

シラバスからの関連用語収集手法の検討と検索システムへの応用

芳鐘 冬樹[†] 井田 正明[†] 野澤 孝之[†] 宮崎 和光[†] 喜多 一[‡]

大学評価・学位授与機構[†] 京都大学 学術情報メディアセンター[‡]

1 はじめに

大学における授業計画を紹介するシラバスには、授業の内容を表す多くの専門用語が含まれている。シラバスに含まれる専門用語間の関係を認識できれば、科目の関連性や大学ごとの傾向の分析[1]などに活用できる。そこで、本研究では、ある用語と関連する概念を持つ関連語や、より限定されたスペシフィックな概念を持つ下位語をシラバスから自動的に収集する手法の提案を行う。また、提案手法に基づいて作成したシラバス検索システムについて報告する。

2 関連用語のタイプと収集対象

用語間の関係のタイプには、同義関係、階層関係、関連関係がある。さらに、階層関係は種類関係、全体部分関係、例示関係に、関連関係は同一カテゴリに属する語の関係と、異なるカテゴリに属する語の関係に分けられる[2]。本研究では、これらのうち、同一カテゴリに属する関連語、種類関係の下位語、そして、それらの関連語・下位語および元の用語の同義表現を収集の対象とする。このような対象の制限は、形態的/統語的な言い換えの認識のみに基づく本手法の限界によるものだが、逆に、これらを他と区別して収集できるという利点があると言える。例えば、同一カテゴリに属する関連語だけを収集し、共通するカテゴリを上位語として系統的に整理できれば、単に関連語を集めるよりも精緻な知識が得られるものと考えられる。

3 収集手法

関連用語の自動収集手法に関しては、文書内の共起情報に基づくものなど、様々な研究が存在する[3][4][5]。しかしながら、それらの多くは、関連用語の収集にとどまり、関係のタイプの区別までは行っていない。区別を行っているものも、収集源が辞書の語義文などに限定されていたり、区別の範囲が限定されており[6][7]、

必ずしも十分とは言えない。本研究の手法では、複合語の形態的/統語的な言い換えに着目し、元の用語を言い換えた異形（を加工したものを）、元の用語に關係する用語として抽出することで、言い換えの分類に基づいて、関係のタイプ（同義表現、下位語、その他の関連語）を区別しつつ、関連用語を収集する。

言い換え表現抽出のために、表1に例示したような品詞パターンにより記述される言い換え規則を作成し、それらを言い換え認識システムFastr[8]に実装した。使用した212規則[9][10]は、(i)機能語の付加/削除、品詞変化、主要部入替のみが生じるもの、(ii)修飾語として内容語が加わるもの、(iii)元の構成要素と等位関係を持つ形で内容語が加わるもの、の何れかに分類できる。

表1 言い換え規則の基本例¹

(i)	N1 N2	N1 S1 N2
(ii)	N1 N2	N1 N3 N2
(iii)	N1 N2	N1 SC1 N3 S1 N2

(Nは名詞，Sは助詞，SCは名詞接続助詞)

(i)の規則が適用される異形を、元の用語の同義表現として抽出する。(例：「生体計測」の同義表現として「生体の計測」を抽出)

(ii)の規則が適用される異形を、元の用語の種類関係にある下位語として抽出する。(例：「生体計測」の下位語として「生体物理計測」を抽出)

(iii)の規則が適用される異形から、「並列関係を持つ用語」を分離して、元の用語の関連語として抽出する。(「生体計測」の関連語として、「生体・環境計測」から分離した「環境計測」を抽出)

シラバスから抽出した関連用語は、図1のような2系列のツリー形式に整理する。1つめは、入力用語の主要部をルートとするツリー、2つめ

Automatic Collection of Related Terms from Syllabus Data and its Application to the Retrieval of Syllabi

[†]Fuyuki Yoshikane, Masaaki Ida, Takayuki Nozawa, and Kazuteru Miyazaki

National Institution for Academic Degrees and University Evaluation

[‡]Hajime Kita

Academic Center for Computing and Media Studies, Kyoto University

は修飾部をルートとするツリーである．入力用語を[a]の位置に，入力語の同義表現を[a']の位置に，下位語（およびその同義表現）を[c]の位置に出力する．[b1][b2]の位置には，それぞれ，入力用語の主要部を共通のカテゴリとする関連語（およびその同義表現），対象・目的などを表す修飾部を共通のカテゴリとする関連語（およびその同義表現）を出力する．



図1 用語間関係

4 検索システム

前節で述べた関連用語収集手法に基づくシラバス検索システムを作成した．予め収集・情報抽出したシラバス集合を対象に，ユーザが入力した検索語，およびその関連用語を含むシラバスを検索し，検索語との関係のタイプごとに整理して表示するというものである．本システムでは，見つかった関連用語を図1の形式で系統的に整理したうえで，それぞれの用語についてその用語を含むシラバスへのリンクを貼る（その際，当該科目が属す学科・コースの情報も付す）．検索語が表すトピック一般か，よりスペシフィックなトピックか，あるいは関連トピックと合わせて扱われているかに従って分類することで，ユーザが，検索要求に適ったシラバス（科目）を探しやすくなるとともに，提示される関連用語は，次に検索語を修正する際の参考になり，シラバス検索の利便性の向上が期待できると考えられる．

一例として，「エネルギー問題」を検索語とし，2つの大学（A・B）の工学部のシラバスを対象に検索を行った結果を示す．

A 大学 エネルギー問題<2>

関連語：環境(の)問題<10>，資源問題<1>

下位語：エネルギーの需給問題<1>，

エネルギー環境問題<2>

B 大学 エネルギー問題<3>

関連語：資源問題<2>，人口問題<1>

<>は，その用語と「エネルギー問題」の両方を含むシラバスの数を表す．例えば，この結果から，エネルギー問題は，A 大学では環境問題とともに扱われることが多く（化学系・機械系など多様な学科で扱われている），B 大学では資源問題・人口問題とともに扱われる傾向がある（地球工学科で扱われている）ことが分かる．

本システムは，(1)シラバス検索と(2)科目間の関連性の図示（および，それに基づく大学ごとの傾向の分析）を目的としたものだが，本研究の関連用語収集手法は，(3)シソーラスの半自動構築などへの応用も可能である．ただし，学術論文を対象とした場合[9]に比べ，シラバスでは専門用語のバリエーションが少なく，量的な面では必ずしも満足のいく収集結果は得られなかった．今後，言い換え規則の拡充などにより改良を試みたい．

謝辞 本研究の一部は「科学研究費補助金若手研究(B)16700245」によるものであり，ここに謝意を表します．また，ご協力いただいた大学評価・学位授与機構「評価情報研究会」参加者の皆様に謝意を表します．

参考文献

- [1] 井田，野澤，芳鐘，宮崎，喜多：シラバスデータベースとシステム工学教育，SSI2004（計測自動制御学会 システム・情報部門学術講演会），pp. 107-112，2004．
- [2] 日本工業規格：シソーラスの構成及びその作成方法：JIS X 0901-1991，日本規格協会，1991．
- [3] Chen, Yim, Fye: Automatic thesaurus generation for an electronic community system, Journal of the American Society for Information Science, Vol.46, No.3, pp. 175-193, 1995．
- [4] 永松，田中：コーパスから抽出した係り受け共起情報に基づく類似度と文書検索における評価，情報処理学会研究報告，NL-116，pp. 73-78，1996．
- [5] 小原，山田，絹川，中川：ウェブを利用した関連用語収集，FIT2004，pp. 183-184，2004．
- [6] 鶴丸，井上，日高，吉田：語義文からの階層関係の自動抽出，電子情報通信学会技術研究報告，NLC92-2，pp. 9-16，1992．
- [7] 佐藤，佐々木：ウェブを利用した関連用語の自動収集，情報処理学会研究報告，NL-153-8，pp. 57-64，2003．
- [8] Jacquemin: Fastr: a unification-based front-end to automatic indexing, RIAO'94, pp. 34-47, 1994．
- [9] Yoshikane, Tsuji, Kageura, Jacquemin: Morpho-syntactic rules for detecting Japanese term variation: establishment and evaluation, Journal of Natural Language Processing, Vol. 10, No. 4, pp. 3-32, 2003．
- [10] Kageura, Yoshikane, Nozawa: Parallel bilingual paraphrase rule for noun compounds: concepts and rules for exploring web language resources, ALR-04, pp. 54-61, 2004．

¹ 本研究は，関連用語収集の最初のステップとして，専門用語の大部分を占める「2要素から成る複合語」を対象にしている．