

# 制度文書に基づく業務データベース設計の半自動化手法 —ChatGPT を用いた構造化支援の試み—

New Method to Construct a Database Using Text Generation AI

稲倉 恒法\*, 今井 匠太朗, 高松 邦彦, 松本 清, 森 雅生  
(東京科学大学)

<抄録>

大学の制度文書に基づいて業務データベースの構造を設計する際に、文章生成 AI (ChatGPT) を活用してその構造化作業を半自動化する方法を検討した。圏論的データベースを用いた olog での関係性の記述は、整合性や意味の一貫性を確認するうえで有効であるが、文脈理解と自然文表現の両面で高い労力を要する。そこで、ChatGPT による関係文の候補生成と、それに対する人間の評価を組み合わせることで、自然文による構造記述の効率化を試みた。試行の結果、構文的・意味的に妥当な候補文が数多く得られ、文章生成 AI がデータベース設計の効率化に有効である可能性が示された。

## 1. 初めに

大学における IR (Institutional Research) や教育 DX の推進により、学内の業務に関するデータを整合性のあるかたちで統合・活用することが重要となっている。こうした整合性の高いデータ構造を実現するには、各部署の業務に基づく個別のデータベース (Database, DB) 設計を理解し、それらを分析用のデータウェアハウス (Data Warehouse, DW) と整合させる必要がある。しかし、実際には非専門家による改修やシステム間の連携の難しさから、ER 図が複雑化し、DB と DW の論理構造の整合が崩れるケースが少なくない。

この課題に対して本研究では、圏論的データベース[1, 2, 3]に着目する。圏論は、対象とその間の関係 (射) を一貫した論理構造として扱う数学理論であり、その一形式であるオントロジーログ (Ontology log, olog) は、データベースの構造を自然言語で記述・可視化できる特徴を持つ。olog では、関係性の意味的一貫性が保証されることで、柔軟かつ直観的なデータモデリングが可能となる。

ただし、関係性の記述や検証は手作業では煩雑になりやすい。そこで本研究では、自然言語の扱いに長けた文章生成 AI (ChatGPT) を用いて olog の記述を支援し、制度文書に基づく関係性の自然文表現を AI が提示し、それを人間が評価するというプロセスを導入する。制度情報に対して意味的に整合する構造を半自動的に構築できるかを検証し、その有用性と限界を考察する。

## 2. 方法と手順

Spivak と Winsnesky が提唱している圏論に基づいた形式的知識表現であるオントロジーログ[3] (Ontology Log, 以下 olog) を活用することで、より直観的なデータベース設計が可能となる。概念を対象、その関係を示す述語を射とする olog は圏をなす。olog で表現された対象をデータテーブルとし、データテーブルに張られた

---

\* 責任著者, inakura.t.b0ac@m.isct.ac.jp, inakura@gmail.com

射は概念間の関係を自然言語で記述しているため、データ参照の正確性を文章の意味や解釈の正しさによって判断することができる。このような関係を記述した概念図は、ER図やデータベーススキーマに直接対応するため、専門的なデータベースの知識がなくても設計を行うことができ、柔軟かつ直観的なデータモデリングが可能となる。

ologを用いたデータベース設計では、データ参照を行うすべてのデータテーブル間の関係性を文章化し、その妥当性を検証する必要がある。特に、大規模なデータベースを扱う業務システム(学籍管理システムや人事管理システムなど)では、関係性を記述する文章が膨大になり、手作業による管理は困難となる。そこで、我々は近年急速に発展した文章生成AIを活用し、ologの設計と文章の妥当性の検証を試みた。文章生成AIは一貫性のある文章を効率的に作成するだけでなく、文章の論理性を検証する能力も備えているため、圏論的データベースの設計をより効果的に支援できると期待される。

本稿では、大学における制度文書をもとに、ChatGPTによる業務データベース構造の構築を試みた。圏論的データベースを用いたologの記法に従い、「概念(対象)」と「関係性(射)」を矢印付きの図として表現した。作業の手順は以下の4段階からなり、その方法について概説する。

## 2.1 準備

初めに、ChatGPTに圏論的データベースやologの知識を提供した。圏論データベースやologの参考文献[1, 2, 3, 4, 5]を読み込ませて、それらの解説を求めることで、知識の注入を行った。解説が不足している箇所については、追加で質問を行うことで対話を通じた補完を行い、応答内容の妥当性を確認した。

## 2.2 概念の抽出

東京科学大学の「学修規程」および「東京科学大学の学部における学修に関する細則」をデータベース設計の出発点とした。これらの文書は学修体系を定めており、データベース化に適した概念群とその制度的関係性が明文化されている。これらの文書を読み込ませて、業務データベースにおけるテーブルに相当する「概念」を抽出した。抽出された概念は以下の14個である。[学生], [教員], [学科], [授業科目], [科目区分], [単位数], [履修履歴], [成績], [単位認定申請], [認定単位], [認定根拠], [認定審査記録], [学外学修活動], [学習成果証明書]

## 2.3 関数的関係の抽出

次に、データベースの外部キーにあたる、概念間の「関係性」を抽出した。圏論の要請から、その関係性は関数的なものに限られる。ChatGPTに、制度文書に基づいて文章化できた関係性を列挙してもらい、その妥当性を判断した。例えば、「[履修記録]は[科目]に対応している」は制度文書や業務に合致していて、妥当な関係性である。14の概念の間にある、21本の関数的な関係性を抽出できた。これらの関係性を図にしたものを図1に示す。

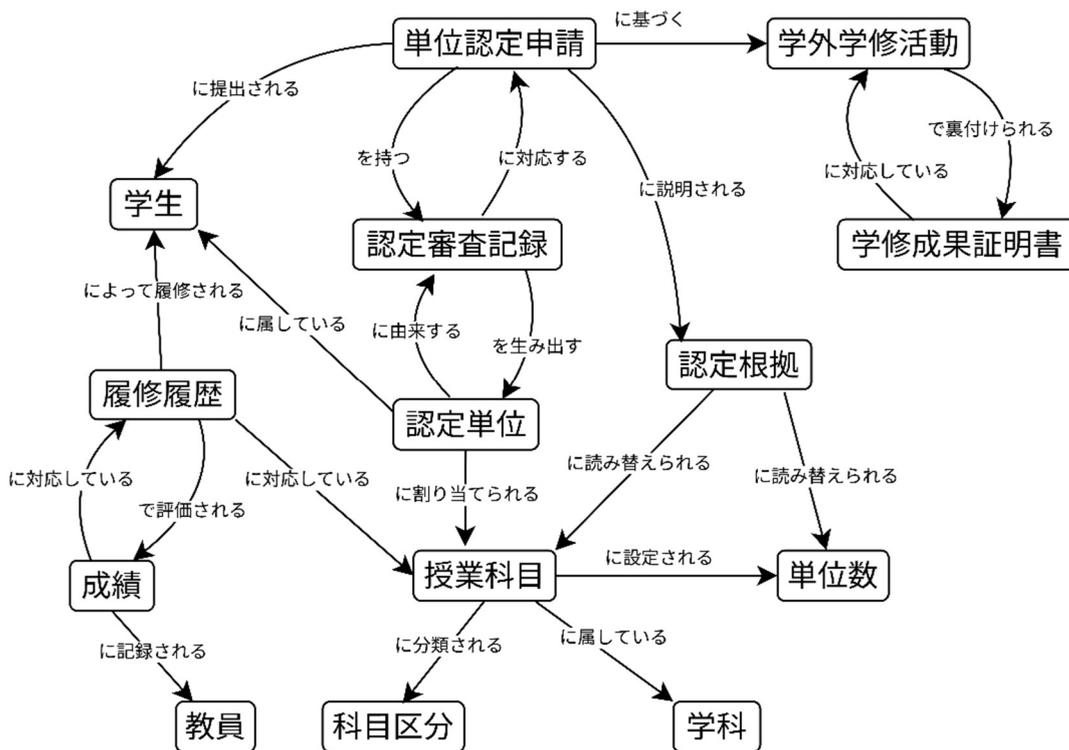


図1: 大学の単位認定業務の olog

## 2.4 意味的一貫性の検証

抽出された関係性(射)を連結することでできる連続射について、その複文を一文で表しても妥当な意味を持つことを確認する[4, 5]。これは、データベースでデータ参照が正しく行われることを保証する。長さ2~4の連続射として32本を検出し、それぞれについて複数の合成文(自然文)をChatGPTに生成させた。元の複文の意味と、合成した一文の意味を比べて、制度文書に基づく意味的一貫性があるかどうかを評価した。評価においては、中間概念が明示されていなくとも、その意味的機能が文中に含意されていれば妥当と見なし、設計者による判断で採用した。

この4段階を経て、圏論的データベース(olog)が完成した。このOlogはそのままデータベーススキーマとなるので、業務データベースの構築が完了した。

## 3. 結果と考察

自然言語の扱いに長けた文章生成AI(ChatGPT)を活用して、大学の制度文書から業務データベースの構築の半自動化を試行した。その結果、以下の五点が確認された。

第一に、制度文書からの概念の抽出、関係性の抽出、意味的一貫性の評価においてChatGPTが有用であった。どの段階でもChatGPTが作業の大半を担い、人間はその結果を判断するだけであった。ただし、これらは、準備の段階でChatGPTに十分な知見を提供できていることが前提となる。

第二に、もっとも煩雑な作業が必要となる意味的一貫性の検証において、ChatGPTは高い能力を発揮した。連続射を表す複文から、構文的にも自然で簡潔な一文となる候補を、短時間でいくつも生成した。データベース

には数多くの連続射があり、それら全てをそれぞれ一文で表して検証することは、人力で行うには膨大で煩雑な作業であるが、ChatGPT は数秒で候補となる文章を提示した。人間は提示された文章が妥当かどうかを判断するだけで良くなり、作業量が著しく軽減した。

第三に、意味的一貫性の検証において、ChatGPT が文章を生成しない場合もあった。これは、連続射が意味的に関係の薄い概念を繋いでいる場合に起きる。ChatGPT が、データの繋がりを形式的に追いかけて文章化しているのではなく、そのデータのインスタンス（実体）を考えた上で、意味のある文章を作れない、と判断したことになる。このような ChatGPT から指摘はデータベースの構造を見直す手がかりとなる。

第四に、最終的に関係性の妥当性を判断するのは人間であり、制度文書の理解や業務文脈の知識に基づく確認が不可欠であるという点も明らかになった。また、文章生成 AI はハルシネーションを起こすので、人間がその出力の正誤を判断する必要があった。

第五に、ChatGPT の出力は再現性が高くない。出力のばらつきを抑えるのは今後の課題である。

#### 4. まとめ

本研究では、大学の制度文書をもとに業務データベースを設計する手法として、文章生成 AI (ChatGPT)を用いた圏論的データベースの半自動化生成を試行した。制度文書から概念を抽出して、概念同士の関係性を文章化し、連続的な関係性の意味的一貫性を検証する、という一連の過程で、ChatGPT が候補を提示して、人間がその妥当性を判断するという役割分担が有効であることが確認された。一方で、最終的な判断は人間の理解と責任に委ねられており、AI の役割はあくまで補助的であることも確認された。

今後は、文章生成 AI の出力のばらつきを抑えることで、データベースを一意的に構築できるかを探る。また、本手法を応用することで、制度文書からの業務プロセス設計にも、文章生成 AI を活用できる可能性を追求していく。

#### <謝辞>

本研究は JSPS 科学研究費「大学 IR の学術的基盤と人材育成および関連技術の確立」課題番号 22H00077 の助成を受けた研究事業です。

#### <参考引用文献>

- [1] D.I. Spivak (2012), Functorial data migration, Information and Computation 217, pp. 31-51.
- [2] D.I. Spivak, R. Wisnesky (2015), Relational Foundations for Functorial Data Migration, arXiv:1212.5303v7 [cs.DB] 24 Jul 2015.
- [3] D.I. Spivak and R.E. Ken (2011), Ologs: a categorical framework for knowledge representation, arXiv:1102.1889v2 [cs.Lo] 7 Aug 2011.
- [4] 稲倉恒法, 今井匠太郎, 高松邦彦, 松本清, 森雅生 (2023), 圏論的データベース構成の IR への応用, 大学情報・機関調査研究集会論文集 第 12 回, pp.170-175.
- [5] 稲倉恒法, 今井匠太郎, 高松邦彦, 松本清, 森雅生 (2024), データの整合性が保証される圏論的データベース, 教育情報学会第 40 回年会論文集, pp.66-69.