

授業理解のための数学ウェブページの提示

Presenting Mathematical Web Pages for Understanding Lectures

山田 奉子*¹ 上田 洋*² 村上 晴美*¹
Kuniko YAMADA Hiroshi UEDA Harumi MURAKAMI

*¹ 大阪市立大学大学院創造都市研究科
Graduate School for Creative Cities, Osaka City University

*² 大阪市立大学大学院工学研究科
Graduate School of Engineering, Osaka City University

Science students have had more and more opportunities to search the Web for mathematical terms used in engineering lectures lately. However, it is difficult to obtain web documents appropriate for lectures. Therefore we propose a system which presents mathematical web pages to understand lectures. We use a syllabus to represent the content of the lectures and present web documents appropriate for mathematical terms in lectures. We also show the result which we had engineering graduate students evaluate the relevance between five syllabuses and these presented web pages.

1. はじめに

工学系等の授業中で扱われる数学概念について学習する場合、従来ならば図書館等で紙メディアを検索してきたが、近年では Web 上で検索する機会も増加してきた。しかし Web 文書は書物のようにバラバラとページを繰ることができず途中で終わっていてもわからなかったりするなど、俯瞰性が悪く全体像が掴みにくい。一方、数学書は昔から分野ごとに定評のある書物があり理論体系として形も整っているが、最初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者にとっては、使にくい場合も多い。

また、せっかく利用価値の高い文書が Web 上にあっても、検索エンジンを用いて数学用語を一つ一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 検索をしようにも、すぐに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。当該の講義や関連する数学に対してすでに知識があれば、的確な少ない数のキーワードを用いて検索エンジンにかけ、良質の Web 文書を得ることも可能であろうが、学習途上の者にとっては難しい。

そこで Web 上の文書の中から、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文書をユーザに提示する手法を提案する。

2. 提案手法

本研究では、授業を特徴づけるものとして電子シラバスを用い、シラバス中の数学用語をできるだけ多く含む Web ページを提示する。

2.1 シラバスについて

シラバスとは、[新村 98]によると、「講義実施要項・講義内容・達成課題・使用テキスト・参考文献・テスト方法などについて記した計画書」となっている。社会情勢の変化につれて授業の中身についての説明が求められ、学内だけではなく外部からもアクセスできる Web 上のシラバスが増えている。受験生や外部評価に利用されることもあるので、内容も充実し実際の授業内容を示す資料となってきている。また、シラバスを用いた研究(例えば[宮崎 05])も見受けられるようになってきた。

実際の電子シラバスについて、Google で検索語を「シラバス 大学」とし、上位から 10 個の大学工学系講座の中から無作為に抽出したものと、本学工学系講座から無作為に 10 個抽出したものと、計 20 個について調べてみた。結果は、行数は 25 行から 115 行、平均行数は 50 行、文字数は 470 字から 1650 字、平均は 1000 字程度であった。

2.2 数学辞書について

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。今回は、旧文部省編纂の「学術用語集 数学編」[文部省 71]を用いた。語彙数が、数学約 1600 語、統計数学約 560 語の、英和と和英の対訳が記載されており、日本語部分を用いてテキストファイルを作成した。作成に当たった留意点を、以下に記す。(1)1語のみは除去(2)複数表記が可能な単語は両方登録(「だ円」と「楕円」、「関数」と「函数」、「2乗」と「二乗」、「n(全角)」と「n(半角)」など)

2.3 システムの概要

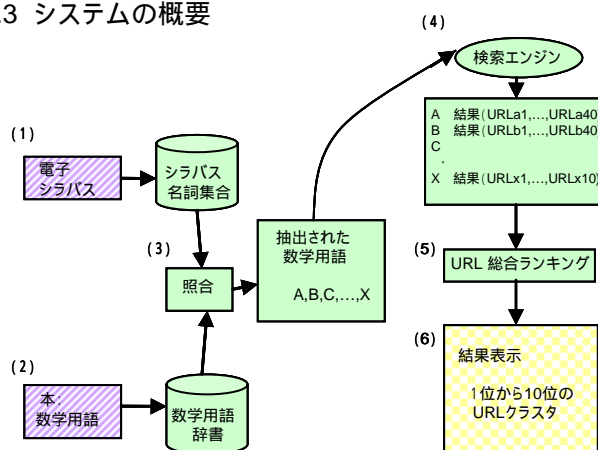


図1: システムの概要

図1にシステムの概要を示し、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る。名詞のみ抽出し、隣接する名詞は複合語にする。(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。得られた URL 集合の中で、URL 中に「blog」「book」「syllabus」「news」等を含むものは、削除する。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。ただし、元の URL が複数のキーワードに対応して複数回出現するのは稀であり、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「/.../」の部分をつき切り捨てた結果が同じものについて、クラスタに分ける。この URL 毎のクラスタ(以下、URL クラスタと表記)について最後に、各クラスタをキーワード数の降順に並べる。(6) 上位 10 件を表示する。

3. 実験

3.1 方法

2章で述べた手法を実装した実験用のシステムを構築し、シラバスと得られた URL クラスタとの関連性を評価した。今回の実験では、1 キーワードにつき 40URL を取得し、クラスタを作るための URL の切り捨てを、右側から「/.../」の部分をつき切り捨てて行った。また、上位 10 件(但し Wikipedia は省く)の表示の際、キーワードが同数のものはコンピュータ出力順とした。シラバスは、本学工学部の電子シラバスより、表 1 の 5 個を用いた。

表 1: 実験用シラバス

シラバス	抽出された数学用語
工業数学 演習	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
電磁気学	単位, 積分法, ガウス, 定理, 微分方程式, グリーン関数, 計算, 系, 発散, 回転, ポテンシャル, ラプラス方程式, 境界値問題, 球面調和関数, 鏡像, 接続, 方程式
電磁気学 演習	単位, ガウス, ベクトル解析, スカラー積, ベクトル積, 行列, ベクトル, ラプラス方程式, ポアソン方程式, 境界条件, 数学
電磁気学	単位, 波動方程式, 境界条件, モード, 分散, ガウス, 一般解, ポテンシャル
物理数学	単位, ベクトル解析, 微分方程式, 偏微分方程式, 解析学, 数学, 関連, ベクトル場, ガウス, 定理, 初期値問題, 拡散方程式, 波動方程式, 境界値問題, 関数, ベクトル, 微分, 発散, 回転, 空間曲線, 積分, 座標変換, テンソル, 解, 軌跡, 応用数学

各シラバスに対応した 10 件の URL クラスタのうちの 1 件目について、「含まれる数学用語」と「URL 数(クラスタの要素数)」を表 2 に示す。

表 2: 1 件目の URL クラスタ

シラバス	1 件目に含まれる数学用語(キーワード)	URL 数
工業数学 演習	複素関数論, 複素関数論, 複素関数, 正則関数, 複素積分, コーシー, ローラン展開, ローラン展開, 留数定理, ベッセル関数	*8
電磁気学	ガウス, グリーン関数, 発散, ラプラス方程式, ラプラス方程式	4
電磁気学 演習	ガウス, ベクトル解析, スカラー積, スカラー積, ベクトル積, ベクトル, ラプラス方程式, ラプラス方程式	6
電磁気学	境界条件, 一般解, ポテンシャル, ポテンシャル	4
物理数学	ベクトル解析, ベクトル場, ベクトル場, ガウス, 定理, ベクトル, 発散, 空間曲線, 座標変換, テンソル, テンソル	6

さらに表 2 中の*部分の 8 個の URL について、実際の Web ページのタイトル(HTML の<title>部分)を表 3 に示す。

表 3: 各 Web ページのタイトル

タイトル
「ときわ台学/複素関数論/複素数・ガウス平面」 「ときわ台学/複素関数論/正則関数」
「ときわ台学/複素関数論/複素関数の積分」
「ときわ台学/複素関数論/コーシー積分定理」
「ときわ台学/複素関数論/コーシー積分表示」
「ときわ台学/複素関数論/ローラン展開」
「ときわ台学/複素関数論/留数定理」 「ときわ台学/複素関数論/リーマン面」

被験者は 10 名で、内訳は本学創造都市研究科 1 名、本学工学研究科 9 名の学生である。シラバスと実際の Web ページを見比べ、シラバスと URL クラスタとの関連性について 1 件ずつ、次の 4 段階で評価してもらった。(3: 非常に関連している 2: 関連している 1: 部分的に関連している 0: 全く関連していない)

3.2 結果と考察

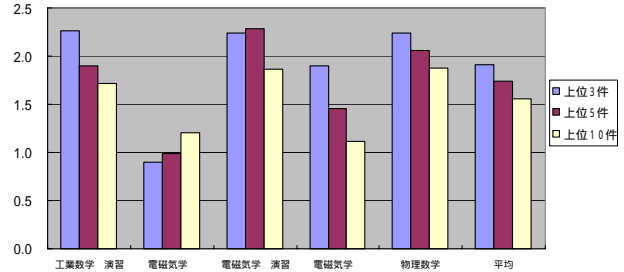


図 2: 実験結果

図 2 は各シラバスに対する 10 人の実験結果の平均を載せたものである。「電磁気学」の評価が極端に悪いのは、2 位と 5 位が書誌情報でその評価点が悪かった影響である。また、表 1 より「電磁気学」の抽出された数学用語数は一見多そうに見えるが、一般的な数学用語や数学に特化していない単語が多く、実際に良いキーワードとなるものが少なかったと考えられる。また、表 2 より、よい評価を得るためには、ある程度多くのキーワード数と URL 数をもったクラスタが得られることが必要であることもわかった。全体的には、上位 3 件の平均が 1.9 であり、まずまずの結果になった。また、上位 3 件、5 件、10 件の結果を比較すると、上位ほど良い結果が出る傾向も見られた。

4. おわりに

シラバスは語数が少なく、重み付け等の処理を一切行わずにシステムを構築したが、今後、抽出した数学用語の重要度を出力結果の順位付けに反映できるようにしたい。また今回数学用語辞書として利用した[文部省 71]は用語が少なく、シラバスに載っている「関数」「関数」等が抽出できなかった。また、逆にその中の統計数学用語に関しては、その単語 1 つでは数学用語と認識されない単語も多く入っており、見当違いの結果が生じる原因ともなった。今後、数学用語辞書についても Web 文書を用いて構築する方向を[中川 03]等を踏まえて考えていきたい。さらに、数学と関連が深いのが、当該シラバスのデータが非常に少なく用語が抽出できない場合に、類似の他のシラバスから用語を抽出する方法も検討していきたい。

参考文献

[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.

[新村 98] 新村出(編): 電子版 広辞苑 第五版, 岩波書店, 1998.

[文部省 71] 文部省(編): 学術用語集 数学編 初版 19 刷, 大日本図書株式会社, 1971.

[中川 03] 中川裕志, 森辰則, 湯本紘彰: 出現頻度と接続頻度に基づく専門用語抽出, 自然言語処理, Vol.10 No.1, pp.27-45, 2003.