

九州大学 WEB シラバス 横断検索システムの開発

稗 大輔*

本研究では、九州大学のWEBシラバスを横断検索できるシステムの開発を目的としている。現状では、九州大学の各学部・学府のシラバスは、一つのサーバに集約されておらず、学部や学科毎のサーバで運用されているものが大部分である。そこで、本研究では、これらのシラバスサーバを横断検索するシステムを作り、WEBシラバスのユーザビリティの向上を図る。

Development of Kyushu University WEB syllabus cross-searching system

DAISUKE HIE*

In this study, I intend to develop the system that can cross-search the WEB syllabus in Kyushu University. Currently, the syllabus system has not been aggregated into a single server for each department of undergraduate and graduate school in the Kyushu University. Most of those servers that are operated by each faculty and department. Then, the present study is to develop the system that makes possible to cross-searches these syllabus servers, and to attempt the improvement of the usability of the WEB syllabus.

1. はじめに

WWW上には、キーワードによる検索が可能なWEBページが数多く存在する。様々な人が、インターネットを利用し、検索システムを使って、求める情報を世界中から集めることが出来る。また、情報量が膨大となった現在では、WEBページから必要とする的確な情報のみを収集した総合検索という考え方が注目されている。そのようなサイトは、特定のテーマに沿った質の高い情報やサービスを提供していることが多く、それらのもつ情報量は、直接参照可能なWEBページよりも情報量が多いと言われている。しかしながら、検索結果のWEBページの多くは、背後のデータベースから動的に生成されるため直接参照できず、統合などによる再利用が容易ではない。そのため、WEBページを文書とみなし、多数の文書から知識を得るWEBマイニングについて様々な研究が行われている[1]。また、ページ群をデータベースのように用いる、データ統合技術についても研究されている[2]。

教育分野においても、情報技術の利用が進んでいる。Eラーニングなどの電子素材や、WEBシラバスなど、多くの高等教育機関でWEBを介した情報提供が行われている。各大学においても、シラバスや、教材を電子化し、公開するところもある。OCW(Open Course Ware)が始めた教材の公開は、日本にも波及しており、日本OCW連絡会では、複数の大学が教材の公開を行っている。

また、一つの教育機関のシラバス群は、その機関が提供する教育全体を表す指標になっている。大学における教育の外部評価や大学評価・学位授与機構における学位授与審査の際にも利用されている。

2. 研究背景

(1) 九州大学のシラバス

九州大学には総合履修選択方式という他学部の授業の単位が認められる履修上の仕組みや、21世紀プログラムという様々な学部の授業を自ら選択し、組み合わせて履修する学部が存在する。しかし、九州大学のシラバスはWEBで提供されているが、学部・学科・学府・コースなどによって、運営されているサイトやユーザーインターフェースなどが様々である。そのため、各学部のサイトで別個に検索するしかなく、横断的な授業名や授業内容を知ることができないものとなっている。そのため、平成21年度から九州大学シラバスという統合シラバスシステム

†九州大学大学院芸術工学部芸術工学専攻
Graduate School of Design, Kyushu University

(<http://syllabus.kyushu-u.ac.jp/>) が運営されているが、そこに統合されているのは、12 学部の中の 5 学部 (一つは学科)・18 学府の中の 6 学府にしか過ぎない。そのため、まだ九州大学のシラバス全体を横断検索できる実用的なサイトとは言えない。表 1 に九州大学シラバスに統合されていない学部・学科・学府のシラバスの一覧形式と詳細ページを学部の授業のみを例として示す。

表 1 学部のシラバス一覧

学部	一覧形式	シラバス詳細ページ
文学部	html	CGI
法学部	html	CGI
経済学部	html	PHP
理学部	html	html
医学部	html	CGI
歯学部	html	PDF
薬学部		PDF
工学部 (建築学科)	html	PDF
工学部 (物質工学科)	html	html
工学部 (地球環境工学科) 建設都市コース		PDF
工学部 (地球環境工学科) 船舶海洋システム工学コース	html	PDF
工学部 (地球環境工学科) 地球システム工学コース	PDF	
工学部 (エネルギー科学科)	PDF	PDF
工学部 (機械航空工学科) 機械工学コース	html	PDF
工学部 (機械航空工学科) 航空宇宙工学コース	html	

(2) 九州大学WEBシラバスの現状

九州大学のシラバスには、九州大学シラバスシステムとそれ以外の各学部・学科・学府ごとにシラバスシステムが、独自のサーバ上で運用されており、本学のホームページから一覧の形で別々のリンクが貼られている (<http://www.kyushu-u.ac.jp/student/education/syllabus.php>)。そのなかには、表 1 で分かるように HTML で書かれたものや PDF、CGI による動的なWEBページなど、様々なものが存在する。また、複数の学部や学科を横断する検索が出来ないだけでなく、情報を見るためだけでもそれぞれ異なるリンクを見なければならない。

(3) ユーザーからの視点

九州大学のシラバスは、学生から見てどのような意見を持っているかについてアンケート調査を行った。アンケート調査の対象は、複数の学部・学科の授業を履修する必要があり、WEBシラバスをよく使うと思われる 21 世紀プログラムの学生で、シラバスのユーザビリティや問題点などを挙げてもらった。図 2～5 に結果を示す。

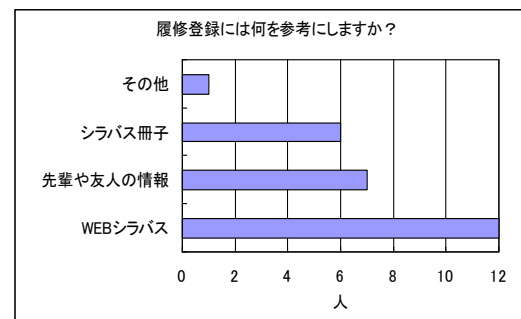


図 2 履修登録時の参考物のアンケート結果

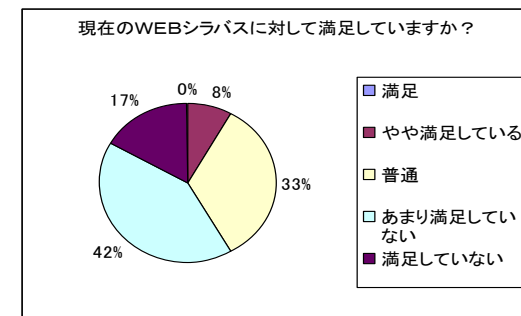


図 3 現在のWEBシラバスのユーザビリティのアンケート結果

アンケートの回答者は、12人であった。12人全員が、図2よりWEBシラバスを参考に行っていることが分かった。また、WEBシラバスのみではなく、冊子や友人の情報を得ていることがわかった。しかし、6割以上が、現状のシラバスのユーザビリティについて満足していないという結果が図3より分かる。次に現状の問題点の結果を図4に示す。

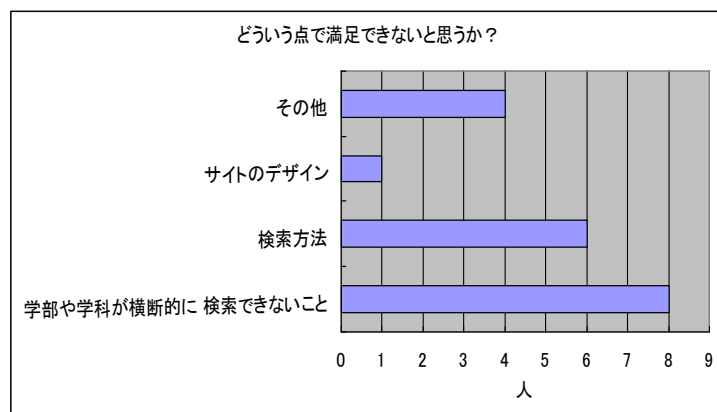


図4 具体的な問題点のアンケート結果

図4では、具体的なユーザビリティの回答を得た。やはり、横断的に検索できないことに不満を感じている学生が、過半数を占めた。

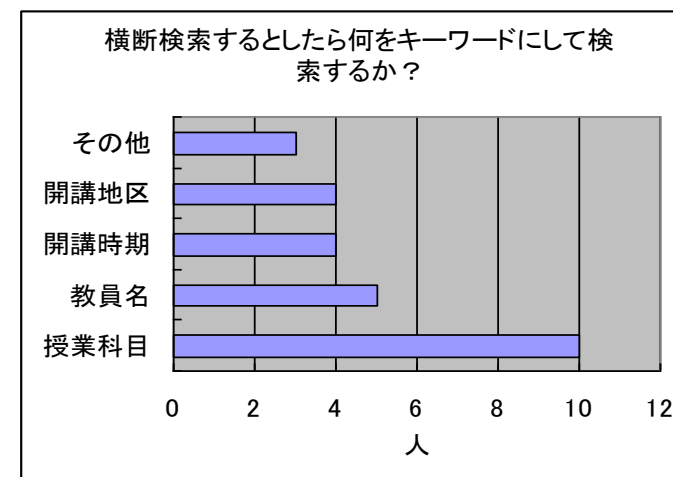


図5 キーワード検索方法のアンケート結果

図5のアンケート結果では、横断検索が実現した場合の、望む検索方法を調べた。それによると、授業キーワードで調べる人が多数を占めた。これは、自分が興味ある分野をまず、検索で調べ、それに合わせて、同じ地区の授業や近い時間帯での授業を調べているということであろう。

以上のアンケートにより、まずは、授業キーワードから、横断的に検索できるシステムを構築することが現状の問題点を改善する一つの方法であると考えた。

3. 関連研究

3.1 横断検索

インターネット上には数十億ページという膨大な数の情報が掲載されているが、ユーザにとってあらゆる分野の情報が必要なわけではなく、自分のニーズに適合した情報が得られることが重要である。検索システムを使って、情報を探すが一般的なWEBの利用法であるが、必要な情報のみを収集した横断検索という考え方が注目されている。横断検索の方法として、3つの研究が進んで行われている。

(1) WEBページの自動収集による情報統合

既存のWEBサイトから関連する情報のみをクロウリングにより自動的に収集し、収集した情報から必要な情報のみを自動分類して抽出するといった方法である。代表

的なものとして、Google(<http://www.google.co.jp/>)がある。Google は、200以上のサーバで動くスパイダが世界中の約33億からなるURLを収集し、全世界11箇所にある2万台のPCで、一日あたり2億件の処理をしている[3]。

しかし、この分野は、情報量が増え続けていることやURLや情報の内容自体が変化することを前提としたWEBページでは、収集したデータの更新や修正を人手で行うことは容易ではない。この状況を改善するための分野がWEBマイニングである。WEBマイニングでは、膨大な情報から計算機を用いて特定の分野の有用な情報のみを取り出す。Google は、収集時に目的に適合する収集条件を基にフィルタリングして、関連するページのみを収集する方法と、対象となるWEBサイトから全ての情報を収集した後、設定した条件でフィルタリングして必要なWEBページのみを抽出している。WEBシラバスにもWEBマイニングの研究が進んでおり、大学のシラバスの自動収集の報告がある[4]。これは、収集時にフィルタリングして条件に合致したページのみを収集する方法を取っている。報告では、htmlの共通語彙から構造を抽出し、シラバス一覧ページとそこからリンクされる個々のページのhtmlを分析して、共通語彙から構造を抽出し、またシラバス一覧ページとそこからリンクされる個々のページを説明するページを評価し、シラバスであるかを判定している。しかし、該当する分野の情報を網羅する割合と、収集した情報が的確である割合を高く出来るかが課題となっており、これらは、相反する条件であるため実現が難しいという点とWEBからのリンク情報から辿れないURLを有するページの収集が出来ないという課題がある。

(2) メタデータを用いた統合検索

メタデータとは、コンテンツの中から特定の情報を抽出するために付与する項目や語彙のことで、「タイトル」や「概要」といった項目が定義され、XMLで記述される。そのため、「タイトル」というメタデータタグのついた語彙のみを検索することにより、その情報につけられた題名からの検索が可能となる。これらの統合検索では、検索システムにメタデータのみが登録され、コンテンツ自体は、外部からアクセス可能であれば、インターネット上のどこのサーバに置かれてもよいという特徴を有する。メタデータを利用した検索は、メタデータに書かれた内容を検索して、メタデータに含まれるURL情報を元にコンテンツへリンクされる仕組みである。教育分野では、LOM(Learning Object Metadata)と呼ばれ、教育情報ナショナルセンターのポータル検索サイトなどで運用されている[‡]。

しかし、この分野の研究は、メタデータの自動付与の精度がまだ低く、コンテンツから自動的に抽出することが難しいため、人的に処理せざるをえず、コンテンツ保有者に負担をかけるという課題がある[1]。

(3) WEBサービスを用いた検索システムの統合

ここでのWEBサービスは、ネットワーク上で動いている異なるサービスを結合することである。ここでの結合は、ハードウェアやプログラム言語の異なるWindows系のシステムやUnix形のシステムなどのOSの異なるシステムを結合することである。それを実現するために、相互運用技術が必要であり、結合するシステム間で、送信側と受信側でデータを交換する仕組みとして、SOAP(Simple Object Access Protocol)とWSDL(Web Service Description Language)が開発されてきた。SOAPは、システム間で構造化されたデータをXMLを用いて情報を交換することを目的としたプログラムであり、WSDLは、データのやり取りの方法を規定するものである。メタデータがXMLで記述されていることにより、(2)で説明したメタデータを利用した検索システムの統合に適した方法である。この方法を用いている教育用コンテンツの総合検索の例としてMERLOT(<http://www.merlot.org/merlot/index.htm>)の総合検索がある。MERLOTは、北米の教材情報共有プロジェクトで、高等教育e-Learning教材のためのGatewayプロジェクトである[5]。Merlotでは、既存の検索サイトを連携する場合に、1箇所をゲートウェイとみなし、そのゲートウェイからそれぞれのサイトにアクセスして、検索結果を統合してゲートウェイ上のサイトに表示している。

しかし、SOAPを利用した場合、個々の検索サイトが検索した結果を統合して表示するため検索時間がかかる。また、個々の検索運用者間で稼働を厭わない連帯が必要であり実現が容易ではない。

4. システム提案

本研究では、WEBページの自動収集によるWEBシラバスの統合を試みる。システムの構成として、情報収集・検索・情報の更新という3つの視点に着目してシステムを開発する。システム構成を図6に示す。

[‡] (<http://www.nicer.go.jp/>)

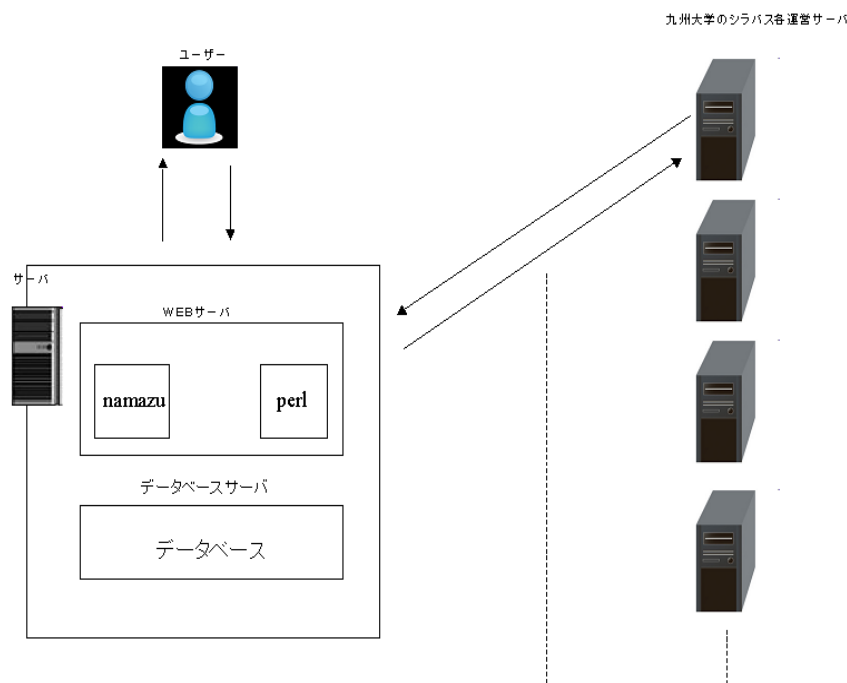


図6 システム構成

(1) 情報収集

情報収集のために、スクレーパと呼ばれるプログラムを制作する。前章で、スパイダは少し説明したが、WEBを横断して、情報収集してくるプログラムのことである。ページ郡、ファイル郡、あるいはその両方から特定の情報を複数取得するものがスパイダと呼ばれ、ファイル郡から特定の情報の一部だけ抽出するのがスクレーパと呼ばれている。今回、スクレーピング（情報収集）と対象とするのは、各学部・各学科・学芸大学の授業科目一覧ページである。授業科目一覧ページをスクレーピングし、データベースに登録する。各学部・学科・学芸大学の授業科目一覧ページのURLは、手動で登録する。

一覧ページは、授業科目と授業科目の詳細ページのリンクが付いているページがほとんどである。そのため、シラバスの詳細ページなどのコンテンツを一つずつスクレーピングするよりは、まず、学部・学科・学芸大学から横断的に授業を検索できる授業科目一覧のページをスクレーピングする方法を選んだ。しかし、一覧がないPDFのみの

学科や、excelでの時間割のみ学科は、プログラムでテキストベースに直し、そこからスクレーピングを試みる。

(2) 検索

横断検索システムの検索方法としては、まずは、授業科目名をキーワードとして検索できるシステムを構築する。ユーザのアンケートにあったとおり、最も使われるキーワードが授業科目名だからである。検索の構造としては、Namazu[6]を用いた全文検索システムを構築する。全文検索とは、必要なデータを取得する部分、取得したデータをまとめて処理するインデксаと呼ばれる部分を検索することをさす。全文検索システムには、2つの機能があり、1つは、インデックス作成を担当する「インデкса」と呼ばれる部分、もう1つは、インデックス生成で作成されたデータを利用して、実際に検索を行う「サーチエンジン」と呼ばれる部分である。このインデксаの機能により、高速検索を実現している。この全文検索を手軽に使えることを目的としたフリーソフトが、Namazuである。このNamazuを用いて、収集した情報から、授業科目のキーワードを検索できるよう開発する。

(3) 更新

各学部・学科・学芸大学のシラバス一覧のページは、約半年、もしくは年に一回更新される。しかし、決まった日にちはなく、いつ更新されるのか分からない。そこで、本システムでは、cronというUnix系のスケジューラを用いる。これは、作業内容をファイルに登録しておくだけで自動的に作業が実行されるものである。cronを用いて、一日一回スクレーピングを試み、新しいデータに書き換える。

5. おわりに

本研究では、学部・学科・学芸大学のシラバスが横断検索できないという現状に対し、実際のユーザである学生に対しアンケートを実施し、早急に横断検索システムを開発すべきであることが分かった。システムの開発に対しては、「3. 関連研究」で説明した3つの関連研究を参考にスクレーピングという自動収集を行う。その際に、課題であった収集の正確性に対し、シラバスというコンテンツから情報を得るのではなく、授業科目一覧からページをスクレーピングすることと、授業一覧ページのURLを手動で指定することにより、確実に収集することを目的としている。まずは、学生が多く使うであろう授業科目キーワードからの検索を開発する。今後は、そのシステムをベースにユーザアンケートで評価を行い、シラバス自体からのキーワード検索や教官名、開講地区、曜日などの検索方法の付与をおこなっていきたい。

謝辞 本研究でアンケートの回答をいただいた 21 世紀プログラムの皆様、九州大学大学院芸術工学研究院藤村直美先生、九州大学大学院芸術工学府安德恭彰氏に謹んで感謝の意を表する。

参考文献

- 1)篠原正典、地蔵真作:Web 上の高等教育用コンテンツの総合検索の課題と効果的検索への提案、メディア教育研究 第一巻 第二号
- 2)中森哲也、大森敬介、廣川左千男:WebDB の Query Form におけるメタデータ自動抽出,日本データベース学会 Letters Vol.5, No.2
- 3)山名早戸:検索エンジンのアーキテクチャ,情報の科学と技術,54(2),pp84-49
- 4)伊藤栄典、島松千春、廣川左千男、篠原正典:Web シラバス統合による教育情報ライブラリ構築
- 5)安武公一、中島英博、井上仁:Merlot と Creative Commons を利用した e-Learning 教材共有のあり方とオープンソース的な活用方法
- 6)馬場肇;改訂 Namazu システムの構築と活用,ソフトバンク パブリッシング
- 7)石橋智美、小山祐太、古賀健一郎、古川優太郎、藤村直美:九州大学におけるシラバスシステムの問題と学生の視点からの提案、平成 20 年度 電気情報関係学会九州支部連合大会
- 8)井田正明、野沢孝之、芳鐘冬樹、宮崎和光、喜多一:シラバスデータベースシステムの構築と専門教育課程の比較分析への応用, 大学評価・学位研究, No.2, 2005.
- 9)Kevin Hemenway,Tara Calishain 著,村上雅章訳:Spidering Hacks,オライリー・ジャパン