

## 工学系シラバスに関連した数学ウェブページ群の提示

## Presenting Mathematical Web Pages Related to Engineering Syllabuses

山田 奉子<sup>\*1</sup>  
Kuniko YAMADA上田 洋<sup>\*1</sup>  
Hiroshi UEDA村上 晴美<sup>\*2</sup>  
Harumi MURAKAMI辰巳 昭治<sup>\*1</sup>  
Shoji TATSUMI<sup>\*1</sup> 大阪市立大学大学院工学研究科  
Graduate School of Engineering, Osaka City University<sup>\*2</sup> 大阪市立大学大学院創造都市研究科  
Graduate School for Creative Cities, Osaka City University

Opportunities continue to grow for science students to search the Web for mathematical terms used in engineering lectures. However, obtaining appropriate web documents to understand lectures is difficult. We therefore proposed a system that presents mathematical web pages to understand lectures. The system displays web pages related to an input syllabus. In this paper, we describe the improvements of the system that provides more useful web page searches. The Improvements include mathematical dictionary, language processing algorithms, stop lists, and ranking algorithms.

## 1. はじめに

工学系の授業中で扱われる数学概念について学習する際に、Web を検索する機会も増加してきたが、ウェブ文書は書物のようにページを繰ることができず、俯瞰性が悪く全体像が掴みにくい。一方、数学書は最初のページから読むことを前提に構成されており、道具として数学を利用する場合、使いにくい場合も多い。また、せっかく利用価値の高い文書が Web にあっても、数学用語を1つ1つキーワード検索にかけた結果の寄せ集めでは、各用語の相互関係もわかりにくく内容の深さの程度もバラバラである。かといって、and 検索では、組み合わせのパターンが増えてしまい、組み合わせ論的爆発に陥ってしまう。

そこで、まとまりがありかつある程度の広がりを持って授業で述べられている数学的内容を補足できるウェブページ群をユーザに提示する手法を[山田 07]で提案した。これは、近年シラバスが、授業内容を最も詳しく示す資料として、大学評価の支援([宮崎 05])にも積極的に使われているなど、シラバスが注目されていることに着目したものである。

今回、システムの改善を行い、以前のシステムより多くの有用な数学ウェブページ群を得られるようになった。改善点は、数学辞書の充実、シラバスからの名詞の抽出方法、不要語のフィルタ処理、得られた数学ウェブページ群のランキング方法である。

## 2. 改善点について

## 2.1 数学辞書作成

シラバスから数学用語を抽出するための辞書として、数学用語のテキストファイルを作成する。前回は、旧文部省編纂の「学術用語集 数学編」を用いた。単語数は 2544 語である。今回、さらに講談社サイエンティク編集の「数学英和小事典」4810 語、青木和彦他編集「岩波 数学入門辞典」4103 語を追加し、総語数 8897 語の辞書を作成した。

## 2.2 シラバスからの名詞抽出

前回、シラバスから名詞を抽出する際に、隣接する名詞は複合語としたが、今回さらに、「 $\pi$  定理」の中で、 $\pi$  が人名の場合、「 $\pi$  の定理」となるので、一語となるようにした。また、

「 $\pi$  法則」では、「の」が入る例と入らない例があり、(例:「オームの法則」と「オーム法則」)それぞれをキーワードとして検索すると違った結果が返ってくるので、シラバスでどちらの表現を使っているか、両方の検索結果が得られるようにした。他に「方程式」「公式」「関数」などもあるが、同様に処理した。また、数学では 2 名連記の定理や公式が非常に多く、英語表記ではすべて「-(ハイフン)」でつながら、日本語表記では他に「無表記」「-(全角マイナス)」「=(全角イコール)」「=(半角イコール)」「 $\cdot$ (中黒)」など、様々な記号が使われている。数学辞典は概ね「-(ハイフン)」であるが、前記学術用語集では続けて並べている。(例 コーシーリーマン)これらが一つの単語として抽出されるようにした。

## 2.3 不要語のフィルタ処理

抽出された数学用語をキーワードとして得られた検索結果には大量の不要な URL が含まれており、これらを除去しないと、検索件数を増やせば増やすほど、 unnecessary 書誌情報やシラバスが検索結果の上位に出現することになる。今回は 10 語程度の簡単な不要語リストであったが、今回、改めて不要な URL を分析し、書誌情報については出版社や書店を表す語、シラバスについては、「kamoku」「kyoumu」などの関連する語を登録し、さらに論文、ソフトウェア、ニュース、オークションなどに関係する語等を含め、111 語の不要語リストを作成し、それらを含む URL を除去するようにした。

## 2.4 ランキングの方法

前回、「数学用語をできるだけ多く含む」ということに重点を置いた結果、ウェブページ群のランキングを数学用語数の降順にした。しかし、出力結果の分析から1つのウェブページに多くの数学用語が入っているものは、シラバスや書誌情報などの不要なものの可能性が非常に高いことが分かったので、ランキングのアルゴリズムを見直し、ウェブページ群の URL 数の降順とし、同数の場合はさらに、数学用語数の降順とした。

## 3. 提案手法

シラバス中の数学用語をできるだけ多く含みかつ解説しているウェブページ群を提示する手法を提案する。

連絡先: 山田 奉子, 大阪市立大学, 〒558-8585 大阪市住吉区杉本 3-3-138, d07tb553@ex.media.osaka-cu.ac.jp

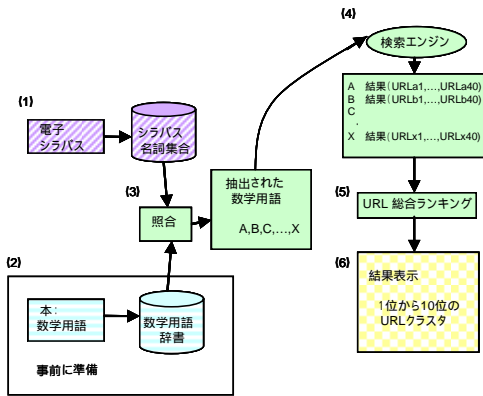


図1: 提案手法の概要

図1に提案手法の概要を示し、図中の番号に従って説明する。  
 (1)電子シラバスをテキストファイルに変換したものを入力データとし、Chasenを用いて形態素解析にかけ、名詞のみ抽出する。  
 (2)数学辞書を作成する。(3)シラバス名詞集合と数学用語辞書を照合することで数学用語を抽出する。(4)得られた数学用語を1件ずつキーワードとしてGoogle Web APIsを用いURL集合を得る。ここでは、(キーワード, URL)のペアが、キーワード毎のクラスタとなっている。(5)(4)で得られたURL集合をすべて集めた集合を作る。その後、(キーワード, URL)のペアを、URL毎のクラスタに並べ直す。そのままではほとんどのクラスタの要素数が1となるので、URLのうち右側から「/.../」の部分いくつか切り捨てた(注:切り捨てる数を以下カット数と呼ぶ)結果が同じものについて、クラスタに分ける。ただし、無関係なページまで纏まらないように、http://.../...の部分は残すようにする。これが、表題の数学ウェブページ群である。(6)上位5群を表示する。

4. 実験

前システムでは、1数学用語につき検索件数40件、カット数1が最適なパラメータであった。また、情報工学系大学院生10名によるアンケートでは、概ね元のシラバスと関連する数学ウェブページ群が得られたという評価であった。

4.1 方法

提案手法を実装した実験用のシステムを構築し、前システムとの相違を調べた。使用したシラバスは前回同様、本学工学部の電子シラバスより、次の5個を用いた。「工業数学 演習」「電磁気学」「電磁気学 演習」「電磁気学」「物理数学」

これらのシラバスを用いて、カット数を1から5まで変え、それぞれ検索件数を10件から10件刻みで100件までとる。

4.2 結果と考察

(1) 結果

表1は辞書の充実と名詞抽出方法の改善により、シラバスからの抽出用語数が増えていることを示している。

表1: 実験用シラバスから抽出された用語数

	工業数学 演習	電磁気学	電磁気学 演習	電磁気学	物理数学
前回	1)13	15	9	8	24
今回	2)17	27	20	14	38

1)の内訳: 単位, 複素関数論, 複素数, 幾何学, 複素関数, 正則関数, 等角写像, 複素積分, コーシー, 定理, テイラー展開, ローラン展開, 留数定理, ベッセル関数

2)で追加された用語: 特殊関数, コーシーの定理, ガンマ関数, ベータ関数

出力結果については、フィルタ処理とランキング方法の見直しによって検索件数を増やしていても、殆ど不要なURLは出現しなくなった。また、カット1では、得られる数学ウェブページ群のURL数がやや少ないが、カット2から5では、それ程大きい変化はなかった。次に前システムとの比較を表2, 表3に示す。カット1で、評価は、数式が含まれるかどうかで 2:良い 1:やや良い 0:よくない の3段階で行った。不要なものが除去されているのがわかる。

表2: 「工業数学 演習」の前システムによる出力結果

取得件数	評価	1位			2位			3位			4位			5位		
		URL数	内容	URL数	内容	URL数	内容	URL数	内容	URL数	内容	URL数	内容	URL数	内容	
40件	2	複素関数論	5	講義要項	2	0	リンク	4	0	幾何学おもちゃ世界	2	2	積分定理: 留数定理	2	2	
20件	2	複素関数論	6	0	講義要項	2	0	リンク	4	2	積分定理: 留数定理	3	1	グラフ式	2	
10件	2	複素関数論	7	0	講義要項	2	0	シラバス	2	2	留数定理	4	0	リンク	4	
40件	2	複素関数論	8	0	講義要項	3	2	複素関数積分	6	0	講義一覧	2	0	書籍情報	5	
10件	2	複素関数論	9	0	シラバス	3	2	講義要項	3	2	講義一覧	6	0	講義一覧	2	
10件	2	複素関数論	9	0	シラバス	3	0	講義要項	3	2	複素関数	6	0	シラバス集	3	
10件	2	複素関数論	9	0	シラバス	4	0	講義要項	3	3	講義要項	2	2	複素関数	6	
10件	2	複素関数論	9	0	シラバス	4	0	講義要項	3	0	講義要項	2	2	複素関数	6	
10件	2	複素関数論	9	0	シラバス	4	0	シラバス集	4	0	講義要項	3	0	書籍情報	4	
100件	2	複素関数論	9	0	シラバス	4	0	シラバス集	4	0	書籍情報	5	0	シラバス集	2	

表3: 「工業数学 演習」の出力結果

取得件数	評価	1位			2位			3位			4位			5位		
		URL数	内容	URL数	内容	URL数	内容	URL数	内容	URL数	内容	URL数	内容	URL数	内容	
170件	2	複素関数論	4	2	複素関数論入門	3	2	関数計算	3	2	ベッセル関数	2	2	複素関数	2	
20件	2	複素関数論	4	2	複素関数論入門	4	2	実関数積分	3	2	関数計算	3	2	ベッセル関数	3	
40件	2	複素関数論入門	5	2	複素関数論	4	2	関数計算	4	2	実関数積分	3	2	ベッセル関数	3	
40件	2	複素関数論入門	5	2	関数計算	4	2	複素関数論	4	2	実関数積分	3	2	ベッセル関数	3	
50件	2	複素関数論入門	7	2	複素関数論	4	2	関数計算	4	2	実関数積分	3	2	ベッセル関数	3	
60件	2	複素関数論入門	7	2	複素関数論	5	2	関数計算	4	2	実関数積分	3	2	ベッセル関数	3	
70件	2	複素関数論入門	7	2	複素関数論	5	2	関数計算	4	2	実関数積分	4	3	ガンマ関数	4	
80件	2	複素関数論入門	7	2	複素関数論	6	2	ガンマ関数	5	2	関数計算	4	2	実関数積分	4	
90件	2	複素関数論入門	7	2	複素関数論	7	2	ガンマ関数	5	2	関数計算	4	2	実関数積分	4	
100件	2	複素関数論	7	2	複素関数論入門	7	2	ベッセル関数	5	2	ガンマ関数	5	2	関数計算	4	

(2) 考察

今回の結果から1位から5位までのウェブページ群を個別に評価するのではなく、総合的に評価する方法もあると気付いた。例えば表3の関数計算ではパラメータを与えると数式とグラフをかくしてくれるが、それだけでは物足りない。しかし、他のウェブページ群と組み合わせると有効に利用できるであろう。また、カット数と検索取得件数を変えても不要なものが殆ど出なくなった結果、ユーザがこれらのパラメータを変更できるようにしたほうが、1位から5位の組合せが色々変わってより有効になる可能性があると考えられる。

5. おわりに

数学辞書の充実、名詞の抽出方法、不要語のフィルタ処理、ランキング方法の改善により、不要なURLを排除して数学に関するウェブページを出力することができるようになった。それを受け、考察で述べたことも含んで新しい観点から被験者による評価実験を行う予定である。さらに、シラバス中に必要な数学が適切な用語を用いて記述されていなければこの手法は使えないので、その場合はもとのシラバスに近い他のシラバス群から必要な数学用語を抽出したものを代替に用いるなど、更なる方法も今後検討したい。

また、この手法は工学系のシラバスに限らず数学用語が抽出できる文書ならば応用が可能である。さらに、辞書を変えれば、他の分野の文書についても応用が考えられる。

参考文献

[山田 07] 山田奉子, 上田洋, 村上晴美: 授業理解のための数学ウェブページの提示, 第21回人工知能学会全国大会, IG2-4, 2007.  
 [宮崎 05] 宮崎和光, 井田正明, 芳鐘冬樹, 野澤孝之, 喜多一: 電子化されたシラバスに基づく学位授与と事業のための科目分類支援システムの試作, 情報処理学会論文誌, Vol.46, No.3, pp.782-791, 2005.