

# 異種の職業情報を統合するメタレベル・ジョブ・データベース・インテグレーション方式

## A Meta-Level Integration Method for Job-Information Databases

高橋 雄介<sup>\*</sup> 清木 康<sup>^</sup>

Yusuke TAKAHASHI Yasushi KIYOKI

本稿では、複合的な領域に関連する課題を解決するために必要となる情報を、既存職種群に関するデータベース群を用いて獲得するメタレベル・ジョブ・データベース統合方式を提案する。本方式は、文脈に応じて関連する職業に関する異種の専門知識を特定し、その統合的利用を可能にするものであり、マルチデータベースシステムの職業データへの応用を実現する方式として位置づけることができる。

本方式により、日々流動的に発生する複雑な課題を解決するために、新しい職業に関する情報を生成することができ、その解決のための指針を提供することが可能となる。

In this paper, we present a method for integrating occupational information with the meta-level system. This method is used to obtain connected occupational information for resolving complicated issues related to several existing occupations. This method is regarded as an application of multi-database system for dealing with various job databases.

By applying this method to occupational data, the new occupational information is obtained, which is to be used for solving complicated subjects that are fluidly generated.

### 1. はじめに

現在、科学・技術の革新や経済のグローバル化に伴い、既存の学問や職業領域に限定したスキルや知識ではなく、それらを横断的に必要とする課題が多く発生している。また、情報流通の量的および速度的な拡大などによって、社会の変化のスピードへの対応が重要となっている。グローバル経済、環境問題、安全保障問題、バイオインフォマティクスなど、新しい複合的な領域における課題が次々に生まれている。

このような状況では、それぞれの課題に柔軟に対応することが重要となる。文脈（解決の対象となる課題）に応じて、柔軟に解決策を構築していくことが望ましい。例えば、遺伝子情報の解析に関わる新規事業を起こすためには、生命科学の専門家のほかに、データベースや統計解析の専門家の知識も必要となる。

本研究の成果は、こうした背景から提案した「教育カリキュラムと職業データベース連携による目的指向型学習支援データベース」[1~2]を補完および拡張するものと位置づけることが

できる。これは、特定の職種について、その人材育成や知識獲得に必要なカリキュラムを動的に生成するシステムであるが、本研究における成果を付加的に利用することによって、「新しい複合的な領域における課題」から、その解決のために必要な人材（関連する職種）の特定や、解決する為に必要な人材育成のためのカリキュラムを動的に獲得することが可能となる。

本方式の適用例として、以下のようなものが考えられる：

- (1) 新規事業やプロジェクトの立ち上げにおける指針を提供することができる。ビジネスの現場において、新規事業を創出したり、新しいプロジェクトを発足させる際に、必要となる人材を特定したり、その人材に求められる専門知識やスキル、基礎的な能力を特定することが可能となる。近年の傾向として、これまでの職業について規定してきたスキルや専門知識などが社会的な変化の早さによりすぐに陳腐化してしまったり、企業や応用領域によって同一の職種でも実際の職務が異なってしまったりすることが多い([3~4])。本方式により、変化に対応しながら、組織を人材という視点から柔軟に構築するための指針を提供することが可能となる。特に、ノウハウや経験の少ないINPO（非営利組織）やNGO（非政府機関）などの中小規模の組織における有用性が高いと云える。
- (2) 複合的な領域に関する課題を解決するフレームワークを提示することができる。複合的な領域に関する課題の解決を目指す際に、関連する職業を抽出することが可能である為、それぞれの個別の職業に関する情報から、具体的に要求される知識やスキル、基礎的な能力を特定することが可能となる。人材育成という観点から、教育機関等における授業に関する情報と連結すれば、大学院の、特にプロフェッショナル・スクールにおけるカリキュラムの生成等に活用可能である。また、カリキュラムの過不足を特定したり、教育の現場と実際の職業の現場を積極的に連結する指針を提供したりすることも可能となる。

本稿では、複合的な領域における課題を問い合わせ語として設定し、その解決のために必要な職業に関する情報を既存職種に関する情報を用いて獲得する方式を示す。本方式の特徴は、異種の知識間の関連性を計量することが可能になることにある。従来、特定の課題という文脈に対して、その課題を定義した知識と、実際に現場で課題解決を行う職業に関する知識は、システムチックに統合されることはなかった。本方式によって実現される職業情報の統合方法は、「課題」に関する情報と「職業」に関する情報の関連性を動的に計量することで、社会的な変化に対応しながら、専門家の手で関連性、つまり解決策を静的に記述し続けるコストを削減することが可能となる。

本方式は、文脈に応じて関連する職業に関する異種の専門知識を特定し、統合的利用を可能にするものであり、マルチデータベースシステムの職業データへの応用を実現する方式として位置づけることができる。マルチデータベースシステムとは、独立に構築されたデータベース群を統合し、単一のデータベースのように共有するシステムである[5~7]。マルチデータベースシステムは、既存のデータベース群を相互利用することにより、それら既存の情報群に新しい応用、および、新しい価値を与えるシステムとして位置づけられる。マルチデータベースシステムの実現方式の詳細は、文献[8~11]に示している。

本方式の本質は、異種の専門家の知識を統合可能であるというところにある。本方式によって統合する専門家の知識は、次の2つである：

- ・ 解決すべき課題に関する知識
- ・ 個別の既存職業に関する知識

解決すべき課題に関する知識は、専門家によって設定された

<sup>\*</sup> 学生会員 慶應義塾大学大学院 政策・メディア研究科修士課程 [yt@sfc.keio.ac.jp](mailto:yt@sfc.keio.ac.jp)

<sup>^</sup> 正会員 慶應義塾大学環境情報学部 [kiyoki@sfc.keio.ac.jp](mailto:kiyoki@sfc.keio.ac.jp)

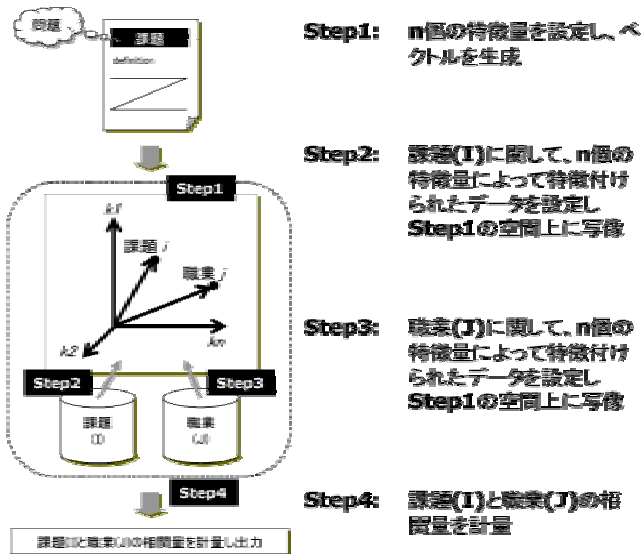


図1 本方式の概要

Fig.1 An Overview of Our Proposed Met

課題に関する定義と位置づけることができる。既存の領域に限定した課題からそれらの複合的な領域に関する課題に至るまで多種多様に存在するが、これらの課題を専門家の視点から定義し、その領域を規定した情報がこれに当たる。

個別の既存職業に関する知識は、具体的な個別職業に関する詳細の定義と位置づけることができる。既に社会に存在する特定の領域の課題を解決する為に設定された職業に関する情報であり、関連する学問領域や要求されるスキルおよび基礎的な能力等の情報との関係も記述している。

既存の学問や職業領域という観点からみると一段抽象度の高い視点から課題を定義するのが、解決すべき課題を定義する専門家である。一方で、特定の職業に関する情報を設定する専門家は、個別の職業に関する情報の専門家ではあっても、それを特定の文脈に応用する分野横断的な抽象度の高い視点は持っていない。この二つの知識には、その専門家の立場の違いから明確な情報の異種性が存在している。個別の領域の専門家と領域間の関連についての専門家の違いである。それぞれの知識は重要なものであるが、より重要なのはこれらの知識を統合的に利用することである。本方式によって、これらの二つの知識の統合が可能となる。

2. 実現方式

本稿において示す、異種の職業情報を統合するメタレベル・ジョブ・データベース・インテグレーション方式は次のとおりであり、4つのステップによって実現する(図1)。

Step-1では、課題および職業に関する情報の関連性を計量するメトリクスとしてn個の特徴量を設定し、ベクトルを生成する。解決したい課題に関する知識(Step-2において設定)と個別の職業に関する知識(Step-3において設定)から抽出した情報の関連性を計量するメトリクスを設定し、設定したメトリクスを適用可能なメタデータの設定方法を設定する。

Step-2では、解決したい「課題」に関する知識を設定し、n個の特徴量によって特徴付けられたメタデータを生成する。生成したメタデータは、Step-1において設定したn個の特徴量を表現するベクトル空間上に写像する。

Step-3では、個別の職業に関する知識を設定し、n個の特徴

id	knowledge
k1	Administration and Management
k2	Biology
k3	Building and Construction
k4	Chemistry
k5	Clerical
k6	Communications and Media
k7	Computers and Electronics
k8	Customer and Personal Service
k9	Design
k10	Economics and Accounting
k11	Education and Training
k12	Engineering and Technology
k13	English Language
k14	Fine Arts
k15	Food Production
k16	Foreign Language
k17	Geography
k18	History and Archeology
k19	Law, Government and Jurisprudence
k20	Mathematics
k21	Mechanical
k22	Medicine and Dentistry
k23	Personnel and Human Resources
k24	Philosophy and Theology
k25	Physics
k26	Production and Processing
k27	Psychology
k28	Public Safety and Security
k29	Sales and Marketing
k30	Sociology and Anthropology
k31	Telecommunications
k32	Therapy and Counseling
k33	Transportation

表1 0\*NETの設定する33種類の学問領域群

Table 1 33 Kinds of Knowledge Set by 0\*NET

量によって特徴付けられたメタデータを生成する。生成したメタデータは、Step-1においてn個の特徴量を表現するベクトル空間上に写像する。

Step-4では、Step-2およびStep-3で抽出したデータ間の相関量を計量する。相関量の計算には、内積やコサインなどがある。これらは、情報検索を行う際に用いられる手法であり、ベクトル間の相関量を計量するのに適している。

3. 実験

本実験では、単純なメタデータの設定により本方式の実現可能性を検証する。

3.1 実験環境

本研究では、以下のとおりデータを設定し、実験を行う：

(a) 既存学問領域に関する情報

既存学問領域に関する情報(Step-1)としては、学問領域全体を構造的に分類している情報を設定した。本実験では、0\*NET(The Occupational Information Network) [12]がその研究成果として設定している情報を用いた。0\*NETでは、既存の専門知識を33種類の学問領域に構造的な分類を行っている(表1)。本稿では、この分類を既知の知識として利用した。

0\*NETとは、米国政府により提供されるデータベースで、職

BI	k2, k7, k20
DE	k3, k4, k10, k17, k19, k26, k30
NG	k7, k17, k19, k24, k30

表2 問い合わせキーワードと関連する学問領域

Table 2 Query Keywords and Related Educational Issues

業の特性と利用者の属性の関係を包括的に体系化している。定期的に行われる現場でのアンケートに基づいて定量化されたデータをもとに、950種類を超える各職業に要求される専門知識、技術・技能、基礎能力、動機などを定義している。本実験においては、33種類の学問領域(Knowledge)を、「解決すべき課題に関する情報(Step-2)」および「職業に関する情報(Step-3)」を表現する特徴語として設定した。

(b) 解決すべき課題に関する情報

解決すべき課題に関する情報(Step-2)としては、慶應義塾大学湘南藤沢キャンパス(以下「慶應SFC」と呼ぶ)において設定されている「クラスタ」を用いた。「クラスタ」とは、既存の個別学問領域のアプローチを横断的に用いて解決すべき課題を設定したものである。慶應SFCにおける総合政策学部および環境情報学部の学生の科目履修選択を支援するためのガイドラインとして、15種類の「先端的研究領域」が設置されている[13]。この情報の中から、「目的・ミッション」および「将来のイメージ」の部分のテキストを抜粋し、該当テキストに対して形態素解析を行い、キーワードの抽出を行った。本実験では、慶應SFCにおける15の「クラスタ」の中からBI(バイオインフォマティクス: 遺伝子・ゲノム・細胞から生命をコンピュータで解明する)、DE(開発と環境: 自然環境と融合した国土・地域開発を計画する)およびNG(ネットワークガバナンス: ネットワーク環境を実践基盤として社会問題に取り組む)の3つを選択した。「クラスタ」に関する詳細については、文献[13]を参照されたい。

(c) 個別の職業に関する知識

職業に関する情報(Step-3)として、O\*NET[12]の提供する個別の職業に関する統計情報を設定した。本実験においては、O\*NETがそれぞれの職業毎に設定している個別の専門知識の重要度を用いた。この重要度は0-100の値で出力されるものであり、この値をメタデータとして付与している。

以上の異種の知識を統合するにあたり、本実験における問い合わせには、内積のプログラムを構築し、利用した。内積は、問い合わせキーワードと検索対象データ間の相関量を計算する一手法である。問い合わせキーワードとして「解決すべき課題と既存学問領域の関係を記述した情報」を、検索対象データとして「個別の職業と既存学問領域との関係を記述した情報」を設定し、「解決すべき課題」と「個別の職業」という異種の知識の相関量を計量する。具体的には、O\*NETの設定する33種類の直交した専門知識(knowledge)によりベクトル空間を生成し、それぞれの知識を同一のベクトル空間上に写像し、内積によって相関量を求める。

問い合わせキーワードとして設定した「解決すべき課題」に関する知識(Step-2)は、複合的な課題に関して定義したドキュメントに形態素解析を行い、関連する専門知識に関するキーワードを抽出した。本実験では、慶應SFCにおけるクラスタに関するドキュメントを用いて、そのドキュメント内に出現する単語群より、「専門知識(knowledge)」に関連した語句を抽出し、メタデータを設定した。関連するキーワードの出現頻度によって専門知識毎に重み付けを行った。

検索対象データとして設定した「職業」に関する情報(Step-3)は、O\*NETの提供する統計情報を設定した。この情報は、個別の

rank	value	related occupation
1	0.744176458	Zoologists and Wildlife Biologists
2	0.722452338	Microbiologists
3	0.716959774	Agricultural Technicians
3	0.716959774	Biological Technicians
5	0.697330737	Animal Scientists
6	0.682190138	Cardiovascular Technologists and Technicians
7	0.678226683	Nonfarm Animal Caretakers
8	0.677938684	Environmental Scientists and Specialists, Including Health
9	0.675320074	Biologists
10	0.674648349	Nuclear Medicine Technologists

表3 実験結果(BI)

Table 3 Experimental Results for BI

rank	value	related occupation
1	0.815265439	Child Support, Missing Persons, and Unemployment Insurance Fraud Investigators
2	0.807048382	Bailiffs
3	0.799027617	Correctional Officers and Jailers
4	0.794276969	Transit and Railroad Police
5	0.787811353	Criminal Investigators and Special Agents
6	0.775964506	Insurance Adjusters, Examiners, and Investigators
7	0.770293186	Government Property Inspectors and Investigators
8	0.768480595	Police Detectives
9	0.767488379	Fire Inspectors
10	0.759238635	Sheriffs and Deputy Sheriffs

表4 実験結果(DE)

Table 4 Experimental Results for DE

職業における33種類の専門知識の重要度を、0-100の値で設定している。この値を用いてメタデータを設定し、ベクトルを生成した。

問い合わせキーワードとして設定した「解決すべき課題」に関する情報(Step-2)も検索対象データとして設定した「職業」に関する情報(Step-3)も、それぞれベクトルを正規化し、ノルムを「1」に合わせた。単位ベクトル化を行うことにより、「課題」と「職業」という異種の情報に関する情報量を同一のものとし、互いに比較可能なものとした。これにより、それぞれの持つ情報量は同一であるという前提を置くことができるため、それぞれのベクトル間の関係のみを計量可能である。

3.2 実験手法

本実験では、「実験環境(3.1)」において設定したデータを用いて、BI(バイオインフォマティクス)、DE(開発と環境)、NG(ネットワークガバナンス)の3つの問い合わせキーワード(Step-2)を設定した(表2)。

3.3 実験結果および考察

実験結果は、表3(BI)、表4(DE)および表5(NG)において、それぞれ上位10位までを示す。

ここから以下のことが云える：

- (1) 問い合わせキーワードとして設定した課題に関して、関連性の高い情報が上位に検索された。
- (2) 問い合わせキーワードとして設定した課題に関して、関連性の低い職業が下位に検索された。

例を挙げるなら、「BI(バイオインフォマティクス)」を問い合わせキーワードとして設定した場合(表3)の実験結果では、関連する学問領域として「k2(Biology)」、「k7(Computers and Electronics)」および「k20(Mathematics)」を設定した(BIの特

rank	value	related occupation
1	0.717737968	Criminal Investigators and Special Agents
2	0.705089301	First-Line Supervisors/Managers of Police and Detectives
3	0.699759813	Child Support, Missing Persons, and Unemployment Insurance Fraud Investigators
4	0.69614053	Judges, Magistrate Judges, and Magistrates
5	0.695817239	Urban and Regional Planners
6	0.685081179	Correctional Officers and Jailers
7	0.676302319	Government Property Inspectors and Investigators
8	0.675310211	Private Detectives and Investigators
9	0.670433998	Insurance Adjusters, Examiners, and Investigators
10	0.66804919	Transit and Railroad Police

表5 実験結果(NG)

Table 5 Experimental Results for NG

微量として, K2, K7, k20を「1」に, それ以外を「0」にしたが, 検索結果の中で, Biology, Computers and Electronics, Mathematicsを専門的に利用するか, その専門知識に直接的に関係する職業と問い合わせとの相関量が高く計量され, 上位に検索された. 一方で, それ以外の職業に関しては, 関連性の低い職業として下位に出力されている. 「DE(開発と環境)」(表4), 「NG(ネットワークガバナンス)」(表5)においても同様の結果が得られた.

以上の実験結果から, 次のことが導き出せる:

- (1) 本方式を適用することにより, 課題を定義した文書があれば, 関連する職業を特定することが可能となる.
- (2) 課題および職業に関する情報に対して, メタデータとしての確な特徴量を設定することができれば, 相互の相関量を計量し, 異種の知識の統合的使用が可能となる.

本システムにより人材, 職業などの情報を統合的に扱うことが可能となり, 実現可能性を示すことができた. 今後の課題として次の三つを挙げる:

- (1) 本方式の有効性の検証
- (2) 課題に関して定義した文書から, その特徴量をメタデータとして抽出する方法の開発

#### 4. 終わりに

本研究では, 異種の職業情報を統合するメタレベル・ジブ・データベース・インテグレーション方式を示し, 実験により実現可能性を示した.

今後の展望としては, 本方式を教育データベース[1~2]と連携させることによる, 人材育成およびキャリア開発支援への適用が考えられる. また, ジョブマッチングマルチデータベース[14]との連携も考えられる.

#### [謝辞]

最後に, 本稿執筆にあたって有益な助言をいただきました, 慶應義塾大学大学院政策メディア研究科の吉田尚史氏, 関子泰三氏, 倉林修一氏および金澤元紀氏に感謝します.

#### [文献]

- [1] 高橋雄介, 清木康: “教育カリキュラムと職業データの連携による目的指向型学習支援データベースの構成方式”, 情報処理学会研究報告Vol.2002, No.41, pp.81-88 (2002).
- [2] 高橋雄介, 清木康: “教育カリキュラムと職業データベース

連携による目的指向型学習支援データベースの実現”, 電子情報通信学会技術研究報告Vol.102 No.208, pp.121-126 (2002).

[3] 原井新介: “キャリアコンピテンシーマネジメント”, 日本経団連出版 (2002).

[4] 高橋俊介: “キャリア論-個人のキャリア自律のために会社は何をすべきなのか”, 東洋経済新報社 (2003).

[5] Bright, M. W., Hurson, A. R. and Pakzad, S. H.: “A Taxonomy and Current Issues in Multidatabase System”, IEEE Computer, Vol.25, No.3, pp.50-59 (1992).

[6] Litwin, W., Mark, L. and Roussopoulos, N.: “Interoperability of Multiple Autonomous Databases”, ACM Comp. Surveys, Vol.22, No.3, pp.267-293 (1990).

[7] Sheth, A. and Larson, J.A.: “Federated database systems for managing distributed, heterogeneous, and autonomous databases”, ACM Computing Surveys, Vol.22, No.3, pp.183-236 (1990).

[8] Kiyoki, Y. and Kitagawa, T.: “A metadatabase system supporting interoperability in multidatabases”, Information Modeling and Knowledge Bases, Vol.5, pp.287-298 (1993).

[9] Kiyoki, Y., Kitagawa, T. and Hitomi, Y.: “A fundamental framework for realizing semantic interoperability in a multidatabase environment”, Journal of Integrated Computer-Aided Engineering, Vol.2, No.1, pp.3-20 (1995).

[10] Kiyoki, Y., Hosokawa, Y. and Ishibashi, N.: “A Metadatabase System Architecture for Integrating Heterogeneous Databases with Temporal and Spatial Operations”, Advanced Database Research and Development Series Vol.10, Advances in Multimedia and Databases for the New Century, A Swiss/Japanese Perspective, pp.158-165, World Scientific Publishing (2000).

[11] 細川宜秀, 石橋直樹, 八代夕紀子, 清木康: “マルチデータベース環境における時間的・空間的関連性評価によるデータ結合方式”, 情報処理学会論文誌: データベース, Vol.40, No.SIG 8(TOD4), pp.95-111 (1999).

[12] O\*NET OnLine URL: <<http://www.onetcenter.org/>>

[13] 湘南藤沢事務室編集: “SFC ガイド2003”, 慶應義塾大学 (2003).

[14] 教育学術データベース等の開発事業: “キャリア開発・ジョブマッチングマルチデータベースの開発” 報告書”, 慶應義塾大学清木研究室 (2001).

#### 高橋 雄介 Yusuke TAKAHASHI

慶應義塾大学大学院政策・メディア研究科修士課程在学中. 2003 慶應義塾大学総合政策学部卒業. データベースシステムの研究に従事. 日本データベース学会学生会員.

#### 清木 康 Yasushi KIYOKI

慶應義塾大学環境情報学部教授. 1983 慶應義塾大学大学院工学研究科博士課程修了, 工学博士. データベースシステム, 知識ベースシステム, マルチメディアシステムの研究に従事. ACM, IEEECS, 電子情報通信学会, 情報処理学会, 日本データベース学会各会員.