科目の関係性に基づく
シラバス分析手法に関する研究

筑波大学
図書館情報メディア研究科

2014年11月

関陽一

目次

図目次

表目次

第 1 章

序論

1.1　背景

情報技術の発展により，インターネットで情報を容易に公開できる環境が整ってきた．その中で，大学の情報公開が進み，Web サイト上でシラバスを公開する大学が増えてきた．公開されたシラバスは，インターネット上で閲覧または入手することができる．シラバスは，学部・学科の履修について説明する情報で，各科目の情報，時間割表，学部・学科の概要，履修計画に関する情報などが含まれている．各科目の情報は科目単位で記述されており，科目ごとの独立が重視されている．

本研究では，科目ごとのシラバスを科目シラバスと称し，科目シラバスに着目する．科目シラバスは，科目名，担当教員名，開設日時，概要，授業計画，履修条件，参考図書，オフィスアワーなど，その科目を履修する学生に必要な情報が記載されている．シラバスの主な利用者は，学部・学科に所属する学生であり，彼らが履修のために閲覧している．一方で，これら以外にもシラバスの利用者が存在している．例えば，大学の学部・学科への入学を考えている受験生である．彼らは，自身が興味・関心を持つことが学べるかを知りたい．このような要求に対して，シラバスは重要な情報源となる．しかし現在のシラバスは，上述のように，科目単位で記述されている．受験生にとっては個々の科目よりも全体の様子を把握することの方が重要と考えられる．また，シラバスはテキストで記述されており，科目数が百を超えることもあるため，それらを読んで学部・学科の全体像や科目間の関係を把握するには膨大な時間が必要である．他にも，受験生は入学候補である複数の学部・学科の中からどれが自分の興味・関心に近いかを知りたい．このような要求に対しても学部・学科の全体像や科目間の関係が重要であるが，複数の学部・学科に渡って数百の科目シラバスのテキストを読む必要がある．

他の利用者の例は，企業の採用担当者である．採用しようとしている人の出身大学の学部・学科は，その人のスキルを知るための重要な情報の一つである．しかし，学部・学科の中には似たような名前のものや，学部・学科名だけでは内容が把握しづらいものが存在している．採

用担当者がより具体的に学部・学科を理解するために，シラバスは重要な情報源となる．しかし，現在公開されているシラバスの形式では，特徴を把握するのは容易でない．

さらなる利用者の例は，大学の教員である．学部・学科のカリキュラム全体を考え，自身が担当する科目を設計するには，他の科目との間の関係を考慮する必要がある．これも同様に，シラバスを読んで関係性を把握するには膨大な時間が必要である．

そこで，本研究では，科目間にある関係性に基づいて全体像の把握や科目間の関係を明らかにするためのシラバス分析手法を提案する．提案法では，まず，組織ごとに個別の形式で公開されているシラバスをオープンデータ形式に変換し統一的に扱えるようにする．大学で公開されているシラバスを収集し，テキストで記載されたシラバスの情報を抽出する．情報をオープンデータの形式に変換して蓄積し，複数の組織に対して統一的なアクセスが可能な環境を構築する．次に，シラバスのデータから特徴量を抽出する．シラバスのテキスト情報を単語（形態素）に分解し，複合語の作成，不要な語の削除を行う．科目間にある包摂関係に基づいて単語リストを形成し，それを基に科目ごとに単語の重みを計算し，特徴量とする．次に，特徴量を用いてシラバスの分析と可視化を行う．クラスタリングによって特徴が類似した科目のまとまりを抽出し，デンドログラムを用いて可視化する．また，科目間の類似度に基づいて科目間の関係性を抽出して，ネットワーク図を用いてを可視化する．

1.2　本論文の構成

本論文では，まず 2 章で本研究に関連する先行研究について述べ，本研究の位置づけを明らかにする．3 章では，科目間の関係性に基づいてシラバスを分析する手法を提案する．シラバスから情報を抽出し，その情報から科目の包摂関係を考慮した特徴量を抽出する．これを基にしてクラスタリングを行い，デンドログラムを用いて可視化する手法と，特徴量から科目間の類似度を算出し，科目間の関係性の抽出を行い，ネットワーク図を用いて可視化する手法を説明する．次に 4 章で，3 章で挙げた分析手法を実際のシラバスに適用し，その結果について評価と考察を行う．また，包摂関係を考慮した特徴量の効果を確認するため，科目間の関係性を考慮しない特徴量を用いた場合の分析結果との比較を行う．最後に 5 章で，本研究の成果をまとめて，今後の展望を述べる．

第 2 章

関連研究

本研究は，個々の大学・学部が公開しているシラバスを収集し，統一的に扱えるようにした上で，科目の包摂関係に基づいた特徴量を抽出し，シラバスの全体像や科目間の関係を分析・可視化する手法を提案している．以下では，本研究に関係する先行研究を概観し，本研究の位置づけを明らかにする．

2.1 Web からのシラバス収集とテキスト抽出に関する先行研究

山田ら [5] は，Web 上に公開されたシラバスを効率良く収集するエージェントを提案している．このエージェントは，シラバスの公開サイトがシラバスリンク集ページと個々の科目シラバスのページから構成される点に着目し，リンク集ページを決定木を用いて判定し，リンク集ページからリンクが張られているページを優先的に収集する．個々の科目シラバスのページの判定にも決定木を用い，高い精度で収集されることを確認している．

伊東ら [1] は，同一組織の科目シラバスは共通の構造で記述されることが多いことに着目し，HTML 形式の科目シラバスを対象に，共通部分（テンプレート）を特定しそこから項目名と項目値を抽出する手法を提案している．科目シラバス間で共通するタグの並びをテンプレートとし，テンプレート内の特定の位置にあるテキストが科目シラバス間で同じ場合には項目名，異なる場合には項目値として抽出している．

2.2 シラバスデータの構造化や管理に関する先行研究

井田ら [6] は，シラバスを構造化した XML スキーマとそれに基づくシラバスデータベースの構築・利用を提案している．XML スキーマを用いることでリレーショナルデータベースからデータ構造を分離させて，柔軟に構造変更できるようにしている．XML 化することで，入力データの妥当性検証，XPath を用いた検索を可能にしている．また，HTTP アクセスに対

応したサービス，XML Web サービスを提供している．特に，XML Web サービスでは外部プログラムから利用できるようにしている．これらによって，シラバスデータの統一的な扱いと保守性の向上が可能としている．

2.3　シラバス分析・可視化に関する先行研究

野澤ら [3] は，シラバスに出現する単語の分布に基づいてクラスタリングし，クラスタに対してカリキュラムがどのように帰属するかの分布を可視化し，複数のカリキュラム間で特徴を比較・分析するシステムを提案している．このシステムは，主に，学位評価の第三者機関に向けて作られており，複数のカリキュラムを横断的に把握する作業の負担軽減を目指している．更に，上で提案したシステムのクラスタリング手法を，専門用語とシラバスによって作られる二部グラフの分割によるクラスタリング手法へと改良し，先の提案で課題となっていた応答性と対話性を改善している [2]．対象カリキュラムを変更して再分析する等の繰り返し処理が行われる時の実行時間から感じるストレスを改善するとともに，学位評価者が持つ分析の観点に沿ってクラスタ分割を決定できるようにしている．

堀ら [4] は，カリキュラムの特徴を基にレーダーチャートを作成し，そこに学生が作成した時間割の特徴を要約した結果を表すシステムを提案している．このシステムは，カリキュラム全体に対して自身が作成した時間割がどのような傾向にあるかを示すものであり，足りない分野を補ったり中心的に学んでいる分野を強化するといった，学生が時間割を作成することを支援できるとしている．

高橋ら [8] は，履修科目の推薦に，先行する刺激が後続する刺激に影響を与えるプライミング効果のメカニズムをネットワーク構造で近似化した活性伝搬モデルを導入し，ユーザーの興味を基に最も脳が活性化する科目の組み合わせを提示するシステムを提案し，関連のある科目群を取り出せることを確認している．このシステムは，ユーザーが合わせて履修すると有効な科目の組み合わせを示すものであり，学生の履修計画を支援できるとしている．

2.4　シラバスからの知識抽出とその利用に関する先行研究

芳鐘ら [7] は，複合語の形態的／統語的言い換え手法を用いて下位語・同義語・関連語を抽出し，これらを基に検索語拡張を行った上でシラバスを検索するシステムを提案し，用語の AND 検索と比較してその有効性を確認している．検索拡張に用いられた用語は下位・同義・関連の関係性がわかる表示形式で確認することができ，なぜその検索結果が得られたかを検証すること，および，利用者が自身の要求に適った検索語を選択するのを支援できるとしている．

2.5　本研究の位置づけ

以上述べた研究をはじめ，シラバスを対象とする研究が数多く知られている．本研究では，以下の点を特徴としている．個々の学部・学科で公開されている異なる形式のシラバスを収集し，オープンデータの考えに基づいて統一的にアクセスできる環境について提案している．受験生や企業の採用担当者などの利用者に対して，シラバス内の情報を用いて，科目の内容に基づき，シラバスの全体像や科目間の関係を分析・可視化する手法を提案している．科目の包摂関係を考慮した特徴量を導入していることも大きな特徴となっている．

第 3 章

シラバス分析手法の提案

3.1　全体構成

提案するシラバス分析手法の全体構成を図 3.1 に示す．図の破線で囲まれた部分が大学で公開されているシラバスである．大学や学部・学科ごとに独自の形式で公開され，HTML を使って Web ページとして書かれている，PDF のファイルとして置かれているなど様々な形式がある．

このような異なる形式のシラバスを統一的に取り扱うために，オープンデータ形成部がある．収集したシラバスから抽出されたテキストは，科目名，概要などの項目名を表すタグを定義した上で構造化され，サーバに格納される．このサーバがシラバスデータベースとして機能し，統一的なアクセスを受け付ける．

利用可能となったシラバスのデータから特徴量を取り出すのが特徴量抽出部である．科目の内容を表す項目のデータを問い合わせて取得し，テキストデータを形態素へ分割し，科目の包摂関係を考慮した上で形態素の重みを算出して，特徴量を取り出し，特徴量データベースへ格納する．

このデータを用いてシラバスを分析し，その結果を視覚的に表現するのが分析・可視化部である．クラスタリングを行った結果をデンドログラムで表すクラスタリング分析，類似度に基づく関係性を抽出した結果をネットワーク図で表すネットワーク分析を行う．

3.2　オープンデータ形成部の構成

近年，シラバスを公開する大学が増えている．また，シラバスをデータとして使い，分析して，有意な情報を抽出・提示する研究が行われている．分析に先立って，データの元となるシラバスの収集とシラバスからのテキスト抽出が必要である．現在公開されているシラバスは，表現形式が組織によって異なっているため，大学・学部・学科の形式に合わせた収集・テキス

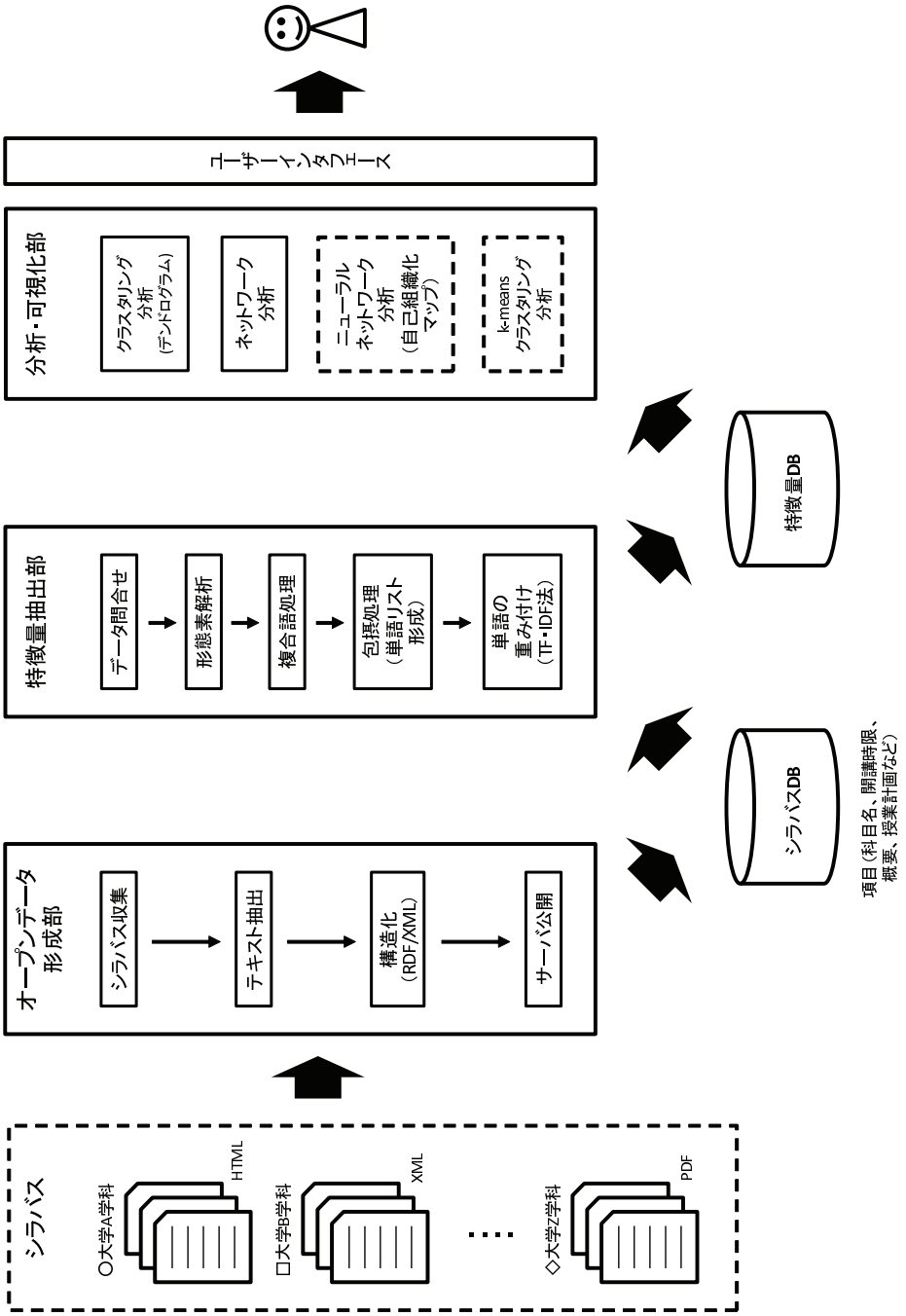
シラバス
○大学A学科
HTML
□大学B学科
XML
◇大学Z学科
PDF
オープンデータ
形成部
シラバス収集
テキスト抽出
構造化
(RDF/XML)
サーバ公開
特徴量抽出部
データ問合せ
形態素解析
複合語処理
包摂処理
(単語リスト
形成)
単語の
重み付け
(TF・IDF法)
分析・可視化部
クラスタリング
分析
(デンドログラム)
ネットワーク
分析
ニューラル
ネットワーク
分析
(自己組織化
マップ)
k-means
クラスタリング
分析
ユーザーインタフェース
特徴量DB
シラバスDB
項目(科目名、開講時限、
概要、授業計画など)

図 3.1　全体構成図

ト抽出の方法が必要である．これらの作業が，シラバスを分析することに主眼を置いている研究者にとって手間であり負担となっている．一方で，行政情報や企業の活動情報といった様々な情報を利活用することを目的として，統一した仕様に基づいて情報を記述し Web 上で公開するオープンデータ化の流れが広まっている．Web 上で情報を統一的に扱う仕組みとしてセマンティックウェブが，オープンデータを記述する一般的な形式として RDF（Resourse Discription Framwork）が知られている．そこで本研究では，セマンティックウェブを用いて，複数の組織から統一的にシラバス収集とテキスト抽出を行い，負担を軽減するシラバスデータ環境を提案する．

シラバスの収集では，ホームページの構造に合わせた Web クローラの作成，手動でシラバスを保存する手間がある．シラバスの収集は，一般に，Web を通して文書を自動でダウンロードするプログラムである Web クローラを作成して行われる．大学・学部・学科ごとにホームページの構造が異なっている場合，構造に合わせた Web クローラを用意する必要がある．ホームページによっては，Web クローラで一括収集することができないため，手動で Web ページを保存する必要がある．

シラバスからのテキスト抽出では，ファイル形式の変換や科目シラバスの構造に合わせて抽出プログラムを作成する手間がある．シラバスは，組織によってファイル形式が異なっており，その形式には HTML，XML，PDF などが存在している．このため，ファイル形式を必要に応じて別の形式へ変換することが必要である．科目シラバスからテキストを抽出は，一般に，抽出用のプログラムを作成して行われる．科目シラバスの記述のされ方（科目シラバスの構造）は組織によって異なっており，構造に合わせたプログラムを作成する必要がある．プログラムでは，タグ名や属性値，規則性のある構造を利用したり，正規表現を用いたりなど，対象となるシラバスの構造に応じた方法が求められる．

これに対して，本研究で提案するシラバスデータ環境では，セマンティックウェブを用いて，統一的な方法によるシラバス収集とテキスト抽出を行う．セマンティックウェブでは，セマンティックウェブの仕様に基づくデータが，セマンティックウェブの仕様に基づくリポジトリであるセマンティックウェブデータリポジトリ内に格納される．シラバスの収集は，セマンティックウェブデータリポジトリへアクセスできる場所であるエンドポイント URL へ，セマンティックウェブデータリポジトリへの問合せ言語で記述した問合せ文を送信することで行う．テキスト抽出は，セマンティックウェブデータの意味を表す語彙すなわちタグを指定して必要な情報を取得することで行う．異なる組織が別々の語彙を使用していたとしても，語彙や語彙間の関係を定義する仕組みであるオントロジーによって語彙同士が同じ意味か違う意味かを把握する仕組みを用意することができる．セマンティックウェブの仕組みを利用することで，大学・学部・学科といった組織間で統一された形式を申し合わせしてデータを同じ形式に整えるようなことをする必要がなく，一方で，利用者が統一的な方法でデータを入手できる環境が構築できる．これにより，シラバスの分析を行う研究者やその他のシラバスデータの二次

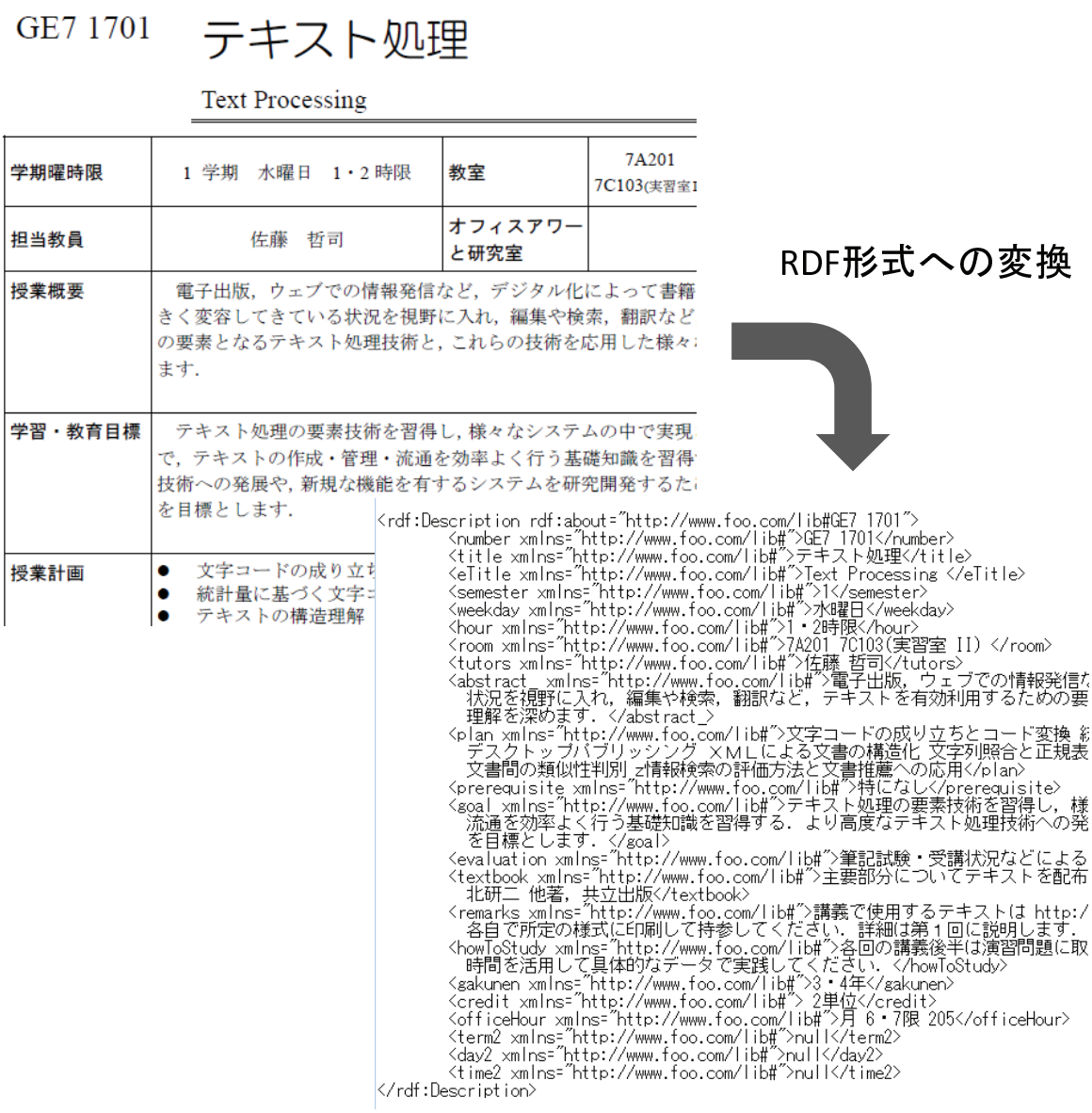

図 3.2　科目シラバスを RDF/XML 形式へ変換する例（知識情報・図書館情報学類）

的な利用者のシラバス収集やテキスト抽出にかかる負担が軽減され，彼らが主眼を置く作業により資源を集中するすることができると考えられる．

以下では，これまで述べた本研究のシラバスデータ環境に基づいて，筑波大学情報学群の情報科学類，知識情報・図書館情報学類，情報メディア創成学類のシラバスを元にシラバスデータ環境を構築した．シラバスを収集し，テキスト抽出を行い，データを構造化し，セマンティックウェブデータリポジトリへ格納し，外部からアクセスを受けられるよう設定を行った．知識情報・図書館情報学類の科目シラバスを，RDF を XML で記述した RDF/XML 形式へ変換する例を図 3.2 に示す．

シラバスの収集では，Web クローラによる一括収集と手動によるダウンロードを行った．

情報科学類，情報メディア創成学類のシラバスは，HTML 形式で 1 科目の内容が 1 ページに記述されている．学類のホームページの構造に合わせた Web クローラを作成し一括収集した．知識情報・図書館情報学類のシラバスは，1 個の PDF 形式のファイルの中に全科目のデータが含まれており．ホームページから手動でダウンロードした．

テキスト抽出では，データの形式を変換したうえで，学類ごとのシラバスの構造に合わせた方法で，科目ごとに各項目のテキストを取り出した．PDF 形式である知識情報・図書館情報学類のシラバスは，科目シラバスが表形式で記述されていることから，PDF を作成・編集するソフトウェア Adobe Acrobat[*1]を用いて HTML 形式へ変換した．これにより，HTML 形式の構造を利用できる情報となり，テーブルの n 番目の要素に講義概要があるなどの規則性に基づいてテキスト抽出を行った．情報科学類のシラバスは，「各週授業計画」などの項目名が書かれている構造を利用し，正規表現を用いてテキスト抽出を行った．情報メディア創成学類のシラバスは，要素のタグの中に id 属性があり，title，plan などの項目名が書かれた属性値が存在するため，目的としている情報がどこのタグの値にあるかを知ることができる．これを利用し，HTML データを取り扱うための各種機能を提供する HTMLParser[*2]ライブラリを用いてテキスト抽出を行った．

構造化では，1 個の科目を表現するために必要な語彙とデータ構造を学類ごとに定義し，RDF/XML 形式で記述した．語彙はタグの表記であり，科目名を title，授業計画を plan，などと定めた．ただし，組織によって記述されている項目の種類が異なるので，語彙の数が学類ごとに異なっている．また，組織によって記述形式が異なるため，語彙の意味合いが同じでも語彙の表記が異なっているものがある．データ構造は，rdf:Description タグの下に語彙を表す title や plan などのタグが並ぶ入れ子構造である．rdf:Description タグでは，rdf:about 属性を記述し，属性値に科目の URI を記述した（本研究では科目番号を URI に利用している）．title タグでは，xmlns 属性を記述し，属性値にリポジトリの URI を記述した．title タグの値には科目名として抽出してきたテキストを記述した．同様に各項目に対してもタグと値を記述して，1 科目分のデータを作成した．ここまでの過程をシラバス内の全ての科目に対して行った．1 学類分のシラバスデータを 1 個の RDF ファイルへ記述した．

セマンティックウェブデータリポジトリへデータを格納し，外部からアクセスを受けられるよう設定を行った．セマンティックウェブデータリポジトリに Sesame[*3]を使用した．Sesame はセマンティックウェブ形式のデータを格納し，セマンティックウェブデータに対する問合せ言語 SPARQL[*4]によるリクエストに応えるオープンソースシステムである．Sesame は，Web

[*1] http://www.adobe.com/jp/products/acrobat.html
[*2] http://htmlparser.sourceforge.net/
[*3] http://openrdf.org/
[*4] http://.w3.org/TR/rdf-sparql-query/

サーバである Apache Tomcat[*5]に配備することで，Tomcat のサブシステムとして動作する．Sesame の GUI を利用して，情報科学類，知識情報・図書館情報学類，情報メディア創成学類のそれぞれ 3 個のリポジトリを作成し，RDF ファイルを読み込み，データをリポジトリへ格納した．Sesame を外部から SPARQL リクエストを受けられるように設定した．

シラバスデータの利用者は，セマンティックウェブデータリポジトリに対してデータ問合せを行い，シラバスのデータを取得する．Seame API または Java[*6]標準ライブラリの HttpURLConnection クラスを用いて，Sesame で定義されているエンドポイントの URL を指定して SPARQL 文を送るプログラムを作成する．SPARQL 文中に利用者が必要なデータと対応している語彙を指定する．Sesame からのレスポンスに RDF/XML 形式で記述されたシラバスデータ含まれており，Sesame の API を使って指定した語彙ごとにテキストを抽出するプログラムを作成する．

本研究のシラバスデータ環境によって，3 学類が組織ごとにシラバス公開しているが，シラバスデータの利用者はエンドポイント URL への問合せと語彙の指定という統一的な方法でシラバス収集とテキスト抽出が可能となった．

3.3　特徴量抽出部の構成

シラバス内の各科目は独立して存在しているわけではなく，多くの科目が別の科目と関係性を持つ．この関係性の内，履修の前後関係に基づく科目の包摂関係に着目した．シラバスでは，「履修要件」，「前提知識，他科目との関連等」，「予備知識・前提条件」などのある科目を履修するうえで事前に履修することが必要または望ましい科目が指定されている．この時，本研究では，ある科目を当該科目，事前に履修することが必要または望ましい科目を事前履修科目と称し，当該科目と事前履修科目との間にある関係を包摂関係と称する．包摂関係が存在する時，当該科目は事前履修科目の内容を含むと考えられる．これを反映させた特徴量を抽出する．特徴量の抽出には，データの問合せ，形態素解析，複合語処理，形態素のフィルタリング，包摂関係に基づく形態素リストの形成，形態素の重みの算出を行う．

分析の元となるシラバスのデータを取得するために，オープンデータ形成部で提案したセマンティックウェブに基づくシラバスデータ環境へ問い合わせを行った．Sesame API を利用し，3 つの学類のリポジトリからシラバスのデータを得るプログラムを作成した．プログラムでは，それぞれの学類のエンドポイントの URL を指定して，SPARQL の get 文を送る．get 文中では，全項目の中から科目の内容を表す項目である科目名や授業計画などのテキストを得るために，title や plan などの語彙を指定する．科目の内容を表す項目としては，科目名，英

[*5] http://tomcat.apache.org/
[*6] https://java.com/ja/

語科目名，概要，授業計画，教科書，参考資料，キーワードを用いる．また，包摂関係を抽出するために，事前履修科目に関する項目を表す語彙も指定する．レスポンスから項目ごとにテキストを取得する．

テキストの内容に基づく分析の場合，文章を単語に分解して行う．日本語では，単語の最小単位である形態素へ分かち書きされる．文章を形態素へ分解するソフトウェアである形態素解析機には MeCab[*7]を使用した．自然言語処理ライブラリである Sloth Lib[*8]を使用して，MeCab にテキストデータを入力し，形態素の配列を得た．MeCab で解析されたデータには，語形変化がテキスト中に出現した形のままである形態素に加えて，語形変化を取り除いた形態素の原型，品詞などの情報が含まれる．同じ意味を表す語形変化の異なる単語同士は同一の単語として扱われるのが望ましいため，形態素の原型を使用する．

専門用語など複数の名詞から成る複合語は，形態素解析の結果，複数の名詞に分けられてしまう場合がある．複合語は科目の内容を表す重要な語の一つであるため，そのままの形を維持する必要がある．そのため複合語化処理を行った．形態素解析後，形態素の品詞を文頭から順に確認し，名詞が連続して出現している場合に，それらの名詞を結合し，複合語を作成した．

文章の中には動詞，形容詞，助詞などの内容と関係が薄いまたは無い単語である機能語が含まれている．文章の内容に基づく分析を行う場合，機能語が不要になるため，これを取り除く必要がある．そこで形態素に対して，品詞に基づくフィルタリング処理を行った．MeCab が生成した品詞情報を利用して，名詞と MeCab によって品詞が判定されなかった語である未知語を残し，それ以外の品詞を取り除いた．未知語には，アルファベットやカタカナの専門用語や新語という内容を表す語が含まれているため残すこととした．

品詞に基づくフィルタリングを行った後も，一部に文章の内容と関係のない形態素が残されている．これを取り除くために，単語に基づくフィルタリングを行った．品詞に基づくフィルタリングの後に残った形態素を目視し，記号を不要な語として取り除いた．

特徴量の作成に際して，個々の文書のみを用いて作成するだけでなく，文書間の関係性を利用して作成できる場合がある．科目シラバス間の事前履修に基づく関係性を利用して，事前履修科目の特徴を当該科目へ反映させる包摂関係処理を行った．シラバスには，「前提となる科目」，「事前に履修することが必要な科目」，「事前に履修することが望ましい科目」といった当該科目を履修するうえで事前に履修することが必要または望ましい科目が指定されている項目（事前履修科目欄）がある．当該科目に事前履修科目欄があり，そこに事前の履修が必要または望ましい科目が記載されている場合，当該科目はその関係性から，事前履修科目の内容を含んでいると考えられる．そのため，事前履修科目のテキストが暗黙的に当該科目のシラバス内に記述されているとみなすことができる．これを特徴量として表すために，事前履修科目の形

[*7] http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html
[*8] http://www.dl.kuis.kyoto-u.ac.jp/slothlib/?FrontPage

態素リストを当該科目の形態素リストへ結合させる．個々の科目の事前履修科目欄にシラバス内の科目名が記述されているかを調べるためにマッチング処理を行った．事前履修科目欄のテキストに科目名含まれている場合，その科目名を記録した．事前履修科目欄と考えられる項目が複数ある場合，それらに書かれたテキストを一つに統合し，事前履修科目欄とした．記録した科目名を元に事前履修科目の形態素リストを取得し，当該科目の形態素リストに加えた．ただし，当該科目の講義内で行われる内容は当該科目のものがメインとなるため，当該科目のテキストの影響が大きく，事前履修科目のテキストの影響が小さいと考えられる．そこで，当該科目と事前履修科目の重みを調整した．当該科目と事前履修科目の間の重みを 2:1 とし，当該科目の形態素リストにさらに当該科目の形態素リストを加えて，形態素の出現頻度を 2 倍にした．これに伴って，事前履修科目を持たない科目の形態素リストにさらに事前履修科目を持たない科目の形態素リストを加えて，形態素の出現頻度を 2 倍にした．こうして，2 倍の量の当該科目の形態素リストと事前履修科目の形態素リストを結合した包摂関係処理済みの形態素リストを作成した．

本研究では内容に基づく分析を行うため，特徴量の要素は形態素の重みである．重みに形態素の出現頻度を用いると，多くの科目で出現し，かつ 1 科目内でも何度も出現する一般的な語の重みが大きく作用する．このような語の作用を小さくし，科目の特徴を表す語の作用を大きくする必要がある．そこで重みには TF・IDF 値を用いた．TF・IDF 値の場合は，科目の特徴を表す語の重みがより大きくなり，一般的な語の重みがより小さくなるよう計算される．TF は，1 科目内における各形態素の出現頻度である．IDF は，1 学類内における各形態素が出現する科目シラバスの数である．Sloth Lib ライブラリに包摂関係処理済みの形態素を入力し，形態素の TF 値と IDF 値を得た．これらを掛け合わせて TF・IDF 値を得た．

以上の処理で作成された特徴量は，文書を形態素の重みで表現した文書ベクトルである．以下で行う分析で用いるために，科目ごとに形態素の重みを特徴量データベースへ格納した．

3.4　分析・可視化部の構成

分析・可視化部では，特徴量抽出部で作成した特徴量を特徴量データベースから取得し，下記の特徴量ベクトル間の距離を用いた手法で分析し，その結果を可視化することで学部・学科の全体像や科目間の関係を表現する．

クラスタリング分析では，クラスタリングを行い，結果をデンドログラム（樹形図）で可視化した．クラスタリングは，特徴が似ている要素同士をひとつの集合（クラスタ）にまとめ，全体がいくつの特徴的な部分集合から構成されるかを把握する手法である．クラスタリングには，最も特徴が近い要素同士をクラスタにまとめる処理を全体がひとつのクラスタになるまで行う階層型クラスタリングを用いた．階層型クラスタリングでは，個々の要素を表す特徴ベクトルの間の距離をベクトル間距離尺度を基に算出し，その結果からクラスタ間の距離をクラス

タ間距離尺度を用いて算出し，近い要素またはクラスタをひとつの上位クラスタにまとめる処理を繰り返していく．

階層型クラスタリングには，統計分析ソフト R[*9]を用いた．特徴量を特徴量データベースから取得し，R の入力形式へ変換し，R に入力した．ベクトル間距離尺度には基本的な距離尺度であるユークリッド距離を用いることとし，R の dist 関数を使用して科目間の距離を算出した．クラスタ間距離尺度には最短距離法，最長距離法，群平均法，ウォード法を用いることとし，科目間の距離を基に，R の hcluster 関数を使用してクラスタリングを行った．

クラスタリングの結果の可視化には，デンドログラムを用いた．デンドログラムは，クラスタリングの結果に基づいて順にクラスタにまとめられた要素またはクラスタ同士を線で結ぶ．そのため，全体がひとつのクラスタになるまでの階層的な過程が視覚的に表現される．デンドログラムの図では，縦軸が要素またはクラスタ間の距離を表しており，次の結合点までの距離が近いほど特徴が似ている．デンドログラムの任意の位置で水平な直線を引くと，そのクラスタリング時点でのクラスタ数を知ることができる．デンドログラムへの可視化には R を用いた．クラスタリングの結果を R の plot 関数に入力し，デンドログラムを作成した．

次に，ネットワーク分析では，科目間の関係性を抽出し，ネットワーク図で可視化した．ネットワーク図では，要素すなわち科目がノード（頂点）と呼ばれる点で表現され，要素間に関係性があるときノード同士はエッジと呼ばれる線で結ばれる．そのため要素全体がどのような結びつきからできているのか知ることができる．ネットワークの中には，多くの要素との間で結びつきを持つ中心的なノードが現れることがある．また，ネットワークの中には，密なネットワーク同士をつなぐ橋渡し的なノードが現れることがある．これらの特徴は次数中心性および媒介中心性と呼ばれ，これらに着目することでネットワーク上で重要な役割を果たすノードを知ることができる．

ネットワーク図の作成に先立ち，類似度に基づく科目間の関係性を抽出した．特徴量を特徴量データベースから取得し，これを Sloth Lib ライブラリへ入力し，特徴ベクトル間の角度であるコサイン類似度を算出して，科目同士の類似度を取得した．本研究では，科目をノードで表し，ある科目に着目した時に類似度が最も高い科目との間に関係性があるとして両科目の間にエッジを張る．ただし，任意の順番で類似度最大を調べている時にすでにその関係性が抽出されている場合（例えば，科目 A が科目 B と類似度最大となることが分かっている時に，科目 B の類似度最大を調べると科目 A となる場合），すでにエッジが生成された科目を除外して，未だエッジが引かれていない科目との間で類似度が最大の科目に対してエッジを張ることとした．

類似度最大に基づいて関係性を抽出すると，二つの科目の間で互いにエッジを引く合う関係がいくつも出現する．これをネットワーク図で表すと，ネットワーク全体が二つの科目のみか

[*9] http://www.r-project.org/index.html

ら成る孤立したサブネットワークをいくつも含む形で描写される．全体の結びつきをとらえるために，二つの科目のみから成る孤立したサブネットワークがなくなるよう，この方法を用いた．

抽出した関係性を，ネットワーク分析ソフト Pajek[*10]の入力形式に変換した．これを Pajek へ入力して，サブネットワークごとに整理して配置する kamada-kawai モデルの Separate Components をレイアウトとして選択し，ネットワーク図を作成した．

[*10] http://vlado.fmf.uni-lj.si/pub/networks/pajek

第 4 章

評価と考察

以下では，包摂関係特徴量に基づいたクラスタリング分析結果，および，ネットワーク分析結果について評価・考察を行う．また，包摂関係特徴量を用いた際の効果について明らかにするため，科目単独特徴量を用いて分析した結果との比較も行う．

4.1　評価に使用するシラバス

本研究で使用したシラバスは，筑波大学情報学群に属する情報科学類，知識情報・図書館学類，情報メディア創成学類のものである．各シラバスに含まれる科目数は，情報科学類が 100 科目，知識情報・図書館学類が 106 科目，情報メディア創成学類が 82 科目である．

各学類のシラバスにおける事前履修科目数の分布を表に示す．情報科学類が表 4.1，知識情報・図書館学類が表 4.2・情報メディア創成学類が表 4.3 である．表は，情報科学類を例にすると，事前履修科目欄に事前履修科目が含まれない科目シラバスの数が 52 件，事前履修科目欄に事前履修科目が 1 件含まれる科目シラバスの数が 27 件ということを示す．事前履修科目ののべ数は，各科目シラバスに含まれる事前履修科目の数を合計したものである．

4.2　評価に使用する特徴量

本研究では，科目シラバスからの特徴量抽出に関して，包摂関係を反映させて作成される特徴量である包摂関係特徴量を提案している．シラバスには，「履修要件」，「前提知識，他科目との関連等」，「予備知識・前提条件」などの項目があり，その科目を履修するうえで事前に履修することが必要または望ましい科目が記述されている．本研究ではこれを利用し，当該科目と事前履修科目の間の関係を包摂関係と称している．包摂関係があるということは，当該科目のシラバス内に事前履修科目のシラバスのテキストが暗黙的に記述されているとみなすことができる．ここから包摂関係を反映させた単語リストを形成し，包摂関係特徴量を作成した．包

表 4.1　事前履修科目数の表（情報科学類）

事前履修科目の数（階級）	科目の数（度数）
0	52
1	27
2	11
3	5
4	2
5	2
6	1
7	0
8	0
9	0
10	0
科目の合計	100
事前履修科目ののべ数	88

表 4.2　事前履修科目数の表（知識情報・図書館学類）

事前履修科目の数（階級）	科目の数（度数）
0	76
1	14
2	6
3	2
4	4
5	3
6	0
7	1
8	0
9	0
10	0
科目の合計	106
事前履修科目ののべ数	70

表 4.3　事前履修科目数の表（情報メディア創成学類）

事前履修科目の数（階級）	科目の数（度数）
0	50
1	19
2	5
3	4
4	2
5	0
6	0
7	1
8	0
9	0
10	1
科目の合計	82
事前履修科目ののべ数	66

摂関係特徴量を用いることで，履修の前後関係が反映された分析が可能になると考えられる．

また，包摂関係特徴量を用いた際の効果について明らかにするための比較対象として，包摂関係の処理を行わないで作成された特徴量である科目単独特徴量を使用した．科目単独特徴量は，特徴量抽出部の処理から包摂関係を反映させた単語リストの形成を除いたものである．

4.3　クラスタリング分析に使用するクラスタ間距離尺度

クラスタリングで用いられるクラスタ間距離尺度にはいくつかの種類がある．本研究では，似ている要素がまとまったクラスタが複数作られる結果，つまり，似ている要素同士は近づくが（内的結合）他のクラスタとの距離が適度に離れる状態（外的分離）が望ましい．そこで，クラスタ間距離尺度について比較・検討を行った．比較・検討に用いたクラスタ間距離尺度は最短距離法，最長距離法，群平均法，ward 法である．情報科学類のシラバスをそれぞれのクラスタ間距離尺度を使ってクラスタリングし，その結果を可視化した図から最適なクラスタ間距離尺度を決定する．

最短距離法（図 4.1）では，形成されたクラスタに対して近い距離にある科目が順に統合され新たな上位クラスタを形成していく数珠つなぎのような形状が初期から見られる．この状態では学部・学科の持つ特徴を把握するのが難しく，特徴を抽出するにはクラスタ間に適度な距

離が必要である．そのため全体的に数珠つなぎのような形状が見られる最短距離法は，本研究でのクラスタリング分析に適していない距離尺度であると考えられる．

群平均法（図 4.2）でも，最短距離法と同じように数珠つなぎのような形状が見られた．最長距離法（図 4.3）では，これまでの二つの距離尺度と比べて一部で外的分離性のあるクラスタが形成されているが，全体的には数珠つなぎのような形状が混在しているのが見られる．これらから，群平均法と最長距離法も本研究でのクラスタリング分析に適していない距離尺度であると考えられる．

ward 法（図 4.4）では，デンドログラムを上から見ると，全体がまず 2 個の大きなクラスタに分けられている．その片方のクラスタでは，さらに 3 個の大きなクラスタに分けられている．ここから，情報科学類が全体で 4 個の大きな科目のまとまりから成ることが把握できる．

最短距離法，最長距離法，群平均法を用いた場合に見られる数珠つなぎの形状や，ward 法を用いた場合に全体が大きなまとまりから成ることが把握できる結果は，知識情報・図書館学類や情報メディア創成学類のシラバスをクラスタリングした際にも同様の傾向が見られた．以上から，本研究におけるクラスタリング分析のクラスタ間距離尺度として ward 法が適していることが明らかとなった．

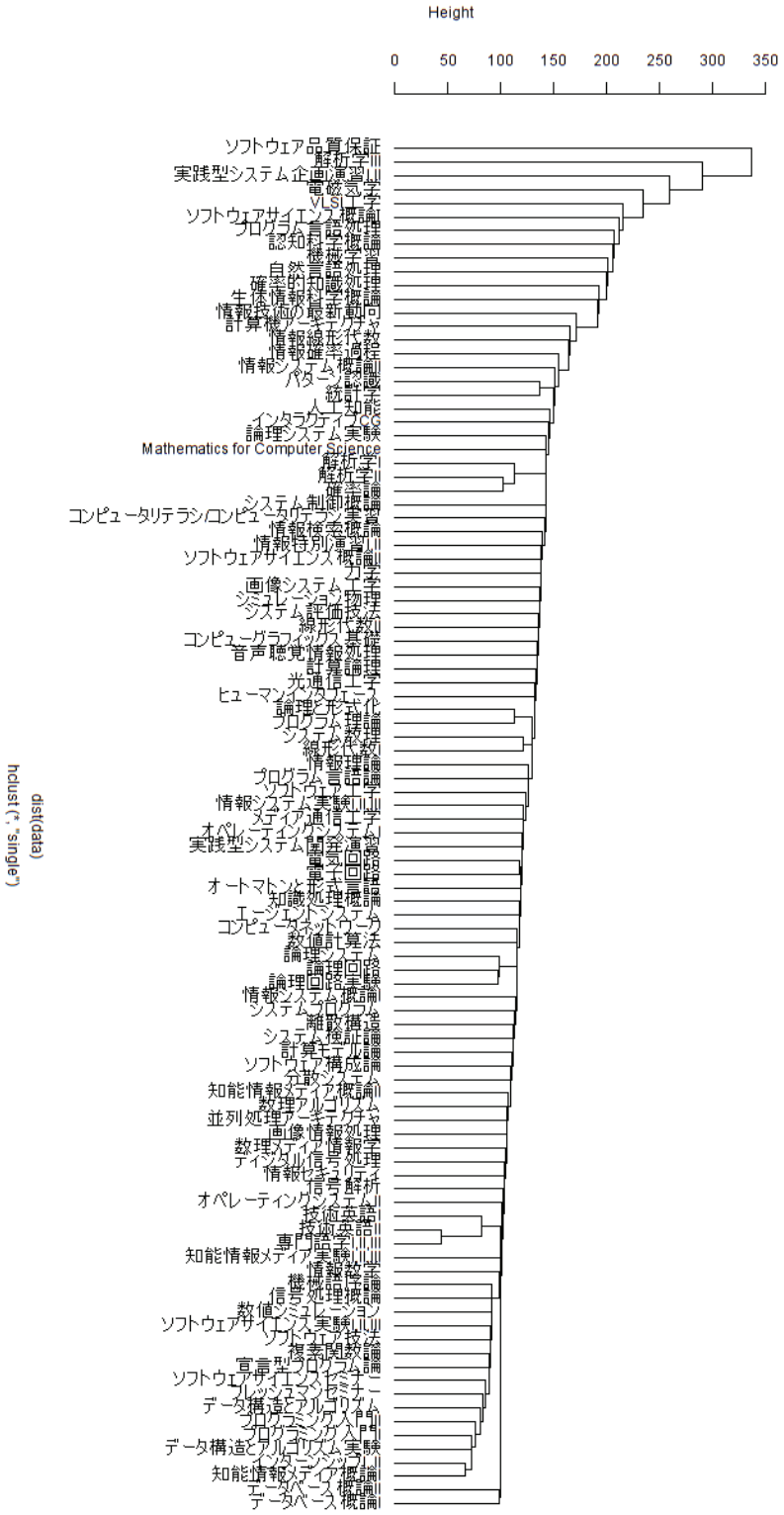

図 4.1　最短距離法を用いたクラスタリング結果（情報科学類）

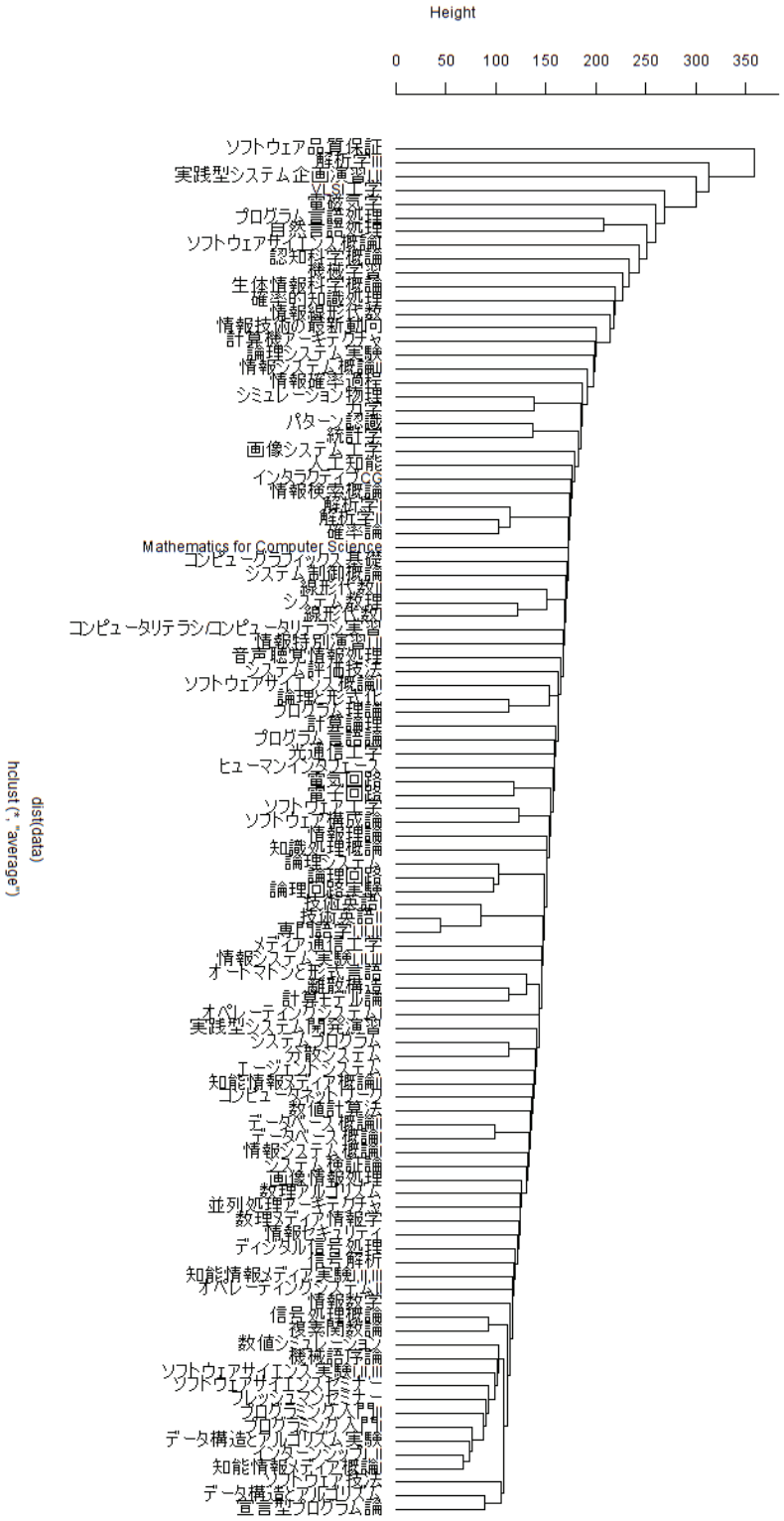

図 4.2　群平均法を用いたクラスタリング結果（情報科学類）

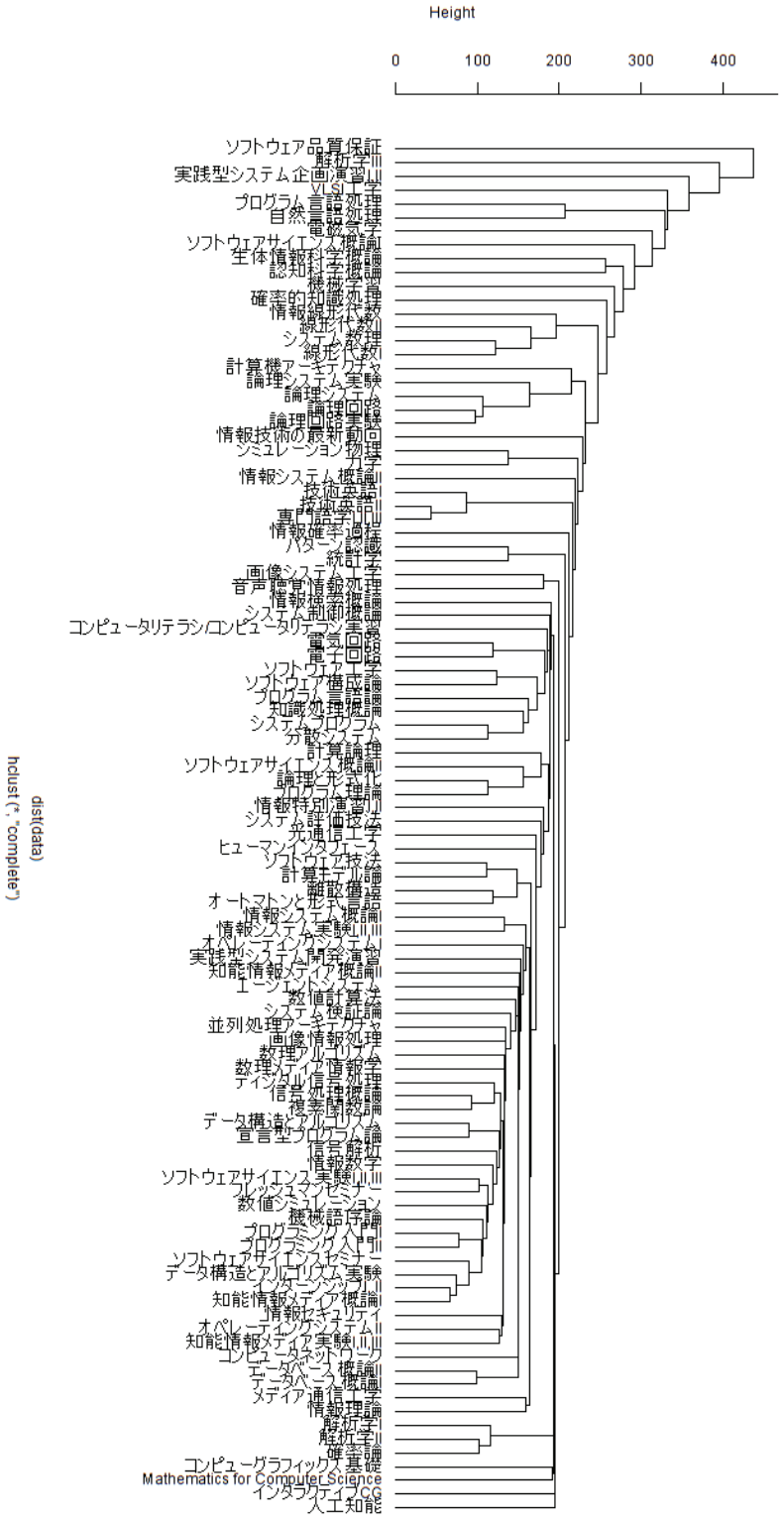

図 4.3　最長距離法を用いたクラスタリング結果（情報科学類）

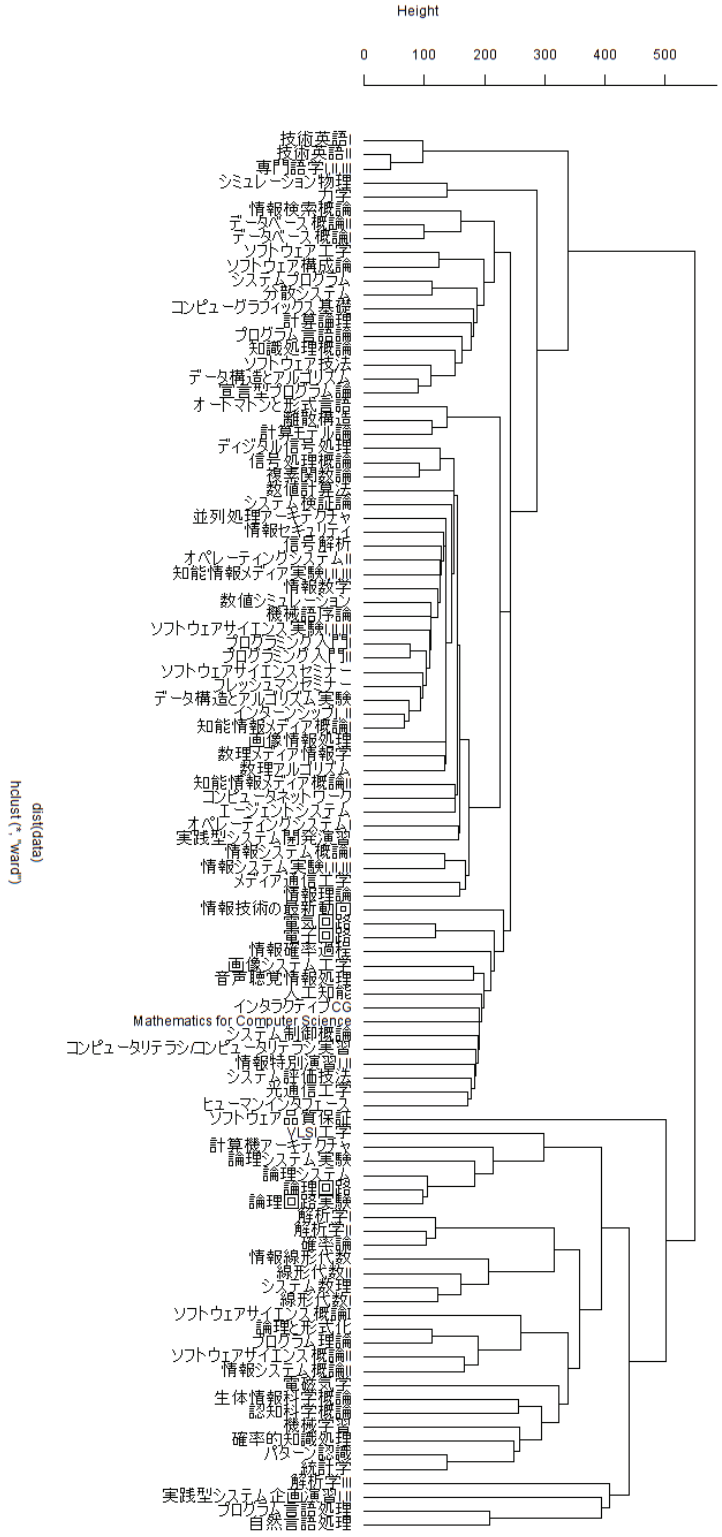

図 4.4　ward 法を用いたクラスタリング結果（情報科学類）

4.4　クラスタリングとデンドログラムを用いた可視化の評価・考察

包摂関係特徴量を用いてクラスタリングを行い，デンドログラムで可視化した．情報科学類の結果を図 4.5，知識情報・図書館学類の結果を図 4.7，情報メディア創成学類の結果を図 4.9 に示す．

デンドログラムを観察した結果，シラバスがいくつかの大きな科目のまとまりから成ることが見られる．デンドログラムを上から見ると，情報科学類（図 4.5）では，全体がまず 2 個の大きなクラスタに分けられ，その片方のクラスタではさらに 3 個の大きなクラスタに分けられ，結果，全体が 4 個の大きな科目のまとまりから成っている．知識情報・図書館学類（図 4.7）では，全体が 2 個の大きな科目のまとまりから成っている．情報メディア創成学類（図 4.9）では，全体が 2 個の大きなクラスタに分けられ，その片方のクラスタではさらに 2 個の大きなクラスタに分けられ，結果，全体で 3 個の大きな科目のまとまりから成っている．これらは，シラバス全体のまとまりを把握するのに役立つと考えられる．また，特徴の似た科目から成る小規模クラスタが多く見られる．これらには包摂関係にある科目同士から構成されるクラスタも含まれている．これらは，科目内容に関係性があるもの同士を知るのに役立つと考えられる．また，個々の科目がまとまっていき一つのシラバスを構成する全過程が見られる．特に，任意の箇所で水平に直線を引くことで，その時点でのクラスタ数を知ることができる．これは，シラバスの全体像や科目間の関係を把握するのに役立つと考えられる．

中規模のクラスタでは，複数の学問分野の科目やクラスタが混在しているのが見られる（図 4.5，図 4.7，図 4.9）．学問分野とは，例えば，知識・情報図書館学類であれば，図書館系，情報処理系，社会学系が考えられる．情報メディア創成学類では，情報処理系，デザイン・コンテンツ系が考えられる．情報処理系をさらに詳しく見ると，プログラミング系，通信系，音声系，画像系，機械系，などが考えられる．他にも，英語系，数学系，物理学系などが考えられる．クラスタリングの初期段階で内容が似ている科目がまとまった後，いくつもの学問分野の科目や小さなクラスタがまとまって中規模なクラスタが作られている．学部・学科がどのような特徴から成るのかを把握するために，中規模のクラスタにおいても同じ学問分野の科目がまとまることが望ましい．しかしながら，包摂関係特徴量だけではこのようなまとまりを形成することができなかった．

クラスタリング分析における包摂関係特徴量の効果を確認するために，科目単独特徴量を用いた結果との間で比較を行った（図 4.5，図 4.6，図 4.7，図 4.8，図 4.9，図 4.10）．

科目単独特徴量の図（図 4.6，図 4.8，図 4.10）では，数珠つなぎの形状である部分が複数見られる．包摂関係特徴量の図（図 4.5，図 4.7，図 4.9）では，この部分が一部分離して，異な

るクラスタ群にまとめられた数が増えている．これにより，特徴的なまとまりが科目単独特徴量の図よりも把握しやすくなったと考えられる．これは，包摂関係にある科目群が単語を共有することでまとまりやすくなり，同時に他のクラスタとの距離感が広くなったためだと考えられる．

左側の軸の数字に着目し，クラスタ間の距離について考える．科目単独特徴量の図（図 4.6，図 4.8，図 4.10）では多くの箇所で次の結合点までの長さが短いことが見られる．これは，クラスタ間の距離が近いためである．包摂関係特徴量の図（図 4.5，図 4.7，図 4.9）では，次の結合点までの長さが長くなった箇所が増えている．これにより，科目単独特徴量の図よりも適度な距離感でクラスタが形成されるようになったと考えられる．このクラスタ間の距離は左の軸の目盛りの幅が広がったことからも確認できる．情報科学類については，科目単独特徴量（図 4.6）では 0〜300 であるのに対して，包摂関係特徴量（図 4.5）では 0〜500 である．知識情報・図書館学類については，科目単独特徴量（図 4.8）ではほとんどの科目が 0〜約 200 の間に含まれるのに対して，包摂関係特徴量（図 4.7）ではほとんどの科目が 0〜約 700 の間に含まれる．情報メディア創成学類については，科目単独特徴量（図 4.10）では 0〜300 であるのに対して，包摂関係特徴量（図 4.9）では 0〜600 である．ここから，包摂関係特徴量がクラスタ間距離の分離性に影響を与えていると考えられる．

クラスタリングの初期段階では，距離の近い要素が 2〜3 個まとまって小規模なクラスタを形成している．科目単独特徴量の図（図 4.6，図 4.8，図 4.10）と比べて，包摂関係特徴量の図（図 4.5，図 4.7，図 4.9）は，特徴の似ている科目から成る小規模なクラスタの数が増えている．これは，包摂関係にある科目群が単語を共有することでまとまりやすくなったためであり，包摂関係特徴量が科目のまとまりに影響を与えていると考えられる．

内容が似ている科目がまとまって出現するかどうかについても違いが見られる．科目単独特徴量の図（図 4.6，図 4.8，図 4.10）では離れた位置に出現するが，包摂関係特徴量の図（図 4.5，図 4.7，図 4.9）では，まとまって出現する箇所がある．情報科学類を観察すると，科目単独特徴量（図 4.6）において，データベース概論 I，データベース概論 II，情報検索概論がそれぞれ離れた位置に出現している．包摂関係特徴量（図 4.5）おいては，データベース概論 I，データベース概論 II，情報検索概論が 1 個の小規模クラスタにまとまっている．また，科目単独特徴量（図 4.6）において，技術英語 II，専門英語 I,II,III が 1 個の小規模クラスタにまとまっているが技術英語 I が離れた位置に出現している．包摂関係特徴量（図 4.5）においては，技術英語 I，技術英語 II，専門英語 I,II,III が 1 個の小規模クラスタにまとまっている．知識情報・図書館学類では，科目単独特徴量（図 4.8）においてコンピュータシステムとネットワークがグリッドコンピューティングと離れた位置に出現しているが，包摂関係特徴量（図 4.7）においては 2 科目が小規模クラスタでまとまっている．また，科目単独特徴量（図 4.8）においてメディア社会学とメディア社会文化論が離れた位置に出現しているが，包摂関係特徴量（図 4.7）においては 2 科目が小規模クラスタでまとまっている．情報メディア創成学類で

は，科目単独特徴量（図 4.10）において CG 基礎とインタラクティブ CG が離れた位置に出現しているが，包摂関係特徴量（図 4.9）においては 2 科目が小規模クラスタでまとまっている．また，科目単独特徴量（図 4.10）において信号とシステム，音声情報処理，音楽・音響情報処理が離れた位置に出現しているが，包摂関係特徴量（図 4.9）においては 3 科目が小規模クラスタでまとまっている．これらの小規模クラスタが包摂関係にある科目で構成されていることから，包摂関係を考慮することで内容の似た科目が小規模クラスタにまとまりやすくなると考えられる．

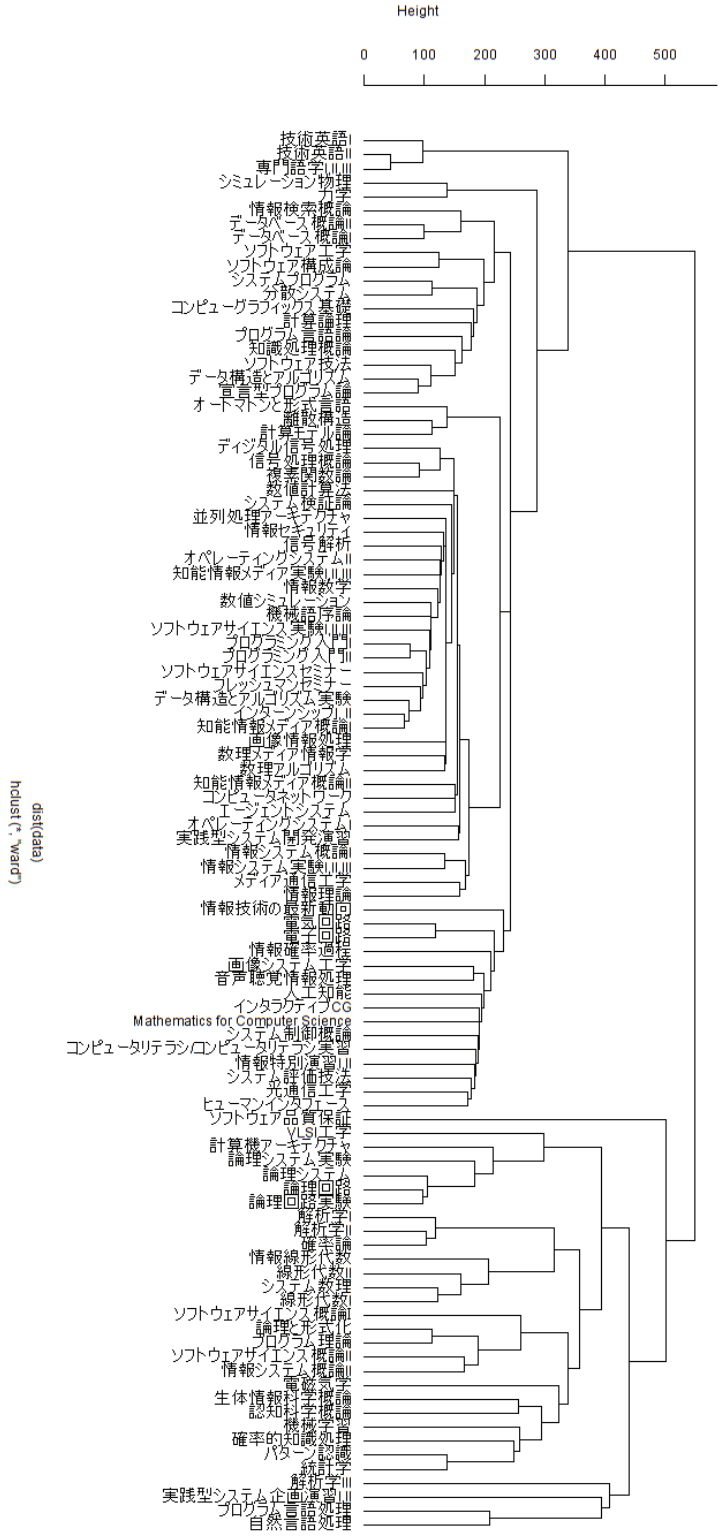

図 4.5　包摂関係特徴量を用いたクラスタリング結果（情報科学類）

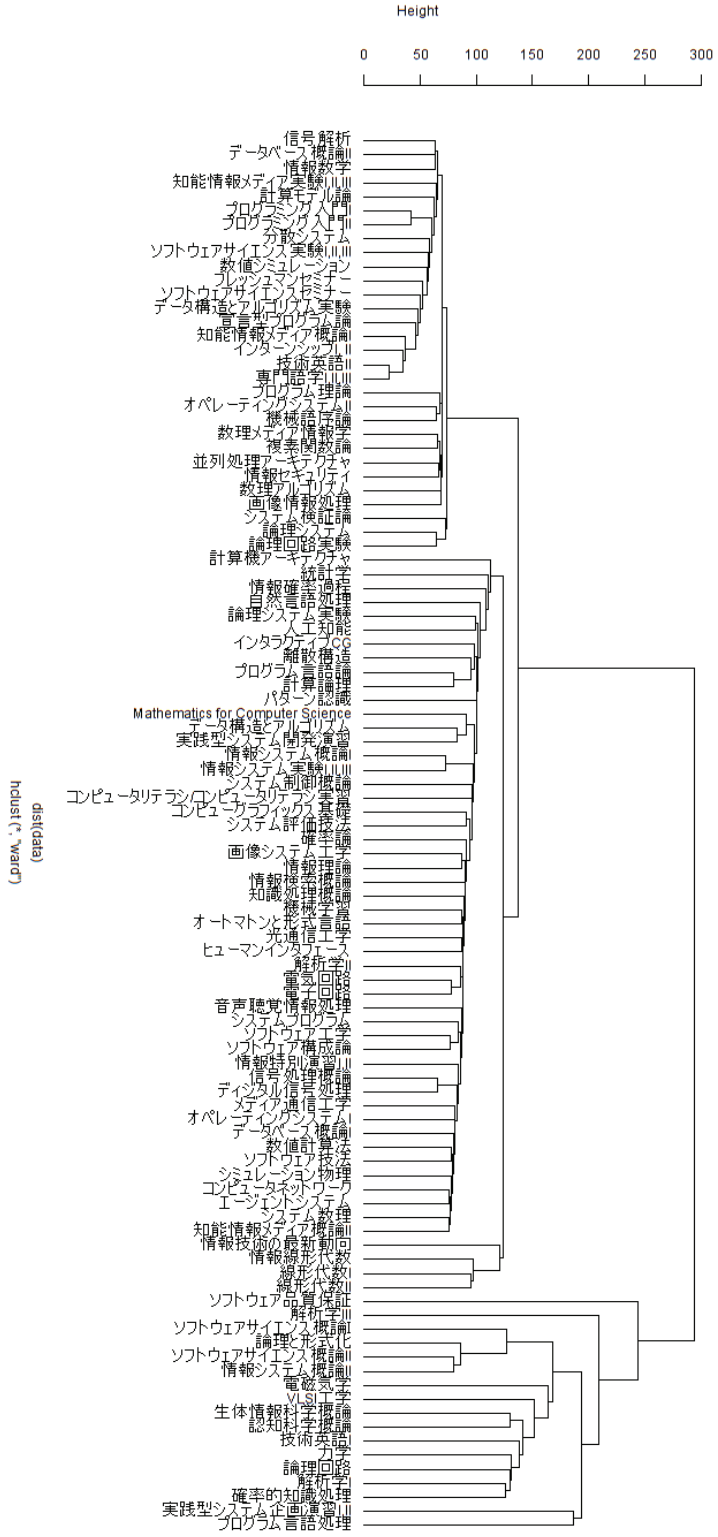

図 4.6　科目単独特徴量を用いたクラスタリング結果（情報科学類）

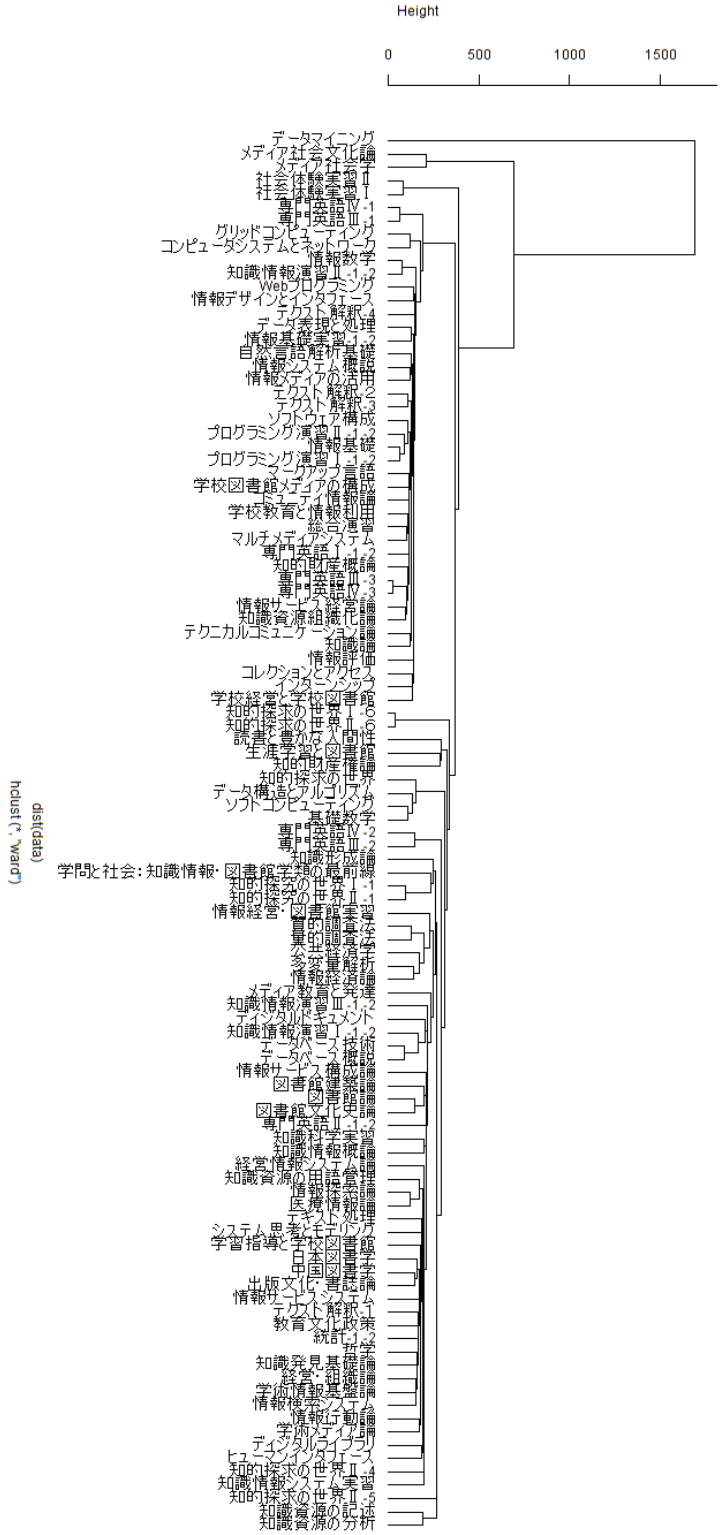

図 4.7　包摂関係特徴量を用いたクラスタリング結果（知識情報・図書館学類）

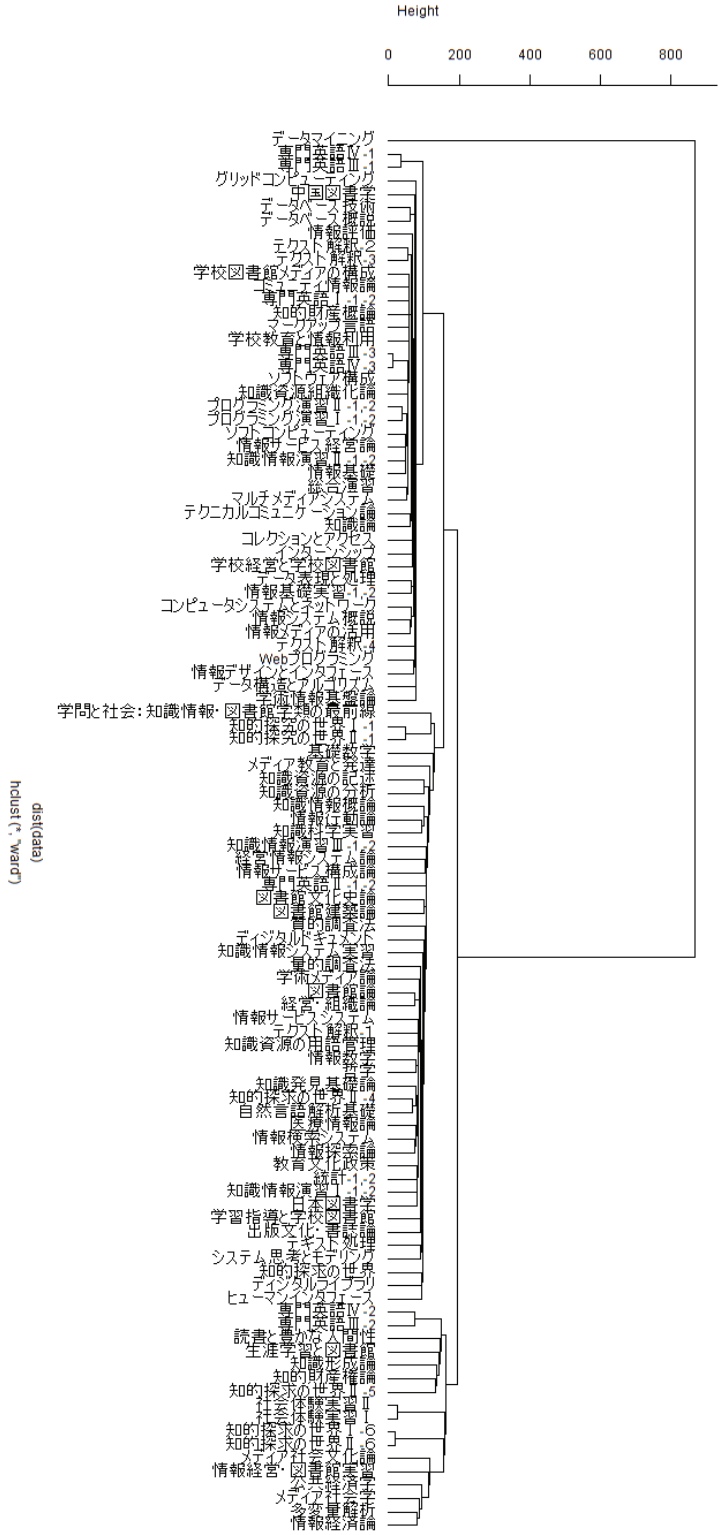

図 4.8　科目単独特徴量を用いたクラスタリング結果（知識情報・図書館学類）

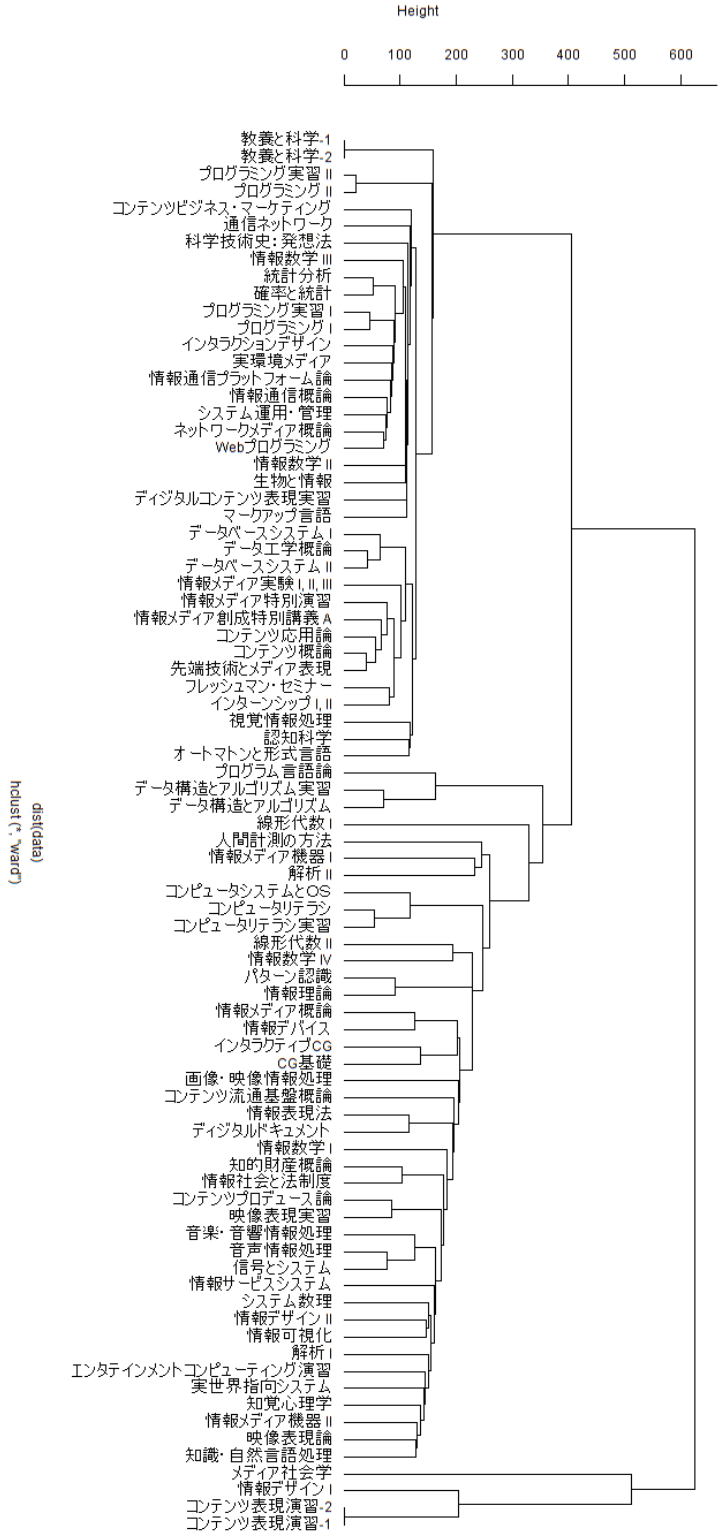

図 4.9　包摂関係特徴量を用いたクラスタリング結果（情報メディア創成学類）

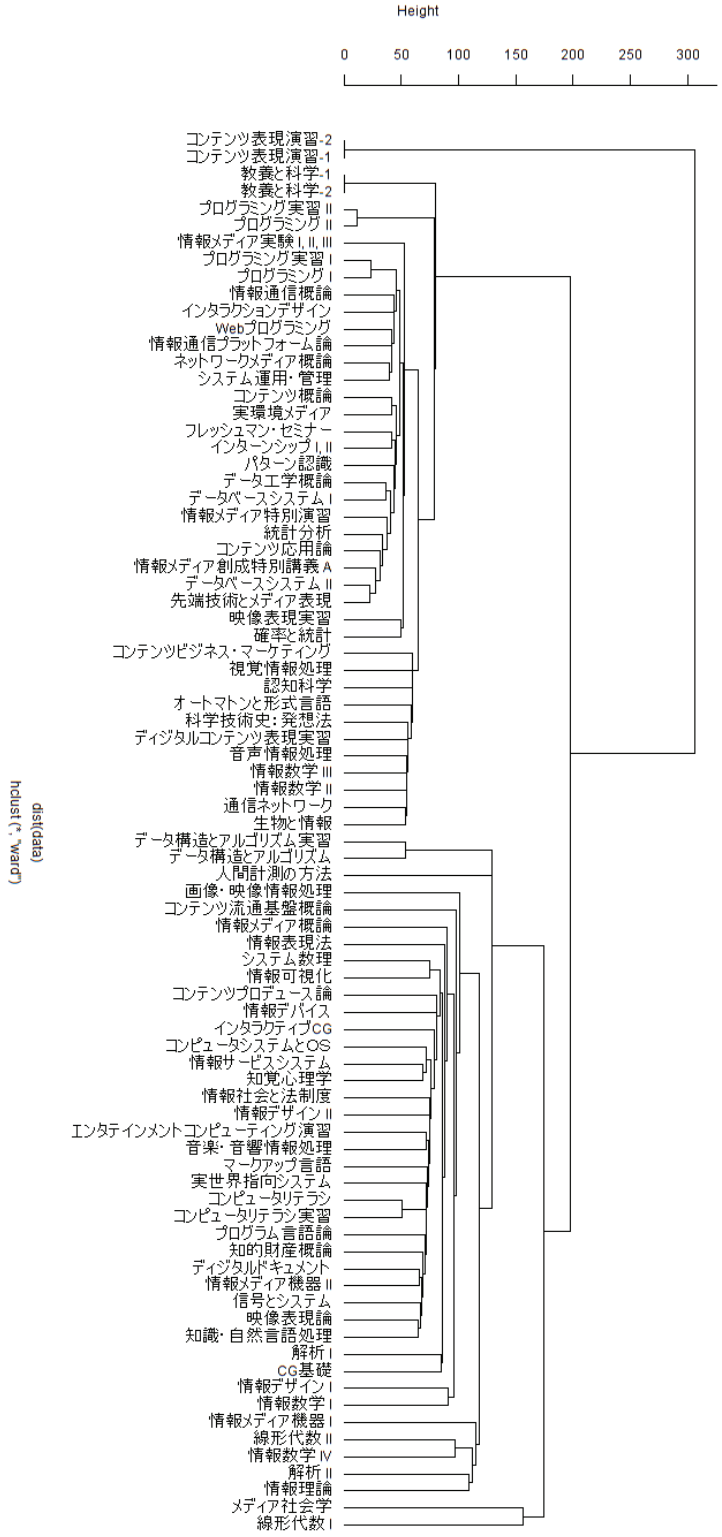

図 4.10　科目単独特徴量を用いたクラスタリング結果（情報メディア創成学類）

4.5　関係性の抽出とネットワーク図を用いた可視化の評価・考察

包摂関係特徴量を用いて科目同士の類似度を算出し，科目間の関係性を抽出し，ネットワーク図で可視化した．情報科学類の結果を図 4.11，知識情報・図書館学類の結果を図 4.13，情報メディア創成学類の結果を図 4.15 に示す．

ネットワークは複数のサブネットワークに分けられ，情報科学類が 8 個（図 4.11），知識情報・図書館学類が 6 個（図 4.13），情報メディア創成学類が 6 個に分けられた（図 4.15）．サブネットワークは内容の似た科目を中心に構成されている．情報科学類（図 4.11）では，英語の科目で構成されたサブネットワーク，情報メディアに関する科目を中心に構成されたもの，論理回路に関する科目を中心に構成されたものが見られる．知識情報・図書館学類（図 4.13）では，英語の科目を中心としたもの，図書館に関する科目を中心としたもの，情報処理に関する科目から構成されるもの，知識資源に関する科目を中心に構成されるものが見られる．情報メディア創成学類（図 4.15）では，数学の科目から構成されたもの，プログラミングの科目を中心に構成されたもの，情報処理の科目から構成されたもの，情報メディアに関する科目を中心に構成されたものが見られる．ネットワークが分けられたのは，三角形や四角形にエッジが連結される箇所ができたからであると考えられる．このサブネットワークは，シラバスがいくつの類似した科目のまとまりから構成されているかを把握するのを助けると考えられる．

多くのノードと連結されている中心的なノードについて見る．情報科学類（図 4.11）では，データ構造とアルゴリズム，確率論，情報システム概論 II，データベース概論 I，線形代数 I が特に多くの科目と連結している．知識・情報図書館学類（図 4.13）では，図書館論，メディア社会文化論，メディア社会論，情報基礎，基礎数学が特に多くの科目と連結している．情報メディア創成学類（図 4.15）では，情報可視化，統計分析，プログラム言語論が特に多くの科目と連結している．これらは，学部・学科の中心的な学問分野や学びの内容を説明をしている．このように，中心的なノードに着目することで，シラバス全体の内容の把握を促進させることができると考えられる．

サブネットワーク単位で中心的なノードについて見る．中心的なノードは，サブネットワーク内の中心的な学問分野や学びの内容を説明している．情報科学類（図 4.11）では，情報メディアに関する科目を中心に構成されたサブネットワークの中心的なノードは，知能情報メディア概論 II である．また，論理回路に関する科目を中心に構成されたサブネットワークの中心的なノードは，論理回路である．知識・情報図書館学類（図 4.13）については，図書館に関する科目を中心としたサブネットワークでは図書館論，情報処理に関する科目から構成されるサブネットワークでは情報基礎，知識資源に関する科目を中心に構成されるサブネットワークでは知識資源組織化論が中心的なノードである．情報メディア創成学類（図 4.15）では，プログラミングの科目を中心に構成されたサブネットワークではプログラミング I，情報メディ

アに関する科目を中心に構成されたサブネットワークでは情報可視化が中心的なノードである．サブネットワーク単位で中心的なノードに着目すると，科目のまとまりの内容を把握することに役立つと考えられる．

ネットワーク内で密な部分同士をつなぐ橋渡し的なノードについて見る．知識情報・図書館学類（図 4.13）では，データマイニングを境に社会学に関する科目のまとまりと情報処理の科目を中心にしたまとまりに分けられる．データマイニングは両方の学問分野に関連性があり，シラバス内で重要な役割を持つ科目であると考えられる．なお，情報科学類や情報メディア創成学類のシラバス分析結果からは，橋渡し的なノードは見られなかった．

ネットワーク分析における包摂関係特徴量の効果を確認するために，科目単独特徴量を用いた結果との間の比較を行った（図 4.11，図 4.12，図 4.13，図 4.14，図 4.15，図 4.16）．包摂関係特徴量を用いた結果，サブネットワークの数は，情報科学類（図 4.11，図 4.12）と情報メディア創成学類（図 4.15，図 4.16）で増加した．これは包摂関係を反映させたことによって，同じ事前履修科目を持つ科目同士の類似度が高まったためだと考えられる．これによって三角形や四角形にエッジが連結される箇所が増えて，それを単位にネットワークが切れたと考えられる．情報メディア創成学類における科目単独特徴量から作成したネットワーク図（図 4.16）では，サブネットワークの数が 3 個ある．そのうちの 1 個は非常に大きなネットワークで複数の学問分野の科目が混在しており，特徴の把握が難しい．情報科学類における科目単独特徴量から作成したネットワーク図（図 4.12）では，英語の科目が他のネットワークに含まれている．包摂関係特徴量を用いた方が，内容の似た科目から成るサブネットワークが構築されやすいと考えられる．中心的なノードについて見る．情報科学類における科目単独特徴量から作成したネットワーク図（図 4.12）では，プログラミング入門 I，システム数理，ソフトウェア概論 II，機械語序論，知能情報概論 II である．これらは，シラバスの中心的な学問分野や学びの内容を説明する重要な科目である．情報メディア創成学類における科目単独特徴量から作成したネットワーク図（図 4.16）では，データベースシステム I，情報可視化，プログラミング I である．データベースシステム I は自身が属する非常に大きく複数の学問分野の科目が混在したサブネットワークの内容を説明するには不十分である．結果としては，中心的な科目として相応しい場合が多く，一部相応しくないものもあった．

クラスタリング分析では個々の科目が一つのシラバスにまとまっていく全過程を見られること，ネットワーク分析ではまとまりの中の中心的な科目を見られることが，それぞれの分析手法における独自の点として特に挙げられる．このことから，互いの特徴は相補的に機能し，これらの図を参考にすることでシラバスの全体像や科目間の関係を把握することが促進されると考えられる．

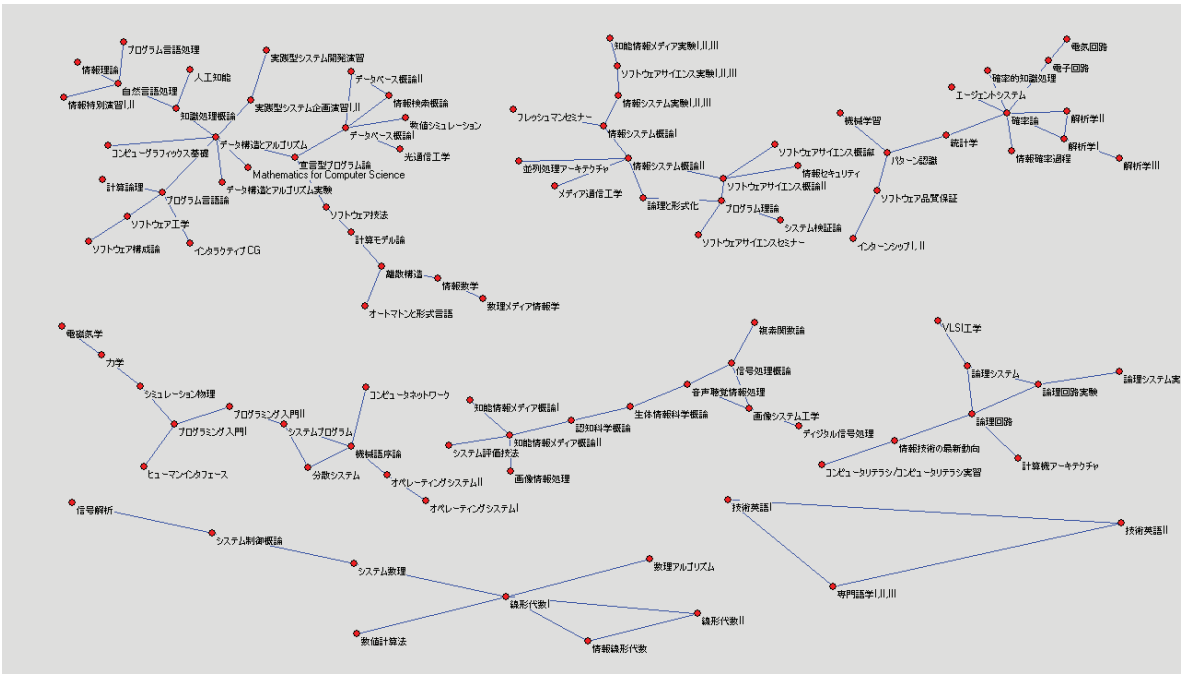

図 4.11　包摂関係特徴量を用いて関係性を抽出した図（情報科学類）

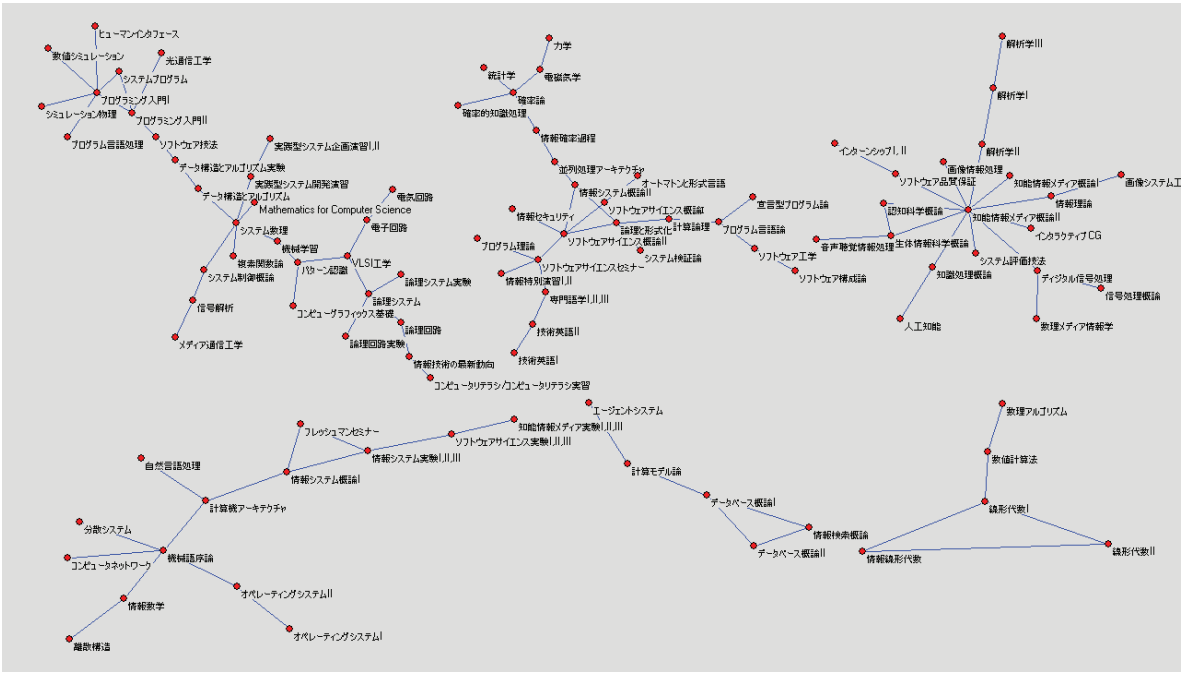

図 4.12　科目単独特徴量を用いて関係性を抽出した図（情報科学類）

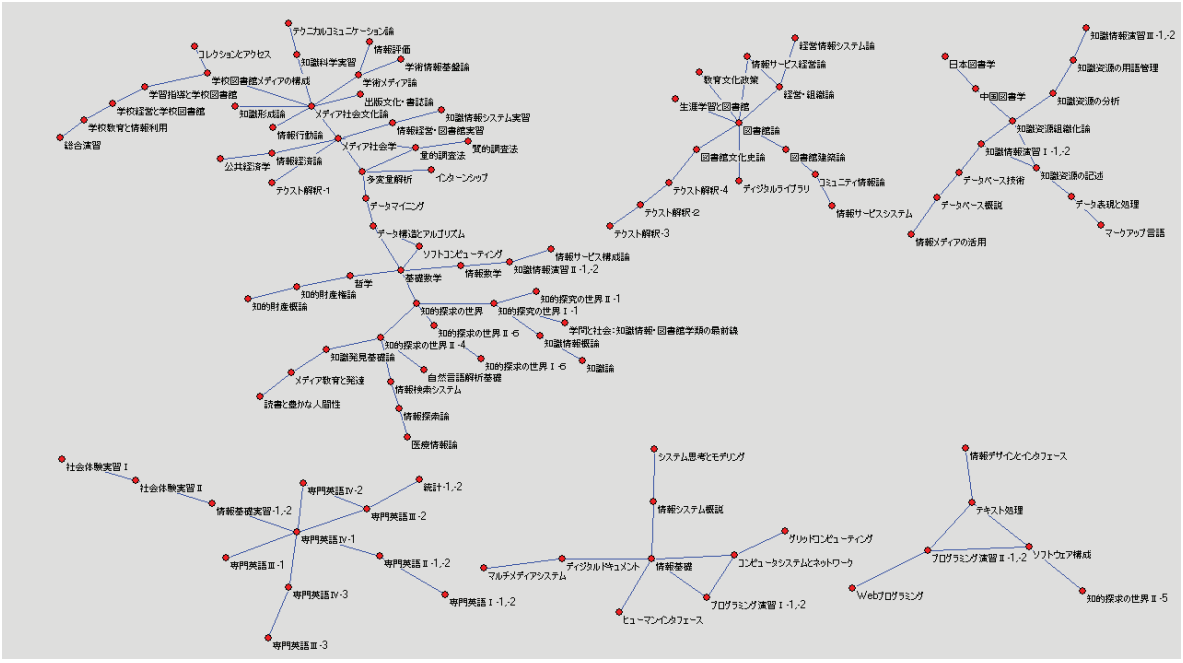

図 4.13　包摂関係特徴量を用いて関係性を抽出した図（知識情報・図書館学類）

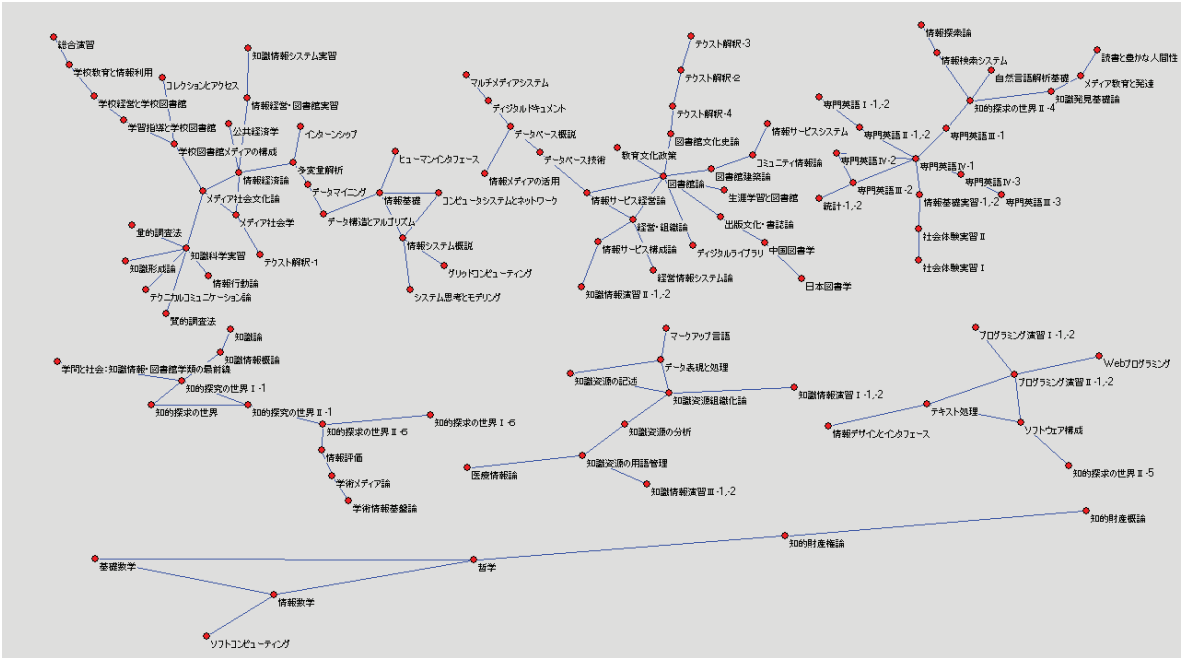

図 4.14　科目単独特徴量を用いて関係性を抽出した図（知識情報・図書館学類）

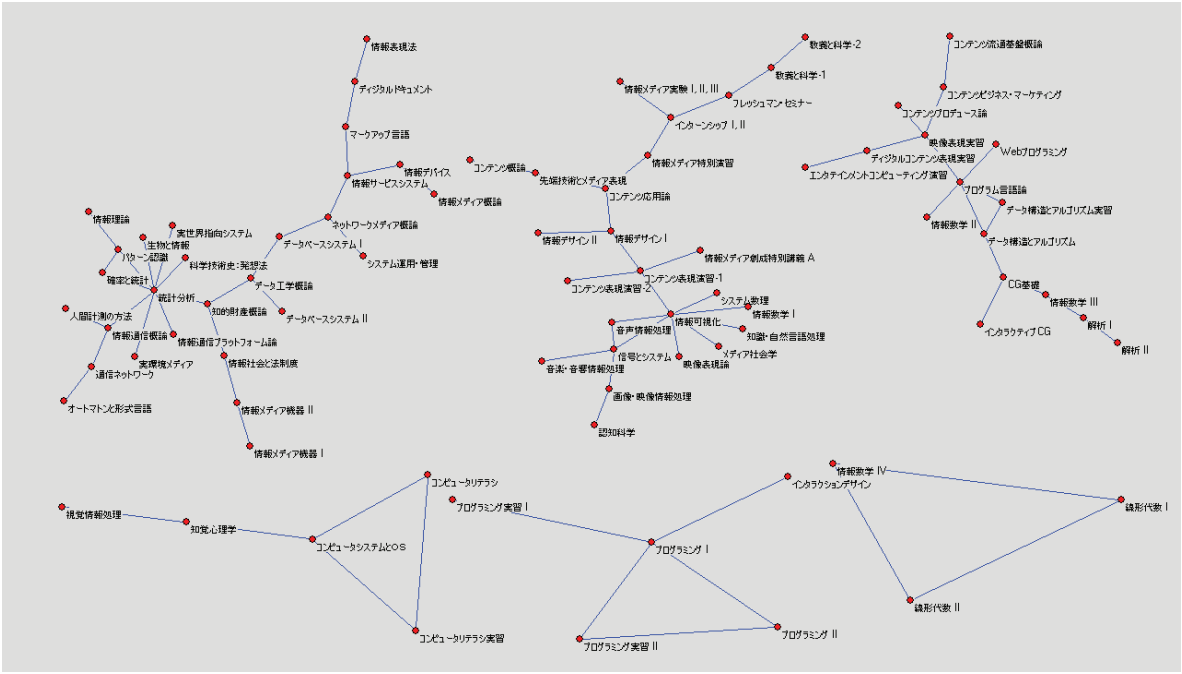

図 4.15　包摂関係特徴量を用いて関係性を抽出した図（情報メディア創成学類）

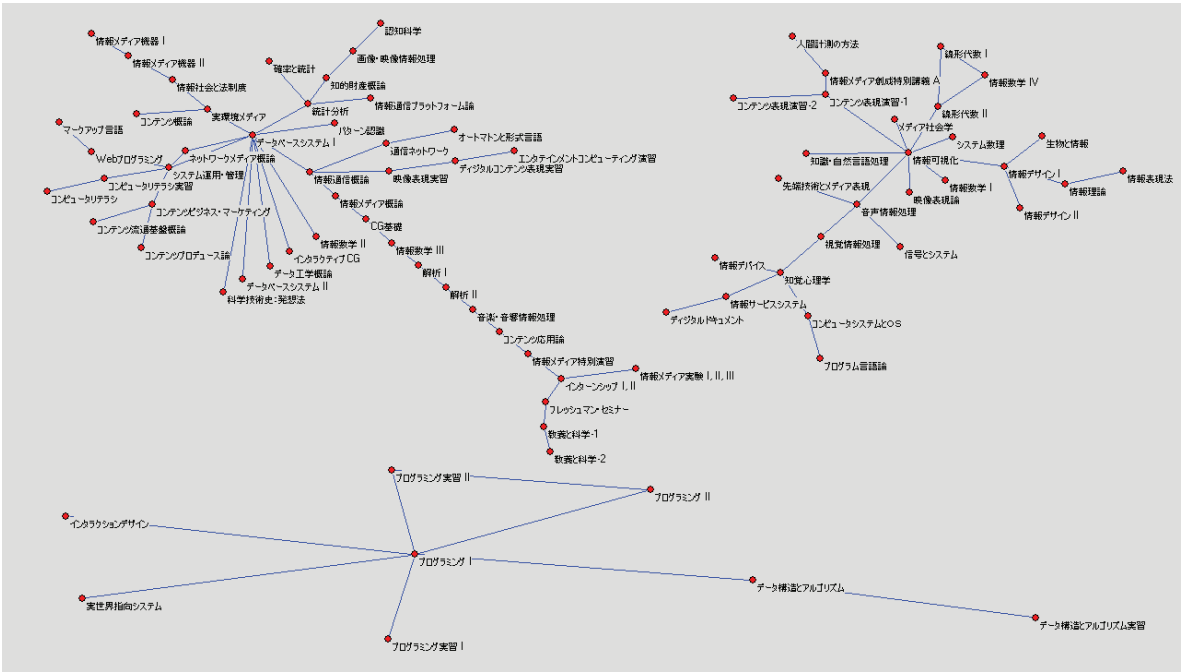

図 4.16　科目単独特徴量を用いて関係性を抽出した図（情報メディア創成学類）

第 5 章

結論

本論文では，科目間の関係性に基づいてシラバスを分析する手法に関する研究を行った．以下で本論文の内容を簡潔にまとめ，今後の展望について述べる．

3 章では，シラバス分析手法の提案として，全体構成とその構成要素であるオープンデータ形成部，特徴量抽出部，分析・可視化部について述べた．外部で公開されている大学のシラバスは組織ごと独自の形式で作成されている．これを統一的に取り扱うためにオープンデータ形成部がある．シラバスを収集し，そこに書かれたテキストを抽出し，データの構造を定義して RDF/XML で記述した．構造化されたデータをセマンティックウェブ要素技術を利用したサーバに格納し，サーバを外部からアクセス可能な状態にし，これをシラバスデータベースとした．

特徴量抽出部では，テキストの状態であるシラバスデータを本研究の分析手法で必要なデータへ変換した．テキストを形態素へ分割し，複合語を作成し，品詞情報や個別の形態素を指定して分析に不要な語を取り除いた．シラバスには事前履修に関する項目があり，ここに科目が書かれているということは，事前履修に関する科目のシラバスに書かれたテキストが暗黙的に含まれていると見なすことができる．このような包摂関係を反映させた形態素リストを作成した．形態素の重みとなる tf・idf 値を計算し，特徴量データベースへと格納した．

分析・可視化部では，シラバスを分析し，その結果を視覚的に表現した．クラスタリング分析では，科目をクラスタリングし，その結果をデンドログラムを用いて可視化した．ネットワーク分析では，科目間の類似度を基に関係性を抽出し，その結果をネットワーク図を用いて可視化した．

4 章では，提案したシラバス分析手法の評価と考察を行った．クラスタリング分析では，まず，クラスタ間距離尺度の比較を行って ward 法の有効性を確認した．次に，デンドログラムを基にクラスタリング分析の評価と考察を行った．特徴が類似している科目から成る多くの小規模クラスタが作られること，個々の科目が階層的にまとまっていく過程が得られることを確認した．また，包摂関係特徴量の効果を確認するために，科目単独特徴量を用いた場合との比

較を行った．包摂関係特徴量を用いることで，類似した科目間の距離が小さくなり，クラスタ間の分離性が向上することを確認した．ネットワーク分析では，ネットワーク図を基に評価と考察をおこなった．特徴が類似している科目を中心に複数のサブネットワークが形成されること，多くの科目と結びつく中心的な科目が存在することを確認した．また，包摂関係特徴量の効果を確認するために，科目単独特徴量を用いた場合との比較を行った．

これらの結果から，本研究で提案するシラバス分析手法を用いることで，シラバスの全体像や科目間の関係を把握する支援ができると考えられる．

今後の展望として，まず，特徴量の改善が考えられる．本研究で導入した特徴量は，クラスタリング分析において，小規模のクラスタを形成する段階では内容が似ている科目を近づけた．しかし，中規模のクラスタを形成する段階では学問分野を表すようなまとまりが得られなかった．そこで，教員名などのシラバス内の他の関係性に着目したり，学問分野に関係があるテキストなどをシラバス外部から導入して新たな特徴量を計算し，それを基にクラスタリング分析やネットワーク分析を行うことが考えられる．

次に，自己組織化マップなどのニューラルネットワークに基づく分析手法の導入が考えられる．自己組織化マップでは，ニューラルネットワークに基づいて多次元の特徴量を持つ科目の関係性を二次元平面上に可視化する．平面上での科目の分布や隣接する科目に着目することが，シラバスの全体像や科目間の関係性を把握するのに有効と考えられる．

さらに，異なる学部・学科間のシラバスを比較することにより力点を置いた分析を行うことへの拡張が考えられる．本研究では，学部・学科ごと個別にクラスタリング分析を行ったが，これを複数の学部・学科を横断する分析へと拡げる．二つの学部・学科の科目が含まれるクラスタに着目することで二つの学部・学科の内容の共通点を，一つの学部・学科の科目のみで構成されるクラスタを発見することでその学部・学科が独自に持つ特徴的な点を抽出できると考えている．

謝辞

本論文は，筆者が筑波大学大学院図書館情報メディア研究科博士前期課程に在籍中の研究成果をまとめたものである．同研究科の佐藤哲司教授には主指導教員として，常に熱心に丁寧に指導していただきました．佐藤先生のご尽力のおかげで，研究を進め修士論文として形にすることができました．多大なる感謝を心より申し上げます．

図書館情報メディア研究科の諸先生方には，講義等で快く質問に答えていただいたことをはじめ，様々な指摘や助言をいただきました．大変感謝しております．特に，鈴木伸崇先生には副指導教員としてご指導いただくとともに，色々と気にかけていただきました．心より感謝致します．また，芳鐘冬樹先生には，シラバスに関する研究をされてきた経験に基づくお話を聞かせていただきました．心より感謝致します．

佐藤研究室の皆さんには，様々な協力，指摘，助言をいただきました．ありがとうございました．

参考文献

[1] 伊東栄典, 山田信太郎, 松永吉広, 廣川佐千男. 国内 web シラバスにおけるレコード抽出に関する一考察. 知識ベースシステム研究会, Vol. 57, pp. 59–64, 2002.

[2] 野澤孝之, 井田正明, 芳鐘冬樹, 宮崎和光, 喜多一. シラバス-専門用語の相互クラスタリングを用いたカリキュラム分析システムの改善. 知能と情報 (日本知能情報ファジィ学会誌), Vol. 17, No. 5, pp. 569–586, 2005.

[3] 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.

[4] 堀幸雄, 中山堯, 今井慈郎. カリキュラムの特徴抽出と時間割の要約生成. 情報知識学会誌, Vol. 20, No. 2, pp. 201–206, 2010.

[5] 山田信太郎, 松永吉広, 伊東栄典, 廣川佐千男. Web シラバス情報収集エージェントの試作. 電子情報通信学会論文誌. D-I, 情報・システム, Vol. J86-D-I, No. 8, pp. 566–574, 2003.

[6] 井田正明, 野澤孝之, 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, Vol. 2, pp. 85–97, 2005.

[7] 芳鐘冬樹, 井田正明, 野澤孝之, 宮崎和光, 喜多一. キーワードの関連用語を考慮したシラバス検索システムの構築. 知能と情報 (日本知能情報ファジィ学会誌), Vol. 18, No. 2, pp. 299–309, 2006.

[8] 高橋和麻, 堀幸雄, 今井慈郎. シラバス解析と活性伝播モデルに基づく履修支援システム. 電子情報通信学会技術研究報告. ET, 教育工学, Vol. 108, No. 247, pp. 57–62, 2008.