

教員データに対する多面的検索システム(eラーニング2, 夏のデータベースワークショップDBWS2005)

廣川 佐千男
九州大学情報基盤センター

関 隆宏
九州大学大学評価情報室

安元 裕司
九州大学大学院システム情報科学府

Yamada Yasuhiro
Graduate School of Information Science and Electrical Engineering, Kyushu University

<https://hdl.handle.net/2324/1301129>

出版情報：情報処理学会研究報告．データベース・システム研究会報告．2005（68），pp.665-672，2005-07-14．Information Processing Society of Japan (IPSJ)

バージョン：

権利関係：

教員データに対する多面的検索システム

廣川 佐千男* 関 隆宏† 安元 裕司‡ 山田 泰寛‡

概要ベクトルモデルに基づく文書検索と、XML に代表される半構造化文書の検索技術の融合は、検索技術についての研究として最もホットなテーマである。本稿では、教育研究活動概要や研究成果一覧などの複数の項目からなる教員データを具体的対象として、それぞれの項目を検索の観点として捉える多面的検索方式を提案する。各文書のそれぞれの項目を独立した文書として捉え、個別インデックスを構成することにより、項目毎のクラスタリングを実現する。利用者は、複数の項目から任意の二項目を選択することにより、2つの観点をを用いたクラスタリング結果を見ることが出来る。さらに、各項目のクラスタから抽出した特徴的キーワードを2つの軸に表示することにより、単純な可視化ではなく、可視化画面においてそれぞれのセルを二種類のキーワードで解釈することができる。このキーワードは、検索結果全体の概観を与えるだけでなく、検索絞込みにも利用でき、対話的な検索の効率が向上できる。

Faceted Search Engine for Activities of Academic Staff

Sachio Hirokawa* Takahiro Seki† Yuji Yasumoto‡ Yasuhiro Yamada‡

Abstract. The vector model information retrieval system and XQuery are two key technologies for information retrieval. Integration of this two methods is one of the big challenges in next generation search engines. This paper proposes a faceted search engine for research and development activities of university professors. The targets are the documents of their activities which consists of outline of research, classes they teach, listing of publications etc. Each components of documents are considered as separate documents and have different indices. These indices realize several different point of view for search. The search results are shown as a two dimensional map. Each axes designates a point of view, and search results are clustered according to each component. This map gives a birds-eye view as well as keywords for narrowing query in each component.

1 はじめに

小説のような文学的な文章とは違い、事務的な文書は一つの文書であっても論理的な構造を持ち、その構造を表す複数の項目から構成される。また、一つの組織内では、それぞれの記録内容に応じて書類は同一様式に統一され、それらが多量に蓄積されて

いる。将来的にはXML文書として蓄積されるようになると予想される。このような半構造化文書に対する検索技術は重要な課題と認識され、構造とコンテンツの両方の観点から様々な検索方式の提案がなされている [19, 13, 4, 3, 14, 6, 9, 10, 5]。

XQuery に代表される構造化文書の検索においては、文書の各項目についての検索条件を表すことで検索要求を表現し、得られるXML文書のリストから必要な項目部分を抽出することで検索結果を構成する仕組みを与えている。一方、現在の検索エンジンの多くは、文書をキーワードの集合 (bag of words) と見なし、tf*idf のようなスコアを用いる

*九州大学情報基盤センター・Computing and Communications Center, Kyushu University

†九州大学大学評価情報室・Office for Information of University Evaluation, Kyushu University

‡九州大学大学院システム情報科学府・Graduate School of Information Science and Electrical Engineering, Kyushu University

ベクトル・モデルにより高速な検索を実現している。キーワードとして表される検索要求にマッチする文書が多量にあっても、文書スコアのランキングを用いて上位の結果のみを表示することで使い易さを実現している。検索結果の表示方法には、リスティング、ランキング、クラスタリング、ディレクトリ表示などの工夫がなされてきた。これらの技術は大量の検索結果を少ない画面上に表示することで、利用者の検索作業の効率を上げることが目的である。

ベクトル・モデルによる単純文書検索と、XML検索という二つの異なる検索技術をどのように融合するかというテーマについて、現在、多くの提案がなされている [7, 8, 1, 2, 3, 10, 11]。しかしその多くは、入力(検索要求)における半構造的な利用であり、検索結果の出力について、半構造的なランキングを融合する研究は少ない。

一つの文書が複数の項目を持ち、それぞれの側面から絞込みを行なうために、複数の独立したカテゴリ体系を設定するファセット分類と呼ばれる手法が注目されている。本稿では、項目に応じた複数の観点からの検索結果の表示することにより、ファセット分類を検索エンジンとして実現する方式を提案する。具体的には、各文書のそれぞれの項目を独立した文書として捉え、個別インデックスを構成することにより、項目毎のクラスタリングを実現する。利用者は、複数の項目から任意の二項目を選択することにより、2つの観点を用いたクラスタリング結果を見ることができる。さらに、各項目のクラスタから抽出した特徴的キーワードを2つの軸に表示することにより、単純な可視化ではなく、可視化画面においてそれぞれのセルを二種類のキーワードで解釈することができる。このキーワードは、検索結果全体の概観を与えるだけでなく、検索絞込みにも利用でき、対話的な検索の効率が向上できる。

ここではまず、検索の絞込みという観点からベクトル・モデルによる検索の問題を分析する。ここでは、キーワードなどの検索条件に基づいて、検索結果を1次元で表示したり、クラスタリングによって膨大な文書データを複数のカテゴリに分類した表示が一般的に使われている。しかし、これらの技術では、指定されたキーワードなどの検索条件に基づく検索結果や、クラスタリングによって分類された結果が単に表示されるだけであるた

め、キーワードに関する文書データの傾向を容易に捉えることができない。その結果、検索条件を何度も繰り返し指定しなければ膨大な文書データを絞り込むことができない。また、検索条件を何度も指定する過程で変化する検索条件の指定内容によって検索結果が大幅に異なってしまうこともあり、所望の文書データを容易かつ正確に検索することを困難にしている。検索システムの利用者が所望の結果を得るまでに、次の2ステップを繰り返すと考えられる。

1. 検索システムに検索キーワード(あるいは検索文)を入力し、検索結果を得る。
2. 1で得られた各結果をユーザ自身が見てその良し悪し、つまり所望の検索結果であるか否かを判断する。

検索システムに求められているのは、第1ステップで利用者が求める結果「だけ」が結果として与えられることである。しかし、利用者の検索要求が曖昧だったり、ほんの2~3個のキーワードで求めるものを厳密に特定することにはそもそも限界があり、第1ステップだけで完璧な検索を終了することは不可能である。そのように考えると、第2ステップが重要である。第2ステップを容易にする、あるいは支援する検索システムが強く求められているといえる。ランキングやクラスタリングは第2ステップ効率化といえる。

本研究が対象とする教員データの文書は、事務的な文書の例と考えられる。事務的な文書は、小説などとは違い、文章が線形に並ぶだけでなく、各々の項目から構成される論理的構造を持つという特徴がある。利用者は、その項目ごとに多面的に捉える。しかし、既存の検索システムでは、文書を一面的にしか捉えていない。つまり、第2ステップで、項目に応じた複数の観点から判断することはできない。XMLのように構造を持つ文書については、多面的検索への支援が必要である。本稿では、この多面的検索システムを新たに提案する。多面的検索システムは複数の項目がデータにあれば、実装可能である。多面的検索システムにおいて、理論上は検索結果を n 次元で表示することができるが、人間が見るといふことを考慮すると2次元表(マップ)で表示するのが適当である。本稿ではそ

の実装例として、筆者らが開発している教員データベース [15] の所属・職名・研究紹介部分・研究業績部分を属性とする教員マップ作成システムについて述べる。

2 多面的検索システム

既存の検索システムは文書を一つの重み付きのキーワード集合とみなし、一つの観点からの検索結果を表示する。しかし、構造的文書であれば、利用者は各項目ごとに複数の視点を持っている。視点を切替えて検索結果を眺め分析し、検索の絞り込みや新たなキーワードによる検索のやり直しができるれば、検索の効率向上が期待できる。最も簡単な多面的検索は2方向から見ることである。検索結果を2方向からの表示方法として2次元表が挙げられる。教員データの場合、大学の基礎情報として全教員を所属と職からなる2次元表として示すことが多いが、この2次元表はその最も単純な例である。この場合、結果となる2次元データ(あるいは2次元表)をあらかじめ作成しておき、それにアクセスさせるのが情報検索における普通の方法である。ここで注意したいのは、2次元データ(2次元表)の作成には時間と労力がかかる点と、「静的に」データ編集を行っている点である。文書項目、つまり対象の属性が複数、例えば n 個、ある場合には、理論上は n 次元表を返す多面的検索システムが考えられる。しかし、人間が容易に認識できるのは高々 $n=3$ までであり、より現実的には2次元までである。そこで、提案する多面的検索システムでは、複数の項目から二つの項目を選択することにより2次元表を返す。

多面的検索システムによる検索は、2方向から(見かけ上は)同時に検索して2次元表の形で検索結果を表示するというものである。これは、2つの観点(属性)から検索することにより、絞り込み作業を容易にすることに大きな特徴がある。特にデータ数が多ければ多いほど有効になると考えられる。また、検索から離れた場面でも、複数の属性を持つデータを分析するとき2次元表はしばしば用いられる。したがって、検索結果を2次元表で示すことにより、複数の属性を持つデータの分析を動的に行えるという利点もある。

一つ項目を決め、検索結果の文書の該当項目を抽出した部分文書を考え、クラスタリングすることで縦軸についての配置が決まる。もう一つの項目を横軸として、検索結果をクラスタリングすることで、一つの文書を縦横の2つの観点から分類することができる。このクラスタリングには、前述の「職種」によるクラスタリングのように「静的」なものだけでなく、ベクトルモデルによる文書の類似度を用いた「動的」クラスタリングもある。

3 GETA とクラスタリング

本節では、多面的検索システムの実装で利用した GETA とクラスタリングについて述べる。

汎用連想計算エンジン (Generic Engine for Transposable Association; 以下、GETA と呼ぶ)¹ は、文書検索における頻度付き索引データを行列化したもの (Word-Article-Matrix; WAM と呼ばれる) を対象に、行と行あるいは列と列 (具体的には文書間および単語間) の類似度を高速計算するツールで、国立情報学研究所で開発された。GETA には大規模文書における連想検索、文書分類、単語間類似度計算などの大規模文書分析に必要な要素技術がサポートされている。したがって、GETA の利用により、疑似文書データを基にしたクラスタリングを高速に行うことができるので、「動的に」クラスタリング結果を得ることが可能になる。

一般に文書集合を複数のカテゴリに分類する処理はクラスタリングと呼ばれる。文書検索におけるクラスタリングには2つの可能性があると考えられる。一つは各文書がどのクラスタに属しているかあらかじめ決まっているもので、もう一つは検索要求によって属するクラスタが変化するものである。本稿ではこれらをそれぞれ「静的クラスタリング」、「動的クラスタリング」と呼ぶことにする。

また、クラスタリングの手法は階層型と非階層型に大きく分類されているが、GETA においては前者が実装されている。そして階層型クラスタリングのアルゴリズムとしては、単一リンク法、完全リンク法、Ward 法が代表的なものである [18] が、本システムでは確率的クラスタリングを採用

¹<http://geta.ex.nii.ac.jp/>

する。なお、これらはすべて GETA で実装されている。確率的クラスタリングを採用したのは、他のクラスタリングアルゴリズムと比べてクラスタごとの構成要素数のばらつきが小さかったためである。確率的クラスタリングについては [12] を参照のこと。

4 教員マップ作成システム

本節では、多面的検索システムの実装例として、教員マップ作成システムの構築について述べる。構築した教員マップ作成システムは九州大学の教員マップ作成を行うものであり、これにより検索要求に対する教員の分布状況が2つの軸から明らかになる。本システムの構成は図1の通りである。

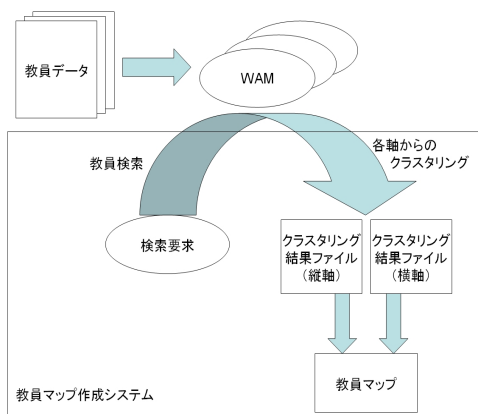


図 1: システムの構成

4.1 データの作成手順

ここで用いるデータは2005年5月9日現在の「九州大学研究者情報」²中の「所属名」「職名」, 「研究・教育・社会活動概要」と「研究業績」である。なお、このデータは全部で2193人分ある。「研究・教育・社会活動概要」は、各教員が自分の研究内容や教育内容、社会連携活動について述べたもので、基本的に日本語で書かれている。「研究業績」は、各教員が「九州大学研究者情報」で公開を認めた学会発表や原著論文、著書のタイトルを用いている。外国語で書かれた論文等は外国語のま

²<http://hyoka.ofc.kyushu-u.ac.jp/search/>

ま扱う。これらのデータ抽出が可能になっているのは、「九州大学研究者情報」のデータ源になっている「九州大学大学評価情報システム」のデータ形式としてXMLを採用しているからである [15]。これらのデータを形態素解析器「茶筌」³にかけ、名詞と品詞が分からなかった語（外国語や専門用語がほとんどである）だけを残し、それらの出現頻度を求め、WAMを作成する。

4.2 システムの内容

3節で述べたように、クラスタリングには「静的クラスタリング」と「動的クラスタリング」の2種類がある。本システムでは、「所属名」と「職名」に基づくクラスタリングが「静的クラスタリング」で、「研究・教育・社会活動概要」と「研究業績」に基づくクラスタリングが「動的クラスタリング」である。従って、2つの軸の特性として、クラスタリングの種類に応じて次の3つの組み合わせが可能となる。

- (a) 両軸ともに「静的クラスタリング」
- (b) 片方の軸は「静的クラスタリング」、他方の軸は「動的クラスタリング」
- (c) 両軸ともに「動的クラスタリング」

ここで、「静的クラスタリング」の軸の項目は常に決まっているが、「動的クラスタリング」の軸の項目は検索要求によって変動する。本システムで片方の軸を「職名」、他方の軸を「研究業績」とすれば、これは上記 (b) の場合に該当する。したがって、「職名」の軸はたとえば「教授」「助教授」「講師」「助手」というあらかじめ決まっているカテゴリが表示されるのに対し、「研究業績」の軸は検索要求にともなって変化する。なお、本システムでは「動的クラスタリング」の軸の項目としては、各クラスタの特徴的な語10語が示されている。

そこで、本システムにおけるそれぞれの場合のマップ作成手順を示す。

- (a) 両軸ともに「静的クラスタリング」
 1. 全教員が各軸のどのクラスタに属するか求める。

³<http://chasen.aist-nara.ac.jp/chasen/doc/chasen-2.2.1-j.pdf>

2. 各マスに該当する教員名や教員数を求める。
- (b) 片方の軸は「静的クラスタリング」、他方の軸は「動的クラスタリング」
1. 検索キーワード(あるいは検索文)を入力する。
 2. 2つの軸(項目の構成要素)とクラスタ数を決定する。(「静的クラスタリング」を行う軸の方はクラスタ数の指定は無視される。)
 3. 「動的クラスタリング」を行う軸の属性に基づいて教員検索を行う。
 4. 上記3の結果をもとに「動的クラスタリング」を行う軸の方向からクラスタリングを行い、その結果をファイルに記入する。
 5. 上記3の結果をもとに「静的クラスタリング」を行う軸の方向からのクラスタリングを行い、その結果をファイルに記入する。
 6. 上記4と5により得られるファイルから各マスに該当する教員名や教員数を求める。
 7. 各マスの教員データから「動的クラスタリング」を行う軸の方向から見た場合の重要語を取り出す。

(c) 両軸ともに「動的クラスタリング」

1. 検索キーワード(あるいは検索文)を入力する。
2. 2つの軸(項目の構成要素)とクラスタ数を決定する。
3. 2つの軸の属性に基づく検索を行い、少なくとも一方に出現する教員を求める。
4. 上記3の結果をもとに縦軸方向からのクラスタリングを行い、その結果をファイルに記入する。
5. 上記3の結果をもとに横軸方向からのクラスタリングを行い、その結果をファイルに記入する。

6. 上記4と5により得られるファイルから各マスに該当する教員名や教員数を求める。
7. 各マスの教員データから縦軸方向に見た場合の重要語ならびに横軸方向に見た場合の重要語を取り出す。

4.3 画面例

ここでは「学習」を検索要求とし、縦軸に「研究・教育・社会活動概要」について4つのクラスタ、横軸に「研究業績」について4つのクラスタを生成することにより作成される教員マップ作成に関わる画面を図2から図5に示す。

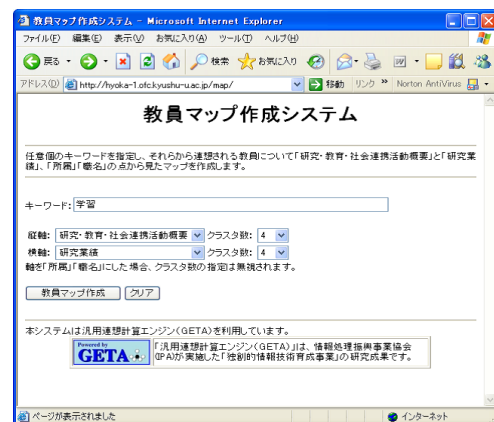


図 2: 入力画面

5 比較実験

本節では同じ教員データを用いた教員検索システムの比較を通じて、多面的検索システムの特徴を考える。比較対象となるシステムは

- (a) 「九州大学研究者情報」の詳細検索
- (b) 「九州大学教員連想検索システム」
- (c) 本システム

の3つである。

(a) は全文検索型で完全一致に基づく検索を行い、検索結果がすべて表示される。(b) は連想検

研究領域	キーワード	件数	キーワード	件数	キーワード	件数	キーワード	件数	総数
言語	翻訳	4		4		4		1	13
システム	制御	1		2		5		4	12
教育	活動	8		0		4		2	14
心理	研究	2		3		1		4	10
総数		15		9		14		11	49

図 3: 教員マップ (数字データ)

研究領域	キーワード	件数	キーワード	件数	キーワード	件数	キーワード	件数	総数
言語	翻訳	4		4		4		1	13
システム	制御	1		2		5		4	12
教育	活動	8		0		4		2	14
心理	研究	2		3		1		4	10
総数		15		9		14		11	49

図 5: 教員マップ (重要語データ)

番号	部局名	職名	氏名
1	理事 (副学長)	理事 (副学長)	I0001
2	理事 (副学長)	理事 (副学長)	I0002
3	理事 (副学長)	理事 (副学長)	I0003
4	人文科学研究院	教授	I0004
5	人文科学研究院	教授	I0005
6	人文科学研究院	教授	I0006
...
1158	産学連携センター	助手	I1158
1159	産学連携センター	助手	I1159

表 1: 「九州大学研究者情報」による検索結果の要約

研究領域	キーワード	件数	キーワード	件数	キーワード	件数	キーワード	件数	総数
言語	翻訳	4		4		4		1	13
システム	制御	1		2		5		4	12
教育	活動	8		0		4		2	14
心理	研究	2		3		1		4	10
総数		15		9		14		11	49

図 4: 教員マップ (教員名データ)

索, すなわち語の連想関係を利用した検索を行い, 関連の高い順に検索結果が表示される。詳細は [16] 参照のこと。(c) は 4 節で述べた通りである。(a) で「研究・教育・社会活動概要」を検索対象文書とし、「教育」を検索要求としたときの検索結果の要約を表 1 に示した。なお, 個人名は匿名にしている (以下同様)。なお, 検索された人数は 1159 人であった。

このように検索結果が多い場合, 絞り込みを行わないと望んでいるであろう検索結果を得ることはできない。(b) で検索対象文書を「研究・教育・社会活動概要」とするときの「教育」の検索結果 (上位 10 人) の要約を表 2 に示す。

順位	関連度	氏名	所属
1	1.000	A0001	人間環境学研究院 助手
2	0.916	A0002	人間環境学研究院 教授
3	0.881	A0003	人間環境学研究院 教授
4	0.863	A0004	人間環境学研究院 助教授
5	0.860	A0005	高等教育総合開発センター 助教授
6	0.857	A0006	人間環境学研究院 助教授
7	0.852	A0007	言語文化研究院 教授
8	0.851	A0008	歯学研究院 助手
9	0.848	A0009	人間環境学研究院 助手
10	0.843	A0010	言語文化研究院 助教授

表3から分かるのは、「教育」という検索要求が2つの軸から分割され、いくつかの分野が見えてくることである。したがって、材料工学関連の教員をさがしているのであれば、A1とB5が交わっているマスの教員を見ればよさそうであるとか、医療関連の教員をさがしているのであれば、A4とB2が交わっているマスの教員を見ればよさそうだと分かる。

表2: 「九州大学教員連想検索システム」による検索結果の要約

	A1	91	50	43	32	107	323
A2	35	33	27	17	9	121	
A3	67	30	52	28	11	188	
A4	41	112	56	13	22	244	
A5	41	54	12	6	15	128	
総数	275	279	190	96	164	1004	
	B1	B2	B3	B4	B5	総数	
A1	研究, 教育, 工学, 活動, 力学						
A2	教育, 研究, 耳鼻咽喉科, 民法, 括約筋						
A3	研究, 教育, 文化, 社会, 活動						
A4	研究, 臨床, 細胞, 治療, 疾患						
A5	研究, 森林, 土壌, 環境, 生物						
B1	教育, 看護, 日本語, ドイツ, 文化						
B2	cell, 研究, protein, mouse, gene						
B3	子ども, 教育, inducer, アメリカ, elegans						
B4	なし, 刑務所, yatoi, wheelchair, caveolin						
B5	解析, 合成, 特性, 構造, plasma						

表3: 教員マップの要約

連想検索システムの特徴の一つとして、検索結果がランクづけされた形で返ってくることで、表2では省略されているが、連想される根拠になった語や検索対象文書における特徴語を出力することができるが挙げられる。したがって、ある程度の絞り込み作業が容易になっているといえるが、各教員の文章の長さが異なったりしているため、長い文章を書いた教員についてその分野のオーソリティであるにもかかわらず上位に検索されないという問題点が指摘されている。

(c)で縦軸を「研究・教育・社会活動概要」、横軸を「研究業績」とし、それぞれの生成クラスタ数を5とするときの「教育」の検索結果の要約をそれぞれ表3に示す。

6 まとめと今後の課題

本稿では、複数の項目を持つ半構造化文書に対し、任意に選んだ二つの観点から分析の分析を実現する多面的検索システムの手法を提案した。具体的な例として、所属、職名、研究概要、研究業績などの複数項目からなる教員データを対象として、この方式を実装した教員マップ作成システムについて述べた。マップ作成手順は各軸の特性が「静的クラスタリング」でも「動的クラスタリング」でも本質的には同じである。ここでは、検索要求に合致する教員だけをクラスタリングして教員マップを作成することにしているが、連想検索の考え方を使って検索要求から連想される教員をクラスタリングして教員マップを作成することも考えられる。さらに、巨大なデータについてマップを作成するとき、各マスに該当するデータ数が増えることも考えられ、各マスについて再度マップを作成するということが可能である。これについても今後取り組むべき課題であるといえる。

提案手法による検索の効率についての定量的な評価実験は今後の課題である。また、定性的な評価については、各種クラスタリングの手法を検討する必要がある。本稿では教員データを対象としたが、今後、他の様々な半構造化文書についても本手法の適用可能性を検討する予定である。

参考文献

- [1] S.Cohen, J.Mamou, Y.Kanza, Y.Sagiv. XSEarch: A Semantic Search Engine for XML, *The 29th International Conference on Very Large Databases (VLDB)*, 2003.

- [2] L.Guo, F.Shao, C.Botev, J.Shanmugasundaram. XRANK: Ranked Keyword Search over XML Documents, *SIGMOD2003*, 2003
- [3] R.Kaushik, R.Krishnamurthy, J.F.Naughton, R.Ramakrishnan. On the integration of structure indexed and inverted lists, *SIGMOD 2004*, 2004.
- [4] R.W.P.Luk, H.V.Leong, T.S.Dillon, A.T.S.Chan, W.B.Croft, J.Allan. A survey in indexing and searching XML documents, *Journal of the American Society for Information Science and Technology archive Vol.53, No. 6*, pp.415-437, 2002
- [5] J.Pehcevski, J.A.Thom, S.M.M.Tahaghoghi, A-M.Vercoustre. Hybrid XML Retrieval Revisited, *Information Retrieval, vol. 8, no. 4*, pp.571-600, 2005
- [6] J.Pehcevski, J.Thom, A-M.Vercoustre. Enhancing Content-and-Structure Information Retrieval using a Native XML Database, *TDM'04*, 2004
- [7] T.Schlieder, H.Meuss. Querying and ranking XML documents, *Journal of the American Society for Information Science and Technology 53 (6)*, pp.489-503, 2002
- [8] Cong Yu, Hong Qi, H.V.Jagadish. Integration of IR into an XML Database, *First Annual Workshop of the Initiative for the Evaluation of XML Retrieval (INEX Workshop 2002)*, 2002
- [9] F.Weigel, H.Meuss, F.Bry, K.U.Schulz. Content-Aware DataGuides:Interleaving IR and DB Indexing Techniques for Efficient Retrieval of Textual XML Data, *ECIR2004*,pp.378-393, 2004
- [10] F.Weigel, H.Meuss, K.U.Schulz, F.Bry. Content and Structure in Indexing and Ranking XML, *Proceedings of the 7th International Workshop on the Web and Databases (WebDB) 2004*, 2004
- [11] F.Weigel, K.U.Schulz, H.Meuss. Ranked Retrieval of Structured Documents with the S-Term Vector Space Model Advances in XML Information Retrieval, *Proceedings of the 3rd Workshop of the Initiative for the Evaluation of XML (INEX)*, LNCS 3493, 2005
- [12] 岩山真, 徳永健伸, 確率的クラスタリングを用いた文書連想検索, 自然言語処理, Vol 5. No.1, pp.101-117, 1998
- [13] 絹谷弘子, 波多野賢治, 吉川正俊, 植村俊亮, XML 文書の文書構造と内容を用いた部分文書の抽出手法, IPSJ トランザクション「データベース」Vol.43, No.SIG02, 2002
- [14] 森康弘, 吉川正俊, 波多野賢治, 特徴ベクトルの要素間の影響を考慮した XML 文書検索手法の提案, DEWS2004, 2004
- [15] 杉本典子, 金丸玲子, 池田大輔, 竹田正幸, 井上 仁, 廣川佐千男. 九州大学自己点検・評価関連情報システム, 情報処理学会 第 41 回デジタル・ドキュメント研究会資料, 2003
- [16] 関隆宏, 廣川佐千男. 教員連想検索システム, 情報処理学会 第 48 回デジタル・ドキュメント研究会資料, 2005
- [17] 田中省作, 関 隆宏, 石野 明, 金丸玲子, 杉本典子, 竹田正幸, 廣川佐千男, 大学経営における大学評価システムの活用, 情報処理学会第 67 回全国大会予稿集, 2005
- [18] 徳永健伸, 情報検索と言語処理, 東京大学出版会, 1999
- [19] 渡邊正裕, 波多野賢治, 吉川正俊, 植村俊亮, 中村 均, 構造化文書を対象とした文字列検索とベクトル検索の統合について, 研究報告「データベースシステム」No.22, 2000