

標準語彙に基づく農業データの連携と統計への活用

Integration and Statistical Processing for Agricultural Data Based on Core Vocabulary

朱成敏 *1
Sungmin JOO武田 英明 *1
Hideaki TAKEDA法隆 大輔 *2
Daisuke HORYU竹崎 あかね *2
Akane TAKEZAKI吉田 智一 *2
Tomokazu YOSHIDA

*1 国立情報学研究所

The Japanese Society for Artificial Intelligence

*2 農業・食品産業技術総合研究機構

National Agriculture and Food Research Organization

Agricultural data from IT systems such as farming information or agricultural activities is expected to be used for statistical processing. However, data in agricultural IT systems is nonetheless not easy to integrate since the terms to describe data are not unified. In this paper, we propose automatic statistical processing system for agricultural data based on agricultural activity ontology as core vocabulary for data integration agricultural IT systems.

1. はじめに

農業分野における統計調査は生産計画の立案や生産量の予測、作業の最適化など国の食糧生産に関わる重要な意思決定の指標である。そのため農林水産省を始め地自体単位や地域社会など様々な単位での統計調査を行い、その調査結果を公表している。現在は調査対象の農家に調査票を配布し、記入してもらったデータを収集して集計を行っている *1。この場合、農家は調査票にデータを記入するために、また調査を行う団体はデータを収集するために時間と努力が必要となる。さらに調査の目的や調査単位によって様々な統計調査を行う場合、その負担はより大きくなる (図 1)。

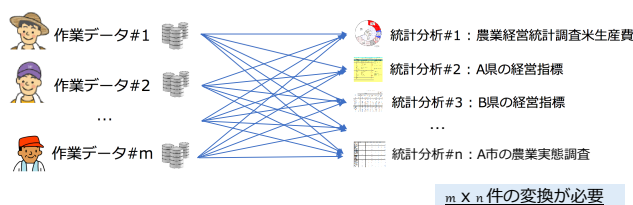


図 1: 既存の統計調査による負担。

そこで、本稿では農業現場に導入されている農業 IT システムのデータを用いた統計調査の新しい枠組みを提案する。それぞれ異なる農家の農業 IT システムのデータを連携・統合し、その集計結果を標準語彙を用いて様々な統計調査に変換することが可能となる (図 2)。これにより既存の調査方法より統計調査に対する負担が軽減されることとともに正確で且つ効率の良く行うことが可能となる。本稿では農業 IT システムのデータを連携・統合するための基準データとして農作業基本オントロジーを用い、自動集計システムの試作を行う。そして、異なる農業 IT システムに記録された実際の作業データを用いて作物別労働時間を算出することで、標準語彙に基づく統計調査の自動化への可能性を確認する。

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

*1 農産物生産費統計の場合, <http://www.maff.go.jp/j/tokei/kouhou-you/noukei/seisanhi-nousan/gaiyou/index.html#9>

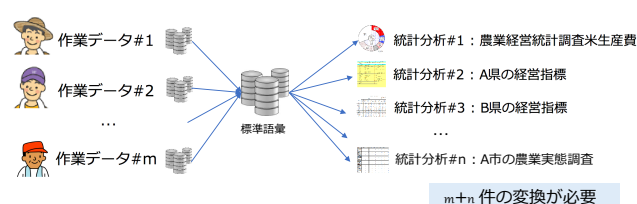


図 2: 標準語彙を用いた負担の軽減。

2. 農業データの統計における標準語彙

本章では農業分野の統計調査を行うため必要な農業データの連携・統合における標準語彙について述べる。

2.1 農業分野の統計調査

農林水産省は毎年収穫量や生産費など、農業分野における統計調査を毎年行い、その結果を公表している *2。また、都道府県や自治体単位でも経営指標や実態調査を公表している。これらの統計調査に必要なデータは対象の農家の協力の下で集計されている。例えば、農業経営統計調査では農家の経営状況を把握するために現金出納帳、作業日誌、経営台帳を調査対象の農家に配布し、回収されたデータ集計を行う。そして、集計の際には農家の独自の作業名や表記を統計調査の作業一覧から対応する項目を選び、整理された作業名で集計を行う。この場合、農家独自の作業を統計調査の作業一覧に対応させるために作業の把握と農家単位のデータ毎に用語の整理が必要となる。このように既存の農業分野の統計調査は農家に対しては配布資料の形式に合わせてデータを作成しなければならないことや、農家の独自の表記と統計調査用の項目に対応させるために語彙の整理が必要ことが問題点である。これらの問題点は農家が用いている農業 IT システムの作業データから直接集計を行うことで、より効率的に改善することができる。また、農業 IT システムに記録されている作業データを集計システムと連携してリアルタイムで統計結果を参考することができ、効率的に農家の営農状況を把握することが可能となる。しかし、この場合もそれぞれの農業 IT システムから集計された作業データ間、作業名や記録単位などデータ項目を整理しなければならない。そして、そのためには語彙の基準情報となる標準語彙が必要である。

*2 農林水産省, <http://www.maff.go.jp/j/tokei/sokuhou/yotei/>

2.2 農業分野における標準語彙

現在、農業分野では様々な農業 IT システムが普及されており、センサーネットワーク、ロボット、POS など多くの分野にわたる。また、これらの農業 IT システムからは膨大なデータが発生されており、データを分析することで政策の策定や生産における状況の把握など意思決定における重要な判断資料として考えられる。しかし、それぞれの農業 IT システムは IT ベンダーによる独自の方針でデータが定義され、異なるシステム同士ではデータの連携が困難である。異なるシステムのデータを連携するためにデータ項目の標準を定義しなければならない。ここでデータ項目の標準として用いられるものが標準語彙である。

日本では農業分野における標準語彙として内閣府 IT 総合戦略本部は「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(本格運用版)(案)」[内閣府 15a]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(試行版)(案)」[内閣府 15b]を提示し、それぞれ農作業と農作物の標準語彙とした。これらは農業 IT システムの運用において基本となる語彙を提供するための階層構造を持つ語彙体系であり、分類における柔軟性が特徴である。しかし、階層構造の曖昧さと概念の定義に不明瞭な部分があり、農業 IT システムに実際にデータ項目として用いるためには改善の必要がある[朱 16a]。

2.3 農作業基本オントロジー

筆者らは IT システムのデータ連携・統合を目的とする農作業における標準語彙として農作業基本オントロジー (AAO, Agriculture Activity Ontology) を開発してきた。農作業基本オントロジーでは農作業を目的、行為、対象、場所、手段、道具、時期、作物の 8 つの属性を用いて定義を行い、定義された属性の値から概念である農作業の包含関係から階層構造を構築した。また、概念同士の関係性を明確にするために記述論理を用いて設計し、概念の推論も可能である[朱 16a]。農作業の名称は概念名である見出し語と表記である同義語、英名を分離することで様々な別名に対応することができる。2017 年 1 月 5 日に公開された最新版である ver1.41 では 410 語が収録されている。表 1 はその更新履歴を表す。

表 1: 農作業基本オントロジーの更新履歴.

バージョン	公開日	語彙数	属性数
0.94	2015-05-12	185	2
1.00	2015-11-02	301	7
1.10	2016-02-12	330	7
1.31	2016-04-22	367	7
1.33	2016-09-23	374	8
1.41	2017-01-05	410	8

農作業基本オントロジーは共通農業語彙 (CAVOC)*3 にて公開されている。CSV または RDF/Turtle 形式のデータと同義語変換など農作業基本オントロジーを用いた関連 API も提供しており、誰もが自由に使うことができる*4。本研究では作業データの語彙変換のために農作業基本オントロジーが提供しているデータと同義語変換 API を用いて、農業データの自動集計システムを構築する。

*3 <http://cavoc.org>

*4 <http://cavoc.org/aa0.php>

2.4 本研究の位置付け

本稿では統計分析における農業 IT システムのデータを連携・統合するデータ集計のために標準語彙を用い、その結果を様々な統計調査に対応させることで農業分野の統計調査を効率良く行う枠組みを提案する。農業分野の標準語彙を用い、異なる農業 IT システムのデータを統合し、統計に必要な集計まで自動化する可能性について確認する。構築する自動集計システムは農業経営統計調査の調査対象作物であるイネと対象外作物であるエンバクの作業データを対象とし、調査対象の多様性や公表されない作物の統計調査に対するニーズに対応できる標準語彙の構築する。

3. 農業データの自動集計

本章では農作業基本オントロジーに基づく農業データの自動集計システムについて述べる。そして、実データを用いて検証を行う。

3.1 集計におけるデータ項目の整理

複数の作業データの集計を行うために本研究では農作業基本オントロジーが提供している農作業名称を標準語彙として用いる。まず、作業データから作業名を抽出し、農作業基本オントロジーの作業名称と対照を行う。ここで一致しない場合は農作業基本オントロジーの作業名の同義語変換 API を用いて同義語から見出し語となる作業名称を用いる。農家独自の作業名や地域によって異なる作業名など農作業基本オントロジーに収録されていない場合は作業データから詳細情報と農作業基本オントロジーに収録されている農作業名称の属性の値を参考し対応付けを行う。図 3 は米の生産費統計調査におけるデータ項目の整理を表す。

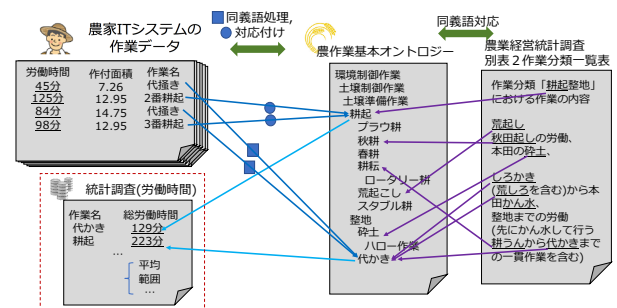


図 3: 農作業基本オントロジーに基づく米生産費のデータ集計例.

3.2 農業データの自動集計システム

複数の農業 IT システムの作業データからデータを集計し、統計を行うために本研究では農業データの自動集計システムの試作を行った。自動集計システムを用いて実データを対象に農作業別労働時間を作物毎に算出する。複数の農業 IT システムから CSV 形式の作業データを取得し、データ項目の調整を行う。データの調整は作業データを圃場、作業者、作物、作業、機資材に関する情報に分類して、データ項目とコラムを指定する。例えば、作物情報には作物名と品種を、作業情報には作業名、日付、労働時間となる。そして、前節に述べたように作業名を農作業基本オントロジーに収録されている標準語彙に変換する。標準語彙に変換された作業のデータから労働時間を抽出し、作物別平均労働時間の算出を行う(図 4)。

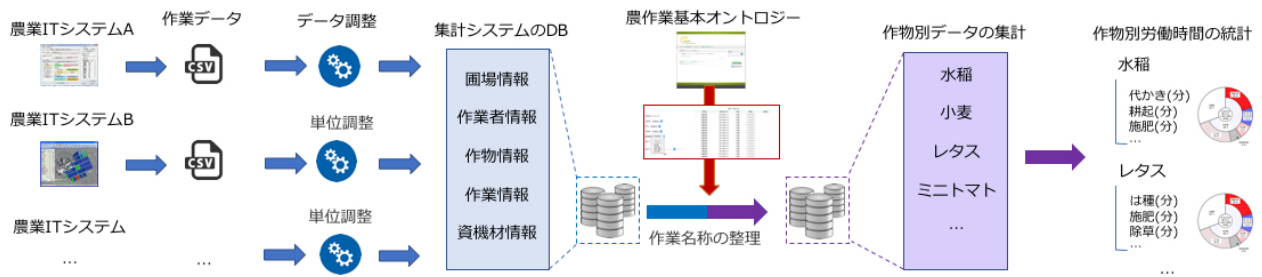


図 4: 農業データの自動集計システム。

3.3 自動集計システムの試作

本節では前節の述べた自動集計システムを試作し、実際の作業データを用いて作物別労働時間を算出する過程について述べる。

3.3.1 データの入力

農業 IT システムから作業データを CSV 形式で抽出し、システムのデータ項目別に格納するために圃場、作業者、作物、作業、機資材情報に合わせてデータ項目を指定する (図 5)。農業基本オントロジーで農作業を定義する属性を考慮してデータ項目を定義した。本研究では研究目的での利用許可を得た 2 件の実データを集計対象とする。1 つ目は農研機構による作業計画・管理支援システム (PMS)*5 から収集したイネ作の作業データ 2,247 件であり、2 つ目は A 社の営農管理システムから収集したエンバク作の作業データ 392 件である。



図 5: 作業データを格納するための項目指定。

3.3.2 作業名の整理

本システムは 2 つの方法を用いて作業名を標準語彙に変換する。まず、農業基本オントロジーが持つ同義語を用いて作業名を整理する。これは共通農業語彙が提供している農業標準語彙への変換 API を用いた。これによりイネ作の作業データ数の 2,247 件のうち 1,939 件が、エンバク作の作業データ数の 392 件のうち 371 件が農業基本オントロジーの作業名に対応付けられた (図 6)。

一方、同義語による処理に失敗した残りのデータは農家法人が独自で定義した作業が多く見られる (表 2)。この場合、作業名はデータ項目の特徴を考慮し手動でマッピングを行う (図 7)。集計システムはデータからデータ項目を抽出し、マッピングの判断資料として提示をする。管理者は提示された項目を選択し、農業基本オントロジーの農作業名と指定することによってマッピングが行われる。

作業名(作業者数)	総労働時間	作業者平均	1日平均
スタプル耕(3名)	103時間 53分	34時間 37分	01時間 26分
ブラウ耕(1名)	12時間 30分	12時間 30分	04時間 10分
代かき(1名)	44時間 59分	44時間 59分	08時間 59分
堆肥散布(2名)	119時間 42分	59時間 51分	03時間 31分
整地(3名)	116時間 09分	38時間 43分	01時間 36分
施肥(1名)	101時間 47分	101時間 47分	05時間 59分
田植え(1名)	06時間 59分	06時間 59分	06時間 59分
畦畔修繕(1名)	05時間 17分	05時間 17分	00時間 13分
畦畔除草(5名)	165時間 16分	33時間 03分	01時間 08分
稲刈り(3名)	276時間 29分	92時間 09分	02時間 25分
追肥(3名)	147時間 00分	49時間 00分	02時間 27分
除草剤散布(2名)	12時間 57分	06時間 28分	02時間 09分
未指定作業 308 件	294時間 56分		

(a) イネ

作業名(作業者数)	総労働時間	作業者平均	1日平均
かん水(4名)	07時間 40分	01時間 55分	00時間 03分
は種(3名)	07時間 44分	02時間 34分	00時間 14分
清掃作業(5名)	65時間 40分	13時間 08分	00時間 56分
耕起(3名)	20時間 26分	06時間 48分	00時間 27分
覆土(3名)	08時間 01分	02時間 40分	00時間 17分
鎮圧(3名)	09時間 32分	03時間 10分	00時間 17分
未指定作業 21 件	10時間 50分		

(b) エンバク

図 6: 同義語 API によるマッピング結果。

3.4 データの集計

農業基本オントロジーの作業名とマッピングされた作業データから労働時間と労働者の情報を集計し、作業別の平均労働時間を表示する。また、その結果を CSV 形式で出力し、ダウンロードすることも可能である (図 8)。この集計結果をそれぞれの統計調査の用語や形式に合わせて編集することで様々な統計調査に対応することが可能となる。

4. 考察と今後の課題

農業データの自動集計のために本研究で構築した集計システムは CSV 形式を用いて農業 IT システムとのデータ交換を行う。しかしながら、農業 IT システムの中ではデータを交換する機能が搭載されていないシステムも確認されており、シス

*5 <http://www.aginfo.jp/PMS/>

表 2: 同義語処理に失敗した作業の一覧.

データ	作業名	計
イネ	1 番耕起, 2 番耕起, 3 番耕起, その他播種	4 語
エンバク	刈り払い, 作付前灌水, ハンマーナイフ	3 語



図 7: 属性による作業名のマッピング.

テムベンダー間データ交換における方針について協議が必要である.

経営指標や実態調査など地域自治体の統計調査ではそれぞれ独自のデータ項目があり、これらの表示形式で集計結果を作成されるためには農作業基本オントロジーとの対応関係を調査する必要がある [竹崎 17]. 今後、公表されている統計調査を中心に対応関係を把握し、修正する予定である.

手動マッピングの結果は事例として蓄積された場合、特定の標準語彙を推測する判断資料としても期待される. 集計システムが事例から用された条件を学習し適切な標準語彙を提示することで、集計を完全に自動化することが可能となる. また、マッピング履歴の分析により農家や作物による農作業が持つ特徴を発見し、生産性との関連性を回帰分析を用いて確認することも今後の課題である.

5. おわりに

本稿ではまず、農業 IT システムから発生された作業データを集計し、統計を行うために必要とされる事項について考察した. そして、農作業における標準語彙を集計における基準項目として用いることで集計の自動化に対する可能性を確認し、試作した自動集計システムを通じて標準語彙に基づく統計調査の新しい枠組みについてその可能性を確認した. 自動集計システムの実現によりリアルタイムで情報を把握することが可能となり、環境要因の変化や異常の探知など国の農業政策の修正や農家の作業方針において有用な意思決定の資料として用いられることが期待される.

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.

作業名(作業者数)	総労働時間	作業者平均	1日平均
スタブル耕(3名)	103時間 53分	34時間 37分	01時間 26分
プラウ耕(1名)	12時間 30分	12時間 30分	04時間 10分
代かき(1名)	44時間 59分	44時間 59分	08時間 59分
堆肥散布(2名)	119時間 42分	59時間 51分	03時間 31分
播種(1名)	06時間 57分	06時間 57分	06時間 57分
整地(3名)	116時間 09分	38時間 43分	01時間 36分
施肥(1名)	101時間 47分	101時間 47分	05時間 59分
田植え(1名)	06時間 59分	06時間 59分	06時間 59分
畦塗修繕(1名)	05時間 17分	05時間 17分	00時間 13分
畦畔除草(5名)	165時間 16分	33時間 03分	01時間 08分
稲刈り(3名)	276時間 29分	92時間 09分	02時間 25分
耕起(3名)	287時間 59分	95時間 59分	02時間 31分
追肥(3名)	147時間 00分	49時間 00分	02時間 27分
除草剤散布(2名)	12時間 57分	06時間 28分	02時間 09分
未指定作業 0 件	00時間 00分		

ダウンロード [CSV]

(a) イネ

作業名(作業者数)	総労働時間	作業者平均	1日平均
かん水(4名)	07時間 56分	01時間 59分	00時間 03分
は種(3名)	07時間 44分	02時間 34分	00時間 14分
モア(1名)	05時間 02分	05時間 02分	02時間 31分
刈取り(2名)	05時間 32分	02時間 46分	01時間 23分
清掃作業(5名)	65時間 40分	13時間 08分	00時間 56分
耕起(3名)	20時間 26分	06時間 48分	00時間 27分
覆土(3名)	08時間 01分	02時間 40分	00時間 17分
鎮圧(3名)	09時間 32分	03時間 10分	00時間 17分
未指定作業 0 件	00時間 00分		

ダウンロード [CSV]

(b) エンバク

図 8: 作業別労働時間の最終集計.

参考文献

- [内閣府 15a] 内閣府, 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 (本格運用版) (案), <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/shiry04.pdf> 2017 年 3 月 9 日 参照.
- [内閣府 15b] 内閣府, 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 (試行版) (案), <http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/shiryo/shiry017.pdf> 2017 年 3 月 9 日 参照.
- [朱 16a] 朱成敏, 武田英明, 法隆大輔, 竹崎あかね, 吉田智一, 農業 IT システム間データ連携のための農作業基本オントロジーの構築, 電子情報通信学会知的環境とセンサネットワーク研究会, 75, 2015.
- [朱 16b] 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.
- [竹崎 17] 朱成敏, 武田英明, 法隆大輔, 竹崎あかね, 吉田智一, 農作業基本オントロジーを基盤とする都道府県発行の技術・経営指標データ連携の可能性, 第 31 回人工知能学会全国大会, 2E3-OS-36a-2, 2017.