Vol.2014-CE-123 No.11
2014/2/9

科学技術論文の二層構造化法

加藤俊弥[†1]　　管村昇[†2]

研究活動を進める上で，関連研究の動向を知るための文献調査は重要な作業であり，それらを支援するための論文検索支援システムが多数提案されてきた．それらのシステムでは，重要な論文やキーワードを提示するに留まっている．どの論文がどういった部分において関連しているかをより具体的に提示するため，本研究では，論文とキーワードをそれぞれ論文層と構成技術層の二層に分けて同一空間上に配置することを提案し，提案法を実装したシステムの評価を行った．

Two-Layer Structuring Method for the Technical Papers

TOSHIYA KATO[†1]　SUGAMURA NOBORU[†]

In promoting research activities, literature search to know the trends of related research is an important work, retrieval support systems for supporting them have been proposed. In these systems, it presents only the important keywords and papers. So In order to present specifically the relevant parts of the papers, In this paper, we propose that keywords and papers be placed in the same space be divided into two layers of construction techniques layer and paper layer, and we were evaluated in the system that implements the proposed method.

1. はじめに

研究活動を進める上で，関連研究の動向を知るために科学技術論文などの文献の検索･調査は重要な作業である．しかし，それら科学技術論文は，複数の専門用語から構成されていることが多く，ある論文を理解するためには複数の専門用語について理解する必要があり，それら専門用語について理解するためには，さらに別の文献を読まなくてはならない場合が多い．これは，特に新たな研究分野に参入する研究者や，初めて研究活動を行う学生にとって大きな負担となる．

その負担軽減のため，論文検索支援システムが多数考案されてきた[1][2][3]．しかしながら，これらのシステムでは，研究経験からの想起が必要なキーワードベースの検索が多い，ある特定の専門用語を理解するために，どの文献を読めば良いか直感的に理解しにくい，文献の重要度の指標に，発表年に依存しやすい被引用件数を利用することが多いなどの課題がある．

本研究では、これらの課題の解決策として，文献を科学技術論文に絞り，キーワードではなく論文の引用情報を入力とし，同一空間上に論文と専門用語，またそれらの接続関係を提示，論文の重要度に専門用語への接続数を利用することを提案する．本稿では，提案手法のための構成技術の抽出方法とそれを実装した検索システムの概要と実験について論じ，手法の有効性を検証する．

1.1 従来の検索支援システムについて

従来の検索支援システムとしてこれまでに，キーワードを入力し検索，抽出された論文をシステム側が分析し，新たなキーワードや重要な論文を提示するシステムが多く提案されてきた[1][2]．しかし，キーワード入力の場合，入力するキーワードによって結果が大きく異なり，適切なキーワードの選択には事前にある程度の知識が必要である[4]．そこで，論文自体を入力とし，その論文と類似・関連した論文を検索するシステムが提案された[3]．論文自体を入力とするため，利用者の事前の知識による影響が少なくなる一方，最初に入力する論文をどのように見つけるかという課題がある．

また，論文の重要度を表す指標として，被引用件数が一般的であるが，古い論文ほど有利に作用することが多い．また，対象となる複数の論文において，被引用件数に差がない場合，重要度を表す指標として利用できない．

1.2 提案する検索支援システムについて

本研究では，従来の検索システムの課題を解決した上で，重要な論文やキーワードを提示するシステムの考案を目的とする．提案するシステムに必要な要件は以下の通りである．

- キーワードや論文に変わり知識に依存しにくい入力
- 重要な論文やキーワードの提示
- 被引用件数以外の重要度の指標の設定

キーワードや論文に変わり事前の知識に依存しにくい入力として，論文の引用情報を入力とする方法を提案する．ソーシャルネットワークなどのネットワークでは，スモールワールド性[5]や六次の隔たりなどと呼ばれる，どのユーザからでもホップ数を 6 前後経由することですべてのユーザに行き着くという考え方がある．論文でも同様に，どのような論文の入力であっても引用のネットワークを経由することで，少なくとも関連する重要な論文には行き着けるのではないかと考えた．実際にはユーザは論文を一つと，

†1†2 工学院大学大学院
Kogakuin University

その引用情報を何次まで参照するかを入力する．また検索結果も，入力する論文を変更せずに，どれだけ引用情報を参照するかを調節することで変更が可能である．図 1 は，論文 A より引用情報を 1 次まで参照した場合と 2 次まで参照した場合の概念図を示している．

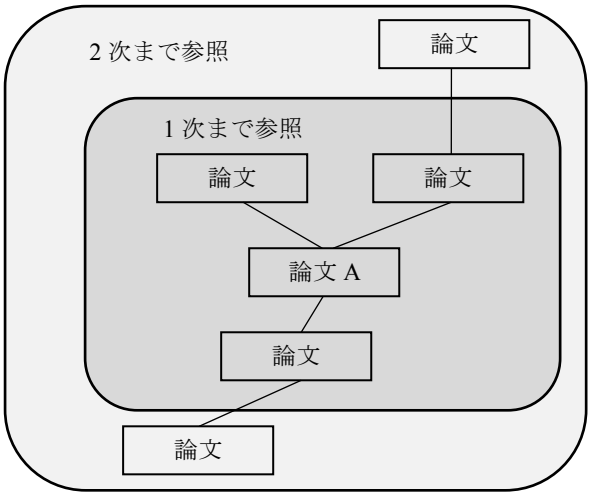


図 1 論文 A を中心とした引用ネットワークと参照次数の概念図

重要な論文やキーワードの提示については，まず本研究では，科学技術論文は暗黙的に図 2 のような多層構造をしていると捉える．最上位の層に論文があり，その論文を複数の構成技術が支えている．その構成技術を複数の学問分野が支えているという構造である．この構造を明確に提示することで，重要な論文とキーワードの間の意味的な繋がりを明示できる他，その背後にある学問分野への繋がりも明らかに出来ると考える．

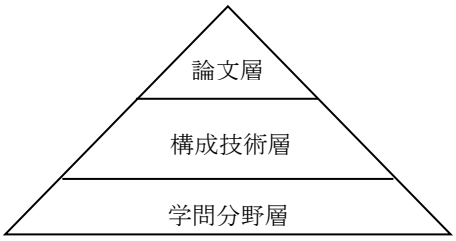


図 2 科学技術論文の多層構造

本研究では，多層構造化の前段階として，図 3 のように論文層と構成技術層の二層に絞り，実際の科学技術論文を構造化し提示をする．論文層は入力した引用情報で繋がった論文群であり，構成技術層は論文層の全論文を分析して得た専門用語群である. 論文内に専門用語が出現する場合，論文と専門用語を接続する．これらを同一空間上に配置することで，例えばある論文 A は構成技術 A,構成技術 B によって構成されていることがわかり，また構成技術 A は，論文 B も構成していることが直感的に理解出来る．このように，どの論文同士がどのような点で似ているかを具体的に提示することができる．

被引用件数以外の重要度の指標としては，構成技術層への接続する数を提案する．構成技術層は，論文層全体を分析して得た専門用語群であり，それら専門用語を多く含む論文はその研究分野にとって重要な論文であると考えられ

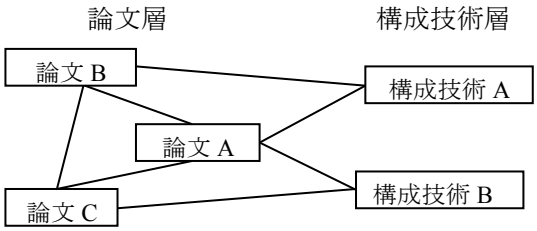


図 3 複数の科学技術論文による二層構造

るためである．また被引用件数と比較すると，新しい構成技術に対しても接続できる点で，新しい論文ほど有利に作用するため，被引用件数では見つけることの出来なかった重要な論文が見つかる可能性も考えられる．

2. 構成技術層の形成

2.1 構成技術の抽出方法の候補

構成技術の抽出方法は，著者キーワードを流用する方法と，アブストラクトから形態素解析を用いて抽出する方法の 2 つが考えられる．前者の場合，著者が論文に合わせて検索用途などを想定して決めた内容であるため，無関係な用語がノイズとして入る可能性はないが，対象論文が著者キーワードの設定されている論文に限られてしまう. また，検索用途のため，詳細な手法の名前などを含まない場合もある．後者の場合，対象論文数は前者と比較すると格段に多くなるが，無関係な用語がノイズとなる他，抽出されない専門用語も出てくると予想される．

本研究では，無関係な用語が抽出されてしまう点などは将来的に改善が可能と考え，対象論文数の多いアブストラクトからの自動抽出を採用した．

2.2 重要度スコア付き専門用語の抽出

文章から専門的な用語を抽出する方法はすでに研究されており，それら先行研究よりスコア付きで専門用語を抽出することが可能な「TermExtract」[6]を利用することとした．「TermExtract」は，専門用語の多くが複合語であることを利用し，形態素解析で品詞単位まで分解した単語を再合成し複合語とした上で，複合語を構成する単名詞の連結回数と，複合語自体の出現頻度を用いて重要度としてスコアを算出する．また，「TermExtract」は，“HMM”などの複数の英単語の頭文字から構成される用語については“HMM”で一つの単語と判定されてしまうため，本研究で利用する際にはアルファベット大文字の数もスコアに追加することとした．実際に文献[7]のアブストラクトを重要度スコア付きで専門用語を抽出した例を表 1 に示す．

2.3 一般性の高い用語の削除(下位フィルタリング)

表 1 を見ると,降順で概ね上位のものほど専門性が高く,下位のものほど一般的であることがわかる．よって，上位 20 位以下を削除する．

表 1 文献[7]のアブストラクトを TermExtract で分析した結果

順位	専門用語	スコア	順位	専門用語	スコア
1	音声認識	11.87	16	能力	1.41
2	言語モデル	7.14	21	除外	1
2	音響モデル	7.14	21	役割	1
4	研究動向	4.12	21	観点	1
5	中心技術	3.46	21	立場	1
6	認識候補	2.89	21	論文	1
6	パターン認識	2.89	21	環境	1
6	認識対象	2.89	21	エントロピー	1
9	モデル	2.83	21	原理	1
10	音声認識能力	2.81	21	人間	1
10	音声認識システム	2.81	21	限界	1
12	次	2	21	多量	1
12	隠れマルコフモデル	2	21	最後	1
14	音声サンプル	1.93	21	改善	1
15	改良研究	1.86	21	削減	1
16	情報理論	1.41	21	処理	1
16	実用化	1.41	21	雑音	1
16	探索空間	1.41	21	機械	1
16	特徴パラメータ	1.41			

2.4 マクロ用語とミクロ用語の分別

構成技術層に，研究分野自体を表すような専門用語が含まれている場合，論文層の多くの論文がその専門用語に接続することが予想でき，接続関係を明示する必要はないため，分別する．

表 1 では，最上位に“音声認識”という研究分野自体を表すような専門用語が出現し，下位に行くほどより該当の論文の中身を表現する専門用語が出現する傾向があることがわかる．しかし，必ずしも最上位に研究分野自体を表す専門用語が出現するわけではないため，下位フィルタのように順位に基づいたフィルタリングをするのは難しい．そこで，引用で接続された他の複数の論文のアブストラクトから抽出した専門用語群をフィルタとして用いて差分を取ることとした．引用で接続された他の論文も，同様に最上位付近には研究分野自体を表現する用語が出現するため，フィルタとして用いることが出来る．また，差分をとる前に，専門用語の順位を保存するため，それぞれの専門用語に 20-順位分のスコアを加算した．差分を取った結果を表 2 に示す．表 2 の結果から，研究分野自体を表す専門用語は負のスコアとなったことがわかる．“隠れマルコフモデル”や“音響モデル”なども負のスコアとなるが，これも“音声認識”と同様に引用で接続された論文において頻出する用語であり，論文と専門用語の接続関係を明示する必要はないため問題ない．この専門用語群から正のスコア上位 10 位の専門用語と,負のスコア下位 2 位の専門用語を抽出し，それぞれをミクロ用語とマクロ用語と呼ぶ．

表 2 表 1 に対し 2 つのフィルタを適用した結果

順位	専門用語	スコア	順位	専門用語	スコア
1	言語モデル	25.14	11	改良研究	6.86
2	研究動向	20.12	12	情報理論	5.41
3	中心技術	18.46	13	探索空間	3.41
4	認識候補	16.89	14	特徴パラメータ	2.41
5	パターン認識	15.89	15	音声認識システム	0.03
6	認識対象	14.89	16	音響モデル	−1.17
7	モデル	13.83	17	隠れマルコフモデル	−4.59
8	音声認識能力	12.81	18	実用化	−17.05
9	次	10	19	音声認識	−96.9
10	音声サンプル	7.93			

2.5 表記ゆれ吸収のためのクラスタリング

フィルタリングにより,各論文につき 10 のミクロ用語が抽出される．この抽出されたミクロ用語に対して，表記のゆれを吸収するために，クラスタリングを行う．クラスタリングには，レーベンシュタイン距離を用いて各用語間の距離を算出し，その距離の近いものをクラスタとしてまとめる．そのクラスタを 20 クラスタ抽出した．レーベンシュタイン距離とは，2 つの文字列に対して挿入・除去・置換をし，どちらかの文字にするために要した回数をコストとして数え，そのコストが少ないほど近い距離の文字列であるとするものである．また，本研究で利用する際には，比較する用語の最大の文字数で正規化した値を用いた．クラスタリング後のミクロ用語を表 3 に示す．

表 3 専門用語のクラスタリング結果

音声認識法,音声認識能力,発声音声認識,音声情報,認識,音声認識実験,音声認識率,音声認識処理,音声認識技術,音声認識方法,音声特徴,音声区間,音声認識方式,音声入力,音声対話,音声言語,音声シフト,音声認識研究,音声認識対象,音声認識誤り,音声認識情報,数字音声認識実験,音声認識関連技術,音声 DB,音声合成,精度認識,音声操作,音声認識器,音声モデル,認識手法,音声認識 LSI,音声分析手法,音声パワー,音声信号,雑音下音声認識実験,音声強調,音声処理,語彙音声認識,表情認識,実環境音声認識,環境音声認識,型音声認識,高性能音声認識,連続音声認識向き,音素認識,連続音声認識
雑音混入ロンバード音声認識手法,雑音混入ロンバード音声,ロンバード音声コーパス,ロンバード音声
音声スタータ,音声インターフェース,規模音声データベース
ロバスト音声認識システム,音声対話システム,混合主導型音声対話システム,音認識システム,音声処理システム
不特定話者単語音声認識技術,孤立単語音声認識,不特定話者認識,特定話者認識,日英シームレス音声認識技術,特定話者単語認識実験,特定話者音声認識,不特定話者用音声認識技術
対話型音声インターフェース,ヒューマンインタフェース,音声言語インタフェース,インタフェース,ユーザインタフェース
擬人化音声対話エージェントシステム,擬人化音声対話エージェント
音声雑音除去手法,雑音除去実験,雑音除去法,音声自動要約手法
連続音声認識コンソーシアム
音声モデル化,言語モデル,変化モデル,雑音モデル,モデル化,音響モデル学習,音響モデル学習法,学習モデル
音声入力インタフェース機能,インタフェース技術,マウス型インタフェース装置
発音分析法,分析窓,分析,特徴分析,発音,発音評価,発音教示,分析手法,分析方法,回帰分析,発音学習,発音分析
単語音声,連続音声,環境音声,単語音声実験,単語 HMM,実音声,言語音認識,学習者音声,母語音声 DB,単語単位,講演音声,雑音音声,単語発声,平常音声,単語区間
音声サンプル,隣接サンプル点間制約,複数サンプル点,音声ダイアル
対数パワースペクトル,スペクトル,パワースペクトル,パワースペクトル領域,ランニングスペクトル
系列パターン認識技術,パターン認識,認識技術,系列パターン,標準パターン,入力パターン,伸縮パターン
音声対話機能,音声対話コーパス,音声対話装置
キーワード音声入力,キーワード検出,キーワード 30 単語,キーワード抽出,カーナビ音声入力
雑音ロバスト音響モデル,HMM 音響モデル,雑音ロバスト性
音声ディクテーション,音声アプリケーション,日本語ディクテーション基本ソフトウェア

3. 自動抽出した構成技術の妥当性の検証

3.1 著者キーワードとの対比実験

3.1.1 目的

論文の内容を表現している著者キーワードとアブストラクトからの自動抽出した専門用語を比較し，著者キーワードがどれだけ含まれているか調査し，構成技術として抽出された用語が妥当であるか検証する．

3.1.2 方法

CiNii(http://ci.nii.ac.jp/)に掲載されている文献から，アブストラクトと著者キーワードの記載のある文献を無作為に 132 論文用意した．この 132 論文の 1 次までの引用ネットワークを入力とし，構成技術を抽出した．入力の元となった論文に含まれている著者キーワードが，自動抽出された構成技術にどの程度含まれているか計算した．

3.1.3 結果と考察

抽出された構成技術中に含まれていた著者キーワードの割合を図 4 に示す．すべての著者キーワードが構成技術に含まれた論文は 132 論文中 6 論文で，著者キーワードの半分(50%)以上が含まれた論文は 132 論文中 27 論文であった．著者キーワードが 1 つ以上構成技術に含まれた論文の数を合計すると，132 論文中 77 論文となり全体の 58%であった．全体として割合が低くなった原因を調べたところ，著者キーワードをそのままアブストラクトに含んでいた論文が少ないことが原因であるとわかった．一方で，著者キーワードと一致した構成技術について調べたところ，マクロ用語の一致度が比較的高く，研究分野自体を表す用語を適切に抽出できていたことがわかった．

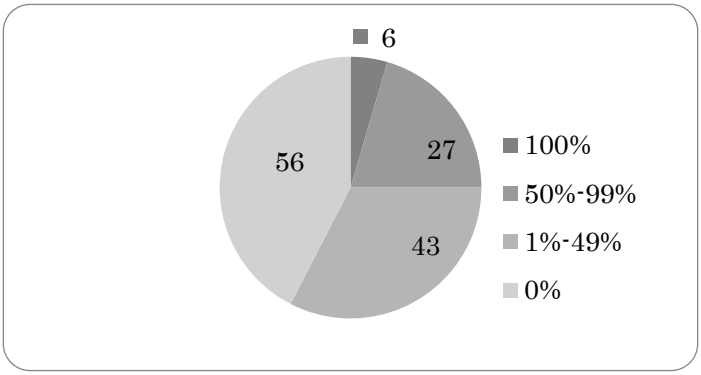


図 4 構成技術中に含まれる著者キーワードの割合ごとの論文数

3.2 専門家による手動抽出との対比実験

3.2.1 目的

著者キーワードとの比較では，アブストラクト中に含まれていないキーワード等のために，妥当性が正確に検証できなかった．そこで，アブストラクトから専門家が手動抽出した専門用語と自動抽出した専門用語を比較し，手動抽出した専門用語内にどれだけ自動抽出した専門用語が含ま

れているか調査する．

3.2.2 方法

専門家 3 名にそれぞれ論文 5 枚，論文 7 枚，論文 15 枚のタイトル,著者，発表日，著者キーワード，アブストラクトを見せ，手動での専門用語の抽出を行わせた．この際，論文 1 つにつき特に重要な用語を最大で 2 つ選び，それ以外に意識すべき用語を無制限で選ぶよう教示を行った．手動での抽出結果と，論文 5 枚，論文 7 枚，論文 15 枚それぞれからの自動での抽出結果を比較し，自動抽出した専門用語が，手動抽出した専門用語にどの程度含まれているか検証した．

3.2.3 結果と考察

結果は,手動で抽出された 124 用語中 61 用語が自動で抽出された．また重要な用語，意識すべき用語の区別なく，どちらも約 50%の用語が自動で抽出できていた．実際に，論文 7 つから手動抽出された専門用語を表 4 に示す．表 4 中の下線の引かれた用語は自動でも抽出されたものである．“サーバ・クライアント方式”などの中黒“・”含んだ複合語は,“･”の部分で分割されてしまっており,「TermExtract」で形態素解析後の複合語を作成する段階で抽出がされていなかった．他の用語については，重要度スコアを元にしたフィルタリングで削除されていたものが多かった．専門家による手動抽出された専門用語の 50%を自動で抽出できたため，精度として必要な条件は満たしていると考える．また実験中，名詞ではなく，動詞などを選出したいという意見もあった．

表 4　専門家による構成技術の手動抽出の結果(下線は自動抽出されたもの)

重要な用語
入力作業効率,マルチモーダルインターフェース,ヒューリスティック関数,A^*ビーム探索アルゴリズム,音声入力,マルチモーダル作図システム S-tgif,音声認識システム,クライアント・サーバ型,住所入力システム,ソフトウェア音声認識,音声対話技術,仮想空間
意識すべき用語
住所入力,操作時間,入力時間,ビーム探索,探索手法,HMM-LR 音声認識,探索アルゴリズム,インタフェース,入力手段,コマンド操作回数,操作労力,HMM-LR 法,連続音声認識システム,マルチモーダル入力ツール,作図ソフト,帳票処理,顧客管理,商品配送,住所候補,サーバ・クライアント方式,住所階層,評価実験結果,単語音声認識機能,認識率,仮想空間技術,ヒューマンインタフェース,音声言語,試着イメージ,画像合成,

3.3 自動抽出された専門用語のノイズの検証

3.3.1 目的

自動抽出された専門用語中に論文の内容との関連性や重要度の低い用語がどれだけ含まれているか検証する．

3.3.2 方法

専門家による手動抽出との対比実験と同じ論文を用いて，専門家 3 名に自動抽出された構成技術内に，専門用語以外のノイズとなる用語を抽出させた．

3.3.3 結果と考察

ノイズと判断されなかった用語は 148 用語中 115 用語で，全体の 77.7%であった．専門家による手動抽出との対比実験では，自動抽出された用語が含まれていた割合が約 50%であったことからも，手動では抽出されなかったが，関連性や重要度の低いとも判断されなかった用語が約 30%あることがわかる．複数の論文に出現し，重要度の低くはないが，手動では抽出されなかった用語が抽出できたことは利点であると考えられる．

4. 開発したシステムについて

4.1 利用した言語とライブラリ

基幹システムは HTML と JScript と JavaScript による HTML Application(以下 HTA)を用いて開発を行った．JScript とはマイクロソフト社製のスクリプト言語で，JavaScript と類似しており，さらに Windows 上でファイル操作が可能な言語である．HTA として開発した経緯は，将来的にサーバー上で動作しブラウザから操作できるようになることを想定しているためである．JavaScript のライブラリとして，jQuery と Raphaël と Dracula と Table Sorter を用いた．jQuery は JavaScript をより容易に記述するために設計されたライブラリで，Raphaël と Dracula は SVG 形式のネットワークグラフを描くために用い, Table Sorter は HTML で組んだ表(Table)にソート機能を付加させるために用いた．CSS では，Normalize.css という見栄えの標準化のためのライブラリを用いた．他に構成技術の抽出用に Perl で動作する TermExtract も利用した．

また各種ライブラリは，本研究で利用する際に内容を書き換えた．TermExtract は，連続での分析を可能とするために，分析終了時に結果をテキストエディタに表示する機能を削除した．Dracula は，ノードの色や形，ラベルの有無，ノードの配置方法，マウスオーバー時の処理を変更した．特にノードの配置方法については，力学モデルを用いた方法で実装されていたが，縦軸方向を固定し，縦軸を時間軸として利用できるように変更した．

4.2 基本構成

ソフトウェアの基本構成は図 5 のようになっており，予め用意した論文関連情報データベースには JSON 形式でタイトル・アブストラクト・著者・発表年月・引用文献・被引用文献が格納されている．

4.3 利用方法と機能紹介

ユーザは以下の手順を用いて，システムを利用する．

1) ユーザは予め自分の研究に関連する分野の論文を用意する．この論文の選択自体は重要ではなく，関連する研究分野であれば良い．

2) この論文の引用情報を CiNii におけるアドレスで入力する．このアドレスをキーに内部データベースで引用情報の検索が行われる．分析が完了次第，図 6 のように結果が表示される．
3) 各ノードの詳細情報や表を確認し，重要な論文やキーワードを理解する．
4) ユーザが論文層に重要な論文が表示されていないと判断した場合は，入力する論文を選びなおす必要はなく，引用を参照する数を増やすだけでよい．これにより，抽出される論文が増え，構成技術も変化する．逆に論文層に論文が多すぎる場合には，引用を参照する数を減らせばよい．

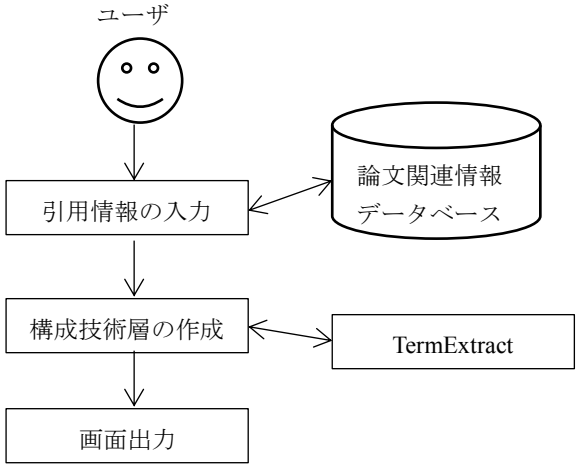
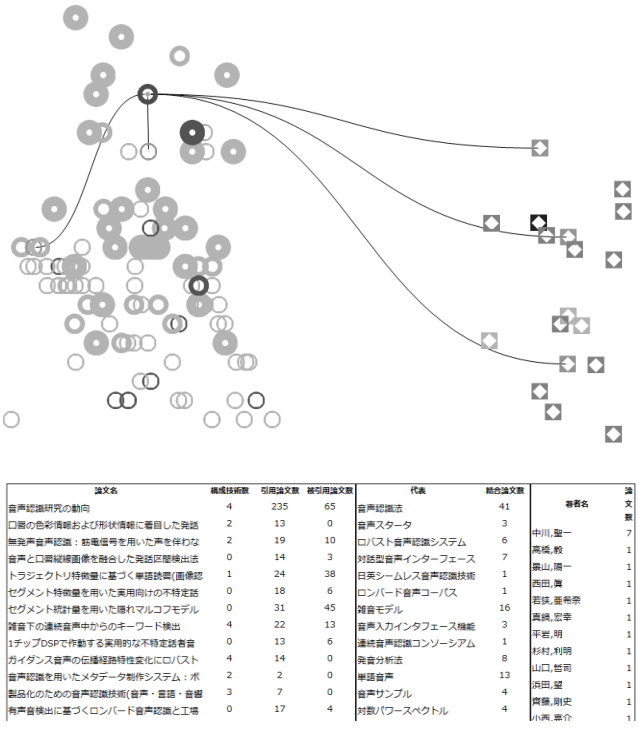


図 5 本システムの概略図



論文名	構成技術数	引用論文数	被引用論文数
音声認識研究の動向	4	235	65
口唇の色彩情報および形状情報に着目した発話	2	13	0
無発声音声認識：筋電信号を用いた声を伴わな	2	19	10
音声と口唇縦線画像を融合した発話区間検出法	0	14	3
トラジェクトリ特徴量に基づく単語読唇(画像認	1	24	38
セグメント特徴量を用いた実用向けの不特定話	0	18	6
セグメント統計量を用いた隠れマルコフモデル	0	31	45
雑音下の連続音声中からのキーワード検出	4	22	13
1チップDSPで作動する実用的な不特定話者音	0	13	6
ガイダンス音声の伝播経路特性変化にロバスト	4	14	0
音声認識を用いたメタデータ制作システム：ポ	2	2	0
製品化のための音声認識技術(音声・言語・音響	3	7	0
有声音検出に基づくロンバード音声認識と工場	0	17	4

代表	結合論文数
音声認識法	41
音声スタータ	3
ロバスト音声認識システム	6
対話型音声インターフェース	7
日英シームレス音声認識技術	1
ロンバード音声コーパス	1
雑音モデル	16
音声入力インタフェース機能	3
連続音声認識コンソーシアム	1
発音分析法	8
単語音声	13
音声サンプル	4
対数パワースペクトル	4

著者名	論文数
中川,聖一	7
高橋,敏	1
景山,陽一	1
西田,眞	1
若狭,亜希奈	1
真鍋,宏幸	1
平岩,明	1
杉村,利明	1
山口,哲司	1
浜田,望	1
齋藤,剛史	1
小西,章介	1

図 6 出力後の画面

図 6 では，左側は論文層のノードであり，右側は構成技術層のノードである．論文層ノードは，構成技術層ノードと接続数により色が変わり，被引用論文数によって大きさが変わる．また，別途，論文情報，構成技術情報，著者情報が表としてまとめられている． 図 7 にノードにマウスを乗せた時の処理を示す. 各ノードにマウスを合わせると，そのノードから伸びているエッヂが表示される他，連結しているノードの色が変わり，マウスを合わせているノードの詳細情報が表示される．ノードの詳細は，論文ノードではタイトル，著者，発表年月，アブストラクト，引用件数などが表示され，構成技術ノードでは，そのノードに含まれている構成技術群の一覧が表示される．

図 8 に表にマウスを乗せた時の処理を示す．表の各行にマウスを合わせた場合も同様に，エッヂの表示と連結ノードの色変更，詳細情報が表示される．

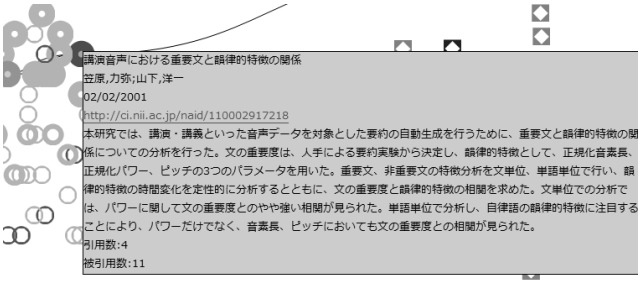
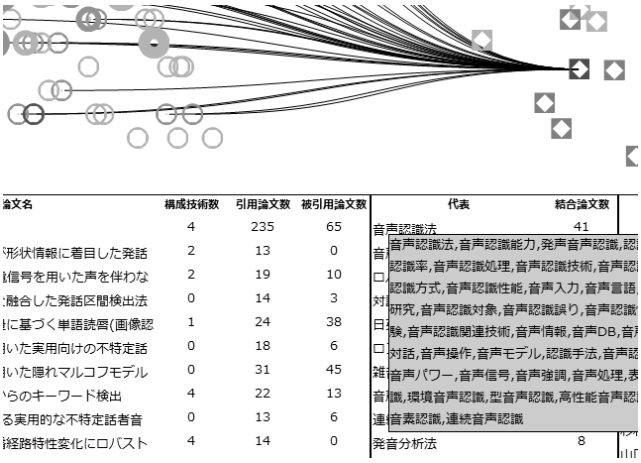


図 7 ノードにマウスを乗せたときの表示



論文名	構成技術数	引用論文数	被引用論文数
	4	235	65
形状情報に着目した発話	2	13	0
信号を用いた声を伴わな	2	19	10
融合した発話区間検出法	0	14	3
に基づく単語読唇(画像認	1	24	38
いた実用向けの不特定話	0	18	6
いた隠れマルコフモデル	0	31	45
らのキーワード検出	4	22	13
る実用的な不特定話者音	0	13	6
経路特性変化にロバスト	4	14	0

代表	結合論文数
音声認識法	41
音声認識法,音声認識能力,発声音声認識,認	
認識率,音声認識処理,音声認識技術,音声認	
認識方式,音声認識性能,音声入力,音声言語	
研究,音声認識対象,音声認識誤り,音声認識	
験,音声認識関連技術,音声情報,音声DB,音	
対話,音声操作,音声モデル,認識手法,音声認	
音声パワー,音声信号,音声強調,音声処理,表	
識,環境音声認識,型音声認識,高性能音声認	
音素認識,連続音声認識	
発音分析法	8

図 8 表にマウスを乗せたときの表示

5. ユーザビリティの検証

システムの評価として，研究経験の浅い学生 6 名を対象に，検索システムを一通り操作した後，改善点や今後利用したいかを尋ねるアンケートを実施した．ユーザビリティに関して，表 5 のような意見を得た．また，このようなシステムを利用したいか，使い勝手が改善されたら利用したいか，利用する必要はないか，を聞いたところ，4 名が利用したいと答え，2 名が使い勝手が改善されたら利用したいと答えた．

表 5 本システムの使い勝手に関するアンケート結果

良かった点
論文と構成技術の繋がりが視覚的に理解しやすい
論文の発表年を意識しやすい
被引用数の多い論文以外も探しやすい
引用次数を操作するだけで，検索結果が変わる
悪かった点
マウスを乗せたときの詳細情報が見難い
小さなディスプレイだと操作しにくそう
ノードが重なったとき，操作がしにくい
画面上にもっと説明があるとよかった

6. 結論及び今後の課題

重要な論文やキーワード等を提示するための検索支援システムの構築を目指した．その際の提示方法として，論文の多層構造化を提案した．本研究では，多層構造化の内の論文層と構成技術層に絞り，論文層と構成技術層の二層構造による研究を行った．検索の際の入力に引用情報を用いることで，事前の知識の影響を受けずに，関連する重要な論文を検索できると考え，用いた．構成技術の抽出は，アブストラクトに「TermExtract」を用いて得た専門用語とスコアをフィルタリングして行った．構成技術の妥当性評価実験では，著者キーワードとの比較では，著者キーワードを 1 つ以上含む場合は全体の 58%であり，最低限の精度を確認し，また専門家による手動抽出された構成技術の約 50%を自動抽出できることを確認した．また自動抽出された構成技術の 77.7%は専門性のある用語と判断され，構成技術層の形成のための自動抽出は必要な精度を備えているといえる. これらの機能を実装したソフトウェアを HTA で開発を行った．使い勝手に対する否定的な意見はあったものの，システムとそのコンセプトには肯定的であった．今後の課題として，本研究では二層構造までであったため，学問分野層も含めた構造化することが最も重要であると考える．その他に構成技術抽出時の精度を高めるため，フィルタリングの方法の変更・フィルタリング時のパラメータの再設定もする必要がある.

謝辞 本研究を進めるに当たり，蒲池みゆき准教授と橘完太准教授には，貴重なご意見や実験に協力いただいたことを深く御礼申し上げます.

参考文献

[1]杉本 雅則, 小山 照夫, 掘 浩一, 大須賀 節雄, 絹川 博之, 間瀬 久雄. 文書間の関連性を可視化することによる文献検索システム, 情報処理学会研究報告, 1996-NL122-3, p15-p22(1996)

[2] 謝 英双, 三末 和男, 田中 二郎. キーワードの頻度推移と文献の被引用数を視覚化した文献検索ツール. 情報処理学会全国大会, 2010

[3]鈴木 雅人. リッチインターフェースを備えたグラフィカル論文検索支援システム,情報処理学会研究報告,2008-HCI-127,p87-p91(2008)

[4]R.N.Oddy : Information Retrival through Man-Machine Dialogue, Journal of Documentation, 33(1),pp.1-14,(1997)

[5]mixi のスモールワールド性の検証 http://alpha.mixi.co.jp/2008/10643/ (2014/1/13 アクセス)

[6]TermExtract http://gensen.dl.itc.u-tokyo.ac.jp/ (2014/1/13 アクセス)

[7]中川 聖一, 音声認識研究の動向, 電子情報通信学会論文誌, Vol.83-DII, No.2(2000)