

# ナレッジグラフ構築のモデル化に関する考察

## A Building Model for Knowledge Graph

朱 成敏<sup>1\*</sup> 小出 誠二<sup>2</sup> 武田 英明<sup>1,3</sup> 小柳 佑介<sup>4</sup> 西野 文人<sup>4</sup>  
Sungmin Joo<sup>1</sup> Seiji Koide<sup>2</sup> Hideaki Takeda<sup>1,3</sup> Yusuke Koyanagi<sup>4</sup> Fumihito Nishino<sup>4</sup>

<sup>1</sup> 国立情報学研究所

<sup>1</sup> National Institute of Informatics

<sup>2</sup> オントロノミー合同会社

<sup>2</sup> Ontolonomy, LLC

<sup>3</sup> 総合研究大学院大学

<sup>3</sup> SOKENDAI University(The Graduate University for Advanced Studies)

<sup>3</sup> 株式会社富士通研究所

<sup>3</sup> Fujitsu Laboratories Ltd.

**Abstract:** Knowledge graphs enable data integration, knowledge discovery and knowledge processing. Knowledge graphs have various methods for building and have various characteristics depending on the purpose of use. This paper discusses about the building model for the knowledge graph and the process of building the knowledge graph and the points to be noted.

## 1 はじめに

2012年、Googleは人や場所、ものに関する様々な情報を連携した「ナレッジグラフ」を検索における基盤技術として導入し、画期的な検索結果の多様性の確保と品質向上を目指した[1]。DBpediaやFreebase、CIA World Factbookなど、既存知識との連携により膨大なデータを対象に具体的な意味関係を表すことができた。これによりナレッジグラフはAIにおける実用的な基盤技術として注目され、高度な知識処理に用いられるようになった。また、様々な分野の知識を表現するナレッジグラフが開発され、知識を必要とする産業現場へ適用されている。

ナレッジグラフの構築は対象となる分野の知識を収集・構造化するプロセスと、データ化やサービスと運用システムの開発などの機械処理を可能とするプロセスが融合した高度なプロセスによって行われるため、構築コストが高まる可能性がある。例えば、高度な知識処理を目的とする場合は、高度な論理的演算が可能な知識構造の設計とそれをICTシステムで有効に活用できる表現手段が必要となる。そのため、ナレッジグラフの構築における様々な関連事項を明確化し、適切な構築方針を策定する必要がある。

そこで、本研究ではまず、ナレッジグラフに関連す

る既存研究を概観し、ナレッジグラフの定義とその特徴的要素について考察する。また、ナレッジグラフの構築に必要な条件について検討し、その条件を基準とした構築形態による分類を行う。最後にナレッジグラフの利用事例から利用形態による分類を行い、それぞれの分類の関連性と構築に必要な関連事項について議論する。

## 2 グラフに基づく知識表現

知識をグラフとして表す手法は人間の認知行動や連想思考を直感的に表現できるメリットがあり、多様な分野において知識表現の手段として活用されてきた。本章では情報技術分野を含め、これまでの様々な分野におけるナレッジグラフについて概観し、ナレッジグラフの特徴的要素について考察する。

### 2.1 意味ネットワーク

意味ネットワークは概念と概念をリンクで結びつけることで意味を表す手法であり、計算機科学や言語学、認知科学などの分野で意味の表現と解析を目的として用いられた。本章では1990年代までの意味ネットワークに関する研究について述べ、その特徴について考察する。

\*連絡先：国立情報学研究所  
〒101-8430 東京都千代田区一ツ橋 2-1-2  
E-mail: joo@nii.ac.jp

Fillmore は語と語の間が持つ意味関係と文の意味構造を表すために単語間の意味的關係を表す格 (case) を提案し、言語学における意味の表現手段としてグラフを用いた。動詞と文意を表す 6 つの深層格を組み合わせる各文法 (case grammar) を用いて文が持つ意味を表した [2]。Schank は認知科学の側面から言語や表記が異なる文章も概念が持つ意味に注目すると同じ意味を持つことに着目し、概念を用いて文の意味を説明する概念依存 (Conceptual Dependency) 理論を提案した。基本行為 (primitive acts) のパターンと関係の分類 (primitive conceptual category) を用いて文章の意味をグラフ化した [3]。計算機科学では Quillian は言語理解システムで用いられる語彙を定義するために意味ネットワークを導入した。Quillian は概念を定義する Type と Type を表示する Token の 2 つのノードと subclass のリンクを用いて下位概念を表すことで情報の意味ネットワークを構成した [4]。

上記の意味ネットワークに関連する研究は情報を概念同士の関連性を用いてグラフ化することで人間の思考を直感的に表すことができた。情報の表現や解析、新たな発見に繋がるネットワークの構築は既存の知識を可視化する出発点となった。しかし、これらの研究は統語論 (syntax) と意味論 (semantics) が区分されていない問題点がある。統語論の観点から構築された意味ネットワークは文章の表現に依存する傾向があり、適用範囲が狭く、論理的基盤としての運用が困難な場合がある [5]。リンクで表す概念間の関連性が人間の主観によって解釈が異なる場合もあり、意味の曖昧さを改善するためにクラスとインスタンスを区分する必要がある [6]。後節では統語論に注目した場合と意味論に注目した場合の意味ネットワークについて述べる。

### 2.1.1 統語論に注目した意味ネットワーク

統語論に注目した意味ネットワークは自然言語処理分野やエキスパートシステムの関連分野で活発に行ってきた。Hoede らは 1980 年代後半から「Knowledge Graph Project」というプロジェクトを行い、医学と社会科学のような専門分野のテキストをグラフ化する研究を行なった [7]。最初は因果関係に着目し、テキストから「causes」と「by」を抽出し、関係を表すリンクとして用いた。その後は OAV(object - agent - verb), PAR(part of), AKO(a kind of) のリンクを、最終的に EQU(equality), ORD(ordering), SUB(subset relationship), CAU(causality), ALI(alikeness), PAR (attribute part), DIS(disparateness), SKO(information dependency) の 8 件のリンクと各フレームを用いて知識をグラフ化した。

これらの研究は意思決定を支援するエキスパートシステムの基盤知識を目的としたが、統語論的側面を重

視していたため論理処理を行うための機能改良が必要となった。そこで階層的な構造を持つ単語の集合である Word Graph を構築し、文意を把握するための手段として用いた [8][9]。また、専門知識について文章単位で意味構造を表したグラフ (彼らはこれを Knowledge Graph と呼ぶ) を別途に用意し、文章の意味関係を把握する手段として用いた [10]。これらの研究は論理演算などの高度な機能を遂行するためには統語論に注目した意味ネットワークだけでは限界があり、意味論的な要素を必要としていた。

### 2.1.2 意味論に注目した意味ネットワーク

Brachman は意味ネットワークを実装レベル (Implementational level), 論理レベル (Logical level), 概念レベル (Conceptual level), 言語レベル (Linguistic level), 認識論レベル (Epistemological level) 5 つの階層で役割を明確にし、既存の意味ネットワークの意味的曖昧さについてこれらの役割が混在していることが原因だと指摘した [11]。そして、認識論レベルの意味ネットワークである KL-ONE を提案した [12]。KL-ONE の構成要素は概念と Role であり、Role には名称以外に、数について制限を持つ (T-Box)。また、Role の関係性を表す (Structural Description) ができ、これらを論理的基盤として推論を行う (A-Box) [13]。

一方、Sowa は述語論理を基盤とした知識表現システムである概念グラフ (Conceptual Graph) を提案した [14][15]。概念グラフは各フレームを用いて意味をグラフ化し、簡潔に表現することができる。また、各概念を階層構造 (Type-Hierarchy) として表すことで述語論理の推論が可能となる。概念の階層構造は多重継承や継承の制限を表現でき、KL-ONE が脆弱した概念の継承を補完した。KL-ONE と概念グラフのような意味論に注目した意味ネットワークは認知科学的側面に注目し、推論のような知識の論理的な利活用に適合する特徴がある。しかし、表現の柔軟性には差があり、概念グラフの場合、格フレームとリファレントを用いて概念と関連性の定義を柔軟に表現することができるが、KL-ONE は記述の便利さや直感的な理解が比較的制限的である。こういった特徴は自然言語への変換において妨げとなる [16]。

## 2.2 エキスパートシステムにおける統語論と意味論

これまで意味ネットワークに関連する研究の一部はエキスパートシステムの開発における基盤技術として行われてきた。エキスパートシステムの開発には自然言語の処理と論理的な推論が必要となり、統語論的な

側面と意味論的な側面の両方を有効に活用できる意味ネットワークの構築が最も重要である。前節で述べたように、性能向上のためには統語論と意味論の両方の特徴を取り入れた意味ネットワークが必要となった。また、統語論に注目した場合は知識の記述と表現において自由度が高く、直感的に理解を促す傾向があり、意味論に注目した場合は推論のような論理的に高度な知識処理を可能とする傾向があった。

エキスパートシステムに関する研究の成功は推論能力だけではなく、実世界に存在する知識を知識ベース化に着目したことに原因があった [17]。しかし、当時の知識ベースの知識量には限界があり、特定分野のような狭い世界を対象とする問題点があった。標準となる形式や運用方針がなく、それぞれ独自で開発されたため知識ベース間の連携が困難であった。これらの問題点がエキスパートシステムに関する研究が沈静化した原因だとは言い難いが、膨大な知識と表現形式の標準化が知識処理において重要な条件だったと考えられる。

## 2.3 Semantic Web と Linked Data

インターネットの発展と同時に Web の情報量は急速に増加し、ユーザは Web ブラウザを通じて様々な分野の情報を容易に取得することができるようになった。さらにインターネットの世界に存在する膨大な情報をより効率的に活用する仕組みとして Tim Berners-Lee は Semantic Web を提唱した [18]。Semantic Web はタグを利用した意味関係の定義、ハイパーリンクを用いたリソースの連携などの環境により Web の情報をデータとして機械処理が可能な形にした。また、XML や RDF, OWL のような情報の記述形式における標準を提示することで Web 上に存在するデータを容易に利用できるようにした。

そして、Semantic Web で提案された関連技術を用いて Web に存在する様々な情報を構造化し、連携して大規模なデータとして利活用する仕組みとして Linked Data が提案された [19]。Linked Data は 1) 識別子の使用, 2) HTTP URI による参照とアクセス, 3) RDF や SPARQL などの標準的な形式で情報提供, 4) 他の情報源へのリンクによるアクセスと情報発見の支援の 4 つの原則が提示された。Semantic Web の概念で構造化された知識の新しい表現形式である Linked Data は統語論的要素と意味論的要素を融合した膨大な知識ベースであり、機械によるアクセスと処理も容易である。すなわち、これまでの意味ネットワークとエキスパートシステムが持つ問題を同時に解決することができた。

## 3 ナレッジグラフ

ナレッジグラフという概念は Google により実世界のオブジェクトの検索を可能とするものとして紹介されて以来、様々な研究や産業分野に急速に普及された。また、AI の基盤技術として様々な分野に用いられながらナレッジグラフは多様な形で用いられることになった。利用場面によっては作業工程のグラフやオントロジー、Linked (Open) Data など様々なものがナレッジグラフとして呼ばれており、近年は「グラフ化した知識」といった表層的な意味で認識されている。また、Linked Data の形式のデータを意味する場合も多い [20]。そこで、本章ではナレッジグラフの構築をモデル化するためにまず、ナレッジグラフの定義とその特徴的要素について考察する。

### 3.1 ナレッジグラフの定義

Ehrlinger は様々なナレッジグラフに関する研究を調査し、それぞれの研究で用いられたナレッジグラフの意味を紹介した [21]。2013 年に Pujara らは Web から知識を抽出し、相互に関連付けをしたものをナレッジグラフとし、Web からナレッジグラフを生成する研究を行った [22]。Paulheim は 2016 年にナレッジグラフの洗練に関する研究にて「(i) 実世界のエンティティと相互関係をグラフと説明し、(ii) エンティティのクラスと関係をスキーマで定義し、(iii) 任意のエンティティ同士に潜在的な関連付けができ、(iv) 様々なドメインに対応する」の 4 つの条件を満たすものとして定義した [23]。Journal of Web Semantics は 2016 年のナレッジグラフ特集を紹介しながら「ナレッジグラフは、エンティティとそのセマンティックタイプ、プロパティ、およびエンティティ間の関係を表した大きなネットワークである」と説明した [24]。Färber らは 2018 年に発表した Linked Data の質に関する研究にてナレッジグラフを RDF グラフとし、RDF トリプルの集合として定義した [25]。これらの定義から考察してみると、それぞれの表現は異なるが、ナレッジグラフの定義は知識となる情報をグラフ化 (またはネットワーク化) したものの ([20], [23], [24]) と知識となる情報をデータ化したもの ([21], [22]) として大別される。

前章で述べた述べたように、これまでのグラフ基盤の知識表現では統語論と意味論に注目して説明することができたが、ナレッジグラフの場合は知識のグラフ化とそのデータ化に注目している傾向がある。これは Semantic Web の概念が普及され、機械における知識基盤の環境を大きく変化させたからだと考えられる。膨大な情報をインターネットを通じて用いることができ、またこれらの情報を標準形式や関連ツールによって容易に問い合わせることができるようになった。すなわ

ち、システムがインターネットを通じて知識の活用ができるように共通の形式が基本条件となっていることがナレッジグラフにおける特徴的要素である。もう一つの特徴は意味論的構成要素である。ナレッジグラフは知識となるエンティティをグラフ化して表したもものとして認識されている。これは前章で検討した意味論に注目した知識表現に比べて、ナレッジグラフは簡潔な構造で知識を表現していることを意味する。

以上の考察から本研究ではナレッジグラフについて次のように定義を行う。

1. ある実世界の知識をエンティティ間の関連性で定義したものであり、
2. それを機械処理が可能な形式にしたもの。

そして、定義の1)から「知識のグラフ化」を、定義の2)から「機械処理」を特徴的要素とする。

## 3.2 ナレッジグラフの特徴的要素

本節では前節で定義したナレッジグラフの項目について考察し、ナレッジグラフ構築との関連性について述べる。

### 3.2.1 知識のグラフ化

前節ではナレッジグラフを定義する条件の一つとして「ある実世界の知識をエンティティ間の関連性で定義したもの」を提案した。これはエンティティそのものに対する定義、またその関連性を定義することで知識間の関係を明確にしながらグラフ化することを意味する。そして、共通の形式を用いることで他のナレッジグラフと連携することで知識のグラフは拡張することもできる。この発展段階を整理すると、「1) エンティティそのものを定義している、2) 他のエンティティとの関連性を定義している、3) 外部のナレッジグラフとの関連性を定義している」となる。本研究では「知識のグラフ化」における構築レベルの1), 2), 3)とする。

### 3.2.2 機械処理

機械処理は本研究が提案するナレッジグラフの2つ目の特徴的要素である。グラフ化した知識を表す方法には様々な形式があり、目的や利用場面、運用されるシステム環境などによって適合する形式を選択することができる。そこで、本研究ではナレッジグラフを機械処理を行うガイドラインとして、「1) 機械可読性がある、2) 識別が可能な形式を持ち、インターネットでのアクセスができる、3) 外部のナレッジグラフと連携し

ている」の3つの発展段階を提示する。まず、機械処理を行うためには機械可読が第一条件である。狭い範囲での利用目的によってはオフライン環境での運用も可能であるが、ナレッジグラフの公開や外部情報との問い合わせなど広い範囲での利活用の場合は、インターネットを基盤とする運用が必要である。この場合はそれぞれの知識に対して識別が可能な形式で公開することが必要となる。また、共通の形式を利用することで他のナレッジグラフと連携することでより規模の大きなナレッジグラフを構築することでできる。本研究では、この3つの発展段階を「知識のグラフ化」における構築レベルの1), 2), 3)とする。

## 3.3 考察

本章ではナレッジグラフの定義と特徴的要素について述べた。ナレッジグラフの関連研究は「グラフ化した知識をICTシステムで利用する」という認識を共有しながら様々な形式や運用システムの開発などを提案しながら発展してきた。こういったナレッジグラフが持つ意味の柔軟性はWebや知識基盤など、既存の関連技術と融合し、様々な形で発展してきた。本節では既存の関連技術との関係性について述べる。

### 3.3.1 ナレッジグラフと Semantic Web

Semantic WebはWeb上の膨大なデータに対し、タグを利用した意味関係の定義、ハイパーリンクによるリソースの連携などのナレッジグラフの基盤となる技術を提供した。Semantic Webというパラダイムの登場はナレッジグラフの大きな背景となり、普及を促進させたきっかけである。Semantic Webは膨大なデータのcWebであり[26]、特定ドメインや目的によって構造化された知識とは言い難い。ナレッジグラフはSemantic Webの概念を基盤とした表現された知識として考えられる。

### 3.3.2 ナレッジグラフと Linked Open Data, オントロジー

Linked Open Data(LOD)が提示している4つの原則は本研究が提案するナレッジグラフにおける機械処理の条件を満たしており、ナレッジグラフを機械処理を行うために必要な表現形式として適合している。そのため、既存のLODの中でエンティティの定義やエンティティ間の関連性を定義したものはナレッジグラフともみなされる。現在、DBpedia<sup>12</sup>を中心に様々な

<sup>1</sup><https://wiki.dbpedia.org/>

<sup>2</sup><https://ja.dbpedia.org/>

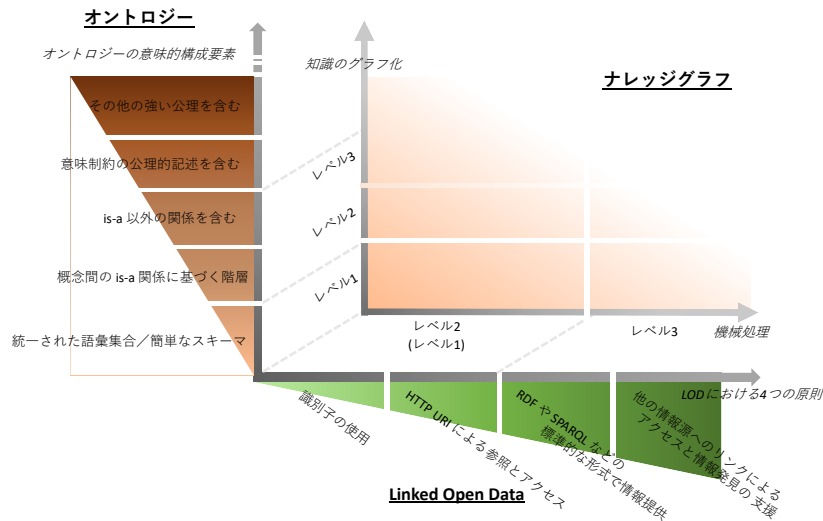


図 1: オントロジー, Linked Open Data とナレッジグラフの特徴的要素.

LOD が連携されており, 膨大なナレッジグラフとしての役割を果たしている.

オントロジーは知識を体系化するために最も有効な手段である. 概念間の関連性を定義することがナレッジグラフにおける「知識のグラフ化」の過程と類似している. 古崎ら [27] はオントロジーを分類する意味的構成要素として「0) 統一された語彙集合/簡単なスキーマ, 1) 概念間の is-a 関係に基づく階層, 2) is-a 以外の関係を含む, 3) 意味制約の公理的記述を含む, 4) その他の強い公理を含む」の 0 から 4 までの条件を提示した. 本研究で提案するナレッジグラフの「知識のグラフ化」の構築レベルは, オントロジーの意味的構成要素の 0), 1), 2) に該当している. このようなナレッジグラフとオントロジーの意味論的な類似性は以下の 2 つを示している. まず, オントロジーは RDF のように機械処理ができる形式で公開することでナレッジグラフとしての利用が可能となる. そして, ナレッジグラフの構築において意味論的構成要素を厳密にすればオントロジーのような高度な知識処理が可能となる.

LOD とオントロジーはそれぞれナレッジグラフにおける機械処理と知識のグラフ化の構築に関係しており, ナレッジグラフの研究や開発において重要な技術分野である (図 1).

## 4 ナレッジグラフの分類

前章ではナレッジグラフの定義について考察し, ナレッジグラフの特徴的要素を提案した. 本章では特徴的要素を用いてナレッジグラフの構築形態による分類と利用形態による分類を行い, 既存の利用事例を通じて構築におけるガイドラインとして提案する.

### 4.1 構築形態による分類

前章で提案した 2 つの特徴的要素を持つ 3 つの構築レベルをそれぞれの軸として表すと 9 つに分類することができる (図 2). 本研究では, これを「構築形態による分類」とする. それぞれの分類は構築における異なる特徴を持ち, こういった特徴は様々な利用目的を果たすために達成する構築条件ともなる. この条件を達成するために必要な関連技術や適合するデータの形式を検討し, 構築方針を決めることができる.

		知識のグラフ化 構築レベル		
		① エンティティそのものを定義している	② 他のエンティティとの関連性を定義している	③ 外部のナレッジグラフとの関連性を定義している
構築形態レベル	① 機械可読性がある	分類 1	分類 2	分類 3
	② 識別可能な形式を持ち, インターネットでのアクセスができる	分類 4	分類 5	分類 6
	③ 外部のナレッジグラフと連携している	分類 7	分類 8	分類 9

図 2: ナレッジグラフの構築形態による 9 つの分類.

### 4.2 利用形態による構築条件

ナレッジグラフは知識処理に関連する様々な分野で用いられており, 利用場面によって多様な形式で運用されている. 本研究では知識の利用形態に注目してナレッジグラフを 3 つの分類に区分する. そして, 開発事例に対し, 前節で提案した構築形態による分類を用いてそれぞれの利用形態に必要な構築条件について検討する.

#### 4.2.1 知識の定義

エンティティの定義はナレッジグラフにおける最も基本的な役割である。そのため、共通語彙の構築や情報の標準化のようにある分野の知識を定義する手段として有用な表現手段である。そして、エンティティ間の関連性を定義することで異なるコンテンツ間の連携が可能となる。さらに、外部のナレッジグラフと連携することで語彙が持つ意味の多様性や関連情報の問い合わせにも対応できる。

その例として農作業と農作物における標準語彙として開発された農作業基本オントロジー [28] と農作物語彙体系 [29] が挙げられる。農作業基本オントロジーは概念となる農作業に対する定義と、定義に用いられた属性の包含関係を用いて概念間の関連性を定義した。農作物語彙体系は植物学的な基準を用いて農作物名を定義して分類を行った。そして、語彙の定義以外にも外部のナレッジグラフや語彙との連携されており、それぞれのエンティティは識別可能な形式で公開されている。すなわち、「知識のグラフ化」は構築レベル3を、「機械処理」は構築レベル3に該当するナレッジグラフである(図3)。

「知識の定義」を利用目的とする場合は、「知識のグラフ化」の構築レベル1「エンティティそのものを定義している」が基本条件である。さらに上位レベルの条件を満たすと様々な機能の拡張が可能となる。農作業基本オントロジーと農作物語彙体系は「知識のグラフ化」の構築レベル3を達成することで、本来の利用目的である共通語彙として役割とともに関連情報の検索や推論のような知識処理も可能となった。

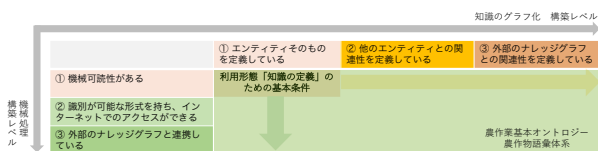


図 3: 利用形態「知識の定義」の基本条件と開発事例

#### 4.2.2 知識の獲得

エンティティ間の関連性を定義し、知識をグラフ化すると、必要とする情報とともに関連情報や意味に基づく検索など高度な検索が可能となる。そのため、ナレッジグラフは情報の閲覧や検索において有効な情報、かつ豊富な関連情報の検索を可能とする。最初に Google が検索にナレッジグラフを導入した事例も代表的な「知識の獲得」の例である。

「知識の獲得」を目的とする開発された事例として LODAC Museum[30] が挙げられる。LODAC Museum

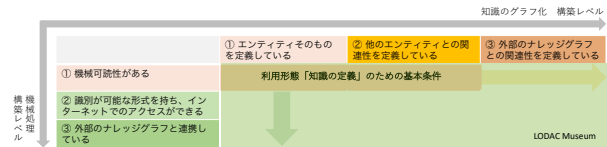


図 4: 利用形態「知識の獲得」の基本条件と開発事例

は日本国内の博物館における様々なコレクション情報を収集し、LOD化したものである。また、DBpediaの膨大な情報と連携することで博物館情報を統合的に活用することを可能とした。それぞれの情報はURIを持ち、RDF形式で公開されている。これにより外部のシステムがLODAC Museumの情報を自由に問い合わせることができる。もう一つの事例としてYokohama Art Spot<sup>3</sup>がある。LODAC Museumが提供する博物館の情報とイベント情報(ヨコハマ・アート LOD)、質問応答など異なる分野のデータをRDFストアとSPARQLエンドポイントという共通の形式とツールを用いて様々な情報が獲得できるようにした。LODAC Museumは「知識のグラフ化」は構築レベル3を、「機械処理」は構築レベル3に該当すると考えられる(図4)。

ナレッジグラフの利用形態が「知識の獲得」の場合は、獲得の対象となるエンティティに対する定義とエンティティ間の関連性が定義が条件であり、「知識のグラフ化」の構築レベル2が要求される。さらに外部ナレッジグラフと連携することで問い合わせする範囲が拡張され、有効な知識を獲得することができる。LODAC MuseumはDBpediaとの連携することで博物館のコレクション情報から作品や作者、所蔵館など様々な関連情報の検索を可能とし、高度な知識の獲得を可能とした。

#### 4.2.3 知識の拡張

ナレッジグラフは他のナレッジグラフとの関連性が定義できれば、連携することができ、多様な知識を統合的な利活用を可能とする。また、共通の形式やツールなどの他のナレッジグラフやICTシステムから参照できる環境を提供すれば、Web上の膨大な知識源の一部となる。このように他のナレッジグラフとの連携によって新しい知識の発見や推論などの高度でかつ広範囲な知識を利活用する利用形態を「知識の拡張」とする。

DBpediaは「知識の拡張」を目的とする代表的な例である。DBpediaはWikipediaの記事について項目や説明、記事間の関連性を抽出し、グラフ化した知識として提供している。また、RDFの形式で構造化データを公開しており、様々なナレッジグラフやデータセットから参照されている。現在は膨大な知識源となるナレッジグラフとして、またナレッジグラフのハブとし

<sup>3</sup><http://lod.ac/apps/yas/>

ての役割を果たしている。DBpediaは「知識のグラフ化」と「機械処理」の両方、構築レベルが3である。

利用形態が「知識の拡張」の場合、他のナレッジグラフとの連携を想定し、RDFによる構造化やSPARQLエンドポイントといった共通の問い合わせ環境が必要となる。そのためにはそれぞれの構築レベル3の条件を満たす必要があると考えられる。

## 5 考察

本章では本研究で提案したナレッジグラフの特徴的要素と構築レベルを実際の構築プロセスに適用して関連事項を検討する。

### 5.1 構築プロセスにおける特徴的要素

ナレッジグラフの構築モデルに関する先行研究を通じてナレッジグラフの構築プロセスを「1)知識の観察、2)知識の体系化、3)知識のICT化、4)知識の社会化」の4つの過程でモデル化することを提案した[31]。本節ではナレッジグラフの構築に直接に関わる「2)知識の体系化」と「3)知識のICT化」のプロセスを中心に構築プロセスにおける特徴的要素について述べる。そして、実際の開発事例を通じてプロセスのコストについて考察する。

「知識の体系化」は前段階である「知識の観察」の「問題の認識」から提示された問題を解決するためのナレッジグラフを設計するプロセスである。このプロセスには調査を通じて分野の知識を明確にする「知識の明確化」と知識の定義や知識間の関連性を定義することでナレッジグラフを構築する「ナレッジグラフの構築」の2つのプロセスで構成される。このプロセスは本研究が提案するナレッジグラフの特徴的要素の一つである「知識のグラフ化」を実際に行うプロセスである。対象となるエンティティの数と「知識のグラフ化」の構築レベルによって構築コストが大きくなるのが要素される。

「知識のICT化」は「知識の体系化」で構築されたナレッジグラフをデータ化し、データを用いて知識の利活用が可能な関連サービスを開発する2つのプロセスがある。「データ化」では機械処理を行うために適切な形式で開発を行う。そのため、「知識のICT化」のプロセスではナレッジグラフにおける「機械処理」の構築レベルにつれてコストの増加が考えられる。

図5は先行研究で提示したナレッジグラフの構築プロセスに、今回提案した特徴的要素の構築レベルを追加した図である。また、先行研究を通じて「知識の体系化」と「知識のIC化」のプロセスには意味論的観点が必要であることがわかった。したがって、ナレッジ

グラフの構築にも意味論的観点が大きく影響を与えると考えられる。

### 5.2 適正レベルに関する考察

本研究では利用形態による分類について基準条件となる構築レベルを提示した。ナレッジグラフを用いた標準語彙の開発は利用形態として「知識の定義」と分類され、それぞれの構築レベル1であるエンティティの定義と機械可読性を基本条件とした。本節では基本条件の以上を満たした場合について開発事例を通じて考察する。

農作業基本オントロジーは農作業名における標準語彙として開発され、ナレッジグラフとして公開された。農作業基本オントロジーは農作業に対して10個の属性を用いて定義をした。また、属性の値が持つ包含関係により農作業間の関連性を定義し、階層構造で表した知識体系である。それぞれの農作業はURIを持ち、関連する外部の情報体系とも連携されている。データはOWL形式で公開されている。農作業基本オントロジーは「知識のグラフ化」はレベル3、「機械処理」はレベル2の条件である。そのため、構築の目的であった標準語彙としての利活用の以外に、WikipediaやAGROVOCのような外部情報を含め、農作業に関する様々な情報システムとしての利活用も提供している。しかしながら、情報の修正や追加など更新を行う場合、既存のエンティティとの関連性に対して確認を行い、エンティティ間の整合性を判断する必要がある。場合によっては階層構造の変更と論理的検証を行う必要もあり、更新作業に多くの時間が必要となった。

構築の目的と利活用の方法を想定して提示基準を参考にして構築すれば、コストの調整が可能だと考えられる。農作業基本オントロジーに比べて簡単な構造を持つ農作物語彙体系の場合は、農作業基本オントロジーより「知識の体系化」のプロセスは約5割、「知識のICT化」のプロセスは約3割の期間でプロセスを完了することができた。

## 6 今後の課題

本研究では主に研究目的で開発されたナレッジグラフを対象とした。産業分野のナレッジグラフ、特に利益を目的とするナレッジグラフの開発では構築期間や予算など、より厳密なコストと条件が構築に影響を与えていると考えられる。今後、様々な産業現場におけるナレッジグラフの開発プロセスを調査し、関連する要因についても明確化したい。

今回は構築形態による分類と利用形態による分類を行い、構築に必要な基本条件を提示した。より多くの

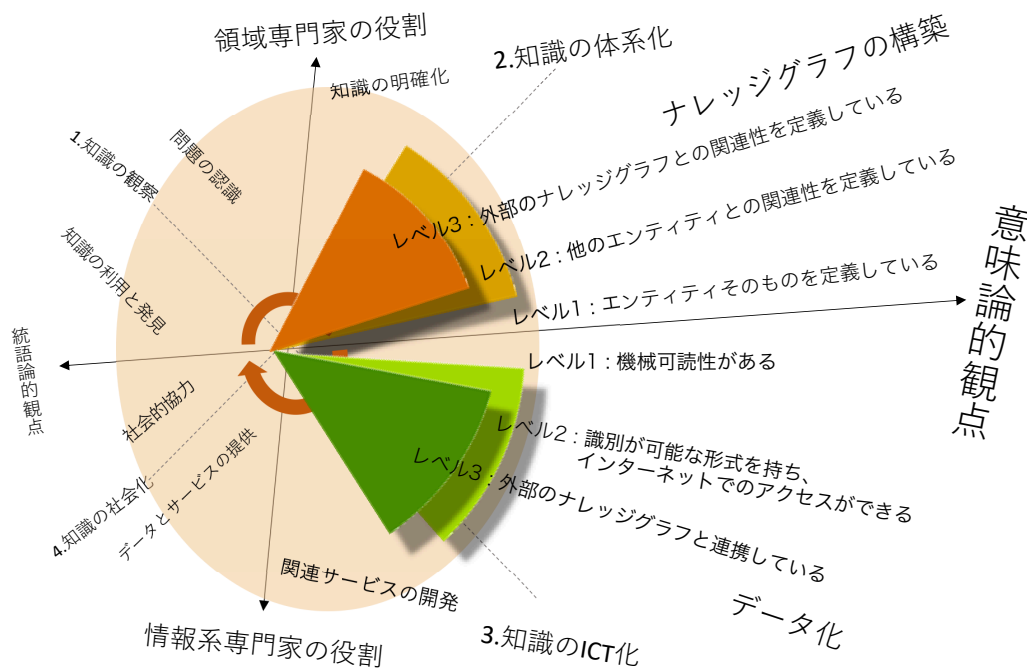


図 5: ナレッジグラフの構築プロセスにおける特徴的要素.

開発事例に適用し、分析することでナレッジグラフに対する評価モデルへの発展も考えられる。また、ナレッジグラフの構築目的によって構築プロセスにかかる時間などのコストが異なるため、既存ナレッジグラフの分析は開発期間の予測における判断指標として用いられると考えられる。今後はナレッジグラフの構築に関わる様々な要素を整理し、構築と評価における実践的なモデルを提案したい。

## 7 おわりに

本研究ではナレッジグラフの意味について考察し、その定義からナレッジグラフの構築における特徴的要素を提案した。それぞれの特徴的要素を構築レベルとして表すことで構築形態における分類を、様々な利用事例から利用形態による分類を提示することができた。今後は多くの事例を分析することでナレッジグラフの構築における総合的な評価モデルを提案していきたい。

## 参考文献

[1] A. Singhal, Introducing the knowledge graph: things, not strings, <<https://googleblog.blogspot.com/2012/05/>>

introducing-knowledge-graph-things-not.html> 2020年2月2日参照.

[2] Fillmore, C. J. : Toward a modern theory of case, *Modern Studies in English*, pp. 361-375 (1966)

[3] Schank, Roger C. : Conceptual dependency: A theory of natural language understanding, *Cognitive psychology* 3.4, pp. 552-631 (1972)

[4] Quillian, M. R. : Semantic Memory, in M. Minsky (ed.), *Semantic Information Processing*, pp. 216-270 (1968)

[5] Woods, William A. : What's in a link: Foundations for semantic networks, *Representation and understanding*, Morgan Kaufmann, pp. 35-82 (1975)

[6] Brachman, R. J. : What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks, *Computer*, (10), pp. 30-36 (1983)

[7] Vries, P. H. D. : Representation of scientific texts in knowledge graphs, Doctoral dissertation, University of Groningen (1989)

[8] Hoede, C., Liu, X. : Word graphs: The second set. In *International Conference on Conceptual Structures*, Springer, pp. 375-389 (1998).



- [9] Hoede, C., Zhang, L. : Word graphs: The third set. In International Conference on Conceptual Structures, Springer, pp.15-28 (2001)
- [10] Hoede, C. : Structuralistic Linguistics: The case of knowledge graph theory. Memorandum Afdeling TW (2005)
- [11] Brachman, R. J. : On the epistemological status of semantic networks, In Associative networks, Academic Press, pp. 3-50 (1979)
- [12] Brachman, R. J., Schmolze, J. G. : An overview of the KL-ONE knowledge representation system, In Readings in artificial intelligence and databases, Morgan Kaufmann, pp. 207-230 (1989)
- [13] Brachman, R. J., Fikes, R. E., Levesque, H. J. : Krypton: A functional approach to knowledge representation. Computer, (10), pp. 67-73 (1983)
- [14] Sowa, J. F. : Conceptual structures: information processing in mind and machine, Addison-Wesley (1983)
- [15] Sowa, J. F., Way, E. C. : Implementing a semantic interpreter using conceptual graphs. IBM Journal of Research and Development, 30(1), pp.57-69 (1986)
- [16] Biebow, B., & Chaty, G. : A comparison between conceptual graphs and KL-ONE, In International Conference on Conceptual Structures, Springer, pp.75-89 (1993)
- [17] 溝口理一郎, 池田満, 来村徳信 : 対象モデリングの視点から見た知識表現: 哲学と AI における対象世界モデリング, 人工知能学会誌, 18(2), pp. 183-192 (2003)
- [18] Berners-Lee, T., Hendler, J., Lassila, O. : The semantic web. Scientific american, 284(5), pp. 34-43 (2001)
- [19] Berners-Lee, T. : Linked Data, <<https://www.w3.org/DesignIssues/LinkedData.html>>, 2020年2月2日参照.
- [20] McCusker, J. P., Erickson, J., Chastain, K., Rashid, S., Weerawarana, R., McGuinness, D. : What is a knowledge graph. Semantic Web Journal (2018)
- [21] Ehrlinger, L., Wöß, W. : Towards a Definition of Knowledge Graphs, SEMANTiCS (Posters, Demos, SuCCESS), 48, 1-4 (2016)
- [22] Pujara, J., Miao, H., Getoor, L., Cohen, W. : Knowledge graph identification. In International Semantic Web Conference, Springer, pp. 542-557 (2013)
- [23] Paulheim, H. : Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods, Semantic Web Journal, pp. 1-20 (2016)
- [24] Call for Papers : Journal of Web Semantics Special Issue on Knowledge Graphs, <[https://iccl.inf.tu-dresden.de/web/JWS\\_special\\_issue\\_on\\_Knowledge\\_Graphs](https://iccl.inf.tu-dresden.de/web/JWS_special_issue_on_Knowledge_Graphs)>, 2020年6月24日参照.
- [25] Färber, M., Bartscherer, F., Menne, C., Rettinger, A. : Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago. Semantic Web, 9(1), pp. 77-129 (2018)
- [26] W3C : LINKED DATA, <<https://www.w3.org/standards/semanticweb/data>> 2020年6月22日参照.
- [27] 古崎晃司, 来村徳信, 溝口理一郎 : Web2.0時代のオントロジー利用雑感-ライトウエイトからヘビーウエイトまで. 第14回セマンティックウェブとオントロジー研究会, SIG-SWO-A602-06 (2006)
- [28] 朱成敏, 小出誠二, 武田英明, 法隆大輔, 竹崎あかね, 吉田智一 : 農業データの連携における農作業の標準語彙体系の構築. 農業情報研究, 28(3), pp.143-156 (2019)
- [29] 朱成敏, 武田英明, 竹崎あかね, 吉田智一 : 農業ITシステムのデータ連携のための農作物語彙体系の構築. 人工知能学会研究会資料, 44(10), pp.1-8, (2018)
- [30] 松村冬子, 嘉村哲郎, 加藤文彦, 小林巖生, 高橋 徹, 上田 洋, 大向一輝, 武田英明 : LODAC Museum: Linked Open Data による博物館情報の統合と活用, 人工知能学会第26回全国大会論文集, 3C2-OS-13b-9 (2012)
- [31] 朱成敏, 小出誠二, 武田英明, 法隆大輔, 竹崎あかね, 吉田智一 : 農業ナレッジグラフの構築に関する考察による領域ナレッジグラフの構築モデルの提案. 人工知能学会研究会資料, 47(10), pp. 1-10, (2019)