb経由で登録するシステムである。卒業論文の検索システムは登録した抄録データをWeb経由で検索し結果を表示するシステムである。卒業論文検索システムは抄録の検索を行うと共に原文の参照機能を含んでいる。卒業論文データベースはOSを含め基本的に全てフリーのソフトを使用して構成している。

## 2. 抄録と原文の登録

### 2-1 登録方法の推移

学生自身が抄録情報をテキストデータとしてWeb経由で登録を行っている。登録データを1件ずつ1ファイルに順に追記して保存した(統合ファイル)。2001年にシステムを更新し、抄録登録時に統合ファイルを作成するとともに、同時にWebページも作成できるシステムとした。作成したWebページを検索の対象ファイルとした。文字コードはEUCを使用した。2012年よりリレーショナルデータベースであるpostgreSQLを使って抄録の登録を行っている。抄録登録時にデータベースのテーブルにデータをinsert文で1件ずつ追加している。

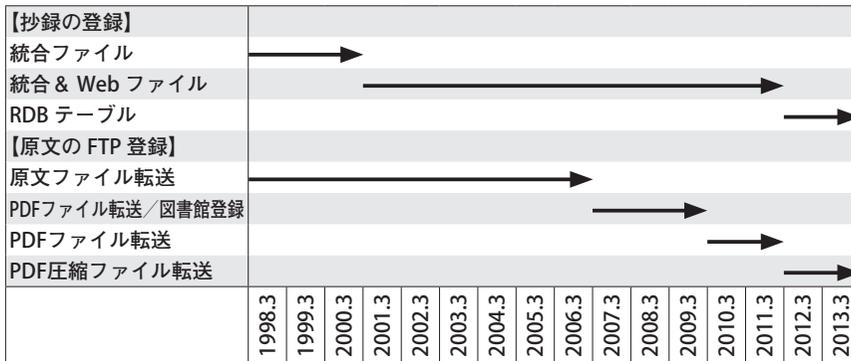
原文登録は必須ではなかったが、当初より登録されたワープロファイルをPDFファイルに変換して蓄積している。PDF化ソフトが高価であったためワープロファイルで登録後まとめてPDF化を行った。2007年より図書館に導入された印刷原稿をスキャンしてPDF化できるシステムを併用した。2010年以降は情報センターのPCに導入されたPDF化機能を利用することにした。図1に抄録と原文の登録方法の推移を示す。

### 2-2 抄録の登録

#### 2-2-1 初期の卒業論文登録システム：1998-2000

1998年に開始した最初の卒業論文登録システムはWebの画面から卒業生が自分で抄録を登録する方式を採用した。最初のシステムは1期の卒業生が、卒業論文「WWWサーバにおける情報検索システムの構築」の中でPerlを使用して作成した登録システムを使用した。データは1人1行として、全体を1ファイルとした統合ファイルを作成した。

検索には検索エンジンを使用したことから検索に対応できるよう統合ファイルの内容を1行ずつ個人別のWebページに書き換えた。1行の内容をそれぞれWebページに変換したことから、登録件数と同数のWebページを作成した。文字コードはEUCを使用した。登録の際にMacintoshおよびWindowsのワープロファイルのデータを、コピー&ペーストする過程で生じ



注) 3月時点の年なので卒業年次は1年早くなる。

図1 抄録と原文の登録方法の推移

たと思われる原因不明の文字化けがかなりの頻度で発生した。

### 2-2-2 統合ファイルと Web ファイルの作成：2001-2011

2001年から抄録登録システムを、Perlを使用して作り直した。従来通り登録データを統合ファイルとして整理するとともに、同時に、学籍番号で個人別のWebページを作成できるようにした。EUCコードを使用したWindowsの文字コードであるShift-JISとの間で文字化けが生じるため、システム内の文字コード全てをEUCコードで処理するように対応を図った。それでも、Word文書のコピー&ペーストが原因と思われる文字化けを防ぐことはできなかった。同時登録送信(複数の学生が意識的に同時に登録を試みるケース)に対応してシステムエラーを回避するためロックファイルを使用しシステムの安定性向上を図った。

### 2-2-3 PostgreSQLを使用した卒業論文登録システム：2012～現在

2012年から抄録の登録はフリーのリレーショナルデータベースであるPostgreSQLを使った卒業論文登録システムに移行した。抄録の登録データは1件ずつinsert文でテーブルに追加される。このテーブルを使用すれば検索にも対応できるが、学外サービスや原文閲覧に対応できないことから検索システムとして使用していない。文字コードはUTF-8とした。UTF-8を採用してから、以前のシステムと比較すると、文字化けのエラー発生件数が格段に減少した。Perlを使用したシステムに比べると管理に要する手数は半減した。

### 2-2-4 登録内容の修正対応

登録システムに登録内容の修正機能を備えていない。登録内容の修正は、修正箇所を管理者にメールで連絡する。件数が少ないため手作業で行った方がシステム維持の労力が少なく済むため、および間違いが生じやすい事例を把握できるためである。送られてきた箇所は後日まとめて修正を行っている。主な修正内容は文字の入力ミス以外に、登録後に登録文章を修正するケース、所属研究室など選択項目を見落としているケース、キーワードを追加するケースなどの依頼が多い。修正依頼はなかったものの抄録に原文テキストそのものを貼り付けた事例もあった。

## 2-2-5 登録件数と登録学問分野

1998年3月から2011年3月の14年間の登録件数を表1に示す。1998年から2011年の14年間で、学部全体で4089件、情報文化学科1635件(40%)、情報システム学科2454件(60%)の登録があった。年平均件数は学部全体で292件、情報文化学科117件、情報システム学科175件となった。学問分野は社会科学2044件(50%)、技術・工学1153件(28%)、人文科学640件(16%)、自然科学252件(6%)の順となった。社会科学の2044件は全登録件数の半数となった。

学問分野を学科別に見ると、学部全体で2044件(50%)を占めた社会科学分野は、情報文化学科の中で1048件(64%)、情報システム学科の中で996件(41%)を占めた。情報システム学科では技術・工学分野が1140件(46%)で最も多かったが、社会科学分野との差は5ポイントであった。図2に学科別登録学問分野を示す。

表1 1998年3月から2011年3月の14年間の抄録登録件数

卒業 予定 年月	合計 件数	情報文 化学科	情報シ ステム 学科	学問分野			
				人文 科学	社会 科学	技術 工学	自然 科学
1998.3	297	117	180	28	153	80	36
1999.3	291	124	167	53	132	74	32
2000.3	303	125	178	38	171	76	18
2001.3	297	118	179	47	141	82	27
2002.3	288	116	172	47	129	95	17
2003.3	278	108	170	35	131	90	22
2004.3	320	122	198	50	148	100	22
2005.3	289	111	178	52	134	85	18
2006.3	290	120	170	48	148	77	17
2007.3	271	112	159	48	147	70	6
2008.3	309	123	186	50	164	88	7
2009.3	294	119	175	50	153	80	11
2010.3	264	106	158	38	145	71	10
2011.3	298	114	184	56	148	85	9
計	4089	1635	2454	640	2044	1153	252
平均	292	117	175	46	146	82	18

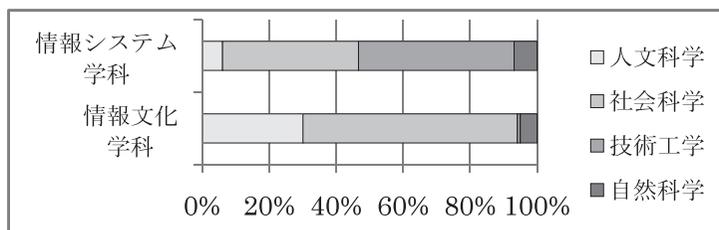


図2 学科別登録学問分野

## 2-3 原文の登録

### 2-3-1 原文ファイルの FTP 転送：1998-2009

原文は当初学生が原文ワープロファイルを FTP 転送する方式で登録した。最終的なファイル形式は PDF ファイルを基準としたが、PDF 化ソフトのアクロバットが高価であったためワード等のファイルで提出されたものを、アクロバットをインストールした数台の PC で PDF ファイルに変換した。PDF ファイルへの変換は学生アルバイトにより対応した。2007 年より情報システム学科は原文ファイル登録が必須となった。

2007 年からは学生が作成した PDF ファイルか、印刷した原稿を図書館でスキャンした PDF ファイルで登録を行った。

### 2-3-2 PDF ファイルによる登録：2010- 現在

2010 年からは情報センターの PC で PDF ファイルの作成が可能になったことから、学生自らが PDF ファイルを作成し登録を行っている。さらに 2012 年からは原文 PDF ファイルとデータベース登録の許諾書のイメージファイルを全て 1 つのファイルに圧縮した ZIP ファイルを作成し登録している。登録方法は全て FTP 転送で対応している。PDF ファイルの作成方法、ZIP ファイルへの圧縮方法および FTP 転送方法はサーバ上に詳しいマニュアル準備して対応している。ZIP ファイル作成は操作が複雑になるため混乱を心配したが、大きな問題は生じなかった。

### 2-3-3 登録件数

1998 年 3 月から 2011 年 3 月の 14 年間の原文登録件数を表 2 に示す。14 年間の全文登録件

表 2 1998 年 3 月から 2011 年 3 月の 14 年間の原文登録件数

卒業予定 年月	入学年月	情報シス テム学科	情報文化 学科学科	合計
1998.3	1994.4	2	0	2
1999.3	1995.4	24	1	25
2000.3	1996.4	71	30	101
2001.3	1997.4	103	46	149
2002.3	1998.4	136	77	213
2003.3	1999.4	114	62	176
2004.3	2000.4	126	28	154
2005.3	2001.4	122	26	148
2006.3	2002.4	136	86	222
2007.3	2003.4	166	28	194
2008.3	2004.4	176	43	219
2009.3	2005.4	170	26	196
2010.3	2006.4	164	25	189
2011.3	2007.4	168	42	210
計		1678	520	2198

出典：http://t-search.nuis.ac.jp/abstract/genbun/

数 2198 件は、抄録の全登録件数 4089 件の 58% に相当する。学科別で見ると情報システム学科では抄録の全登録件数 2454 件に対し原文の登録件数 1678 件は 68% に相当する。情報文化学科では抄録の全登録件数 1635 件に対し原文の登録件数 520 件は 32% に相当する。

#### 2-3-4 登録されたが利用できない原文

ショートカットを登録するケースが毎年発生していたが、情報センターの PC 環境で PDF ファイルが作成できるようになった 2010 年以降は減少した。他に、許諾書が提出されているにもかかわらず、PC にインストールされていないプログラムで作成された原稿など、利用できない原文が含まれていた。実際にデータベースとして閲覧できたのは 2198 件であるが 14 年間の登録確認件数は 2317 件であり、119 件（約 5%）が無効な登録であった。2001 年に学内の PC が Macintosh から Windows に切り替わったが、無効な登録ファイルの中には初期の Macintosh 環境で作成されたファイルが含まれる。

### 3. 抄録の検索と原文の閲覧

#### 3-1 検索方法の推移

初期の抄録の検索システムは検索エンジン SSE を使用した。検索エンジンは Web の 1 ページを 1 件と扱うため、登録した抄録データを 1 件ごとに Web ページに変換して使用した。検索サービスは学内外に行った。2002 年より日本語全文検索システム Namazu を使用した検索システムに更新した。

2012 年からはサーバの情報センターへの移行に伴い抄録の検索システムも NetCommons を使用したシステムへ移行した。

原文の閲覧は図書館で CD を貸出す方式でスタートした。2004 年からは Web ページ上で学術番号から原文 PDF ファイルへリンクを張り、原文を入手できる方式とした。卒業論文という性格上、原文をコピーして利用される懸念が払拭できないことから、盗用を防ぐためテキストを取り出せないよう文書保護のセキュリティレベルを高くした。2012 年からは抄録検索の結果表示画面から、原文 PDF の圧縮ファイルにリンクを張ったことから原文入手が容易になった。

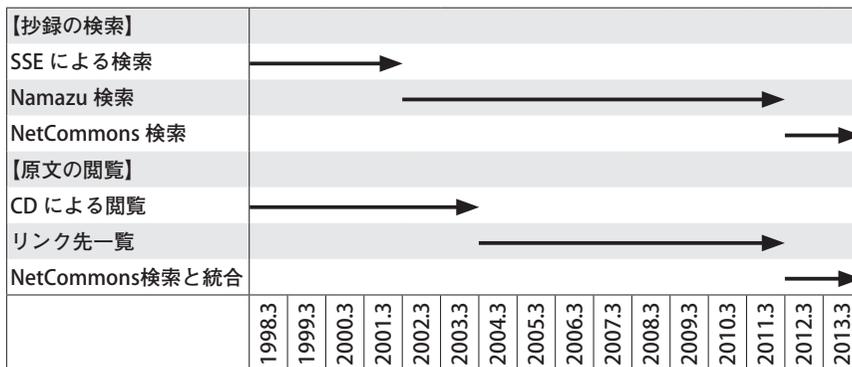


図3 抄録と原文の検索方法の推移

### 3-2 抄録の検索

#### 3-2-1 SSE による抄録の検索

1998年に開始した最初の抄録検索システムは、登録システム同様卒業生が、卒業論文の中で作成した検索システムを使用した。検索システムにはSSEを使用した。SSE (SonyDrive Search Engine) はSONYのWebサイトにある商品ページなどを検索するために開発され、実際のサイト運営に実績があった検索エンジンで、フリーで公開されていた。SSEはOSがUNIXで、ApacheとPerlが稼働する環境で実行可能であった。抄録の検索システムは大学のLinux環境で稼働させた。1人1行として登録された統合ファイルから、1行毎にWebページに変換し、SSEの検索データとした。

#### 3-2-2 日本語全文検索システム Namazu による抄録の検索

2002年から全文検索システム Namazu を使用したシステムを導入した。Namazu はSSE同様に検索用インデックスを事前に作成するが、サイトの管理用ではなくWebページの全文検索用に開発されていることからデータベースとしての使用に適していた。インデックスは新たなデータを追加する毎に作り替えた。大学のWebサーバの管理下で使用できたため大学のサーバを使用して抄録の検索サービスを大学内だけでなく、大学外にも行った。この学外サービスを通して原文の複写依頼に関する問い合わせが来るようになった。Namazuによる抄録検索サービスを2011年まで実施した。2005年に個人情報保護法が施行され、その直後に、公開に関する許諾をとっていなかった卒業生から氏名を検索結果から削除するよう要請があった。卒論データベースに表示される個人名も個人情報保護法の対象になりうることから、2008年以降、氏名は本人が許諾した場合のみ学外に公開することにした。

#### 3-2-3 次世代情報共有基盤システム NetCommons による抄録の検索

2012年からは専用サーバへの移行に伴って、国立情報学研究所が次世代情報共有基盤システムとして開発したNetCommonsを使用した抄録の検索システム(新潟国際情報大学卒業論文データベース)に移行した。基本的な検索は、所属学科検索、所属研究室検索、対象学問領域検索、あるいはそれらの複合検索である。いずれの検索項目もプルダウンメニューから検索内容を選択して検索する。キーワード検索にも対応でき、and, or, フレーズ検索が可能である。個人情報保護法に配慮し、2012年以降、学外へのサービスから個人名を全て削除した。学内の抄録の検索結果画面には従来通り個人名を表示している。

### 3-3 原文の閲覧

#### 3-3-1 CD による公開

1998年から2003年までの6年間はCD-ROMに登録原文データを保存し、CD-ROMを図書館で公開した。カウンターでCD-ROMを借り、図書室内のPCで閲覧するようにした。CD-ROMは毎年登録原文データを追加し内容を更新した。ファイルはMacintoshとWindowsで作成したワープロファイルをWindowsで閲覧できるように整理した。

#### 3-3-2 リンク先一覧による公開

2004年からはNamazuの抄録検索ページの下に卒業年度別に原文ファイルへのリンクを作成

した。ワードファイルとして登録された原文ファイルは一括して全て PDF ファイルに変換した。各年度のページには学籍番号一覧を表示し、学籍番号の右に、論文を構成する全ての PDF ファイルと、許諾ファイルへリンクを張った。必要な卒業論文の学籍番号に該当するファイル名をクリックすることにより、原文 PDF ファイルを閲覧、ダウンロードできるようにした。ファイルが複数に分かれている場合は一定のルールを示してファイル名を付与した。PDF ファイルへの変換とリンクリストの作成は、毎年学生アルバイトで対応した。ファイル名の作成ルールと作成例を以下に示す。

- ・ルール：ファイル名は学籍番号（8ケタ）に姓名の頭文字（英数小文字2文字）を付けたものを基本とし、ファイルが複数存在する場合は、姓名の頭文字の後に A, B, C（大文字）をつける。

- ・例：学籍番号が 12009055 で、氏名が赤塚不二夫で、3 ファイルから構成される場合

12009055afA.pdf

12009055afB.pdf

12009055afC.pdf

### 3-3-3 抄録の検索結果との統合

2012 年から NetCommons を使用した検索システムとなり、抄録検索の結果表示画面から原文ファイルへのリンクをクリックすることにより、原文ファイルを入手できる設定となった。原文ファイルが複数存在すると管理が複雑になり間違いが生じやすいため、許諾ファイルと全ての原文ファイルを 1 つの zip ファイルに圧縮して提供している。登録には従来以上の手数が必要になるが、検索結果から zip 形式に圧縮された原文ファイル一式を入手することができ、利用者の利便性が向上した。原文の著者である卒業生から学内に限った公開の許諾をうけているので、学内からの検索についてのみ原文へのリンクが有効で、閲覧が可能である。学外から検索する場合は原文へのリンク項目が表示されない。

### 3-4 外部からの原文提供依頼

外部へ公開している抄録の検索結果から、原文提供依頼が主に図書館へ来ている。卒業論文の著者は卒業生であり、著作権も卒業生に帰属すると判断し<sup>8, 9)</sup>、これまでは原文提供依頼があると図書館から著者に連絡をとり、本人から原文提供の許諾を得られた場合にのみ依頼者へ原文を提供してきた。非常に手数がかかったことから 2008 年以降、原文登録時に著者から事前に原文提供に関する許諾を得ておき、依頼があった場合に許諾を確認して情報提供を行っている。

## 4. 論文表題に使用された言葉の概要

### 4-1 解析方法

卒業論文の内容を表現する重要な言葉は表題中で使用される場合が多いと考えられる<sup>10)</sup>。そこでこれまでに登録された表題の日本語形態素解析を行った。解析は形態素解析システム茶釜 Version2.0 をウィンドウズ環境で使用した。茶釜で切り出した単語は perl で作成したプログラムで処理を行い頻度リスト等を作成した。

#### 4-2 表題中で使用された言葉の品詞別割合

表題中で使用された言葉の品詞別割合を表3に示す。全語数は5857語となった。卒業論文の主要な概念を示すと思われる品詞である名詞が最も多く4439語(75.8%)を占めた。続いて動詞が365語(6.1%)であった。未知語が764語(13.0%)を占めたがその内容は、Java, Web, Ruby, IT, Basic, PostgreSQLといったアルファベットから成る情報関連の用語や、ジェンダー、ドメスティック、コリアン、ツーリズム、リコメンドシステム、アプリといったカナからなる翻訳語、年代などの数値などよりなる。764語の内、スペースと記号が約1/3を占めたことから、残りの約500語は、アルファベット表記と、カナ表記の言葉であり、名詞の割合が最も高かった。未知語の中にも論文中で使用された重要な概念を示す用語が含まれていることになる。

#### 4-3 卒業論文表題中に使用された名詞

##### 4-3-1 名詞の種類

卒業論文表題中に使用された名詞の種類を表4に示す。名詞に分類された全4439語の中で一般名詞が最も多く2412語(54%)を占めた。名詞の中でも一般名詞は論文が扱う概念を直接表現するのに適していることから、論文中で使用された重要な概念は一般名詞の中に含まれる可能性が高い。

サ変接続名詞は996語(22%)となった。一般名詞が扱う課題の主要な概念を示すのに対し、サ変接続名詞は論文で実施した内容、手段、方法などを示している。一般名詞とサ変接続名詞を組み合わせると、何を(一般名詞)論文で実行したか(サ変接続名詞)を把握できる可能性が高い。固有名詞は600語(14%)となった。固有名詞は論文で扱う国や地域を示していることから注目されている地域の傾向を把握することができる。

表3 論文表題に使用された言葉の品詞別割合

品詞の種類	語数	%
名詞	4439	75.8
未知語_	764	13.0
動詞	365	6.1
記号	100	1.7
接頭詞	54	0.9
助詞	50	0.9
形容詞	35	0.6
副詞	21	0.4
助動詞	17	0.3
連体詞_	10	0.2
接続詞_	7	0.1
他の品詞	4	0.1
計	5857	100.0

表4 卒業論文表題中に使用された名詞の種類

名詞の種類	語数	%
一般	2412	54
サ変接続	996	22
固有名詞	600	14
接尾	222	5
形容動詞語幹	89	2
副詞可能	47	1
数	46	1
非自立	14	0
他の名詞	13	0
	4439	100

#### 4-3-2 一般名詞

一般名詞の一覧を表5に示す。表1に示した4089件が対象になる卒業論文の総数である。実際に使用されている一般名詞から論文が扱う主要なテーマを把握することや、特に使用頻度の高い用語から論文が扱う課題の傾向を知ることができる。

表5には出現頻度の上位150語を示したが、卒業論文で扱う対象分野を検討する際に参考になると思われる。上位20名詞の14年間の頻度の推移を時系列的に線形近似直線で見ると有意な変化は読み取れないが、中心、要因、事例、戦略、データベース、文化、言語が増加傾向を示した。増加傾向を示した7語の中でも、中心、要因、事例、現代の増加傾向が強かった。具体的な事象の概念より、抽象的な状態を表現する言葉が増加傾向を示す可能性がある。

#### 4-3-3 サ変接続名詞

サ変接続名詞の一覧を表6に示す。表題では客観的にかつ簡潔に内容を表現することが求められることから、文末が体言止めになることが多く、サ変接続名詞はこの体言止めに使用される場合が多い<sup>11)</sup>。一方、サ変接続名詞に「する」を加えると動作や動きを表す動詞として使用できることを考えると、卒業論文の中で実現を試みた動きを推定できる。

研究が圧倒的に多いが、考察、分析、影響、調査、作成、構築、比較、利用、教育、提案などが多く使われている。サ変接続名詞の上位10語の15年間の頻度の推移を時系列的に線形近似直線で見ると研究がやや減少傾向にあるが、他の考察、分析、影響、調査、作成、構築、比較、利用、教育は増加傾向を示した。このように研究が減少したのは、より具体的な研究内容を示す言葉が使用されるようになった結果と考えられた。また、作成、構築といったものを作る卒業論文や、調査、利用、提案といった能動的内容を含む卒業研究が増加している可能性が認められた。

#### 4-3-4 固有名詞の一覧

一般名詞が抽象的な概念を示すのに対し、固有名詞は特定の事物を区別するために与えられた名称である。固有名詞は人名、地名、国名などを表すが、卒業論文の表題では、国名が主たる固

表5 一般名詞の一覧（上位150語）

順位	一般名詞	出現頻度	順位	一般名詞	出現頻度	順位	一般名詞	出現頻度
1	問題	340	51	会社	35	101	障害	20
2	システム	247	52	交差点	34	102	人権	20
3	現状	192	53	方法	34	103	政治	20
4	情報	168	54	構造	33	104	日中	20
5	社会	166	55	食品	33	105	能力	20
6	中心	140	56	活性	32	106	暴力	20
7	要因	131	57	株式	31	107	コミュニケーション	19
8	環境	119	58	商品	31	108	テレビ	19
9	県	110	59	地球	30	109	外交	19
10	女性	104	60	都市	30	110	原子力	19
11	事例	95	61	役割	30	111	現実	19
12	戦略	91	62	原因	29	112	視点	19
13	企業	88	63	主義	29	113	自己	19
14	現代	88	64	人	29	114	動向	19
15	課題	87	65	未来	29	115	特徴	19
16	歴史	83	66	格差	28	116	ごみ	18
17	経済	77	67	市場	28	117	ホームページ	18
18	データベース	75	68	食	28	118	価値	18
19	文化	72	69	日本人	28	119	外国	18
20	言語	68	70	スクリプト	27	120	個人	18
21	インターネット	65	71	メディア	26	121	思想	18
22	業界	63	72	状況	26	122	視覚	18
23	制度	63	73	量	26	123	人口	18
24	高齢	62	74	アンケート	25	124	背景	18
25	電子	62	75	モデル	25	125	サーバ	17
26	国際	61	76	適性	25	126	マネー	17
27	地域	61	77	農業	25	127	温暖	17
28	大学	59	78	サッカー	24	128	経路	17
29	ゲーム	58	79	セキュリティ	24	129	子ども	17
30	効果	58	80	デジタル	24	130	少子化	17
31	産業	53	81	技術	24	131	精神	17
32	自動車	53	82	例	24	132	文字	17
33	世界	49	83	コンピュータ	23	133	理論	17
34	マーケティング	48	84	周辺	23	134	音声	16
35	オブジェクト	46	85	犯罪	23	135	価格	16
36	映画	46	86	野球	23	136	家庭	16
37	策	45	87	エネルギー	22	137	観点	16
38	若者	45	88	サイト	22	138	酸素	16
39	政策	45	89	医療	22	139	事件	16
40	ビジネス	43	90	形態	22	140	商店	16
41	音楽	41	91	事業	22	141	青年	16
42	交通	41	92	事故	22	142	年代	16
43	実態	41	93	人間	22	143	方向	16
44	学生	40	94	あり方	21	144	本	16
45	大学生	40	95	基礎	21	145	ソフトウェア	15
46	ネットワーク	39	96	店舗	21	146	ロボット	15
47	民族	38	97	道路	21	147	意義	15
48	スポーツ	36	98	傾向	20	148	公共	15
49	画像	36	99	最適	20	149	児童	15
50	時代	36	100	取り組み	20	150	少数	15

注) 同じ頻度の場合でも便宜上、並び替え順で順位を付けた。

表6 サ変接続名詞の一覧（上位 150 語）

順位	サ変接続名詞	出現頻度	順位	サ変接続名詞	出現頻度	順位	サ変接続名詞	出現頻度
1	研究	240	51	検査	26	101	製造	14
2	考察	196	52	差別	26	102	設定	14
3	分析	184	53	処理	26	103	防止	14
4	影響	169	54	発電	26	104	ショッピング	13
5	調査	150	55	移民	25	105	化粧品	13
6	作成	148	56	介護	25	106	仮想	13
7	構築	119	57	活用	25	107	加工	13
8	比較	109	58	機能	25	108	展開	13
9	利用	108	59	検索	25	109	報道	13
10	教育	99	60	使用	25	110	トレーニング	12
11	提案	98	61	実証	25	111	応用	12
12	関係	93	62	認識	25	112	会議	12
13	意識	92	63	会計	24	113	施設	12
14	経営	71	64	広告	24	114	就職	12
15	プログラム	69	65	選択	24	115	制御	12
16	携帯	68	66	対応	23	116	製作	12
17	シミュレーション	67	67	変遷	23	117	相違	12
18	電話	64	68	改革	22	118	動作	12
19	支援	61	69	緩和	22	119	認証	12
20	対策	60	70	向上	22	120	配信	12
21	評価	60	71	表示	22	121	普及	12
22	開発	57	72	予測	22	122	運用	11
23	解明	53	73	イメージ	21	123	拡大	11
24	展望	52	74	学習	20	124	企画	11
25	変化	52	75	観光	20	125	記憶	11
26	運動	50	76	実現	20	126	競争	11
27	指向	45	77	生産	19	127	自殺	11
28	解析	44	78	通信	19	128	小売	11
29	管理	44	79	導入	19	129	成長	11
30	発展	44	80	プログラミング	18	130	設計	11
31	生活	42	81	慰安	18	131	測定	11
32	検証	41	82	計算	18	132	探索	11
33	サービス	38	83	リサイクル	17	133	著作	11
34	戦争	36	84	経験	17	134	低下	11
35	関連	35	85	購入	17	135	廃棄	11
36	改善	34	86	紛争	17	136	表現	11
37	案内	33	87	保護	17	137	変容	11
38	検討	33	88	意味	16	138	位置	10
39	交流	32	89	喫煙	16	139	共同	10
40	労働	32	90	決定	16	140	共有	10
41	活動	31	91	結婚	16	141	削減	10
42	行動	30	92	統一	16	142	刺激	10
43	計画	29	93	販売	16	143	従軍	10
44	作業	29	94	満足	16	144	信頼	10
45	渋滞	28	95	メール	15	145	総合	10
46	消費	28	96	協力	15	146	増加	10
47	在日	27	97	提言	15	147	統合	10
48	成功	27	98	放送	15	148	変動	10
49	解決	26	99	依存	14	149	保障	10
50	形成	26	100	合併	14	150	歩行	10

注) 同じ頻度の場合でも便宜上、並び替え順で順位を付けた。

有名詞であった。固有名詞の一覧を表7に示す。日本、新潟、中国が最も多く、次いで、韓国、アメリカ、日、朝鮮、ロシア、米、韓、北朝鮮、台湾が多かった。日本と新潟が多いのは情報文化学科と情報システム学科の両方の学科の卒業論文の表題に使用されているためであった。日本を含む表題の17%、および新潟を含む表題の60%が情報システム学科の卒業論文であった。日本を含む表題の83%、新潟を含む表題の40%が情報文化学科の卒業論文であった。中国、韓国、アメリカ、朝鮮、ロシアは情報文化学科の卒業論文が大半を占めた。

図4に国、地域をあらわす固有名詞出現頻度の時系列変化を示す。時系列的な推移を見ると明らかな増加、減少の傾向は認められない。日本が増加傾向を示しているように見えるが線形近似直線で有意な変化ではなかった。

表7 固有名詞の一覧

順位	固有名詞	出現頻度
1	日本	316
2	新潟	308
3	中国	157
4	韓国	92
5	アメリカ	61
6	日	55
7	朝鮮	39
8	ロシア	38
9	米	29
10	韓	25
11	北朝鮮	24
12	台湾	18
13	アジア	16
14	ドイツ	13
15	フランス	12

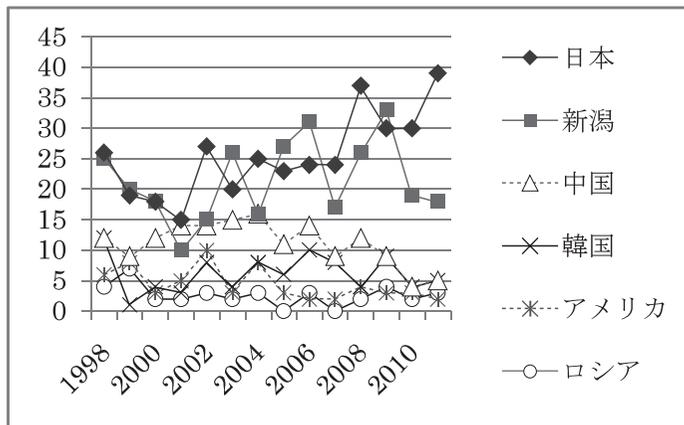


図4 国、地域をあらわす固有名詞出現頻度の時系列変化

## 5. 考察

一般名詞とサ変接続名詞の上位3語を組み合わせると典型的な卒業論文の構成を推定できた。表8に一般名詞とサ変接続名詞の上位3語を組み合わせた概念を示す。

一般名詞で最も出現頻度の高い「問題」とサ変接続名詞で最も頻度の高い「研究」を組み合わせると、「-----問題に関する研究」となる。特定の話題に関する問題を洗い出し、その解決方法を研究するといった構成が最も代表的な卒業論文の概念として推定された。表8のA～Iの概念に相当する、一般名詞とサ変接続名詞を組み合わせた卒業論文の表題例を表9に示す。「-----問題に関する研究」に相当する具体例として、「日韓両国の金融問題の研究」「公共施設予約システムの現状と問題点に関する事例研究」などが該当した。表9に示したように、他の一般名詞とサ変接続名詞の組み合わせにおいても、該当すると考えられる表題が存在することを確認できた。従って、卒業論文の課題や構成を考えるために、よく使用される一般名詞とサ変接続名詞を組み合わせることにより形成される概念を参考にできると考えられた。

表8 一般名詞とサ変接続名詞の上位3語を組み合わせた概念

	研究	考察	分析
問題	A 問題の研究	D 問題の考察	G 問題分析
システム	B システムの研究	E システムの考察	H システム分析
現状	C 現状の研究	F 現状の考察	I 現状分析

表9 一般名詞とサ変接続名詞を組み合わせた卒業論文の表題例

概念	一般名詞とサ変接続名詞の組み合わせ	該当すると考えられる具体的な表題
A	問題の研究	日韓両国の金融問題の研究
		公共施設予約システムの現状と問題点に関する事例研究
B	システムの研究	流通情報システムの研究
		Javaを用いた穴埋め問題作成システムについての研究
C	現状の研究	固定資産課税業務における地理情報システムの現状に関する研究
		移動実態の分析による県境の現状に関する研究
D	問題の考察	高齢化問題と大学生の意識調査の考察
		死刑存廃問題に関する一考察
E	システムの考察	HORBによる分散処理システムの構築に関する一考察
		自然環境の未来を作るシステムの考察
F	現状の考察	製麺業界に関する若干の考察—特に生麺業界の現状と課題について—
		Web上の「ニュース」についての考察 ～現状とその展望～
G	問題分析	2002年サッカーワールドカップにおける新潟スタジアム周辺の交通アクセス問題についてのシミュレーション分析
		国内自動二輪車ユーザーの意識調査 ～スズキへのイメージと問題点の明確化から考察するユーザーニーズの分析～
H	システム分析	製造卸売業における情報システム分析・設計
		(有)ゴトーミルクプロダクトのシステム分析
I	現状分析	古町商店街の現状分析と活性化策の提案
		スポーツ・スポンサーシップの現状の分析

データベースの維持管理の面から考えると、登録する卒業生にとっては卒業要件を満たすために登録しているのが現状で、相当する見返りを得ていない場合が多い。自分の登録した内容は検索する必要がなく、卒業後は使用することもないことを考えると、卒業論文着手時期や卒業論文作成時期にデータベースを活用できる環境を提供する必要がある。

特に卒業論文作成時期に日本語表現や図表の表示を含めたフォーマットの学習、引用文献の整理の仕方や本文中における引用方法の学習などに、先輩の卒業論文を参考にすることは有効である。卒論着手時期に学生自らが、取り組む課題を設定したり、論文の構成を考えたりするのも、複数の先輩の卒業論文を参照することは有効と考えられる。また、過去の先輩の卒業論文を参考

にすることにより、時系列的に課題を把握することも可能になり、時代背景の変化についても気づくことができる。そして、これらの利用は原文情報の利用が基本になる。

抄録の主な利用方法として以下のような場面が想定できる。

- ・論文の内容や分野を決める。
- ・先行卒業論文を調べる。

原文の主な利用方法として以下のような場面が想定できる。

- ・引用文献の整理方法を理解する。
- ・本文中における文献の引用方法を理解する。
- ・卒業論文執筆の際に章立てや、目次を考える。
- ・目標規定文、新規性、主張点など論文の骨子にあたる部分のまとめかたを調査する。
- ・日本語の書き方を理解する。

学生の情報環境は Web 情報の利用が主体となっており、活字情報の利用比率が減少している。従来の図書、雑誌、新聞などの活字情報を利用する環境とともに、デジタル情報の利用環境も重要になっている。卒業論文データベースもスマートフォンなどで原文情報を利用できる環境が求められるようになって考えられる。

## 6. 今後の検討課題

今回形表題に使用されている用語の解析に、形態素解析システムに茶釜を使用した。辞書がカナやアルファベットに対応していないため、本来名詞等に分類されるべき全体の 1 割程度の用語が未知語として処理された。全体の傾向には影響がないと思われるが辞書の充実した環境での解析が好ましい。また、形態素解析とは別に、全表題をデータベースにして中間一致で検索した場合の件数もチェックしたところ、茶釜により切り出された語数と一致しない用語が存在した。データベースの場合は中間一致で検索しており、品詞で切り出すわけではないので不一致が生じるのは避けられないが、その差異に関する考察はできなかった。さらに、表題中に同じ用語が重複する場合は事前に想定できなかったため、一部ではあるが同一表題に同一の言葉を 2 回使用している場合は頻度が 2 として処理されている。

頻度の時系列変化については、有意な増加、減少を認めることはできなかったが、参考のため増加、減少の傾向として示した。20 年、30 年といった長期間のデータが蓄積すれば有意差を認めることができる可能性が高くなる。

卒業論文データベースの登録、維持に関して図書館スタッフの小片氏、佐藤氏からの多大な支援に感謝します。また、NetCommons を使った検索システムの構築に尽力頂いた情報センターの丑田氏に感謝します。

## 参考文献

- 1) 松本 裕治, “形態素解析システム「茶釜」〈特集〉使いやすくなった自然言語処理のフリーソフト: 知っておきたいツールの中身”, 情報処理 41 (11), 1208-1214, 2000-11-15
- 2) 立教大学, “2007 年度修士論文・卒業論文題目一覧”, 応用社会学研究 2008 No.50

247-252

- 3) 千葉大学文学部, “卒業論文・修士論文題目一覧”, <http://www.l.chiba-u.ac.jp/ja/418/m>, 2013-1-16
- 4) 昭和女子大学, “平成二十四年度日本語日本文学科卒業論文題目” 學苑 867, 118-120, 2013-01-01
- 5) 長江 貞彦, 的場 崇志, “Building a New Graduation Thesis Registration and Retrieval System on the Web (Web 上での卒業論文の登録検索システムの構築)”, 近畿大学生物理工学部紀要, 近畿大学生物理工学部紀要 8, 56-63, 2001-05-31, 近畿大学
- 6) 岸野 秀樹, 平瀬 健吾, 菅沼 美佐子, 奥野 修一, “卒業論文データベース化プロジェクトの現状と課題”, 藍野学院紀要, 第 20 卷, 2006, pp.53-63
- 7) 小樽商科大学附属図書館, “卒業論文, 修士・博士論文 タイトル一覧”, 小樽商科大学, <http://www.otaru-uc.ac.jp/htosyo1/siryo/soturou/soturou.html>, 2013-1-16
- 8) 浜田 良樹, “情報系研究室における知的財産マネジメントに関する研究”, 情報処理学会情報処理学会研究報告 . CSEC, [コンピュータセキュリティ] 2005 (122), 37-42, 2005-12-09
- 9) 浜田 良樹, 瀬川 典久, 村山 優子, “大学の卒業論文制作における知的財産権の帰属および取扱についての考察” 電子情報通信学会技術研究報告 . ISEC, 情報セキュリティ 105 (193), 139-145, 2005-07-14
- 10) 田中, 早苗, 大塚, 有里, “本学服飾美術学科における卒業制作の主題に関する研究”, 東京家政大学研究紀要 第 49 集 (1), 2009 pp.71 ~ 78
- 11) 岩越守孝, 増田英孝, 中川裕志, “Web と携帯端末向けの新聞記事の対応コーパスからの文末言い換え抽出”, 自然言語処理. 12 (5), pp. 157-184, 2005-10
- 12) 国際基督教大学図書館, “卒業論文・修士論文・博士論文の検索”, 国際基督教大, <http://www-lib.icu.ac.jp/Thesis/index.htm>, 2013-01-30