

論文

# 自然言語処理と可視化を利用した 履修選択支援システムの実用化

美馬 秀樹<sup>1,a)</sup>

受付日 2019年6月19日, 再受付日 2019年12月26日,  
採録日 2020年2月29日

**概要:** 社会の新たな課題の解決において, 学際的, 分野横断的教育の発展は, 非常に有効な手段となり得るにもかかわらず, 大学等の高等教育において, 社会のあらゆる課題に対応できるカリキュラムを提示できる状況にはない. このような問題に対し, 本論文では, シラバスを対象とし, カリキュラムの検索, 構造化を行うシステム (MIMA サーチ) を提案し, その実用化について述べる. MIMA サーチの特徴は, 自然言語処理技術を活用することで, 与えられたキーワードに関連する講義の関連性を検索結果に応じて即時的に計算し, この結果を基にカリキュラムの全体像の可視化を行うことにある. MIMA サーチを利用した履修支援により, 分野や学科, 専攻によらず, 社会の課題等に対して, 必要な知識をより効率的に習得することにつながることを期待される. MIMA サーチは, 東京大学において, 2006 年よりシラバス構造化, および講義ビデオ (UTokyo Open Course Ware) の構造化システムとして実用化され, 10 年以上にわたりサービスを続けている. 実用化後の実際にシステムを利用したユーザ評価を基に, 本システムの有用性についても論述する.

キーワード: カリキュラム, シラバス, 構造化, 可視化, 自然言語処理, 機械学習

## Practical Realisation of a Course Selection Support System Using Natural Language Processing and Visualisation

HIDEKI MIMA<sup>1,a)</sup>

Received: June 19, 2019, Revised: December 26, 2019,  
Accepted: February 29, 2020

**Abstract:** The recent explosion of scientific knowledge due to the rapid advancements in academia and research has made it difficult for students and educators to gain a comprehensive overview of their syllabi. Additionally, the growing number of interdisciplinary fields of research has compounded the difficulty for students to find subjects that they would prefer from their syllabi. In such a situation, structuring of issue-oriented curricula is expected to be an efficient method to aid suitable subject selection for students and educators alike. In this paper, we introduce an issue-oriented automatic curricula structuring system that integrates automatic term recognition as an issue extraction and similarity calculation as terminology-based document clustering. In the system, retrieved syllabi are automatically classified according to the terms or issues that are included in them. Compared to conventional simple keyword-based syllabus retrieval systems, the proposed system's advantage is its capability to provide efficient support in selecting suitable courses from a large number of syllabi by using visual-based GUIs even to novice users (students). The system is already in operation practically in the form of issue-oriented syllabus structuring at the University of Tokyo's Open Course Ware, school or department of engineering syllabi and University of Tokyo's lecture catalogue. Questionnaire-based evaluations analysing the practical usability of the system have also been also incorporated.

**Keywords:** syllabus retrieval, term extraction, clustering, curricula structuring, visualization, natural language processing, machine learning

### 1. はじめに

近年の科学の著しい発展や, 工学の拡大, 専門分野の深化等を背景に, 知識の量も爆発的に増加している. また,

<sup>1</sup> 東京大学  
The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan  
<sup>a)</sup> mima@he.u-tokyo.ac.jp

このような知識の爆発的増加の結果として、分野の過度な細分化が行われ、結果として、学術知の体系が非常に分かりにくくなっている。一方で、環境、エネルギーのような、地球規模での複雑で多様な問題が顕現化し、学際的、分野横断的知識の活用を促す仕組みの構築がより重要性を増している。たとえば、ものづくりと、環境問題、安全性やデザイン性、さらには、倫理、法令遵守等は、今や切っても切れない関係にある。このように、社会の課題に対応するためには、分野によらず、より「広く」、「深く」知識を習得する必要があるといえるが、必ずしも教育の現場において、そのように講義の履修を支援する仕組みが提供されているとはいえないのが実状である。

他方、近年のICT (Information Communication Technology) の発展により、教育のICTによる支援も拡充してきており、大学シラバスの電子化とも相まって、授業検索システム等が履修選択においても重要なツールとなっている。しかしながら、従来のキーワード入力と、シラバスに含まれる用語とのマッチングの度合い等を基にした単一のランキングによる全文検索のみでは、キーワードに関連する授業を羅列することはできても、依然として講義間の関連によるカリキュラムの構造を把握するのは困難である。一方で、事前履修や事後履修といった講義間の関連を示すメタ情報をあらかじめ与えることにより、講義間の関連を示すことは可能であるが、そのような情報をすべての分野、課題に応じて、事前に人手で与えるのは現実的ではない。たとえば、東京大学工学部では約900の講義が開講されているが、それらに対する十分な情報が得られる状況とはいえない。学生にとっては、a) 自分の興味がどの講義(群)によって教授されるか(カリキュラムの検索)、b) その講義(群)を受けるために習得すべき講義は何か(講義間の関連の明瞭化)、等を得ることが望ましい。また、教官も同様に、講義全体の効率化のために、講義内容の重複や講義の抜けを知る必要があり、総じて、カリキュラムの構造化、およびその可視化が非常に重要となる。

本論文では、このような問題に対し、シラバスを対象とし、カリキュラムの検索、構造化を行うシステム(MIMAサーチ)について述べる。MIMAサーチの特徴、独創性は、自然言語処理技術を活用することで、与えられたキーワードに関連する講義の関連性を検索結果に応じて即時的に計算し、この結果を基にカリキュラム全体像の可視化を瞬時に行うことにある。MIMAサーチにより、分野や学科、専攻によらず、社会の課題等に対して、必要な知識をより効率的に習得することが期待される。MIMAサーチは、まずは東京大学工学部において、学部授業科目約900、大学院科目約700に対し2006年度に実運用が開始された。その後、2012年度からは全学の学部後期課程の授業を対象とした授業カタログシステムへと拡張された。さらに2018年度には学部前期課程科目も追加され、全学の約1万4千の授

業に対する検索可視化システムとして現在に至っている。

以下、システムが想定する履修選択と教育のモデル、およびMIMAサーチの詳細について述べ、MIMAサーチにおけるカリキュラムの分析、構造化に関し、例を交えて報告するとともに、実用化に際する工夫点や、実用化後に行った有用性に関する評価について論述する。

## 2. 大学教育と履修選択のモデル

高度化・学際化・多様化する科学・技術分野で、国際的なリーダーシップを取れる人材の育成、ならびに、幅広い教養教育と、高度の専門教育の結合、つまり「広く」、「深く」を実現する工学教育の実装が喫緊に望まれることは先に述べた。また、工学が特に重視する「社会とのつながり」という観点においても、一般的に社会が人材に求める能力(理解力、判断力、記憶力、観察力、洞察力、想像力、独創性、発想力、責任感、協調性、順応性、主導性)をいかに教育全体で充足するかが課題となる。ここでの教育の特徴は、分野にとらわれず、より広く、より深く、知識やコンピテンシーにかかる能力を習得させるか、に重点を置いていることにある。つまり、中央教育審議会等でも議論される「教育方法と教育内容とのバランス」、「教科教育と教科横断的教育のバランス」等の考え方を含めた「Knowledge, Skills, Character」のバランスのとれた習得を目指すものである[1]。

したがって、カリキュラムの構造化においては、1) 課題に関連する授業の選択、2) それらの(時系列上の)関連性の同定、3) 専門性、コンピテンシーや単位数、時限等の様々な視点による可視化、が重要であり、これらをいかに充足したシステム設計を行うかが課題となる。

一方、医学教育等、資格教育的側面の強いカリキュラムや、その時間的な制約等も含め、カリキュラムにはほぼ選択の余地がない大学や学部が存在することも確かである。しかしながら、分野等による意識の差はあるが、現代社会においては、医工連携のような学際領域、分野横断領域に対する人材育成が必要との認識は共有されているのが通常であり、必ずしも、本論文で議論するような履修支援の余地がないとはいえない。ただし、そのような学部や分野を横断した講義の関連付けや、履修の(厳密な)順序付けを行うためには、本論文での関連性計算のみでは不十分なものは明らかであり、Causal-Effectのような時間軸を含めた関連性推定を行う仕組みも必要となる。生命科学分野での発生のメカニズムの分析等の目的で自然言語処理を利用したCausal-Effect分析の研究が行われてはいるが、このような高度な関連性の推定を高精度に行うことは、現在の技術レベルでは実用的とはいえないため、本論文では、これらは議論の対象外とする。実際のシステムにおいては、シラバスに付与される履修可能な学年の情報や、中教審等でも議論されている授業の体系化(ナンバリング)におけるレベ

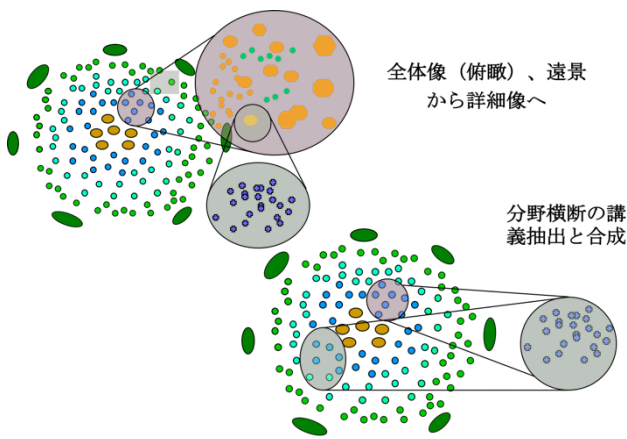


図 1 履修選択操作のイメージ  
Fig. 1 Course selection image.

ルの情報に基づき、そのような履修の順序が考慮されることを期待している。

また、履修選択支援としては、協調フィルタリング等の技術により、授業のリコmendを行うシステム [2], [3] も種々、提案され、すでに実用化されているものもあり、他の学生の履修履歴等を利用した授業リコmendを行うことが可能となっている。しかしながら、実際にこのようなシステムを導入した大学では、リコmendされた科目に過度に学生が集中し、全体の授業の運営に支障を来したため、システムの運用を取り止めた事例も存在する等、コースリコmendシステムの教育現場での実運用に関しては、いまだ議論が必要かと思われる。つまり、このようなシステムの教育現場での実用化に関しては、教育理念や教育現場に特有の問題が存在し、他分野において有用な技術が必ずしも教育現場においては有用とは限らない。他方、本論文で議論する履修選択支援の方針としては、いかに履修者本人の興味や関心を優先できる選択支援を行うか、さらには、いかに専門性、コンピテンシー、レベルをバランス良く、習得させるか、により重点を置くものである。よって、これらの点を考慮した履修選択支援のモデル、およびシステムの構築を行うことを目標とする。具体的には、複数の視点による全体像の把握から、ズームイン、ズームアウトのような操作により、ピンポイントに講義を選択し独自のカリキュラムを構築する一連の操作が可能なシステムの構築を目指すこととした (図 1)。より具体的には、「情報学」であれば、大きく、「情報基盤」、「情報社会」、「情報文化」、「情報機械」等の下位構造に分類されること、およびそれぞれの分類が相互に関連していること、さらに、それらが「ロボット」や「検索エンジン」、「メディアアート」のような詳細な科目により構成されていること等の構造が考えられるが (図 2)、履修科目の選択においては、このような構造が様々な視点から適切に可視化され、上位の分野情報から詳細な科目への選択操作を繰り返す (図 1 の「全体像から詳細像へ」の操作) ことで、ユーザが独自に履修計画を

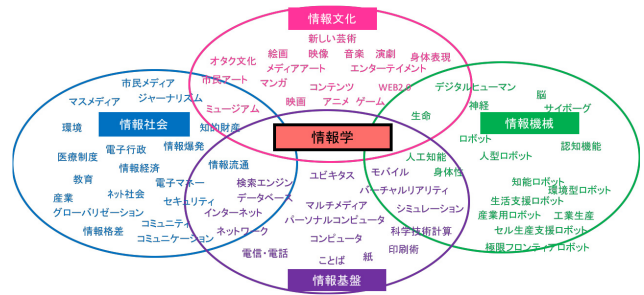


図 2 「情報学」カリキュラム全体像の例  
Fig. 2 Example of curriculum "Information Science".

立てられることが期待される。

### 3. システム構成と実装

図 3 に本システムの構成を示す。

**自動用語抽出エンジン (TermEngine)**：膨大なテキスト情報に対し、概念の包含性を基に用語としての妥当性を判断する C/NC-value 手法を行い、専門用語を自動抽出するモジュールである。C/NC-value 手法の特徴は分野依存性が低く、様々な分野のテキストから高精度に専門用語の抽出が可能となることにある。本システムは日英両言語のテキストの解析に辞書の切り替えなしに対応できる。C/NC-value 手法の詳細に関しては文献 [4] を参照されたし。

**知識類似度計算エンジン (SimEngine)**：TermEngine により認識された専門用語とその頻度の情報を用いて、科目間の関連性を定量的に計算するためのモジュールである。本エンジンは、高度な拡張性を有し、計算手法を独自に定義、プラグインすることで容易に拡張が行える機構を備えている。これまで、関連度の計算に VSM (Vector Space Model) におけるコサイン内積 [5] や平均相互情報量を利用した階層化クラスタリングや、クラスタ間の関連度計算に WARD 法、重心法、群平均法、最短 (長) 距離法等の一般的手法、また、クラスタの認識に Bron-Kerbosch による最大クリーク探索を用いるもの等、様々な手法を実装し、実験評価したが、現状では、計算速度と精度とのバランスをとり、階層化クラスタリング、および関連度の計算に VSM (Vector Space Model) におけるコサイン内積を利用したものを採用している。なお、関連度計算に利用する特徴量に関する詳細は次節で述べる。

**関連の可視化エンジン (VizEngine)**：SimEngine により得られた科目間の関連性を基に、カリキュラムの全体像を可視化し、俯瞰可能とするモジュールである。本可視化手法では、グラフ理論を基に、各科目をノード (点) に割り当て、それらの関連性が大きいほど、ノード間の物理的距離が短くなるよう全体の最適配置によるグラフ描画を行う機構を備える。また同時にノード間のリンク (線) を太くすることで、科目間の関連性が直感的に把握できることを

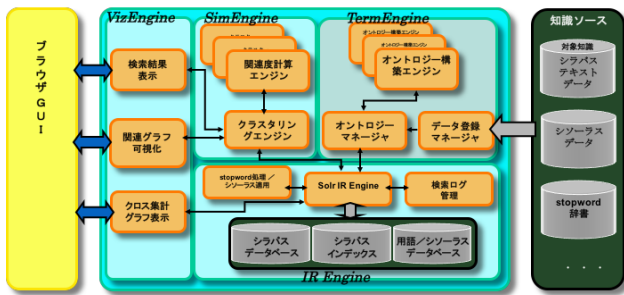


図 3 システム構成  
Fig. 3 System diagram.

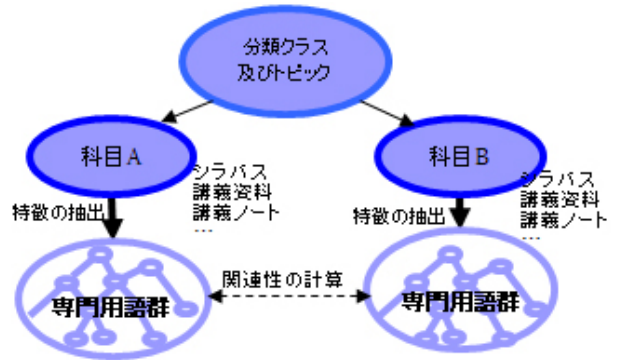


図 4 カリキュラム構造化のモデル  
Fig. 4 A model of curriculum structuring.

特徴とする。

さらに、本システムの大きな特徴として、階層化クラスタリング分析により、階層化された科目分類の各階層を可視化できることがあげられる。ユーザより指定されるクラスターの階層の情報により、その階層でのグループを抽出し、グループに属するノードをさらに大きな円で囲うように可視化を行う（以下、トピックマップと呼ぶ）。また各グループを特徴づける専門用語を抽出し、各グループの代表ラベルとすることで、より直感的に全体の構造を理解しやすい可視化を行っている。

また、現在のシステムでの標準検索エンジン（*IR Engine*）としては、Solrを採用している。

なお、可視化の詳細については、文献 [7] を参照されたし。

### 3.1 カリキュラム全体像の把握とテキストマイニング

上記のように、カリキュラム全体像の把握を可能とする仕組みとして、本システムでは階層化クラスタリングを採用している。これは、野澤らや、井田らによっても示されているが [7], [8], カリキュラムの全体像を把握するという目的において、最適な分析手法の1つと考えられるからである。一般に、階層化クラスタリングを始めとするデータ分析手法によるテキストマイニングでは、蓄積されたテキストに対し、1) 特徴の抽出, 2) 分析, 3) 可視化, により、テキスト全体の傾向を示し、そこからある種のパターンを探し出すことなどといえる。新聞等のテキストを対象に、形態素解析と形態素の頻度統計分析（時系列分析）を利用したテキストマイニングは、文書分類やトレンド分析等の目的で多用されている。また、Webマイニングや論文のマイニングでは、さらにハイパーリンク、言及関係等の静的な関連情報を利用してクラスタリング等と同様の分析を行うことが可能であり、情報検索におけるランキングや、ネットワーク分析における中心性の分析等に利用される。このように、クラスタリングにおいては、主に、①分類対象の内容（テキスト）からの特徴抽出を利用した関連性の分析、②あらかじめ人により与えられた関連性による分析、の二通りが大きく考えられ、もちろん両者を融合した方法も可能である。カリキュラムの分析においても同様のことが考

えられ、たとえば、②として、シラバス入力時に担当教官より、事前履修や事後履修の科目を指定させる等の対応であろう。しかしながら、先にも述べたように、学問の過度な細分化や、授業科目の毎年の再編により、関連性の抽出を人手に過度に依存するのは現実的ではない。さらには、学科や学部を越えた関連性に至っては、ほぼ期待できない。よって、本システムでは、①を基本とした関連性の抽出による分析を行っている（図 4）。ただし、学科や学部等の分野を越えた関連性の抽出においては、「環境」や「デザイン」のような、分野により意味の異なる単語（形態素）が多く存在するため、単語（形態素）のみによる特徴の抽出では不十分である。よって、太田ら [10] と同様、関連性の計算のための特徴量として、より長単位の（専門）用語を併用することとし、システムとしては、このための（専門）用語抽出技術（*TermEngine*）をオフラインで適用し、データベースに格納している。さらに、一般に分類システムでは、分類の根拠となった用語を含めた形でユーザへ情報提示することが重要であるとの報告 [11] 等、科目間の関連を明確にすることに対する要望が多くある。よって、これへの解決策として、階層化クラスタリングにおける各クラスターへのトピック（各クラスターを特徴付ける概念）を付与している。一般に、階層化クラスタリングのみでは、このような処理は不可能であるが、本システムでは、クラスター内のトピック分析として、用語頻度と文書頻度\*1の積に基づいたスコア付けにより、これらのトピックを決定し、可視化の際に活用している。つまり、より重要度が高く（用語頻度が高く）、また一般性の高い（より多くの文書に含まれる）傾向を示すトピックが選ばれるようスコア付けを設定している。なお、上記の関連性計算と同様に、分野を越えた意味の曖昧性を解消する目的により、ここでも用語として *TermEngine* の結果を利用している。

### 3.2 可視化と履修選択操作—全体像から詳細像へ

データマイニングやテキストマイニングにおいては、一

\*1 対象の用語が含まれる文書数

般の折れ線グラフや棒グラフを利用した統計結果、分析結果の提示を行うのも可視化の一部と考えるが、ある対象間の関連性が重要となる場合、ノード（点、Vertex）とそれらのリンク（線、edge）によるグラフ表現（グラフモデル）が多用される。グラフモデルを利用した可視化においては、さらに、

- 関連性をどのように定義するか
- 全体の配置をどのように行うか

が重要となると考えられる。当然、上記の可視化においては、表示装置、計算量の両面から対象の大きさが問題となる。また、分析の軸や表示の情報量を増やす目的で、3次元（3D）化による可視化も多用されるが、この場合も表示装置やマウス等を始めとするユーザインタフェース自体は、通常2次元の情報しか扱えないため、

- 視点の位置や視点の変更の方法
- ピンポイントに対象を選択する方法
- 上記に対するユーザインタフェース・デバイス

が問題となる。

つまり、対象全体の傾向等を読み取るマイニングとは別に、さらに特定の文書（本システムの場合はシラバス）を効率的に探し出す目的でマイニングにおける分析結果や可視化を用いることを考える場合は、可視化したグラフ表現の直感的な理解と、選択対象へのマニピュレーション（たとえば科目の選択操作）に対する直感的な反応が重要であると思われる。よって本システムでは、これらの課題に対し、以下の仕組みを実装している。

#### 1) バネモデル [12] の利用によるノード配置の最適化

ノード間（科目間）のリンク、および関連性の大きさ、それぞれバネ、およびその強さと見なし、物理モデルにより、ノードの最適な位置を計算する。これにより、科目間の関連性が大きければ大きいほど、可視化の画面上でも近くに配置されるため、全体像に対する、より直感的な理解が期待される。また、前述のように、科目間の関連性を明確にするため、クラスタごとに科目群のまとまりを明示し、各クラスタの特徴を示す用語をその科目群のトピックとして表示することとした。さらには、他のクラスタと同じトピックが選択される場合には、スコア順位の低いトピック候補を追記し、クラスタ間のトピックの曖昧性を解消することとした。たとえば、「数学」のキーワードで検索した場合、学部等によっては複数のクラスタで「常微分方程式」のような基礎的なトピックが選択されるが、それぞれ次の候補の用語を追加することで、「常微分方程式、ベクトル解析」、「常微分方程式、線形代数」のようなトピックとなり、クラスタ間の違いが明確になる。

#### 2) 複数の2D表現の切り替えによる操作性の向上

関連するシラバスによるカリキュラム全体像の表現において、2章で議論したような、授業内容のみの関連を縦軸に、受講の学年等の時系列上の関連を含む関連を横軸に、

コンピテンシー等の講義の種類を含む関連を奥行き軸に割り当てる等の3D表現が、履修モデルの表現としては適していると思われる。現に、初期のカリキュラム構造化システムでは、このような表現により可視化を行ったバージョンも存在する。しかしながら、全体像から特定科目への選択の操作においては、いまだ、2Dディスプレイ上におけるマウスやタッチスクリーンによる選択が通常であるため、学生利用者より、視点を回転させながら科目を選ぶ操作には、操作性に問題があるとの評価が多くあった。そのため、複数の軸を切り替え可能な2D表示による可視化を採用することとした。これにより、通常のポインティングデバイスによる直感的かつ分かりやすい操作が可能となった。また、カリキュラム全体像から詳細像への操作を行うため、以下の機能群の設計と実装を行った。

#### a) 任意のクラスタ階層指定による全体像の可視化

階層化クラスタリングを行った結果に対し、利用者によるどの階層のクラスタを可視化するかを任意に指定できるようにした。どの科目群が同じクラスタに属するかが、利用者により指定された階層に応じて即時的に計算され可視化に反映されることとした。これにより、大分類、中分類、小分類といった任意の段階での階層が操作に応じて即時的に把握可能となる。

#### b) 関連する講義の選択と再検索

関心のある任意のクラスタや科目を選択することで、選択されたシラバスに含まれる用語の情報を基にレリバンス・フィードバック検索を行うことを可能とした。これにより、より利用者に関心の高い科目群が検索結果として提示されることが期待される。

#### c) クラスタの選択による絞り込み操作

関心のある任意のクラスタを選択し、ドリルダウン（絞り込み）操作を行うことで、検索結果をさらにクラスタ内の科目群のみに絞り込む操作を可能とした。

#### d) 上記のUNDO/REDO

上記、すべての操作に対し、「元に戻る」もしくは「繰り返す」操作を可能とした。

図5では、「数学」をキーワードにした検索、可視化の例を示す。ここでは、「数学」という言葉に関連するシラバスを選び出すだけでなく、そのシラバス群が、他にどんな用語を使っているかも解析し、重要な用語の重なりを基に科目間の関連性の度合いが計算され、さらに、この関連性の度合いを基準に、全体の構造が可視化されている。共通科目から各学科（標準カリキュラム）への講義のつながりや、さらには大学院で学ぶことへのつながり等が明確に示され、目標設定や、目標に対する講義の組み立て等を効率的に行うことができると期待される。また、図6では、縦軸に学年の情報、横軸に分野（学部）の情報を配した視点による科目の分類の可視化例を示す。MIMAサーチでは、このような仕組みにより、様々な視点によりカリキュラム

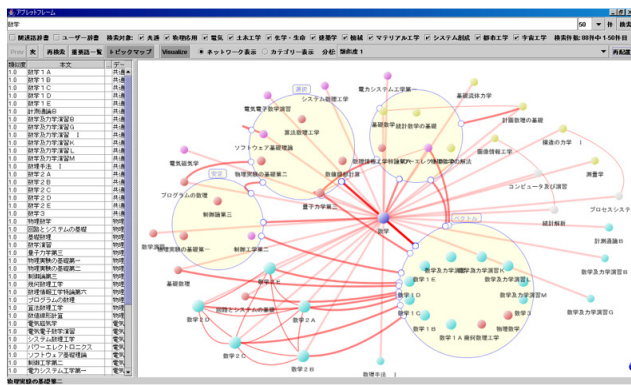


図 5 「数学」に対するカリキュラムの構造化

Fig. 5 Curriculum visualization of “mathematics”.

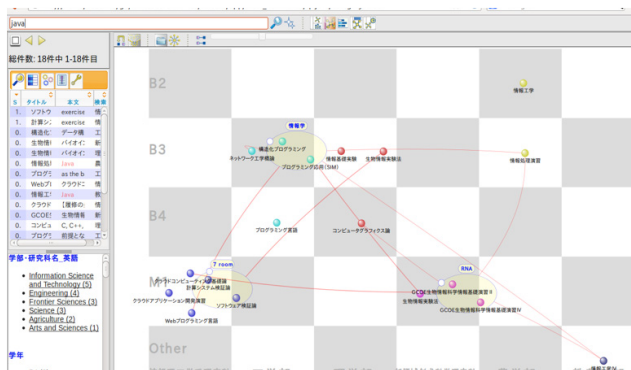


図 6 横断的構造化による講義履修順序の組み立て

Fig. 6 Curriculum visualization sorted by course level.

の構造化、およびカリキュラム間の比較等が可視化可能となっている。

### 3.3 講義の抽出と合成

3.2 節による可視化と操作の組合せにより、全体像から任意の詳細像への絞り込みの操作を、試行錯誤を交えて行うことが可能となる。システムには利用者ごとに任意の科目群を記憶可能な My リスト機能が用意されており、利用者は履修したい科目を My リストに登録することで自分独自のカリキュラムを合成することが期待される。

しかしながら、このような独自のカリキュラムの合成を行うことが可能な一方で、履修選択に特有の状況として、いくら授業内容に関心があったとしても、時間割の制約により、同学期（同ターム）には履修できないということがある。つまり、通常、ある学科等における固定のカリキュラムを履修するのみであれば、時間割が重ならないようにあらかじめ調整されるため、このような状況は考えられないが、学科や学部を越えて履修選択を行う場合は、同学期（ターム）の同時間割に開講される科目を選択する可能性が出てくる。このような状況を選択時に回避し、関連する他の科目で充当する等の操作を行うことを支援する目的で、My リストには、複数の学期（ターム）にわたる時間割への自動マッピングが可能な機能を実装している。

他方、どうしても上記のような科目のコンフリクトの解消が困難な場合も十分考えられる。単位互換制度の活用による大学を越えた履修選択支援や、オープン・エデュケーション教材との併用学習に対する学修支援システムに関して、増田らにより研究が進められている [13]。

## 4. システムの実装と実用化に関する考察

本システムは、先にも述べたように、東京大学工学部シラバス検索、さらには全学授業カタログ、および東京大学 OCW (UTokyo Open Course Ware) でのオープン・オンラインコース検索において、授業の検索や、他大学の授業を含む他の授業との関連性を俯瞰するためのシステムとして実用化され、すでに 10 年以上にわたり実稼働を続けている。また、芝浦工業大学や、新潟大学、大阪大学、東北大学等の複数の大学でも採用され（一部は学内アクセスのみ）、学生等の実際の利用者により高い評価を受けている。現に、MIMA サーチは、東京大学の中期目標の達成状況報告書等でも新たな教育の取り組みに関する成果の 1 つとしてあげられている\*2。また、特許情報の検索や知識発見、企業におけるユーザーサポート事例の構造化等でもシステムの実用化がなされ、すでに多くの企業で利用されている。さらには、日本学術会議での材料・化学分野におけるリサーチマップ作成提案に利用される等、その高い実用性と有用性が示されるところであり、今後の発展により、世界中のオープンエデュケーション・コンテンツを関連付ける等の広範囲な普及が期待される。

一方で、このようなシステムを実用化し、運用維持するためには、安定性、システムバランス、メンテナンス、不具合対応、セキュリティ対応、広報、多言語対応等、様々な困難を解決しながら進める必要があるのも確かである。これらのすべてにわたり詳述するのは本論文のテーマの範囲外であるため、ここでは、シラバスを自然言語処理の処理対象とし、総合大学における一サービスとして実用化する場合に特に問題となることとして、以下 4.1 節、および 4.2 節に述べる。

### 4.1 理系と文系のシラバスの違い

前述のように、分野横断が重要視されているなかにおいて、すでに学問を理系と文系に大きく分けることはナンセンスではあるが、やはり学問に対する考え方の違いがあるのは確かである。たとえば、理工学系のシラバスは比較的、専門用語を中心に記述されることが多く、また書籍の目次のように、講義の 1 回ごとにテーマが決まることが多いため、一般の用語抽出による特徴量の抽出が精度良く動作する対象だといえる。これに対して、文系、特に人文学系のシラバスには、「プラトンの著『ポリテイア (国家)』を第

\*2 <http://www.u-tokyo.ac.jp/content/400004374.pdf> (2019 年 6 月 10 日アクセス確認)

1 巻冒頭から読むことで基礎訓練を積み、古代ギリシア哲学を実際に学んでいく」, 「近世哲学の古典中の古典, Francis Bacon の Novum Organum (The New Organon) を読解していく」といった, 著書のリーディングにより, 概念そのものを学ぶというよりも, いわば学習の姿勢を学ぶという感が強く, これが一連の講義全体に流れるテーマとなる。よって, 単純に用語を抽出して特徴量とするのみでは, 学習内容に至る関連付けという観点では不十分であり, 改善の余地が残される。このような状況に対して, 以下の対応により改善を進め, 実装しているが, いまだ十分とはいえない状況である。

#### ●用語認識の拡張とソーラスの利用

用語認識を人名パターン(未知語を含む)や, 日本語文書内での英語表記等にも対応することで, 必要な情報の抽出精度(主にリコール)を上げるとともに, シラバス間の関連度計算の際にソーラス(同義語, 上位下位語, 関連語の辞書)を利用できるよう, SimEngine を拡張している。しかしながら, たとえば, ゲーテには, ゴエテ, ギューテ, 等の45通りの表記があり, 学派によって標準形が異なるといわれるように, 単純にソーラスにより同様の言葉として扱うことには問題があるとの意見もある。このような指摘にも対応可能なよう, ソーラスの関連度計算への適用には, 学部等のまとまり(まとまりは自由に指定可能)ごとに, その内容を細かく個別に設定が可能な仕組みとしている(現状で, 学派までを別のまとまりとはしていないが)。加えて, 用語抽出のリコールを上げることで副作用的にノイズ(不要語)も多く見られるようになるため, stopword リストも同様に上記まとまりごとに個別に設定可能としている。

#### ●シラバスへのキーワード付与による特徴付け

上記のみでは不十分な場合も考えられ, また, 比較的シラバスに用語の記述量が少ないことを補填する目的により, 最大40個までのキーワードをシラバス入力時に設定できるように, シラバスシステム自体を拡張し, 関連性の計算時には自動用語認識の結果と併用できるように改良している。しかしながら, 「そもそも人文学系の学問はキーワードにより特徴付けられるものではない」等の主張もあり, 上記ソーラスの議論も含めて, やはり, 関連性の計算には改善の余地が多々ある。自然言語処理の向上という観点からも, 記述された内容の深い意味的理解や, 既存のカリキュラム構造からの関連の教師あり機械学習等, 今後の研究に期待したい。

## 4.2 検索ランキング, 関連度の問題

むしろ, これは検索における問題であるが, 「キーワードを指定しているにもかかわらず, なぜ結果に出て来ないのか」であるとか, 「自分の授業はAという授業と関連しているが, なぜBとは関連していないのか」という意見はよ

く聞く(7章で述べるアンケート評価においても多くの記述がある)。主に, シラバスの記述量や内容に起因する問題ではあるが, 検索ランキングや, ランキング上位リストの作成方法(一般に一度に出す検索結果件数に制限があるため), 関連性の計算の仕方等に課題があるといえる。たとえば, 史学の科目における「ヨーロッパの歴史」と数学の一部である「数学の歴史」において, 学際性という観点からは, “歴史”というキーワードで検索結果に同等に出ることがふさわしいとする考えもあろうが, 単純に言葉の頻度に比例して検索ランキングが上がる仕組みにおいては, シラバス中の“歴史”という語の頻度は史学のほうが多いであろうから, 検索結果はほぼ史学の科目で埋め尽くされてしまうことになる。このような, 言葉の頻度のみが検索ランキングに反映される問題を緩和するため, 現状では, 同じ検索スコアのものであれば(検索エンジンへの登録順ではなく), なるべく多様な学部から選択して検索結果上位リストを作成する等の対応を行っているが, 根本的な解決とはいえない。検索結果の多様性向上[6]とともに, 情報検索の一般的な議論に近いが, 分野横断型教育の推進(文系と理系のバランス)や, 教育における可能性, 気づきの促進の観点から, やはりシラバスの検索にどのようなランキングがふさわしいのかは, 今後の大きな研究課題であろう。

これらは, 一般的な自然言語処理における技術的な課題の解決とはいえないが, 著者の経験では, 自然言語処理技術の実用化には, 例外処理を丁寧に行うこと, およびユーザーインタフェースをいかに(見た目や操作性, コンセプトの観点で)バランス良く仕上げるかに多くのリソースが割かれる。実際に, 著者の関わった自然言語処理応用製品では, 理論的な自然言語処理はプログラムの1割程度であり, 残りの9割は例外処理であった。より多様なユーザに, バランスよい精度やユーザビリティを提供するには, やはりそのような(必ずしも理論的ではない)技術の積み上げやノウハウの蓄積が必要なのは, 統計的な言語処理が主流の現在においても同様であろう。これらの観点からも, 当該分野の発展には, やはり, 「教育」や「大学」, 「人材育成」等に関する広い知識やノウハウの収集と, それらの実装が必要であろう。

## 5. 実験と評価

本章では, 主に MIMA サーチのシラバスの構造化(シラバス NAVI)への適用とその評価に関して述べる。なお, MIMA サーチ・システムの技術的評価において, 用語抽出の精度, に関しては文献[7]に, 用語, および形態素によるシラバスの分類に関しては文献[10]にあるため, ここでは割愛する。

以下, 5.1 節では, MIMA サーチ・シラバス構造化システムへの年ごと, および月ごとのアクセス統計により, 議論を進める。また, 以後の節では, 本システムのユーザビ

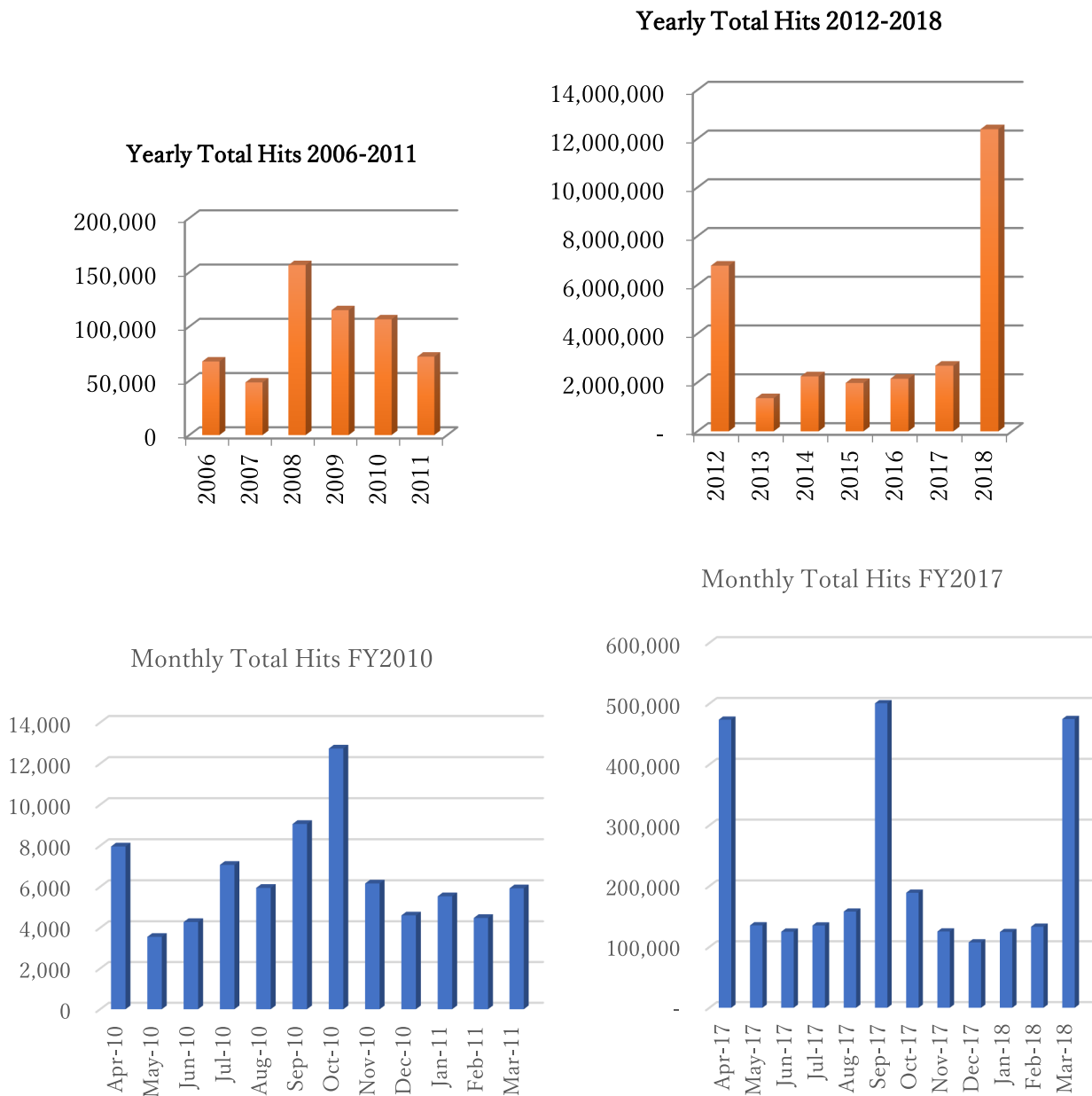


図 7 MIMA 検索への年間、および月間アクセス統計  
 Fig. 7 Access statistics of yearly/monthly total hits.

ティを評価する目的で、フリーアンケート形式による評価を行った結果を示す。アンケート調査では、より実情に即した評価を行うため、エンドユーザ、および情報技術に関する知識をある程度持つユーザの両者に対する調査を行った。5.2節では、エンドユーザへの評価に対し、5.3節では、情報技術に関する知識をある程度持つユーザに対し、それぞれ記述する。

### 5.1 シラバス構造化への適用と運用状況

まず、図 7 (Yearly Total Hits 2006-2011, および 2012-2018) に、2006 年 4 月以来 13 年間のシステムへのアクセス統計を示す。なお、1 章でも述べたが、2011 年までと 2012 年以降では、登録科目数や対象ユーザが異なり、アクセス数のスケールが大幅に異なるため、別グラフとして提

示してある。2006 年から 2011 年の稼働においては、2007 年の公開翌年にいったん、アクセス数が落ち込むが、2008 年に東京大学工学部トップページからのリンクが張られたこともあり、4 倍程度のアクセス数の伸びを記録した、その後の 5 年間は、おおむね毎年 10 万~12 万アクセス、月平均 1 万アクセス程度で推移した。2012 年には全学部 (医学部除く学部後期課程、および大学院) のシラバス約 1 万科目分が追加され、東京大学のトップページよりリンクが張られたため、アクセス数は約 100 倍に増加した。さらに 2018 年には学部前期課程約 2 千科目が追加されるとともに、東京大学教養学部トップページからのリンクが張られたため、大幅にアクセス数が増加している。実用システムにおいては、特に利用者数の上限がほぼ決まっている大学を対象としたシステムということもあり、学事に必須のシ



システムでもない限り、広報等の活動を行わないとアクセス数は年々、減る傾向にあることは、UTokyo-OCWの講義ビデオ・コンテンツ配信等での経験からも分かっている。よって、パンフレット配布や、ガイダンスでのアナウンス等による継続的な広報活動等も非常に重要となる。また、図7 (Monthly Total Hits FY2010, および FY2017) には、1年間 (2010年4月~2011年3月, および 2017年4月~2018年3月) の月別統計を示す。月によりアクセスのばらつきはあるが、FY2010では月3千から1万2千程度、FY2017では月10万から50万程度のアクセス数があることが読み取れる。

特に、毎年、履修選択の時期である4月、および10月 (2016年のクォータ制移行後は9月) にアクセスが伸びる傾向があり、また、アクセス元のほとんどは日本国内からのものであることより、実際に学生が科目の履修選択に活用していることがうかがえる。

## 5.2 エンドユーザへのアンケート調査 (理系教養学部1年生)

本節では、エンドユーザからの評価として、大学1年生 (理系教養学部生) 130名 (自由記述による回答が得られた学生は121名) に対し行ったアンケート調査に関し詳述する。調査では、実際にMIMAサーチによる検索を行ってもらい、「MIMAサーチを使用した感想、および改善すべき点」に関して記述を行ってもらった。アンケートの実施時期は2011年5月である。全体としては、319の評価に関する記述 (1人の人が複数の記述をしている場合も別々の記述としてカウントし、1つの記述が複数の文より構成される場合も1記述としてカウント) が得られたが、その内訳としては、ポジティブな評価に関する151記述、および改善点に関する評価168記述となった。表1に、それぞれの記述に対して、内容を基に分類した結果を示す。

表より、おおむね、関連性の計算による構造化を行い可視化することで、カリキュラムの把握や履修選択の効率化に関する効果があることが見てとれる。なお、上記の151記述は、7割以上の学生から得られたもの (残り3割の学生からは改善点に関する感想のみ) の集計であり、この結果からもMIMAサーチの有用性に関する評価が高いことがうかがえる。

それぞれの内容に関して、具体的には、以下のような記述が得られた。なお、以下での「リスト表示」とは、検索結果を (ランキングに応じて) 羅列表示することを示している。

①に関しては、

「自らの志す1つのテーマに向けて、どのような形で講義の選択を行えばよいかが目に分かる」、「新たな興味の対象を発見できるとともに、その項目を包含する項目たちの全体像が眺められるという点でよいと思う」、「いままで

表1 ポジティブな記述に関する分類  
Table 1 Classification of positive response.

ポジティブな評価の内容	記述数
①構造化のメリット (講義体系の把握の容易さ等) に関するもの	45
②可視化表現の分かりやすさに関するもの	41
③履修選択の効率向上に関するもの	22
④ユーザインタフェースの使いやすさに関するもの	20
⑤その他 (漠然とした感想なども含む)	23
計	151

のようにただシラバスの一覧表が出てくるのでは、見るのが面倒なうえ、それがいったい何についての講義なのか、どのように他の講義と関連しているかがあまり分からず講義名で想像するしかなかったが、MIMAサーチではそういったことが点と線のつながりによって可視化されることで、一目で分かるようになっていて、それでシラバスが一発で出せて、大変便利だと思います」、「シラバス間の関係を構造的に見ることができるのは、履修登録の際だけではなく、専門課程で何を学んでいくか思案する際に非常に役立つと思われる」、「今までは関連があると思っていた講義どうしがあるキーワードによって結び付けられる」のような記述があった。

②に関しては、

「自分の興味のある分野についてどの講義が適切であるのかというのが見てぱっと判断できたので、あると非常に便利だと感じた」、「情報を球にモデル化しその球の距離によって関連性を表現していることにより直感的に関連性の高低が分かり、非常に容易に情報を得ることができる」、「表示方法が選択でき、トピックマップも用意されているため、さまざまな視点から検索結果をみることができるといふことにも感心」のような記述があった。

③に関しては、

「自分の履修したい授業をより早く、確実に見つけることができる」、「一気に多数の講義が調べられるので、見たい情報がすぐにピックアップしやすい」、「履修分野を考える際において、とても便利」のような記述があった。

④に関しては、

「重要語句一覧があり、どのようなキーワードで検索すればよいのか分からない人にも優しく作られていることに感心」、「直感的に使用できる分かりやすい検索システムであり、使い方の説明を読まずとも使用することが可能であった」のような記述があった。

⑤として、

「明確なビジョンを持った学生にとっては利用しやすいシステム」、「関連用語についてなら講義がすぐに表示されるのがおもしろくて様々な言葉をいれて検索しようという気にもなった」のような記述があった。

表 2 改善点に関する記述の分類

Table 2 Classification of responses regarding improvement.

評価(改善点)の内容	記述数
グラフ表現の改善に関するもの	67
他学部や他の教材等(他大学のシラバスを含む)との関連性を知りたい	23
関連の不明瞭さに関するもの	11
学務システムとの連動に関するもの(授業変更, 講の反映等)	13
キーワード検索の不備に関するもの	10
操作性に関するもの	5
動作速度に関するもの	1
その他	38
計	168

これら全体の記述内容より、(従来のリスト表示のみの検索システムと比較して)カリキュラム全体像の把握が可能なことや、その重要性に関する認識、講義間の関連や新たな興味の発見、履修選択への貢献、ユーザインタフェースとしての使いやすさ、等が読み取れ、本システムが目指す、高等教育における履修選択のあり方において、主観的ではあるが、その支援に一定の有効性があると判断できる。

一方で、表 2 に、改善点に関する記述を内容によって分類した結果を示す。

改善点に関しては、やはりグラフ表現の改善を望む声が多くを占めることが分かる。これは、たとえば、「講義どうしがどのように関連があるかということが分かるようにしたほうがよい」、「点がたくさん出てきたときに、線があまりにも多くなってしまっ点同士の関連がよく分からなくなる(点を動かせるけど、それでも分かりづらい)ので、選択した点を結ぶ線の色を変えるか、検索キーワードの点を消せるようにして欲しい」のような記述による。

また、たとえば、「大学内における講義の関連性だけでなく社会でどのように役立つか、その講義で得た知識を将来的にどのように発展応用させることができるのかなども表現できるとさらによくなると思う」、「学部の境界線を越えて講義の関連を示せるようになれば、学部選択の際に、自分の興味の持つ分野に強い学部はどこであるかが一目瞭然」といった、他学部や他の教材と関連性の明確化までも求めるものが意外と多く、講義間のつながりを明確にすることへの学生の要望が見てとれる。

なお、上記の「グラフ表現の改善」に関しては、その内訳(内容による分類)を表 3 に示す。

やはり、表示領域の限界に関する内容が多いと思われる。以上を総じて、i) 可視化方式の改善, ii) 関連性計算, 検索基準の明確化, iii) 他学部(他大学)のシラバスや、他の教材との連動、が今後の急務と思われる。I) に関しては、指摘内容にもあるとおり、3D 化を含むグラフ表現方式の要

表 3 グラフ表現の改善点に関する内訳(1)

Table 3 Breakdown of responses regarding graph-expression improvement (1).

グラフ表現の改善に関する内容の内訳	記述数
グラフの表示領域の限界(文字の重なり等)に関するもの	18
3次元化への要望に関するもの	11
ノードの色分けの不明瞭さに関するもの	11
カテゴリ表示の改善に関するもの	6
リスト表示の必要性に関するもの(グラフ表示領域を広げたい)	1
その他	20
計	67

望も比較的多く見受けられるが、これは前述のとおり、3D 表現に適したポインティングデバイスの普及等とともに検討を開始する必要がある。また、グラフ表現以外の可視化方式の併用による構造化能力の向上の方向性も探る必要がある。

ii) に関しては、すでにマニュアル等で基準を明確に示すことでの対応を行った。これらの改善とともに、グラフ内に関連性スコアや種類の情報を挿入する等の可視化の際のユーザインタフェースの高度化による改善策も検討したい。iii) に関しては、すでに全学(学部後期課程のみ)授業カタログにより、ほぼ実現できている。また、他大学として、MIT の OCW との連動に関してはすでに東京大学 OCW における MIMA サーチで実現し、実稼働を行った経緯があるが、この評価等を元にさらに他大学シラバスとの関連の可視化等についても研究開発を進めているところである。なお、より具体的な感想内容として、

「部分的にズームアップして見る機能があったら便利」  
→ ズームの追加やドリルダウン処理の追加

「分野ごとに分けて表示してしまう方が分かりやすいと思う」  
→ 分野(学部)ごとや学年ごとの軸によるマトリクス表示画面を追加

「どの講義がどの講義を土台にしているのかというような講義の流れが分かる樹形図表現があると分かりやすいと思う」  
→ 同上

「放射線状に展開された多くの講義のうちどれをすでに検索してどれをまだしていないのか、ということが多少分かりにくかった気がしたので、自分が検索したなかで特に興味を持った講義を MIMA サーチ内の別の場所に登録しておけるようにする、などすればさらに使い勝手が良くなる」  
→ My リストにより実現

「完全自動処理でなく教官(各学科の主任だけでも)が講義間の関連性の強さなどをいじれたほうがよい」  
→ (権限) ユーザからのタグ付与や関連の付与が行えるシステムを検討

等、基本的にはアンケートによって得られた改善案に対して、複数の対応案を MIMA サーチ開発チームで議論し、それを優先に実装するかを、費用対効果の観点も含めて決定、といったルーチンをつねに回しているのが実情である。今後も、改善点等に関して、さらにユーザ調査を進め、その効果や全体とのバランス等の評価を基に改良を繰り返す予定である。

他方、注目すべきは、履修選択用の情報として、「授業の曜日時限や単位数の情報が欲しい」や「授業の評価データがあれば選択に役立つ」といった、シラバスの内容以外のデータで履修選択の支援をしたいとする記述数が 3 (3名/121名中) であったことである。これは、後節の、専門課程に進んだ学生に対する評価との比較において、興味深い内容となる。

### 5.3 情報技術に関する知識をある程度持つユーザへのモニタ調査 (工学系学生)

本節では、検索等の情報機器操作や科目、専門分野により高度な知識を持つと思われる工学系の学生 8 人 (うち、大学院生 2 人) に対し、モニタ評価を行った結果について詳述する。本モニタ調査の実施時期は 2008 年 2 月である。対象者のうち 1 人のみ事前に MIMA サーチを使ったことがあるモニタがいたが、残りの対象者は今回の調査で初めて使ったモニタであった。調査では、実際に一定期間 MIMA サーチを利用してもらい、その後、5.2 節と同様、「MIMA サーチを使用した感想、および改善すべき点」に関して記述を行ってもらった。結果として、計 94 の記述 (先と同様、1 つの記述内容が複数の文から構成されるものも含む) が抽出された。

なお、以下での ( $N = n$ ) の表現では、 $n$  個の記述があったことを示しているが、必ずしも  $n$  人が記述したとは限らない。それに対し、人数の情報が評価に有用な場合において適宜、人数についての情報を  $m$  名と記述する。

#### 5.3.1 使用した感想

MIMA サーチ利用の状況として、とりあえず思いつくキーワードを指定し、検索を行った結果、初頭結果として、表示されるノードの数や文字の重なりが目についてしまうという場合が多いと思われた。したがって、レポート内における感想の記述は、長所として (A) 講義の関連の可視化が可能 (4 名) である一方、短所として (B) 情報量の過多 ((A) とは異なる 4 名) があげられている。それぞれ、

「今までの羅列の検索結果とは異なり関連する項目と有機的に繋がっていること、検索語の自由度が高いことなどに新規性を感じた」(長所)

「全体的に文字が多く見にくい。検索件数が多いとなおのこと。デザインとしてまだ最適でないように思える。いろいろなボタンが多いのではないだろうか。トピックマップボタン<sup>\*3</sup>の使い方がわからない」(短所)

のような記述が主なものである。

#### (1) 対象者の方向性

数回検索し慣れてくると、対象者の方向性は 2 群に分かれる。1 つ目は自分の履修科目における (a) 簡便性を図りたい群 (5 名)、2 つ目は (b) 知識補完や新たな講義との出会いを図りたい群 ((a) と異なる 3 名) でありこの 2 群で MIMA サーチへの要望が大きく異なる。(a) 群は学生としての利便的な機能の追加を求めた主観的な意見が多いが、後者の意見は学生自身の自己研磨のみならず学問体系を横断的に関連付けることに対する意見であると考えられる。

#### 【(a) 群の要望】 ( $N = 8$ )

・履修条件の詳細な設定による絞り込みを望むもの

例：履修できない科目や既履修科目の表示、学科、学年、曜日、時限ごとの設定

記述例：「関連する科目をすべて表示されても履修できない」

・履修選択のための情報の追加を望むもの

例：成績の評価法、学生の授業評価

記述例：「シラバスに書かれた以上の受講者側の視点からの情報は欲しい。見学する時間があれば実際に行ったほうがよいのだろうが、見に行きたい授業はあまりにも多く、授業登録までの時間はそれに比べて少ない」

これらは、5.2 節と同様、現在では、My リストと時間割へのマッピングによりある程度は解決できる問題である。また、すでに履修した科目を除くといった要望は、やはりある程度履修選択を経験してきた学生の意見であることがうかがえる。成績の評価法等での分類は、現状のシステムでは解決している (機能が提供されている) 課題ではあるが、既修であるかどうかは学務システムとの連動が必要であり、情報セキュリティ (成績等の個人情報の管理) の観点からも検討が必要であろう。また、学生の口コミや履修履歴情報の履修選択への利用に関しては、2 章のリコメンドでの議論で述べたとおりである。

#### 【(b) 群の要望】 ( $N = 9$ )

・知識補完および未知の学問領域や講義の発見

例：キーワード一覧から検索するシステムの構築、他学科との連携

記述例：「工学のみでなく分野をまたがっているキーワードなので、他学科の情報もほしい」、「新たな授業との出会い促進のため「ランダム表示」や「人気キーワード表示」を取り入れてほしい」

・学問領域における位置付けの把握

例：履修した科目の位置付け、複数科目から得られる知の体系

記述例：「工学部の授業に限定されるのが知の体系を形成

\*3 トピックマップを ON/OFF するためのボタン。

表 4 検索における問題点の分類

Table 4 Classification of responses regarding information retrieval.

キーワード検索における問題点の内訳	記述数
①関連講義名の未検出に関するもの	8
②検索ランキングに関するもの	3
③検索結果件数の量に関するもの	3
④入力方式に関するもの	4
⑤関連ない講義名の検出に関するもの	8
⑥関連講義のリンクに関するもの	2
⑦その他	5
計	33

するうえで残念でならない」

これらも 5.2 節と同様の意見である。キーワードの補完に関しては、現在では、一般的な Suggestion システムを導入し、入力頻度の高いものより順に関連するキーワードを提示する仕組みは実装している。また、授業カタログシステムにより、他の学部の講義との関連も検索、可視化できるようになっている。

● 検索における問題点の指摘

やはり専門分野や授業の知識を比較的多く持つ学生であるからだと思われるが、授業の検索自体に関する問題の指摘が、割合としては多く見られる。モニタが指摘した検索における問題点は、内容により表 4 のように分類される。

なお、科目の検索自体は、5 章でも考察を与えており、また、本論文での本システムに対する評価の目的とは直接関係ないため、詳述は割愛する。

なお、先と同様、現時点では、大規模ソーラス\*4の利用により、①～⑤に関しての対応は済んでいる。また、⑥に関しても、5 章（検索ランキング、関連度の問題）により考察を行っている。

5.3.2 グラフ表現

グラフ表現に関しては全員（8 名）が記載している。モニタが指摘したグラフ表現における問題点の内容は、表 5 のように分類される。

5.2 節での教養学部 1 年生でのアンケート評価との記述内容の差は、カテゴリ表示について記述をしていたモニタ 1 名（N = 2）の「カテゴリ表示の意味が分からない」ということに関するものだけであり、これはマニュアル、ヘルプ等の整備により対応するものとした。そのほかはほぼ、5.2 節の内容と同様である。よって、詳述は割愛する。

5.4 評価のまとめと考察

以上の評価結果より、科目間の関連性の計算による構造化を行い、可視化することで、履修選択の効率化や、カリ

\*4 大規模ソーラスとして、JST 科学技術用語ソーラス、および人文社会 37 万語英和対訳大辞典（同一英単語に対する日本語訳語を同義語と見なす）を利用している。

表 5 グラフ表現の改善点に関する内訳 (2)

Table 5 Breakdown of responses regarding graph-expression improvement (2).

グラフ表現の問題に関する内容の内訳	記述数
カテゴリ表示*5に関するもの	2
グラフ表示領域の限界に関するもの	7
ノードの色分けに関するもの	4
リスト表示の必要性に関するもの	2
3D 化への要望に関するもの	1
その他	7
計	29

キュラム全体像の把握に関するポジティブな評価の内容が多く、多くの学生から得られたこと、また、それらの具体的な記述の内容からも MIMA サーチが履修選択に有効であることがうかがい知れる。このように、関連性の計算による構造化を行い可視化することに関する効果が見てとれる一方で、グラフ表現の改善、特に、表示領域の限界に関する内容が多く見られた。より詳細な記述内容として、アンケート評価、モニタ評価ともに、

- (1) 3D 化等を始めとする可視化手法の改善
  - (2) ユーザインタフェースの改善
  - (3) マニュアル等、説明資料の整備
- への高い要望があった。

これらの内容から、今後の可視化の改善としては、3次元化を含むグラフ表現方式の検討を行うとともに、他の可視化方式の併用による全体俯瞰能力や操作性の向上の方向等を探る必要があることが分かった。さらには、

- (4) 検索、特徴抽出、関連度計算の精度向上

に関しては、やはり、教養学部 1 年次の学生よりは、検索や言語処理にある程度の知識を持つ工学部学生からの要望の割合が多く、より技術レベルの高い観点からの要求として今後も出てくるとと思われる。まずはマニュアル等で基準を明確に示すとともに、グラフ内に関連性スコアや、種類の情報を挿入する等の可視化の際の改善策も検討したいと考えている。より高度な内容となる大学院講義を対象とした場合や、学際領域、境界領域等での学生の評価等の、さらなる評価も望まれるが、総じて、関連性の計算における自然言語処理や、可視化の高度化は今後の研究課題として大きな位置付けとなるであろう。

一方で、教養学部 1 年生と、専門課程に進んだ学生大学（全員 3 年次以上）の比較という観点では、入学当初（教養学部 1 年次）は科目の関連やカリキュラムの全体構造の俯瞰ということに非常に関心が高い印象があるが、専門課程に進んだ学生は、履修選択に対する慣れのせい、空いている時限を埋めることが優先される等の他の要因が履

\*5 図 6 で示した、学部や対象学年（レベル）をカテゴリとして横断表示する機能

修選択にかかわる割合が多いのではと感じる。このような状況に関しての結論は、今後、より多くの被験者に、継続的にモニタ調査をする等して導く必要があるが、大学選択に、興味の内容や、将来の職業選択に対する希望が多く関わっている報告 [15] や、芝浦工業大学で行われている取り組み [16] にあるような、より早い段階で全体を見渡すことに対する意識の変容を促すことで、このようなスキルを身につけさせることが重要なのではと考える [17], [18]。その意味でも、本システム利用に関する全学での広報による周知等を含み、今後も継続的に改善して行くべきことは多いであろう。

## 6. まとめ

以上、本論文では、高等教育、学際的、分野横断的教育における履修選択支援として、シラバスを対象とし、カリキュラムの検索、構造化を行うシステム (MIMA サーチ) について述べた。MIMA サーチにより、分野や学科、専攻によらず、社会の課題等に対して、必要な知識をより効率的に習得し、履修選択に役立てることが期待される。本システムは、東京大学授業カタログ [14] を始め、多くの大学、企業等で実用化されており、すでに 10 年以上にわたりサービスを続けている。これらサービスは、学生等の実際の利用者により高い評価を受けている。現在、東京大学においては、おおむね毎年約 200 万アクセス (月平均約 20 万アクセス) 程度で推移しており、特に、1) 毎年、履修選択の時期である 4 月、および 9 月にアクセスが伸びる傾向が見られる、2) アクセス元のほとんどは日本国内からのものである、ことより、実際に学生が科目の履修選択に活用していることがうかがえる。

ただし、履修選択での実際の利用のされ方に関しては、やはり、学生の学年や意識の違いにより大きな差が生まれると思われる。具体的には、履修に対する目的意識や主体性が明確であればあるほど、システムの利用に対する理解等が深まると考えられる。現在、このような観点から、Personalization をはじめ、利用者の素性 (たとえば成績等) により、どのように履修支援を行うのが望ましいかに関する研究を進めているところである。

## 参考文献

[1] 2030 年に向けた教育の在り方に関する第 2 回日本・OECD 政策対話 (報告): 文部科学省, 中央教育審議会, 教育課程分科会資料 (2015), 入手先 ([http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo3/004/siryo/\\_icsFiles/afiedfile/2015/09/04/1361405\\_6.pdf](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/004/siryo/_icsFiles/afiedfile/2015/09/04/1361405_6.pdf)) (参照 2019-06-10).

[2] 蓑輪 拓, 寺澤卓也: 履修プラン策定のためのレコメンダシステム, 情報処理学会第 76 回全国大会講演論文集, Vol.1, pp.615-616 (2014).

[3] 西森友省, 堀 幸雄, 今井慈郎: 科目履修履歴を用いた難易度別時間割の自動作成, 情報知識学会誌, Vol.21, No.2, pp.185-190 (2011).

[4] Mima, H. and Ananiadou S.: An application and evaluation of the C/NC-value approach for the automatic term recognition of multi-word units in Japanese, *Int. J. Terminology*, Vol.6, No.2, pp.175-194 (2001).

[5] Salton, G.: Developments in automatic text retrieval, *Science*, Vol.253, pp.974-980 (1991).

[6] Abid, A., Hussain, N., Abid, K., Ahmad, F., Farooq, M.S., Farooq, U., Khan, S.A., Khan, Y.D., Naeem, M.A. and Sabir, N.: A survey on search results diversification techniques, *Neural Comput. Appl.*, Vol.27, No.5, pp.1207-1229, DOI: <http://dx.doi.org/10.1007/s00521-015-1945-5> (2016).

[7] Mima, H., Ananiadou, S. and Matsushima, K.: Terminology-based Knowledge Mining for New Knowledge Discovery, *ACM Trans. Asian Language Information Processing (TALIP)*, Vol.5, No.1, pp.74-88 (2006).

[8] 野澤孝之, 井田正明, 芳鐘冬樹, 宮崎和光, 喜多 一: シラバスの文書クラスタリングに基づくカリキュラム分析システムの構築, 情報処理学会論文誌, Vol.46, No.1, pp.289-300 (2005).

[9] 井田正明, 野澤孝之, 芳鐘冬樹: 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, pp.85-97 (2005).

[10] 太田 晋, 美馬秀樹: 課題志向別シラバス自動分類システムの設計と実装, 言語処理学会論文誌, Vol.16, No.4, pp.91-106 (2009).

[11] 宮崎和光, 井田正明, 芳鐘冬樹, 野澤孝之, 喜多 一: 分類候補数の能動的調整を可能にした学位授与事業のための科目分類支援システムの提案と評価, 知能と情報 (日本知能情報ファジィ学会誌), Vol.17, No.5, pp.558-568 (2005).

[12] Sugiyama, K. and Misue, K.: Graph Drawing by the Magnetic Spring Model, *Journal of Visual Languages & Computing*, Vol.6, No.3, pp.217-231 (Sep. 1995).

[13] 増田勝也, 美馬秀樹: 講義の自動分類を利用した可視化によるカリキュラムの比較分析, 第 45 回可視化情報シンポジウム講演論文 (2017).

[14] MIMA Search 東京大学授業カタログ, 入手先 (<https://catalog.he.u-tokyo.ac.jp/>) (参照 2019-06-10).

[15] 山下仁司: 高校生はどのように志望する大学を選んでいるか, ベネッセ教育総合研究所高等教育研究室, 研究室トピックス, 入手先 (<https://berd.benesse.jp/koutou/topics/index2.php?id=2583>) (参照 2019-06-10).

[16] 星野 励, 相場 亮, 工藤一彦: 学生のセルフオリエンテーションシステムの構築—MIMA Search とカリキュラムマップを用いた履修計画立案支援, 工学教育研究講演会講演論文集平成 24 年度, No.60, pp.620-621 (2012).

[17] 三宅なほみ: 学習科学: 協調的な実践科学と理論構築との互恵関係をめざして (〈特集〉「学習支援の新たな潮流—学習科学と工学の相互作用」), 人工知能学会誌, Vol.21, No.1, pp.77-84 (2006).

[18] 土屋衛治朗, 白水 始: 情報環境学部の教育システムの効果について, 日本教育工学会論文誌, Vol.35, No.SuppL, pp.113-116 (2011).



美馬 秀樹 (正会員)

平成 8 年徳島大学工学研究科システム工学専攻修了。博士 (工学)。(株)ジャストシステム研究員, ATR 音声翻訳通信研究所研究員, 英国マンチェスターメトロポリタン大学講師, 東京大学大学院理学系研究科研究員, 同工学系研究科助手, 同工学系研究科特任准教授, 同大学総合教育研究センター准教授を経て, 令和 2 年 4 月より同工学系研究科准教授。IPA 未踏ソフトウェア天才プログラマ認定, The international Daiwa Adrian prize, 英国マンチェスター大学 Senior honorary research fellow, 工学教育賞 (日本工学教育協会), 各受賞。

## 正誤表

下記の箇所に誤りがございました。お詫びして訂正いたします。

訂正箇所	誤	正
38 ページ	受付日 2019 年 6 月 19 日	受付日 2019 年 6 月 18 日