

農業事例ベースのための農業用語辞書の作成

大塚彰 北村實彬

農業研究センター 研究情報部

1. はじめに

農業者の意思決定を支援する仕組みを構築するために、農業研究センターでは農業事例ベースの研究を行っている。作物の生育状態の予測や経営管理など線形モデルを用いて記述可能な問題は様々あるが、農業で扱う耕地生態系は多要素が非線形な相互作用をして複雑なため、系全体をモデルで記述するのは現在のところ極めて困難である。こうした複雑な系を対象とする農業者は新しい状況に対処する場合、過去に起こった出来事（事例）を実際に利用してしてきた。こうした事例を収集・再利用して意思決定を支援する仕組みが農業事例ベースである[1, 2]。

農業事例ベースのプロトタイプとして、Webベースの類似事例検索システムを開発した[3]。検索部はベクトル空間法の一つで Latent Semantic Indexing 法[4]を用いて単語に索引を付ける手法を用いている。このシステムの評価の結果、形態素解析の間違いが適合率を下げるのが分かった。例えば個別検索の例で、幼い苗の意味の“乳苗”が乳と苗に誤解析されたために稲作事例（正解事例）に畜産事例（誤事例）が混ざって検索された。

形態素解析性能が検索システムの性能に与える影響は論文[5]で議論されている。そこでは形態素解析辞書の単語数を減らして検索システムの性能を評価した結果、単語を 13 万語まで減らしても検索性能は変わらないことを明らかにした。しかし農業のように特定の分野で使用される語彙は新聞で 사용되는語彙と異なるため、上記の結果がそのまま適用可能かどうかは明らかでない。

そこで本研究では検索システムの性能を向上させ

るために農業用語を幅広く収集し、形態素解析用辞書を作成する。

2. 方法

農業で使われている技術用語や作物名等を文献[6]–[25]を参考にしながら人手で幅広く収集した。以下で示すように農業の様々な分野を網羅している。

この形態素解析用の農業用語辞書は、形態素解析システム「茶筌」の形態素辞書のうち普通名詞辞書として作成した。したがって農業用語辞書には品詞情報と見出し語、読みが登録されている。

3. 特徴

農業用語辞書の特徴をまとめると次のようになる。

3.1 分野

農学一般、園芸、畜産、土壤肥料、農業気象、病虫害、雑草、食品関連の各分野の用語を収めた。

3.2 作物名

穀物、果樹、野菜、花き等の作物名の主要なものを網羅した。

3.3 品種名

水稻陸稻、大麦、小麦などの麦類、だいず、あずきなどの豆類、かんしょ、ばれいしょなどのいも類、とうもろこし、そばなどの雑穀類、いぐさ、なたねなどの工芸作物、野菜、果樹の主要な品種名を収めた。野菜については平成 5 年時点で種苗登録されている全ての品種を収録した。米、果樹についてもほぼ全ての品種を収録した。

3.4 病虫害名

農林水産省が発生予察の対象としている全ての

病害虫の名前と一部北海道特有の病害虫の名前を収録した。

3.5 雑草名

水田雑草、畑地雑草と水路雑草の主要な雑草名を収録した。

3.6 論文に出現した用語

農林水産技術会議事務局が農林水産関係の国内論文に現れたキーワードを中心に収集した用語集を参考に、属名、科名、化学物質名、植物名、機械名等を収集した。

3.7 総用語数

「茶筌」の形態素辞書との重なりを除いた農業用語辞書の総数は約 57000 語であった。

4. 考察

農業者の意思決定支援の観点から、使用頻度が高いと期待される病害虫、雑草、品種等に力点を置いて用語を収集した。農業の分野で使用される用語のかなりの部分が収集できたと考えているが、今後この辞書を追加したことによる検索精度への影響を定量的に調べる必要がある。

また専門用語については学術事典を参考にしたために、農業者からの問い合わせにおいて使われる語彙と異なる可能性もある。したがって、農業者等が参加するメーリングリストなどを利用して実際に使用されている語を収集する必要があると考えられる。

5. まとめ

農業事例ベースの検索部で使用する形態素解析用語辞書「農業用語辞書」を作成した。総用語数は約 57000 である。これは以下のサイトから公開する。

<http://ss.narc.affrc.go.jp/jyoho/jyoho.html>

参考文献

[1]北村實彬,高度情報処理技術を用いた農業の発展,農林水産技術会議事務局編「農林水産業の高度情報システム」,農林水産技術情報協会,p.231-241,1996.

[2]大塚彰,農業事例ベース,農業情報利用研究会編「農業情報化年鑑 1999」,農文協(印刷中).

[3] Otuka, A. and Ninomiya, S. "Conceptual Retrieval of Agricultural Cases with Latent Semantic Indexing Approach", In proceedings of BioDecision98, Feb. 23-27, Montpellier, France, 1998.

[4]Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., & Harshman, R. 1990. "Indexing by Latent Semantic Analysis", J. American Soc. Info. Sci. 41(6), 391-407,1990.

[5]多田智之,金岡秀信,形態素解析性能の検索システムに与える影響,言語処理学会第4回年次大会発表論文集,p.120-122,1998.

[6]「第2次増訂改版 農学大事典」,養賢堂,1987

[7]「農業気象用語解説集」,日本農業気象学会,1986

[8]「新編農業気象学用語解説集」,日本農業気象学会,1997.

[9]「農作物有害動植物発生予察事業調査実施基準」,日本植物防疫協会,1986.

[10]「北海道病害虫防除提要」,北海道植物防疫協会,1995.

[11]「原色 雑草の診断」,農山漁村文化協会,1986.

[12]「原色・雑草の防除(改訂版)」,家の光協会,1977.

[13]「新編農作物品種解説」,農業技術協会,1984.

[14]「米品種大全」,米穀データバンク,1996.

[15]「米と麦」,創造書房,1983.

[16]「野菜品種名鑑 平成5年版」,日本種苗協会,1994.

[17]園芸学会編「改訂園芸学用語集」,養賢堂,1987.

[18]日本畜産学会編「新編畜産学用語集」,養賢堂,1981.

[19]「果樹遺伝資源保存目録」,果樹試験場,1997.

[20]園芸学会編「園芸学用語集 園芸作物名編」,養賢堂,1979.

[21]「新編日本食品事典」,医歯薬出版,1982.

[22]「食品材料ハンドブック」,培風館,1993.

[23]日本土壌肥料学会編「土壌肥料用語集」,養賢堂,1983.

[24]日本育種学会編「新編育種学用語集」,養賢堂,1994.

[25]「農林水産関係 国内文献検索のための用語集 - 1994-」,農林水産省農林水産技術会議事務局,1994.