

異種の職業情報と教育情報を連結するメタレベルデータベースシステムの構成方式

A Meta-Level Database System Connecting Occupational and Educational Databases

高橋 雄介[▼]清木 康[◆]

Yusuke TAKAHASHI

Yasushi KIYOKI

本稿では、異種の職業情報と教育情報を連結するメタレベルデータベースシステムの構成方式について示す。本システムは、異種の教育コンテンツ、職業データベース、および、個人のキャリア情報の関係を、検索者の目的や状況（コンテキスト）に応じて動的に計算可能にするものである。実際のデータベースに適用し、提案方式の実現可能性を検証する。本システムにより、職業、教育、人材に関するデータベース間の関係の計量が可能になり、人材育成やキャリア開発分野における多大な貢献が期待できる。

In this paper, we present a method to design and construct a meta-level database system connecting occupational and educational databases. This system enables dynamic computation among databases that deal with information on occupation, education and personal career, according to users' contexts such as career objectives and histories. We evaluate feasibility of our method by applying it to real databases. By using our method, dynamic computation is realized among career-related databases, and a great contribution is made for people who require effective support for their personal career development.

1. はじめに

近年、価値観、職業選択、教育機会の多様化といった社会的背景により、個人に特化した学習計画やキャリア開発プランの設計が重要な課題となっている。しかし、多様な職業、教育についての異種の情報資源間の連結メカニズムが存在しない為に、実現が困難である。こうした背景から、個人に特化したキャリア開発支援およびカリキュラムの動的なデザインを対象とした、異種の職

業情報、教育情報、個人の学習履歴及びキャリア情報などの関係を動的に計算可能な計量系の実現は、これらの互いに独立に構築された情報源を相互運用可能な形式で統合、編集し、検索可能にするものであり、高度な情報検索技術の応用として重要である。

本稿では、異種の職業情報と教育情報を連結するメタレベルデータベースシステムの構成方式について示す。本システムは、異種の教育コンテンツ、職業データベース、および、個人のキャリア情報の関係を、検索者の目的や状況（コンテキスト）に応じて動的に計算可能にするものである。本方式は、これまでに提案してきた異種の職業間関係計量方式 [1] 及びキャリア開発支援システム [2, 3] を拡張し、既存の教育コンテンツ、教育における分野や課題、職業情報、個人のキャリアにおける履歴と目標に関するデータベース群の相互の関係を、本システムの利用目的に関する情報に応じて計量するメタレベル計量系を実現する。

本構成方式の特徴として、第1に、個人のキャリア開発支援という視点において、利用目的に応じたメタデータを抽出し、メタデータ間の意味的な関係を定義することで、任意のデータと関係の強いデータ、およびその集合に対して、利用目的に応じた意味を与える点がある。第2に、既存の教育及び職業のデータベース間を個人の目的や履歴に応じた視点で「計算による動的設計」を実現する優れた知識表現がある。

国内外の関連する研究動向として、専門家によって静的に設計・記述されたカリキュラム自体の提案 [4]、および、専門知識に関する知識体系と既存の教育コンテンツ群の対応関係を論理的に記述したカリキュラムの実践 [5]、論理的に記述された関係によって利用者の学習をナビゲーションする方法 [6, 7, 8] がある。これらは共通して、教材間や教材と利用者が論理的に関係づけられている。本方式は、キャリア開発やカリキュラム設計に関連する専門知識体系を用いて、社会や個人の状況に応じた動的なカリキュラムの設計を可能とする。

2. 実現方式

異種の職業情報と教育情報を連結するメタレベルデータベースシステムの実現方式について述べる。

2.1 キャリア開発について表現可能な語彙集合を用いた意味的關係計量ベクトル空間構成

本構成方式では、メタレベルデータベース環境において、ベクトル空間モデル [9] を用いることで、共通の特徴語を持つメタスキーマを構成し、既存のデータベース群に内在する異種性を排除する。本構成方式は、職業と教育に関する既存のデータベース群の上位層（メタレベル）にマルチデータベースシステムとして実現し、そこにキャリア開発支援を対象としたベクトル空間を構成する。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 [10]。ベクトル空間モデルは、文書（検索対象）および検索質問の内部表現を多次元ベクトルで表現し、ベクトル間の類似度を計算することにより、文書の類似検索を実現する。

本構成方式において重要な点は、個人のキャリア開発という特

▼ 学生会員 慶應義塾大学大学院 政策・メディア研究科 yt@sfc.keio.ac.jp

◆ 正会員 慶應義塾大学 環境情報学部 kiyoki@sfc.keio.ac.jp



図1 O*NETTMは、米国労働省、U.S. Department of Labor, Employment and Training Administrationのトレードマークです。

Fig. 1 O*NETTM is a trademark of the U.S. Department of Labor, Employment and Training Administration.

化した領域において、関連する専門知識体系を用いた計量空間を構築する点にある。ベクトル空間モデルでは、応用領域に関する包括的な知識体系を用いた網羅的な検索を実現することが可能となる。また、意思決定支援環境を実現する為には、膨大な情報の中から重要な情報とそうでない情報を効率的に識別する必要がある。ベクトル空間モデルでは、ベクトルの各要素の値の設定方式として、関係がある場合には1を、関係がない場合には積極的に0の値を付与することが可能である。

本構成方式においてベクトル空間を構成する特徴語セットとして、O*NETTM (the Occupational Information Network)¹ (図1)により提供されているContent ModelTMから、専門知識やスキル、基礎能力を構造的に定義したKSAs²を採用する。O*NETTMは、働く人の属性や職業の特徴に関して定めた包括的なデータベースである。Content ModelTMは、O*NETTMの概念的な基盤をなすモデルであり、個人のキャリア開発に関連するデータベース間の関係を計量する為に適している。

2.2 キャリア開発に特化したメタデータ生成

キャリア開発に関する既存データベース LDB_L 内の個々のデータ $\{data_j\}_{j=1\dots n}$ に対して、メタデータ群を付与する ($metadata_{j,ksa} \in KSAs$)。 LDB_L 毎にメタデータ抽出の判断基準 (criterion) を定義し、これに合致する場合メタデータは真値を返す。メタデータ抽出の評価基準として、 LDB_L 毎に、当該KSAsが、ある社会的・教育的課題 (SEI) の解決の為には必要か (Required)、ある職業 (O) に就業する場合に当該KSAsが必要か (Required)、ある学習コンテンツ (EC) で習得可能か (Developable)、ある個人の能力 (キャリア履歴, PCH) として習得済みか (Developed) とする。

個人のキャリア情報 (LDB_{PCH}) は、キャリア履歴や目標の変化に対応するために、時系列データ付きで生成し、ユーザ (User) 毎に5種類のユーザ情報定義ベクトル (UDVs) を定義する (表1)。

個々のデータ毎に生成されたメタデータは、2進重みによりベクトル形式に変換する。

2.3 利用目的に応じた意味的関連評価機構

メタデータ間の意味的関連は、利用目的 (Objective) に応じて定義し、相関量 (cor) が閾値 (threshold) 以上の集合を評価結果として出力する関数 $E_{Objective, k_x}$ を定義する。検索質問と

表1 時系列ユーザ情報定義ベクトル (UDVs) の定義。

Table 1 Descriptions of the UDVs.

ID	定義	データクラス
UDV_{1time}^{User}	個人のスキルや経験	S_1
UDV_{2time}^{User}	興味のある/付加的なKSAs	S_2
UDV_{3time}^{User}	目標 (職業)	S_3
UDV_{4time}^{User}	目標に対して必要なKSAs	$S_3 \setminus (S_1 \cap S_3)$
UDV_{5time}^{User}	将来のキャリアイメージ	$S_1 \cup S_2 \cup S_3$

して指定する特定のベクトルを k_x 、検索対象データを t_y とし、それぞれ同一の特徴語 (w_{x_n} および w_{y_n}) を持つ n 次元ベクトル形式で表現し、ベクトル間の相関関係を計量する。相関量の閾値は $0 \leq threshold \leq \max(cor)$ 内で設定する。利用目的 (Objective) に応じた $E_{Objective, k_x}$ の例を、以下に示す。

カリキュラムに適合する職業の選択肢 教育分野において必要とされるKSAsと、職業において要求されるKSAsをメタデータとして持つデータベース間の相関量計量により、必要とされているKSAsに関して職業を特定する。

$$E_{CareerCandidates, k_x} = \{cor(k_{x,SEI} \cdot t_{y,O}) \mid cor \geq threshold\}$$

社会的・教育的課題に対応するカリキュラム 教育分野において必要とされるKSAsと、学習コンテンツを用いて習得可能なKSAsをメタデータとして持つデータベース間の相関量計量により、習得可能なKSAsに関して学習コンテンツを特定する。

$$E_{Curriculum, k_x} = \{cor(k_{x,SEI} \cdot t_{y,EC}) \mid cor \geq threshold\}$$

職業に必要なKSAsを学習するためのカリキュラム 教育分野において必要とされるKSAsと、学習コンテンツを用いて習得可能なKSAsをメタデータとして持つデータベース間の相関量計量により、開発、習得可能なKSAsに関して学習コンテンツを特定する。

$$E_{CareerOrientedCurriculum, k_x} = \{cor(k_{x,O} \cdot t_{y,EC}) \mid cor \geq threshold\}$$

本構成方式においては連携する各データベースにおける情報量は、ベクトル空間上で相関量を計量する際には、同等であるとする。本来、異種の情報資源群は、それぞれ別々の目的のもとに構築されており、相互に比較、連携の出来ない異種の情報資源である。しかし、メタデータの値の意味としての「要求されるKSAs」と「習得済みのKSAs」、「習得可能なKSAs」は、個人のキャリア開発という視点においては同等の情報源として扱うことが可能であり、実際の類似度計量の際には同一の情報量を持つものとして正規化し、相関関係を計算する。

3. 実験

異種の職業情報と教育情報を連結するメタレベルデータベースの実現方式の実現可能性を検証する為に、次の2つの実験について示す。実験環境を、表2に示す。

¹ <http://www.onetcenter.org/>

² KSAs: Knowledge, Skills and Abilities

表 2 実験に用いたデータベース群 (LDB_L) .Table 2 Databases used for Experiments (LDB_L).

LDB_L	説明
LDB_{SEI}	慶應義塾大学湘南藤沢キャンパス(以下, 慶應SFC)設置のカリキュラムである15の“クラスター”情報から, メタデータを抽出.
LDB_O	O*NET Database(統計調査に基づく職種定義)から, 職種毎の個々のKSAsのImportance(0から100)をベクトルとして利用. 実験1にv.3.1, 実験2にv.5.0を利用.
LDB_{EC}	慶應SFC設置の670の授業における「講義概要」の記述からメタデータを抽出.

表 3 検索質問毎の正解数 N_l ($l = 1$) .Table 3 Number of correct answers (N_l ($l = 1$)).

Cluster	#	Cluster	#	Cluster	#
BI	51	GE	24	MD	33
CM	49	GG	147	NG	162
DE	57	IT	39	PP	74
ED	49	LC	100	SI	46
FE	35	MO	40	WE	60

3.1 実験 1

既存のカリキュラムを再現可能であることを示すために, 提案方式によって得られたカリキュラムと, 専門家によって手動で定義され, 実際に教育機関において運用されているカリキュラムを比較する. 具体的に, 以下の手順によって実施する.

1. 検索質問を定義し, 提案システムに対して入力し, 内積, ノルム, コサイン尺度, 距離を用いて結果を出力する.
2. 以下に定義する評価関数を用いて, 提案システムによる出力と正解データを比較する. 評価関数 EV_l は, 正解数 N の l 倍点での正解授業のカバー率を評価し, クラスター C_j の上位 $N \times l$ 位での再現率の全クラスターにおける平均により求める.

$$EV_l = \frac{1}{15} \left(\sum_{i=1}^{15} R_{l,C_j} \right) \times l$$

ただし,

- 15: 全クラスター数
- R_{l,C_j} : クラスター C_j の $N \times l$ 位での再現率
- C_j : 各クラスターにおける R_l, C_j
- N : クラスター毎の正解数
- l : クラスター毎の正解数 N の倍数

検索質問として, 各クラスターの「将来のイメージ」から代表的な職業 (RO) を設定する (例えば, BI に対して “Biophysicists”). 正解データとして, 「クラスターに必要な授業科目」を用いる (表 3).

$$E_{CareerOrientedCurricula, RO_x} = \{cor(RO_{x,O} \cdot t_{y,EC})\}$$

実験 1 の結果 (図 2) から, 正解数の 3 倍程度以上において, EV の値が 0.8 を越えている. カリキュラム設計時, 特に履修する授業の選択時において, 必要な授業数の 3 倍以上の選択肢を提示することによって, 履修選択等の意思決定を支援するために, 適量な選択肢を提示している. カリキュラム設計の専門家が手動で設計するカリキュラムを再現可能であると云える.

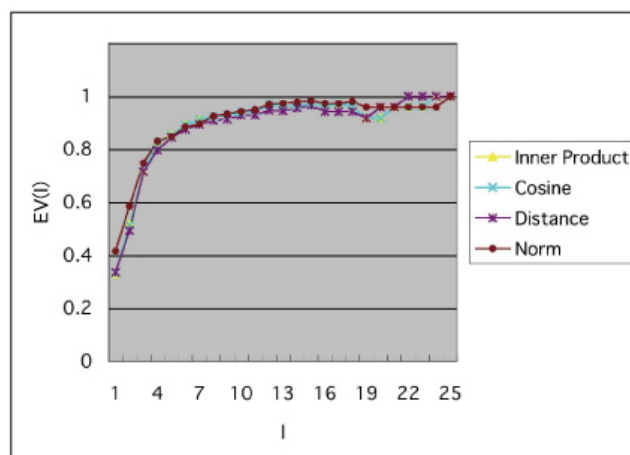


図 2 実験 1 の結果.

Fig. 2 Result of Experiment 1.

3.2 実験 2

様々な検索要求に対する結果を比較することで, 適応可能性を示す. 以下の検索質問を定義し, 出力を評価する. 相関量は, コサイン尺度を用いて計量し, 閾値は 0.6 とする ($0 \leq \text{threshold} \leq 1$).

$$Q2-1 \quad E_{Curriculum, BI} = \{cor(BI_{SEI} \cdot t_{y,EC}) \mid cor \geq 0.6\}$$

$$Q2-2 \quad E_{CareerCandidates, BI} = \{cor(BI_{SEI} \cdot t_{y,O}) \mid cor \geq 0.6\}$$

実験 2 の結果 (表 4,5) は, 様々な検索要求に対する結果を比較することで, 提案方式の適応可能性を示している. Q2-1 において示した BI に関連の強い講義群は, 14/15 の出力が正解であり, 不正解の「多変量解析」も正解と判断可能であり, 既存のカリキュラムを再現し, それ以外の重要な講義も発見可能であることを示している. 他方, Q2-2 では正解データを設定し客観的な評価をすることが困難であるが, 検索質問に対して特に関連の高い職業のみを出力することにより, 職業選択やキャリア設計時の, 明確で重要な指針を提供していると云える. 利用者が検索時に入力する社会的, 教育的課題は, 利用者が本人のキャリアとして興味を持っている領域に対応しており, その領域に関係のある職業を探すことは, キャリア開発の選択肢を取得することに値する. 同様に, 出力される学習コンテンツ群は, キャリア目標を実現するためのカリキュラムに対応していると云え, 具体的なキャリアプランを示している.

4. おわりに

本稿では, 異種の職業情報と教育情報を連結するメタレベルデータベースシステムの構成方式について示し, 実験により実現可能性を検証した. 本システムは, 異種の教育コンテンツ, 職業データベース, および, 個人のキャリア情報の関係を, 検索者の目的や状況 (コンテキスト) に応じて動的に計算可能にするものである. 今後の課題として, 実現方式を用いた実証実験の実施,

表 4 Q2-1 の結果 (実験 2) .

Table 4 Results of Q2-1 (Experiment 2).

相関量	講義名
0.816496	多変量解析
0.816496	多変量解析 A
0.816496	生体情報論
0.816496	多変量分析
0.816496	メタボローム解析実習
0.816496	脳情報科学
0.816496	統計データベース
0.816496	数理生物学
0.707106	バイオインフォマティクスツール演習
0.707106	ゲノム解析プログラミング実習
0.707106	バイオシミュレーション実習
0.707106	バイオインフォマティクスアルゴリズム実習
0.666666	プロテオーム解析実習
0.666666	アルゴリズム論
0.666666	メタボローム解析実習

表 5 Q2-2 の結果 (実験 2) .

Table 5 Results of Q2-2 (Experiment 2).

相関量	職種
0.745690	Zoologists and Wildlife Biologists
0.705599	Microbiologists
0.679068	Environmental Scientists and Specialists, Including Health
0.666754	Mathematicians
0.653262	Biological Technicians
0.653262	Agricultural Technicians
0.646709	Biophysicists
0.644354	Mathematical Technicians
0.640687	Biologists
0.640085	Statistical Assistants
0.609493	Statisticians

様々なキャリア開発関連情報の分析機構の実現を予定している。

謝辞 本研究の一部は、平成 18 年度科学研究費補助金 若手研究 (B)「異種の職業情報と教育情報を連結するメタレベルデータベースシステムの実現」(課題番号: 18700097, 代表: 高橋雄介)により、実施した。また、本研究に関して、金子郁容慶應義塾大学政策・メディア研究科教授兼総合政策学部教授、および、古川康一同大学同大学院教授兼環境情報学部教授から、大変有益なコメントを頂いた。ここに、感謝の意を表す。

[文献]

- [1] 高橋雄介, 清木康: “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”, 日本データベース学会 Letters, Vol.2 No.3, pp.33-36 (2003).
- [2] Takahashi Y., Kiyoki Y.: “A Meta-Level Career-Design Support System for Connecting Educational and Occupational Databases”, IEEE Computer Society Press, Pro-

ceedings of the 2004 International Symposium on Applications and the Internet Workshops, pp.523-530 (2004).

- [3] Takahashi, Y. and Kiyoki, Y.: “The Implementation and Application of a Meta-Level Career-Development Support System”, Proceedings of The 7th IASTED International Conference on Computers and Advanced Technology in Education, pp.558-563 (2004).
- [4] The Joint Task Force on Computing Curricula IEEE Computer Society and Association for Computing Machinery: “Computing curricula 2001”, Journal on Educational Resources in Computing (JERIC), Volume 1, Issue 3es, No. 1 (2001).
- [5] Ramakrishnan, S. and Cambrell A.: “An In-Forming Web-based Environment for a Bachelor of Software Engineering Degree - DoIT”, Informing Science and IT Education Joint Conference (InSITE), Cork, Ireland, pp.1291-1299 (2002).
- [6] Brusilovsky, P.: “Adaptive Hypermedia”, Journal of User Modeling and User-Adapted Interaction, Vol.11, No.1/2, pp.87-110 (2001).
- [7] Gaines, B. R. and Shaw M. L. G.: “WebMap: Concept Mapping on the Web”, Proceedings of Second International WWW Conference (1995).
- [8] Kashihara, A., Suzuki, R., Hasegawa, S. and Toyoda J.: “A Learner-Centered Navigation Planning with Web-bases Learning Resources”, Transactions of the Japanese Society for Artificial Intelligence, Vol.17, No.4, pp.510-520 (2002).
- [9] Salton, G., Wong, A. and Yang, C. S.: “A vector space model for automatic indexing”, Communications of the ACM, Vol. 18, No. 11, pp.613-620 (1975).
- [10] 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).

高橋 雄介 Yusuke TAKAHASHI

慶應義塾大学大学院政策・メディア研究科博士課程在学中。2003 慶應義塾大学大学院政策・メディア研究科修士課程修了。データベースシステムの研究・開発に従事。情報処理学会, 日本データベース学会, 各学生会員。

清木 康 Yasushi KIYOKI

慶應義塾大学環境情報学部教授。1983 慶應義塾大学工学研究科博士課程修了, 工学博士。1984-1996 筑波大学電子・情報工学系講師, 助教授を経て, 1996 慶應義塾大学環境情報学部助教授, 1998 同学部教授。データベースシステム, 意味的連想検索, マルチメディアシステムの研究に従事。ACM, IEEE-CS, 電子情報通信学会, 情報処理学会各会員。