

農業データ連携のためのナレッジグラフに基づく標準語彙の運用

Establishment of Core Vocabulary based on Knowledge Graph for Interoperability between Agricultural Data

朱成敏*1

Sungmin JOO

武田英明*1*2

Hideaki TAKEDA

竹崎あかね*3

Akane TAKEZAKI

吉田智一*3

Tomokazu YOSHIDA

*1国立情報学研究所

National Institute of Informatics

*2総合研究大学院大学

SOKENDAI University(The Graduate University for Advanced Studies)

*3農業・食品産業技術総合研究機構

National Agriculture and Food Research Organization

Agricultural ICT system has been accompanied by generating data such as farming information or agricultural observations. These data is expected to be used for statistics, analysis and prediction for area agriculture. However, vocabulary in the agriculture domain is treated with no standard information, data in agricultural ICT systems is not easy to integrate and maintenance. In this paper, we investigate the existing vocabulary system for agriculture domain. Then we propose the core vocabulary system based on knowledge graph by paying attention to interoperability between agriculture data.

1. はじめに

近年、日本の農業分野は ICT 基盤技術の普及により大きな進歩を遂げている。センサーネットワークによるモニタリングや営農管理システムを用いた圃場管理、栽培の自動化など生産性を従来より高めている。そして、農業 ICT システムから様々なデータが発生されており、そのデータを分析することによって作業の最適化や収穫量の予測など新しい発見が可能となる。それぞれの農業 ICT システムから発生したデータを連携・統合し、膨大なデータとして活用すれば今までより正確で多様な情報を予測することも可能となる。これまで人間の経験と勘で行われてきた意思判断を蓄積されたデータに基づいて行い、より精度の高い意思決定に導く、いわゆる知識基盤の農業が可能となる。

一方、これまでの農業 ICT システムはデータ項目の作成において基準となる情報がないためシステムベンダーは独自の定義を行い、データ項目の設計を行った。現在はそれぞれの農業 ICT システムは独自のデータ項目を持ち、その項目に従ってデータを作成しているため、異なるシステムから発生したデータの場合は連携・統合が困難となった。こういった背景から内閣官房 IT 総合戦略本部は農業分野における情報創成・流通の促進のために農業情報の相互運用性等の確保に資する標準化や情報の取扱いに関する政府横断的な戦略を策定した [内閣官房 17b]。しかし、現状は公的機関による統計調査の項目や食品安全に関する語彙が浸透しているため、農業 ICT システムと実際の現場の語彙管理に差が生じている。こういった語彙間の関連性について対応関係が定義されていないため栽培から収穫、加工、流通までのフードチェーン全般において一貫性のある管理が困難となっている。また、農作業や農作物、資機材のように農業分野におけるそれぞれの語彙間の連携も必要である。

そこで、本稿では農業データの連携・統合のためにナレッジグラフを用いた標準語彙の管理を提案する。ナレッジグラフによる標準語彙の管理は用語の多様性が多く見られ、またそれが一貫性のある語彙管理の妨げとなる農業分野の語彙が持つ特徴

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

を補完することができる。また、農業 ICT システムの運用において機械可読性を向上するデータ形式を提供することができ、農業 ICT システム間の円滑な連携を可能とする。

2. 背景

本章では農業分野における語彙について概観し、その現状と農業 ICT システムのデータ連携における語彙に要求される要素を確認する。

2.1 農業分野における語彙の多様性

農業は食料を確保するために人類が営んできた長い歴史の産業であり、農家や地域単位の集団によって代々受け継がれた栽培方法で行われてきた。そのため、農家や地域独自の語彙が発生し、農業分野には同じ概念に対して複数の名称が存在する場面が多く見られる。そして、農家は農業 ICT システムに独自の名称を記録し、農業データには独自の語彙が多く含まれることになった。すなわち異なる農家の ICT システムから発生した農業データにはそれぞれ農家独自の特徴が反映されており、データの連携や統合が困難となっている。現状ではデータ項目の対応関係を調べ、人手によってその関係を再定義する必要がある。この場合、データの量に比例して対応関係を人が確認しなければならないので、効率的なデータの活用が困難である。

2.2 既存の語彙

前節で述べた農業データの連携における問題を認識し、内閣官房 IT 総合戦略本部は「農業情報の標準化に関する個別ガイドライン」を提示し、農業 ICT システムのデータ標準化を推進している。農作業名称のガイドラインとして「農業 IT システムで用いる農作業の名称に関する個別ガイドライン (第 3 版)」を、農作物名称に関しては「農業 IT システムで用いる農作物の名称に関する個別ガイドライン (第 2 版)」を提示している。農作業名称の項目には農作業の分類単位である「大項目」と「中項目」、「シソーラス」があり、それぞれの農作業と 11 種類に分類された農作物 (群) との関連も表示されている。農作物名称の場合は、分類単位である「大分類」、「中分類」、「小分類」、「シソーラス」、詳細を記載する「品種等」、「収穫部位」、「属性項目」がある。それぞれの項目には「その他」があ

り、項目の選択が困難な場合を想定して柔軟に対応している。

一方、上記の語彙以外にも公的機関による従来の語彙があり、農業現場で用いられている。まず、農作業の場合は統計調査における調査項目が挙げられる。国や自治体単位で収穫量や生産費など経営実態を調査しており、農家の経営状況を把握している。データの集計は農家が作成した作業履歴を収集し、担当部署が統計調査の項目と対応させて集計を行う。そして、農作物の場合は食品安全の確保のために管理管轄の公的機関による語彙があり、3つの語彙が中心となっている。まず、農林水産省が農薬の登録申請に適用対象となる作物名を提示するために発行している「農薬登録における適用作物名」がある。「農薬登録における適用作物名」は農作物を分類する単位である「大グループ名」と「中グループ名」、「作物名」と作物名の関連語彙である「作物名に含まれる別名、地方名、品種名等の例」、そして適用される農作物の部位を定義する「備考」がある。次に残留する農薬を管理するために厚生労働省が発行している「農産物等の食品分類表」がある。収録項目として「食品名」と「食品分類」の二つがある。最後に健康、栄養管理のために食品の成分に関する資料である「日本食品標準成分表」が挙げられる。文部科学省が発行しており、収録されている項目は「食品番号」、「食品名」、「生物種の英名」、「学名」の4つがある。農作物に関する3つの語彙はそれぞれ異なる目的によって発行されたものであり、その目的によっては農作物名は農作物の名称そのものではなく収穫部位を意味する場合もある。

2.3 考察

語彙と語彙を連携するためには基準となる情報が必要がある。「農業 IT システムで用いる農作業の名称に関する個別ガイドライン (第3版)」と「農業 IT システムで用いる農作物の名称に関する個別ガイドライン (第2版)」に共通の情報があれば、それをういて連携することが可能となるが、農作業の対象となる農作物名が農作物の分類と異なり、現状では農作業と農作物のデータ間の連携が困難である。例えば、「農作物別農作業名」に分類されている11分類の中で「米」、「麦類」、「豆類」、「いも類」は「農業 IT システムで用いる農作物の名称に関する個別ガイドライン (第2版)」には分類として収録されてないためこれらの語彙を連携し、農作物に関するデータと農作業に関するデータを連携するためにはその詳細を対応させる必要がある。農作物に関する3つの既存語彙の農作物名称についても同じ問題を抱えており、農業 ICT システム間のデータと連携のためには異なる分野の語彙間の連携も考慮しなければならない。

特に農作物の語彙についても一つ考慮しなければならない事項がある。公的機関による3つの重要語彙はそれぞれ農作物、食品、食品成分としての作物を定義した語彙である。これらの語彙は栽培から加工、流通まで至るフードチェーンにおける食品安全の管理に基準なる情報である。近年食品安全の確保に対する要望が高まっており、ICTシステムによる一貫性のある管理が求められている。そのためには語彙間の連携が必要であるが、対応関係を把握し、整理を行う必要がある。例として「じゃがいも (ばれいしょ)」が挙げられる。一般的にはじゃがいもと呼ばれるが「農薬登録における適用作物名」では「ばれいしょ」が、他で語彙では「じゃがいも」が代表表記になっている。また、「農薬登録における適用作物名」と「農作物等の食品分類表」では「ばれいしょ」と「じゃがいも」の関連が定義されているが、「日本食品標準成分表」の「じゃがいも」には「ばれいしょ」の情報が収録されていない。

以上の考察より農業 ICT システムのデータ連携のために基準となる標準語彙に求められる条件として以下の2つが考えら

れる。

- 明確な構造
- 語彙間の連携

本研究ではこの2つの条件を満たすためにナレッジグラフに注目した。ナレッジグラフを基盤とする語彙システムは語彙の効率的な管理と柔軟な適用、そして ICT システムにおけるデータの基準としての利活用が期待される。

3. ナレッジグラフ基盤の標準語彙

本章では農作業と農作物、そしてこれらを連携するために必要な語彙の設計について述べる。

3.1 農作業基本オントロジー

筆者らは農作業の標準語彙として農作業基本オントロジー (AAO, Agriculture Activity Ontology) を開発し、推進してきた [朱 16]。農作業基本オントロジーは農作業の概念を定義するために10項目の属性(目的, 行為, 対象, 副対象, 場所, 手段, 機資材, 作物, 時期, 作業条件)を定義し、その属性が持つ値の包含関係によって農作業概念の階層構造を構築した知識体系である。また、複数の目的を持つ多義的概念と別名を収録し、農作業名称が持つ意味の多様性にも対応できる。オントロジーは記述論理に基づいて設計されており、階層構造における明確な分類と上位概念と下位概念間の論理的な矛盾がないように構築された。機械による推論も可能であり、新しい概念を追加する際に階層構造の中で最も適切な位置を提示することができ、語彙管理の手間も軽減できる。機械による推論は農業基本オントロジーが持つ論理に基づいて処理を行うため、今後 AI 農業の基盤技術として期待される。

農作業基本オントロジーは現在、共通農業語彙 (CAVOC, Common Agricultural Vocabulary)*1 にて公開されており、農作業名称の URI と機械可読性の高い形式である RDF/Turtle と CSV 形式で公開されている。そして、これらのデータを円滑に利活用ができるよう関連 API も公開しており、誰もが自由に使うことができる。最新版である ver2.01 には438語が収録されており、「農業 IT システムで用いる農作業の名称に関する個別ガイドライン」に収録されている農作業名称と農林水産省の農産物生産費統計調査 [農林水産省 16] における調査項目の農作業名とも交換性を持つ。また、Wikipedia 日本語版*2 の項目と国際機関の FAO (Food and Agriculture Organization) による AGROVOC*3 の項目との対応関係を調査し、収録したため、知識体系と国際標準との交換性を持つ。

3.2 農作物語彙体系

農作物の標準語彙として筆者らは農作物体系 (CVO, Crop Vocabulary) を構築した [竹崎 17]。農作物語彙体系は農作物の生物学特徴、収穫部位と栽培形態、品種としての特徴に基づいて農作物を定義している。公的機関による3つの重要語彙、「農薬登録における適用作物名」、「農産物等の食品分類表」、「日本食品標準成分表」の作物名 (食品名) を中心に構築し、それぞれの語彙と相互交換性を持つ。そして、品種情報を収録するために独自で収集し、整理した品種情報も収録されており、最新版である ver1.12 では1,191語が収録されている。

*1 <http://cavoc.org>

*2 <https://ja.wikipedia.org>

*3 <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>

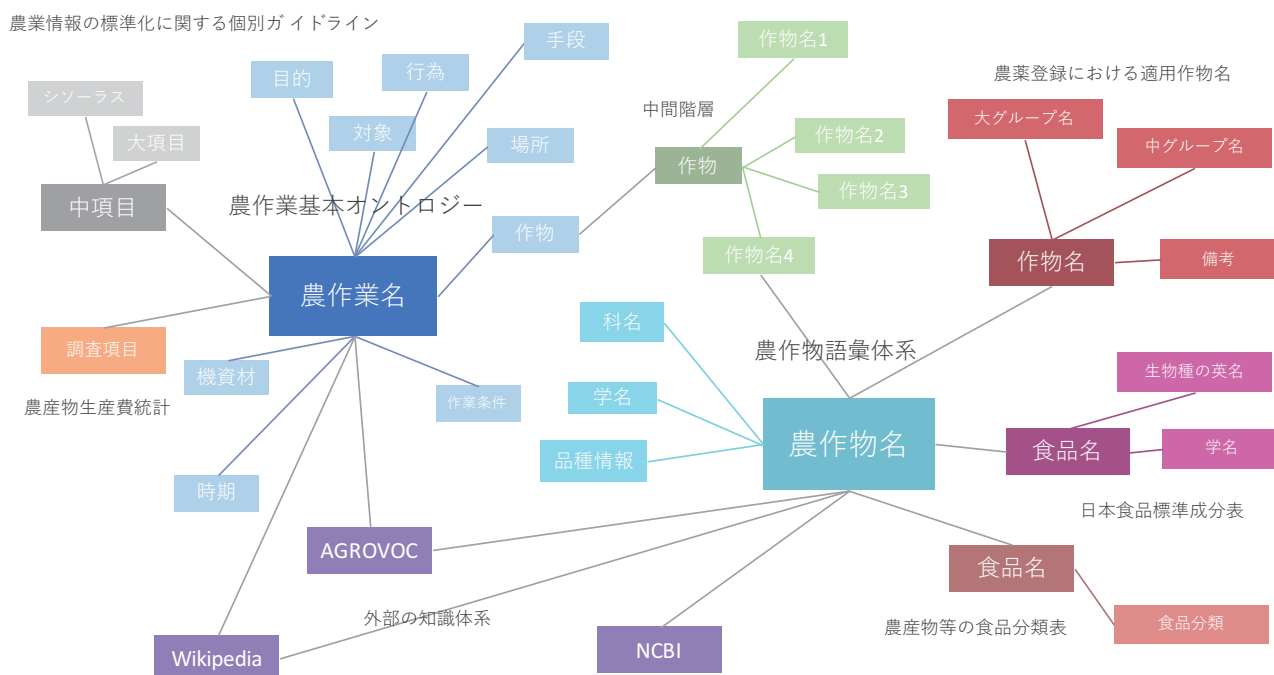


図 1: 農業分野における語彙の連携.

農作物語彙体系も農作業基本オントロジーとともに共通農業語彙にて公開されており、農作物の URI やデータとして SKOS/Turtle, CSV 形式のデータを提供している。また、農作物における 3 つの語彙が持つ重要性を考慮し、その利活用のために農作物の URI と機械可読性を向上させたデータ形式も提供している。それぞれの URI には対応関係がリンクで連携されており、対応関係の把握や詳細の情報の取得が可能である。ともに提供している API とアプリケーションを用いてこれらの語彙を農作物語彙体系が持つ対応関係の情報を用いて横断的な活用も可能となる。例えば、「農薬登録における適用作物名」の「ばれいしょ」も農作物語彙体系の API を用いると「日本食品標準成分表」の「じゃがいも」が持つ生物種の英名や学名の情報を取得することが可能となる。

そして、データの相互運用性を確保するために外部データとの連携も実装されており、内閣官房による「農業 IT システムで用いる農作物の名称に関する個別ガイドライン」の作物名、知識の連携としては Wikipedia 日本語版、国際標準である FAO の AGROVOC、そして生物種のデータであるアメリカの NCBI(National Center for Biotechnology Information)^{*4} の taxonomy ID と連携されている。農作物の語彙のみならず、外部のデータセットや知識体系とも繋がることで農作物全般に対する包括的情報運用が可能である。

3.3 語彙の連携

本節では上記に述べた農作業と農作物の標準語彙として構築された農作業基本オントロジーと農作物語彙体系を連携するために構築した中間階層の語彙体系とその利活用について述べる。

3.3.1 連携のための中間語彙体系の設計

農作業の中には農作物に依存しない一般的な農作業もあるが、特定の農作物でのみ行われる農作業が存在する。例えば、収穫作業の一つである「堀取り」は根菜類に適用される農作業

であり、穀物類などには適用されない。そのために、農作業を記述するためには適用される作物の定義が必要であるが、すべての作物名を農作業の定義に用いることは管理・運用において非効率的である。そこで、本研究では効率的で円滑な連携が行われるように語彙と語彙の間に中間階層となる語彙体系の構築を提案する。「堀取り」を定義する属性の一つである「作物」の値を「根菜類」とし、根菜類について中間語彙体系で定義をする。中間語彙体系には該当する語彙とそれに含まれる詳細を収録する。例えば、「根菜類」の場合、農作物語彙体系に収録されている「ダイコン」、「カブ」、「ジャガイモ」などを対象作物の情報として持つ。これによって農作業と農作物の語彙が繋がり、データの連携も可能となる。図 1 は農作業基本オントロジーと農作物語彙体系を中間階層の語彙体系を用いて連携した全体図である。

このように中間階層の定義し、異なる語彙を連携すると、農作業と農作物、それぞれの語彙は基本情報のみ収録することができ、今後語彙の拡張にも対応することができるなど、その管理における手間が軽減する。それぞれ語彙が持つ URI を利用することで一貫性を持つ語彙の管理も可能となる。また、新規語彙の導入や概念の変更など目的によって中間階層の概念を追加することによってより柔軟な対応と幅広い利活用も可能となる。

3.3.2 中間語彙体系を用いた語彙の利活用

中間階層の語彙体系を用いると農業 ICT システムも横断的な利活用が可能である。図 2 は中間階層を用いて農作業基本オントロジーの農作業名称とそれに適用される農作物を農作物語彙体系を通じて既存語彙の情報を問い合わせる API の例である。農作業基本オントロジーに収録されている「堀取り」の作物に関する属性の値は「根菜類」である。中間階層の語彙体系から「根菜類」に該当する農作物語彙体系の農作物名を問い合わせし、そのリンクから「農薬登録における適用作物」と「日本食品標準成分表」の情報を得ることができる。

*4 <https://www.ncbi.nlm.nih.gov/>

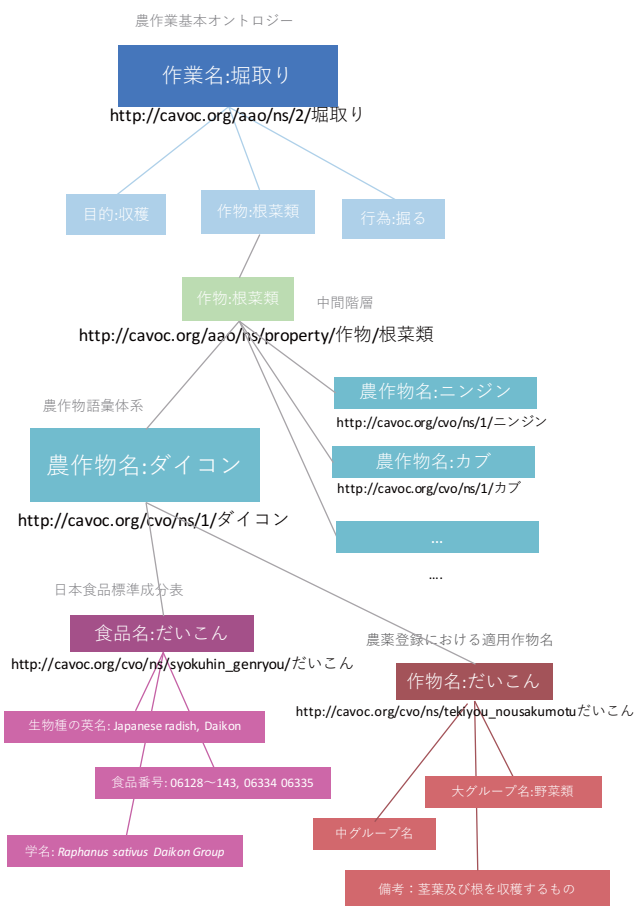


図 2: 語彙の連携の利活用.

4. 考察と今後の課題

今回は農作業と農作物の標準語彙を連携するために中間となる語彙体系を構築した。それぞれの語彙の構造を維持し、語彙を拡張することによって概念を具体化することができた。語彙が持つ構造を明確にしたまま、異なる語彙との連携が可能となり、語彙の横断的な運用についてその可能性を確認した。今後は機資材や土地利用についても具体化し、標準語彙を構築することで知識体系として繋がる語彙体系を構築していきたい。

中間階層となる語彙体系は語彙と語彙の関係を定義するために様々な対応関係の情報を収録する必要がある。対応関係を定義するためには様々な分野、業界の語彙を調査しなければならない。今後は各分野の有識者に依頼し、協議を通じて整理していく予定である。

5. おわりに

本稿では農業 ICT システムのデータ連携のための標準語彙を管理するためにナレッジグラフを用いた語彙体系を提案した。既存の語彙について分析を行い、標準語彙とその管理に求められる条件として明確な構造、語彙間の連携を提示した。ナレッジグラフによる語彙体系の設計と管理は ICT システムでの利活用に適しており、必要に応じて拡張にも柔軟な対応が可能である。今後は農業全般における標準語彙の確立と連携にナレッジグラフを取り組み、包括的な標準語彙の運用と安定的な活用を目指していきたい。

謝辞

本研究（の一部）は、総合科学技術・イノベーション会議の SIP（戦略的イノベーション創造プログラム）「次世代農林水産業創造技術」（管理人：農研機構生物系特定産業技術研究支援センター）の支援を受けて行った。

参考文献

[内閣官房 17a] 内閣官房, 農業情報創成・流通促進戦略に係る標準化ロードマップ (第 3 版), <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310roadmap.pdf> 2016 年 3 月 1 日参照.

[内閣官房 17b] 内閣官房, 農業 IT システムで用いる農作業の名称に関する個別ガイドライン (第 3 版), <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310gl1_besshi.xlsx> 2017 年 3 月 1 日参照.

[内閣官房 17c] 内閣官房, 農業 IT システムで用いる農作物の名称に関する個別ガイドライン (第 2 版), <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/170310gl3_besshi.xlsx> 2017 年 3 月 1 日参照.

[農林水産省 18] 農林水産省, 農業登録における適用作物名, <<https://www.acis.famic.go.jp/shinsei/3986/3986beppyou1.pdf>> 2018 年 3 月 1 日参照.

[厚生労働省 15] 厚生労働省, 農産物等の食品分類表, <<http://www.mhlw.go.jp/file/06-Seisakujouhou-11130500-Shokuhinanzanbu/0000159254.pdf>> 2018 年 3 月 1 日参照.

[文部科学省 15] 文部科学省, 日本食品標準成分表, 資料 2 食品の原料となる生物種の英名・学名, <http://www.mext.go.jp/component/a_menu/science/detail/_icsFiles/afieldfile/2017/06/22/1365334-1-0321r2.xlsx> 2018 年 3 月 1 日参照.

[朱 16] 朱 成敏, 小出 誠二, 武田 英明, 法隆 大輔, 竹崎 あかね, 吉田 智一: 記述論理に基づく農作業オントロジーの設計と応用, 第 38 回人工知能学会セマンティックウェブとオントロジー研究会 (SIGSWO), 06, 2016.

[農林水産省 16] 農林水産省, 農産物生産費統計, <http://www.maff.go.jp/j/tokei/kouhyou/noukei/seisanhi_nousan/> 2018 年 3 月 14 日参照.

[竹崎 17] 竹崎あかね, 朱成敏, 武田英明, 吉田智一: 農業 IT システム間のデータ連携を促進する農作物語彙体系の構築, 電子情報通信学会技術研究報告, 知的環境とセンサネットワーク研究会, vol. 117, no. 310, ASN2017-J27, pp. 98-99 (2017).