

レシピ文書を対象としたオントロジー整備作業の評価

稲葉 真純[†] 古田 和浩[‡] 長野 伸一[†]

(株)東芝[†] 東芝ホームアプライアンス(株)[‡]

オントロジー活用アプリケーションにおいて、オントロジーの構築や更新は避けられない作業である。しかし、従来オントロジーに関する評価指標は明確でなく、構築規模や品質レベルに対する方針を決めることが難しい。そこで、本稿では料理用語を対象としたオントロジーの整備作業の事例を示し、その結果を基にしてオントロジーの規模と精度の指標策定に必要な要件について考察する。

1. はじめに

オントロジー活用アプリケーション^[1, 2]では、オントロジーの品質がアプリケーションの品質に直結するため、オントロジーの構築や更新が不可欠である。しかし、オントロジーは品質評価指標が明確でないという問題がある。我々は、アプリケーション活用で必要とされるオントロジーの品質評価指標を策定するため、レシピ検索システムで用いる料理用語オントロジーを対象として、オントロジーの整備作業を実施した。その結果を基にして、オントロジーの規模と精度について検証する。

2. オントロジーの構築方法

レシピ検索システムは、食材名などのキーワードからレシピ文書を探すアプリケーションである。語彙辞書として料理用語オントロジーを用いることで、同義語検索が可能になる。料理用語オントロジーは、食材名の分量や料理の手順が記されたレシピ文書から、食材名、器具名、動作名を手手で集めて構築する。はじめに、レシピ文書を形態素解析して品詞情報を付与する。動詞と判定された語彙のうち、人間の動きや食材・器具の状態を表す表現を動作名として抽出する。ただし、“する”などのサ変動詞については、係り受け関係にある名詞を動作名とする。例えば、“みじん切りにする”の場合は、“みじん切り”が動作名になる。次に、名詞から食材名と器具名を取り出す。形態素解析器の誤判定や未知語判定は、手手で語彙を補正する。

3. 評価実験

3.1. 実験方法

レシピ検索システムにおけるアプリケーションの精度を計測するため、レシピ文書中に含まれる料理用語と、オントロジー収録語彙とのマ

ッチング精度を算出した。評価用のレシピ文書は整備作業に用いていない 20 件とし、料理ジャンルを分けて、おかずとお菓子のレシピ文書各 10 件ずつとした。整備前と整備後のオントロジーについて精度を算出し、比較評価した。

料理用語オントロジーの語彙数を表 1 に示す。整備前の料理用語オントロジーには、食材名、器具名、動作名の 3 つの語彙分類について合計 5579 語の語彙が収録されている。オントロジーを整備するために、Web 上のレシピ文書約 1000 件を収集し、目視により未収録語 2562 語を抽出して料理用語オントロジーに追加した。

表 1 料理用語オントロジーの語彙数

語彙分類	整備前	整備後	追加数
食材名	5194	6698	1504
器具名	256	448	192
動作名	129	995	866
合計	5579	8141	2562

3.2. 実験結果

3 つの語彙分類それぞれに対する精度評価の結果を示す。まず、食材名の精度を表 2 に示す。整備前に比べて、整備後は適合率、再現率、F 値が向上した。しかし、依然として再現率は低く、語彙数が十分ではないことがわかった。食材名は語彙のバリエーションが多く、再現率の向上には多くの語彙が必要である。料理ジャンルで比較すると、食材名のバリエーションが少ないお菓子は、おかずに比べて適合率が高くなった。

次に、器具名の精度を表 3 に示す。食材名と同様に、整備後は適合率、再現率、F 値が向上した。お菓子の再現率が低い原因は、お菓子作り専用の器具がオントロジーの収録語彙から漏れているためと考えられる。器具名については、料理ジャンルごとに使用する器具が異なるため、様々な料理ジャンルのレシピ文書から語彙を収集する必要がある。

最後に、動作名の精度を表 4 に示す。動作名については、整備後に適合率が低下した。動作

Evaluation of Ontology Development from Recipe Text

[†]Masumi INABA and Shinichi NAGANO, Toshiba Corporation

[‡]Kazuhiro FURUTA, Toshiba Home Appliances Corporation

名には“混ぜ合わせる”や“振り入れる”などの複合動詞が多いため、語彙数が増えたことで、複合動詞に含まれる動詞の一部を誤抽出したことが原因と考えられる。しかし、再現率が向上したため、F値は向上した。動作名では、料理ジャンルによる顕著な差異は見られなかった。

表 2 食材名の精度

料理ジャンル	精度	整備前	整備後
おかず	適合率	62.96	89.47
	再現率	8.95	26.84
	F 値	15.67	41.30
お菓子	適合率	100.00	100.00
	再現率	20.51	32.05
	F 値	34.04	48.54
全体	適合率	76.74	92.68
	再現率	12.31	28.36
	F 値	21.22	43.43

表 3 器具名の精度

料理ジャンル	精度	整備前	整備後
おかず	適合率	40.00	100.00
	再現率	5.00	70.00
	F 値	8.89	82.35
お菓子	適合率	66.67	96.55
	再現率	3.64	50.91
	F 値	6.90	66.67
全体	適合率	50.00	98.25
	再現率	4.21	58.95
	F 値	7.77	73.68

表 4 動作名の精度

料理ジャンル	精度	整備前	整備後
おかず	適合率	100.00	97.67
	再現率	38.99	75.81
	F 値	56.10	85.37
お菓子	適合率	98.11	96.80
	再現率	35.14	81.76
	F 値	51.74	88.64
全体	適合率	99.38	97.35
	再現率	37.65	77.88
	F 値	54.61	86.54

4. 考察

同一のレシピ文書から食材名、器具名、動作名を抽出したが、語彙分類ごとに整備作業の貢献度が異なると考えられる。実験結果から、この仮説を確認する。

3つの語彙分類における整備作業の貢献度を表

5に示す。器具名の語彙追加数は最も少なかったが、F値の変化は大きく、整備作業によって頻出語彙が追加されたと考えられる。追加語彙数に対するF値の変化を貢献度として算出すると、器具名が最も大きくなった。器具名は他の語彙分類に比べて整備が進んでいなかったため、精度が向上しやすい状態だったと考えられる。整備状況の違いから、語彙分類によって整備作業の貢献度が異なることを確認した。

また、同一のレシピ文書から抽出可能な語彙数は、語彙分類ごとに異なることもわかった。出現頻度が低い語彙分類では、語彙を収集するために多くの文書が必要となる。従って、文書から抽出可能な語彙数を予測することは、作業時間を見積もるために重要である。

表 5 整備作業の貢献度

語彙分類	語彙追加数	F 値の変化	貢献度
食材名	1504	22.21	0.01
器具名	192	65.92	0.34
動作名	866	31.93	0.04

語彙分類によって出現頻度や整備状況が異なるため、オントロジーの規模と精度は、文書数や語彙数のみでは予測できない。そこで、文書中に含まれる語彙分類の割合を推測することを検討する。語彙分類の割合を推測するには、レシピ文書に含まれる名詞や動詞などの品詞の数と、品詞中に含まれる語彙分類の割合を計測することが必要である。これらを予め計測しておき、今後整備するレシピ文書に適用することで、抽出可能な語彙数を推測できるのではないだろうか。そして、語彙数の推測値を用いて、将来的にどの程度の精度が見込めるか予測できると考えられる。特に、レシピ文書は料理用語が頻出する手順書という特徴があるため、統計的手法が有効と考えられる。

5. まとめ

今後は、料理用語オントロジーの構築から得られた知見を基に、ドメインに依存しないオントロジーの品質評価指標を検討していく。

参考文献

- [1] 飯田, 他: 評判分析のための感性表現オントロジーメンテナンスツールの開発, 第 18 回セマンティックウェブとオントロジー研究会, 2008.
- [2] 木村, 他: オントロジーを用いた英日料理レシピ変換システム, 電子情報通信学会技術研究報告, Vol.107, No.428, pp. 77-82, 2008.