

日本農業シソーラス (JAT)

農林水産省農林水産技術会議事務局筑波事務所
中央農研 データマイニング研究チーム

増加する流通文書から効率的に情報を入手するためには、文書を解析するコンピューター（計算機）の性能を向上させることが重要である。その方法の一つとして、計算機に言語知識（辞書やシソーラスなど）を与えて、文書の理解や正確な解析を支援することが考えられる。そこで、我々は、農林水産関連分野における検索向上に資する、日本農業シソーラス (JAT ; Japan Agriculture Thesaurus) を整備した。

JAT の概要

JAT の整備は、国際連合食糧農業機関が推進する「多言語シソーラス (AGROVOC http://www.fao.org/aims/ag_intro.htm)」構築事業に参加したことが発端である。AGROVOC は、農林水産業、食品およびその関連分野の用語が収録しているが、日本固有の用語が不足していた。そこで、我々は日本固有の用語を AGROVOC に追加した、我が国独自の日本農業シソーラスを整備した。

JAT は、約 57,000 語を日本語、及び英語で表記し、AGROVOC に準じて階層関係・等価関係で用語を整理している (図 1)。階層関係は、上位語 (広義語)、下位語 (狭義語) で表される関係であり、上位語が「グループ」を示す用語の場合は、その下位語は「メンバー」を示す。図 1 は、“ゴボウ属” の上位語が“キク科”，下位語が“ゴボウ”となる例を示している。等価関係は同じ概念を表す用語間で結ばれる関係のことであり、同義語で表される。例えば“ゴボウ”の同義語は“牛蒡”である。

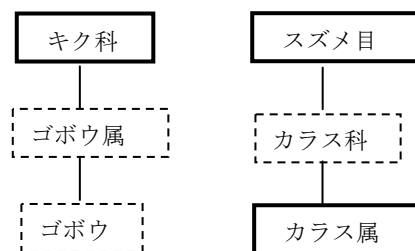


図 1. JAT における階層関係の構築例
AGROVOC の用語 (実線) の下層 (左図)
中間層 (右図) に用語 (点線) を追加

JAT の利用方法

計算機による情報検索では、ユーザーが入力した検索語と、文書から抽出された複数の「索引語」が比較され、一致するかどうか判定される。このため、ユーザーが“コメ”を検索語にした場合、計算機は“コメ”の代わりに“米”を用いた文書を検索できない。そこで、ユーザーが使う語彙と、文書中で用いられている語彙の不一致から起こる検索の漏れを防ぐために、検索語に階層関係や等価関係にある用語を追加することが行われる。例えば、検索語“コメ”で検索した場合、文書 B しか検索されないが、検索語“コメ”に同義語“米”を追加した場合は、文書 B とともに A も検索され、検索漏れが防げる（図 2）。このように、JAT に収録した上位語・下位語・同義語を検索語に追加することで、語彙の不一致による検索漏れが防げると期待できる。

まとめ

開発した JAT は、今後、筑波事務所が保有する情報データベースの検索効率化のために利用していく予定である。また、新たな用途開発も模索しており、研究用途での JAT の活用を推進していきたいと考えている。JAT に関するお問い合わせは、筑波事務所研究情報課 (ric@affrc.go.jp) までお寄せいただきたい。

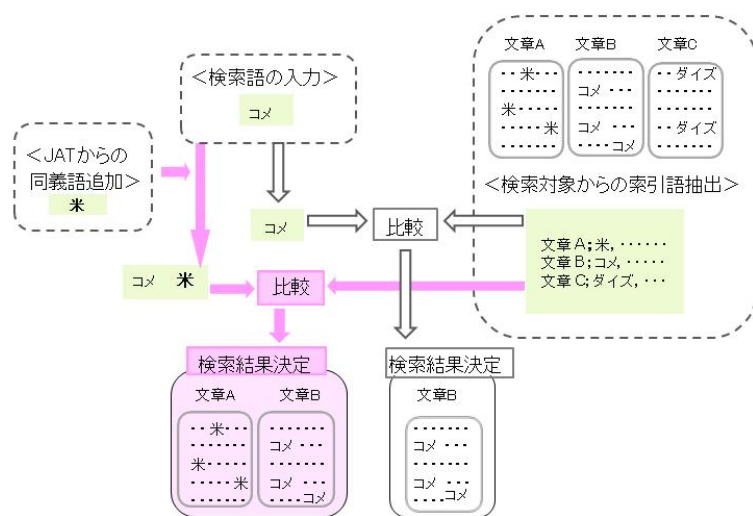


図 2. JAT による検索語への同義語追加は検索漏れを防ぐ