

キーワードの機械的付与と人による付与

——『経済学文献季報』での試み——*

豊岡 文 英**

櫻田 忠 衛***

はじめに、

文献情報のデータベースを構築するとき、我々が想定しなければならないのは計算機全体にかかる負担をより軽くすることと、効率的な検索システムが設計されることである。とくに、検索システムの設計においては、検索のためのアクセスが自由な形で行われるようにどんな項目からでも検索が可能となることが必要とされる。つまり、データの入力時にデータのあらゆる属性が検索項目として活用されるように配慮されていなければならない。しかしながら、そのことによって計算機にかかる負荷が大きくなれば逆に検索効率が悪くなることが考えられ、現状においては検索システムの設計

* この稿は昭和58、59年度文部省科学研究費補助金（試験研究1）を受けた「共同利用データベースによる経済学文献情報システムの研究」（課題番号58830001、研究代表者木原正雄）のプロジェクトの一環として行った作業の結果をまとめたものである。この作業の結果については既にこのプロジェクトの研究成果報告書（1985年3月）に掲載され、第40回資料協議会総会（1985年5月16～17日、伊勢志摩）にて研究報告として発表したものであるが、今回多くの方からのご意見、ご批判をいただくために加筆訂正のうえ本誌に掲載していただくことになった。我々の作業は未だ完成したものではなく目的とするところの端初を開くにとどまっている。この機会に多くの方からのご意見、ご批判をいただきたいと願っている。

未熟な私どもに敢えてこの作業の機会を与えて下さった科研プロジェクト代表者の木原正雄先生と実質的な推進者であられた京都大学経済研究所の前田昇三先生、そして今回の本誌への掲載を快くお引き受けいただいた経済資料協議会出版委員会の菊川秀男氏に心から謝意を申し上げる。

この作業をすすめるにあたっては、資料協議会の方々から多くのご援助をいただいたが、とくにこのプロジェクトのメンバーでもあった図書館情報大学の松井幸子先生と大日本印刷株式会社 CTS 事業部の山崎清一氏には機械処理に関して多大のご援助をいただいた。また、日本科学技術情報センターの荒木啓介氏にはキーワード自動抽出の方法についてご多忙中にもかかわらず直接お会いしてご教示を受けることができた。記して謝意を申し上げます。

** とよおか ふみひで 名古屋大学経済学部

*** さくらだ たたえ 京都大学経済学部

における検索項目の設定がそうした限界をもふまえながら慎重に行われなければならない。

データベースからある文献情報を得ようとするとき、その検索項目としては著者名、著者所属機関、タイトル、掲載誌名、巻号、発行年、発行所、言語、キーワード等が考えられる。

これらの検索項目の中でも研究者が実際にどの項目を利用してデータベースにアクセスするかを調査した報告がある¹⁾。この報告は京都大学大型計算機センターにおける情報検索の実態を調査したもので、同センターで管理する16のデータベースへの研究者のアクセスの方法を調べたものである。それによると検索項目としてキーワード、タイトル、著者の三つを利用したものが全体の67.1%に達しており、16のデータベースのうちでも「よく整理された市販のデータベース INSPEC, INSPECJ, ERIC, JMARC や準市販データベース DESY, XDCBIB」をとり出してみるとそれらのデータベースでは「キーワード、標題、著者項目でほとんどの検索要求を満た」しており、その調査結果から「キーワード、標題、著者項目を検索項目としてレコードを構成し、キーワード値を充実させれば、文献二次情報では9割以上の検索要求を満たすことができる」と結論づけている²⁾。つまり、情報検索のためにデータベースにアクセスするには、検索項目としてはデータのあらゆる属性のうちでもキーワード、タイトル、著者項目の三つが最も多く利用され、重要と考えられている。これは現在の情報検索の方法からみるならば妥当といえる結論であろう。

この三つの検索項目のなかでもとりわけキーワードによる検索は大きな比重をしめる。第1表は渡辺、堀池の調査結果からの抜粋であるが検索項目の利用頻度はキーワード、標題、著者の三つでそれぞれのデータベースにおいて80%以上を越え、その三つのうちでもとりわけキーワードの利用はERIC, INSPECでそれぞれ59.3%, 62.4%というように50%をはるかに越えて群をぬいている。JMARCでは他のデータベースに比べてキーワードの利用が低く、著者項目の利用が高くなっているがこれはこのデータベースが単行書のみを収録していることと、タイムラグが大きいことが原因していると思われる。JMARCにアクセスする研究者は、現在の研究に必要な新しい情報を求めるのではなく、むしろ遡及的な情報を求めるという情報要求のちがによるものであろう。

以上のように現在の情報検索においては、キーワードが検索項目として大きな比重をしめているが、それはキーワードが他の検索項目に比べて検索する情報の内容にま

1) 渡辺豊英；堀池博己，京都大学大型計算機センターにおける情報検索，「情報管理」Vol. 27, No.3, June 1984.

2) 同上。

第1表 データ項目の利用頻度

項目名 DB名	キー 項目	キー ワード	標題	著者	所属	発行所	発行年	内容 梗概	その他	計
ERIC	54 2.4	1,339 59.3	517 22.9	200 8.9	—	—	97 4.3	4 0.2	47 2.0	2,258
INSPEC	1,413 4.4	20,188 62.4	2,983 9.2	5,883 18.2	1,033 3.2	22 0.1	629 1.9	8 0.0	189 0.6	32,348
JMARC	156 3.9	894 22.0	840 20.7	1,506 37.1	—	53 1.3	359 8.9	—	249 6.1	4,057

上段は利用回数，下段は利用率(%)を示す。

渡辺豊英；堀池博巳：京都大学大型計算機センターにおける情報検索，「情報管理」Vol. 27, No. 3 (June 1984) の第4表(210ページ)より抜粋。

でたちいって検索できるという特徴をもち，必要な情報を効率的に検索するための重要な要素となりえているからである。それ故に，これまでに検索項目としてのキーワードに関して数多くの研究がなされてきており，文献情報のデータベースを構築する際にはキーワードをいかに付与するかがそのデータベースが効率的でしかも正確な情報を提供できるかどうかの根幹的な課題とされている。

I 作業の目的

『経済学文献季報』はプリント・システムとオンライン検索システムの二つのサブシステムをもつ経済学情報システムをめざしている。現在，そのデータベースを構築するために経済資料協議会によってデータの入力が行われている。『経済学文献季報』（以下『季報』に略す）が経済学の文献情報を提供するデータベースとして将来オンラインを利用したアクセスに対応するためには，すくなくともキーワードを検索項目とした検索システムが設計されなければならない。

しかしながら，自然科学の分野とはちがって社会科学の分野ではそれぞれの論文に著者あるいは編集者によるキーワードの付与が未だになされていないのが現状である。経済学分野のデータベースの場合，検索項目としてのキーワードを付与しようとするとデータ作成者が経済学に関する専門的知識をもって論文に目を通し，キーワードを付与しなければならないことになる。この作業は，経済学に関する高度な知識と図書館学に関する専門的な知識と豊富な経験とを同時に要するものであり，一朝一夕にしてなせるものではとうていない。

『季報』を「経済学情報システム」として発展させようとした時，当初からこの問題につきあたり，解決の方法をさぐる大きな課題とされた。

我々の今回の作業は、その解決のための糸口を見出すことを目的としたものであり、同時に経済学に関するデータベースとしての『季報』の検索システム設計のための一素材を提供することを課題としている。

II 作業の方法

『季報』は1983年度第3号(111号)より、それまでの電算写植印刷から機械編集処理システムに移行して本格的なデータベース構築のためのデータの入力が始まった。

データの作成は、これまで続けられてきた方法を継承して経済資料協議会メンバーの中で採録誌の分担を行って、それぞれの機関でワークシートに記入する方法をとった。111号と112/113号では人手によるフリータームの件名項目を記入したが、それは各機関にかなりの負担を強いる結果となり、114号以降は件名項目の記入は断念せざるを得なくなった³⁾。

『季報』のデータベースからオンラインを利用して経済学文献情報の検索を行おうとするとき、キーワードを検索項目としなければならないことは先に述べたが、件名項目の記入を断念せざるを得なくなってしまった現在、それに代替される何らかの方法が考え出されなければならない。

我々は、既にデータ入力(件名項目も含まれている)の終わっている『季報』の111号と112/113号を利用して、この二つの号から人手によって記入された件名項目と、「HAPPINESS」(平和情報産業 K. K.)を利用してタイトルから自動抽出したキーワードを比較することによって、『季報』におけるキーワード検索の可能性を探ることとした。

ここでは、時間的な理由と財政的な理由とによって、上述した『季報』2つの号の全てについての分析は不可能になってしまったので、日本文献の雑誌論文でしかも『季報』の特徴⁴⁾を最もよく表現すると思われる「分類2」=「経済理論 経済学」のみの分析にとどめざるを得なかった。

『季報』111号と112/113号の日本文献の全レコード数は4,679件になるが、今回とりあげた「経済理論 経済学」の雑誌論文レコード数は472件で、全体の10.1%にす

- 3) 川原が『季報』の機械編集処理システムへの移行に際して、その労力コストに言及して「電算処理にともなう入力負担増は半ば常識化している。結局のところ、労力コストは累積版の作成その他、蓄積データの反復利用によってしか回収されないであろう」(川原和子、『経済学文献季報』機械編集システム、「大学図書館研究」21号、1982年12月、86ページ)と指摘されたのは正しかった。
- 4) 「経済学情報という名称は経済情報(日経 NEEDS や JOINT 等の蓄積データ)との区別を若干意識している。……KEIS(『季報』のこと=引用者)では従来どおり、主として大学の経済学部や研究所・センター等が生産する学術的情報を採録する」(川原和子、同上、85ページ)。

ぎない。これだけのデータで『季報』全体の傾向をおしはかるうというのは最初から無理があるが、その限界をよく認識したうえで、経済学分野におけるデータベースとしての『季報』構築のためのささやかな材料を提供できるかもしれないという期待をもって今回の作業をすすめることとする。

III 量的な概況

1. 「HAPPINESS」によるタイトルからのキーワード自動抽出

『季報』111号と112/113号のタイトルレコードから「HAPPINESS」を利用してキーワードの自動抽出を行った。その例は次に示すとおりである。

A0096A 0 0062 9311	@経験科学@と@して@の@経済学@ [@ < @ 2 @ > @] @	←タイトル
A0096A 0 0024 9313	経験科学@経済学	←自動抽出されたキーワード
A0099A 0 0092 9301	@経済@と@社会@の@大転換@の@時代@に@/@ライブラリー@・@ポリテックス試論@	←タイトル
A0099A 0 0058 9303	経済@社会@大転換@ライブラリー@ポリテックス試論	←自動抽出されたキーワード
A0170A 0 0062 9301	@動的@システム@に@おける@非線形計画@に@ついて@	←タイトル
A0170A 0 0034 9303	動的@システム@非線形計画	←自動抽出されたキーワード
A0171A 0 0042 9301	@商品生産@と@資本制@/@補遺@	←タイトル
A0171A 0 0030 9303	商品生産@資本制@補遺	←自動抽出されたキーワード
A0175A 0 0052 9301	@BASIC@に@よる@回帰分析プログラム@	←タイトル
A0175A 0 0038 9303	BASIC@回帰分析プログラム	←自動抽出されたキーワード

タイトルを切断処理して、不要語や記号などを除去し、残ったタームをキーワードとしてとり出している。

こうして抽出されたキーワードを量的にみると第2表のようになる。1レコードから平均3～4のキーワードが抽出されている。長いタイトルが付されるとキーワードの数は増えることになる。「分類3」＝「経済学史 経済思想史」は他の分類に比較してキーワード数が多くなっているが、これは「分類3」に長いタイトルが比較的多いということの反映であろう。

第 2 表

		--- フォンメイ ハツ キーフト* ヒント* ヒヨウ (シマツズ, コクセイヒ) --- 2.ウツサツシ ロフク*ン																			
キーフト*コ*ウカイ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
フォンメイ																					
0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	115	0	2	17	28	28	20	10	4	2	3	1	0	0	0	0	0	0	0	0	0
2	472	0	27	98	121	94	58	40	15	14	3	0	0	1	0	0	1	0	0	0	0
3	101	0	2	17	23	17	10	15	15	0	2	0	0	0	0	0	0	0	0	0	0
4	175	0	4	33	27	20	33	25	9	8	6	2	0	0	0	0	0	0	0	0	0
5	174	0	4	37	42	31	20	21	7	9	1	1	0	0	1	0	0	0	0	0	0
6	283	0	15	42	77	55	41	25	15	7	3	0	2	0	0	1	0	0	0	0	0
7	153	0	8	27	38	28	26	14	5	5	1	1	0	0	0	0	0	0	0	0	0
8	773	2	29	114	164	153	136	72	49	28	14	9	1	0	0	1	1	0	0	0	0
9	252	0	1	61	58	45	35	20	12	13	2	3	1	0	1	0	0	0	0	0	0
10	390	0	17	85	79	83	56	39	20	6	4	1	0	0	0	0	0	0	0	0	0
11	393	0	18	84	101	81	43	29	20	7	6	1	2	0	1	0	0	0	0	0	0
12	267	0	11	45	65	63	33	32	12	4	2	0	0	0	0	0	0	0	0	0	0
13	128	0	11	28	31	20	16	10	8	2	1	0	1	0	0	0	0	0	0	0	0
14	19	0	0	4	4	4	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0
15	352	2	15	56	76	67	52	33	26	8	13	1	3	0	0	0	0	0	0	0	0
16	38	0	5	4	14	6	4	4	1	0	0	0	0	0	0	0	0	0	0	0	0
コ*ウカイ	4086	4	169	752	948	803	588	390	220	113	61	20	10	1	3	2	2	0	0	0	0
0	100.0	0.0	0.0	0.0	0.0	0.0	0.0100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	100.0	0.0	1.7	14.8	24.3	24.3	17.4	8.7	3.5	1.7	2.6	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	100.0	0.0	5.7	20.8	25.6	19.9	12.3	8.5	3.2	3.0	0.6	0.0	0.0	0.2	0.0	0.0	0.2	0.0	0.0	0.0	0.0
3	100.0	0.0	2.0	16.8	22.8	16.8	9.9	14.9	14.9	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	100.0	0.0	2.3	18.9	15.4	16.0	18.9	14.3	5.1	4.6	3.4	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	100.0	0.0	2.3	21.3	24.1	17.8	11.5	12.1	4.0	5.2	0.6	0.6	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0
6	100.0	0.0	5.3	14.8	27.2	19.4	14.5	8.8	5.3	2.5	1.1	0.0	0.7	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0
7	100.0	0.0	5.2	17.6	24.8	18.3	17.0	9.2	3.3	3.3	0.7	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	100.0	0.3	3.8	14.7	21.2	19.8	17.6	9.3	6.3	3.6	1.8	1.2	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0
9	100.0	0.0	0.4	24.2	23.0	17.9	13.9	7.9	4.8	5.2	0.8	1.2	0.4	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0
10	100.0	0.0	4.4	21.8	20.3	21.3	14.4	10.0	5.1	1.5	1.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	100.0	0.0	4.6	21.4	25.7	20.6	10.9	7.4	5.1	1.8	1.5	0.3	0.5	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0
12	100.0	0.0	4.1	16.9	24.3	23.6	12.4	12.0	4.5	1.5	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	100.0	0.0	0.6	21.9	24.2	15.6	12.5	7.8	6.3	1.6	0.8	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	100.0	0.0	0.0	21.1	21.1	21.1	26.3	5.3	5.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	100.0	0.6	4.3	15.9	21.6	19.0	14.8	9.4	7.4	2.3	3.7	0.3	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	100.0	0.0	13.2	10.5	36.8	15.8	10.5	10.5	2.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
コ*ウカイ	100.0	0.1	4.1	18.4	23.2	19.7	14.4	9.5	5.4	2.8	1.5	0.5	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0

2. 人手による件名の付与

先述したように『季報』では、機械編集処理システムに移行した最初の2号、111号と112/113号は各機関のワークシート作成者によって件名項目が記入された。この件名項目の記入法は、マニュアルによると「採録論文又は単行書のタイトル中のキーワードを基本とし、(フリータームもふくめて可)計4個までを記入する」とし、具体的な方法については「①日本語はできるだけ一語に成語化すること、②人名は被伝者として人名欄に入っているため、キーワード化する必要がない、③分類中に地域コードの入っていない国名、地名を必ず件名化する、④歴史上の事件名、団体名を件名化する、⑤その他固有名詞でキーワードとなりそうなものはできるだけ件名化する」とされている。

111号と112/113号でマニュアルに従って付された件名の量的な概観は第3表に示さ

第 3 表

--- フ*コンレイ ハ*ツ ケンメイ ヒント* ヒヨウ (シ*ツスク、コウセイヒ) ---											2.フサ*ツシ ロンフ*ン	
ケンメイ	コ*ウケイ	0	1	2	3	4	5	6	7	8	9	10
フ*コンレイ												
0	1	0	0	0	0	1	0	0	0	0	0	0
1	115	3	19	48	24	21	0	0	0	0	0	0
2	472	6	94	156	138	78	0	0	0	0	0	0
3	101	0	25	29	29	18	0	0	0	0	0	0
4	175	2	5	26	49	93	0	0	0	0	0	0
5	174	0	20	52	45	57	0	0	0	0	0	0
6	283	0	30	91	81	81	0	0	0	0	0	0
7	153	0	9	54	44	46	0	0	0	0	0	0
8	773	0	71	214	243	244	1	0	0	0	0	0
9	252	0	14	82	81	75	0	0	0	0	0	0
10	390	2	61	141	108	77	1	0	0	0	0	0
11	393	2	76	143	97	75	0	0	0	0	0	0
12	267	0	34	98	69	66	0	0	0	0	0	0
13	128	0	12	50	30	36	0	0	0	0	0	0
14	19	0	3	5	5	6	0	0	0	0	0	0
15	352	0	40	98	108	106	0	0	0	0	0	0
16	38	1	15	11	9	2	0	0	0	0	0	0
コ*ウケイ	4086	16	528	1298	1160	1082	2	0	0	0	0	0
0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
1	100.0	2.6	16.5	41.7	20.9	18.3	0.0	0.0	0.0	0.0	0.0	0.0
2	100.0	1.3	19.9	33.1	29.2	16.5	0.0	0.0	0.0	0.0	0.0	0.0
3	100.0	0.0	24.8	28.7	23.7	17.8	0.0	0.0	0.0	0.0	0.0	0.0
4	100.0	1.1	2.9	14.9	28.0	53.1	0.0	0.0	0.0	0.0	0.0	0.0
5	100.0	0.0	11.5	29.9	25.9	32.8	0.0	0.0	0.0	0.0	0.0	0.0
6	100.0	0.0	10.6	32.2	28.6	28.6	0.0	0.0	0.0	0.0	0.0	0.0
7	100.0	0.0	5.9	35.3	28.8	30.1	0.0	0.0	0.0	0.0	0.0	0.0
8	100.0	0.0	9.2	27.7	31.4	31.6	0.1	0.0	0.0	0.0	0.0	0.0
9	100.0	0.0	5.6	32.5	32.1	29.8	0.0	0.0	0.0	0.0	0.0	0.0
10	100.0	0.5	15.6	36.2	27.7	19.7	0.3	0.0	0.0	0.0	0.0	0.0
11	100.0	0.5	19.3	36.4	24.7	19.1	0.0	0.0	0.0	0.0	0.0	0.0
12	100.0	0.0	12.7	36.7	25.8	24.7	0.0	0.0	0.0	0.0	0.0	0.0
13	100.0	0.0	9.4	39.1	23.4	28.1	0.0	0.0	0.0	0.0	0.0	0.0
14	100.0	0.0	15.8	26.3	26.3	31.6	0.0	0.0	0.0	0.0	0.0	0.0
15	100.0	0.0	11.4	27.8	30.7	30.1	0.0	0.0	0.0	0.0	0.0	0.0
16	100.0	2.6	39.5	28.9	23.7	5.3	0.0	0.0	0.0	0.0	0.0	0.0
コ*ウケイ	100.0	0.4	12.9	31.8	28.4	26.5	0.0	0.0	0.0	0.0	0.0	0.0

れている。

マニュアルでは件名は4個までつけられることになっているが、全体としては1文献に2～4の件名が付されている。分類別に少し注意してみると「4 経済史」, 「5 経済地理 地域開発」, 「7 各国経済政策 各国経済事情」, 「8 産業」, 「14 人口」, 「15 労働問題 社会問題」など実証的分析が中心になる分野では付与された件名の量が多くなり、逆に「1 人文科学 社会科学」, 「2 経済理論 経済学」, 「16 統計学」などの抽象性の高い理論分野では付与された件名の量が少なくなっている。

IV 自動抽出キーワードと人手による件名の比較分析

「HAPPINESS」の利用によって自動抽出したキーワード(以下、自動キーワードという)と人手によって付された件名(以下、単に件名という)との比較分析を試みる。ここでは先述したように『季報』111号と112/113号のデータのうち、日本の雑誌文献の「2 経済理論 経済学」をその対象にした。

1. 自動キーワードと件名の対応関係

第4表は自動キーワードと件名の対応関係を量的に総括した表である。

第4表 自動キーワード=件名総括表

区 分	個 数	%
イ) 一 致	361	28.5
ロ) 不 一 致	317	25.1
ハ) 非 対 応	587	46.4
計	1,265	100.0

「一致」「不一致」を具体的に例示する。

「一致」

1) 完全一致

(自動キーワード) (件 名)

インフレーション——インフレーション

経済人類学 —— 経済人類学

資本財需給分析 —— 資本財需給分析

所得エントロピー——所得エントロピー

経済開発モデル —— 経済開発モデル

2) 語幹一致

①接尾語の有無

(自動キーワード)	(件名)
マルクス経済学説	——マルクス経済学
スタグフレーション論	——スタグフレーション
貨幣的	——貨幣
価値	——価値論
長期世界不況下	——長期世界不況

語幹一致は「接尾語の有無」だけに限られるものではなく、「接頭語の有無」「単数・複数形による異なり」なども考えられるが、ここでは出現しなかった。

「不一致」

①同一概念であるが異なる単語

(自動キーワード)	(件名)
スラッファ型	——スラッファ理論
アメリカ	——アメリカ合衆国
古典派	——古典学派
わが国	——日本

②概念のねじれがあるもの

(自動キーワード)	(件名)
ME革命	——マイクロ・エレクトロニクス技術
ケインズ理論	——ケインズの均衡
インフレ期待下	——インフレーション
不均衡問題	——不均衡理論
再分配行動	——再分配理論
拡大再生産	——再生産論

③複合語による異なり

(自動キーワード)	(件名)
インフレ予想	——インフレーション
二部門間stagflationモデル	——スタグフレーション
均衡分析	——一般均衡分析
転形	——転形問題
企業	——代表的企業
貨幣	——貨幣的生産理論
雇用失業問題	——雇用

「非対応」

単語としては一致しなかったもの

第4表で示されたように、自動キーワードと件名は「一致」と「不一致」あわせる
と何らかの形で一致しているのが53.6%になり、半数以上の一致がみられる。

2. 件名の分析

次に件名の付与のされ方についてみることにする。

第5表 件名の付与のされ方からみた自動キーワードとの対応関係

	タイトルから	内 容 から	計
一 致	165(62.5)	13(4.9)	178
不 一 致	71(26.9)	61(22.8)	132
非 対 応	28(10.6)	193(72.3)	221
計	264(49.7)	267(50.3)	531

タイトル中のワードからそのままとり出されて付与されたものと、タイトル中には
出てこないワードで内容からおしはかって付与された件名とに区分すると第5表のよ
うになる。件名全体としては、タイトルと内容から付与されたものは半分づつである
が、自動キーワードとの対応関係では当然のことながらタイトルから付与された件名
の方が良く一致し(62.5%)、内容から付与された件名は非対応が多い(72.3%)。

内容から付与された件名のうち72.3%が自動キーワードとは非対応の関係にあるが、
これをいくつかの例をとってみる。

- @発展@の@過程法則@/@否定@の@否定@の@法則@に
- @ついて@ ←タイトル
- 発展@過程法則@否定@法則 ←自動キーワード
- @経済法則@
- @弁証法@ } ←件 名
- @ヘーゲル論理学@ }
- @不確実線形離散時間@モデル@の@一つ@の@安定条件@ ←タイトル
- モデル@安定条件 ←自動キーワード
- @不確実パラメータ@
- @線形離散時間@モデル@ } ←件 名
- @安定性分析法@
- @計量経済学@

@合理的@予想モデル@の@性格@と@問題点@	←タイトル
合理的@予想モデル@性格@問題点	←自動キーワード
@マネタリスト@	}
@マクロ経済学@	
@新古典派@理論@	
@合理的@予想形成仮説@	
@冒頭商品論@の@方法@に@ついて@	←タイトル
冒頭商品論	←自動キーワード
@商品論@	}
@「@資本論@」@の@方法@	
@転形@問題@に@ついて@／@置塩信雄氏@の@所説@に@	
@関連して@	←タイトル
転形@置塩信雄@所説@関連	←自動キーワード
@転形@問題@	}
@生産価格論@	
@平均利潤率@	
@マルクス価値論@	
@資本@一般@と@諸@資本@の@蓄積@／@『@要綱@』	
@「@資本@一般@」@の@理論質@の@中間総括@を@か	
ねて@	←タイトル
資本@蓄積@要綱@理論質@中間総括	←自動キーワード
@資本蓄積論@	}
@剰余価値論@	
@「@経済学@批判要綱@」@	

例示のごとく、タイトル中のワードからだけでは正確にその情報までとりつくことが不可能なものがかかなり多い。内容からの件名付与は正確で効率的な検索のための条件となる。この結果からするならば、自動キーワードのみが検索項目とされることには大きな限界があるといわざるを得ない。

3. 自動キーワードの分析

自動キーワードが検索項目とされるためには、すくなくともタイトル中からの切断が正しくなされて、検索用語として機能することが可能でなければならない。

それをみよとしたのが第6表である。

第6表のそれぞれの区分について具体的な例をあげておく。

イ) 切断が正しくかつ有効なワード

第6表 自動キーワード切断度表（全数内訳）

区 分	個 数	%
自動キーワード		
イ) 切断も正しく有効なワード	375	51.1
ロ) 切断は正しいが、組み合わせなければ有効でないもの	22	3.0
ハ) 切断は正しいが、雑データが残ったもの	9	1.2
ニ) 切断は正しいが、不用語・不適切語	111	15.1
ホ) 切断が短かすぎるもの	143	19.5
ヘ) 切断が長すぎるもの	4	0.5
ト) 切断が不適切で不用語・不適切語	56	7.6
チ) 切断が不適切で組み合わせなければ検索できない	13	1.8
リ) 切断が不適切で雑データが残ったもの	1	0.2
計	734	100.0

価格機構，価値価格，家事労働，技術進歩，計量分析 etc.

- ロ) 切断は正しいが組み合わせなければ有効でないもの
危機，法則，批判，系譜，概念規定 etc.
- ハ) 切断は正しいが雑データが残ったもの
森岡孝二著，故吉田静一，関恒義教授 etc.
- ニ) 切断は正しいが不用語・不適切語
一考察，覚書，視角，研究ノート，再検討 etc.
- ホ) 切断が短かすぎるもの
現実的，歴史的，動学的，国際的 etc.
- ヘ) 切断が長すぎるもの
マルクス恐慌論，賃労働論再論，国家独占資本主義世界体制，Global Uni-
valence 再考 etc.
- ト) 切断が不適切で不用語・不適切語
意義，論理，論的，視，両，四者 etc.
- チ) 切断が不適切で組み合わせがなければ検索できないもの
論 etc.
- リ) 切断が不適切で雑データが残ったもの
択一労働供給モデル

この表からは，タイトルからのキーワード自動抽出の方法をそのまま利用しても50%は，適切なキーワードが得られることがわかる。そして，経済学用語の特性を研究

してのシソーラスの作成や複合語、多重検索処理、不用語・不適切語処理などのシステムの改善が行われるならば、自動キーワードによる検索項目の付与は相当の効力を発揮することになるであろう。

ただ、先述したように自動キーワードの付与はタイトル中からのみという限定をもつために、タイトルが具体的でない場合にはその内容を表現したり、あるいは関連した検索のためのワードが抽出できないという大きな限界をもたざるを得ない。

人手による件名が付与され、それを補完する機能として自動キーワードが利用されるのが最も良い方法ではあるが、現在の『季報』編集における力量を考えたとき、人手による件名の付与のための負担はあまりにも大きく、また、件名付与のための経済学用語のシソーラスの研究もまだ十分に進んでいない現状がある。

こうした現状を考え合わせるならば、自動キーワードによる検索項目の付与はいくつかの限界をもちながらも当面はこの方法によらざるを得ないであろう。