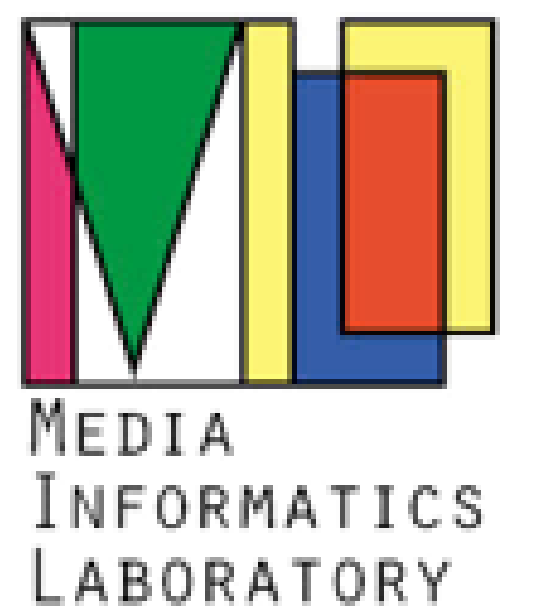


日本食品標準成分表を用いた 食材オントロジの構築とその利用



○窪田 茜* 馬場 睦也** 楠 和馬** 波多野 賢治*

* 同志社大学文化情報学部 ** 同志社大学大学院文化情報学研究科

1. 背景と目的

料理レシピを対象にしたさまざまな言語処理研究が増加している
固有の言い回しや表現の省略が多く存在することや
料理レシピ中で用いる食材名の表記が投稿者によって異なることなどが
料理レシピを十分な精度で解析できない大きな要因となっている [1]

食材名の問題点

・代替食材の提案

じゃがいもの代替食材： さつまいも
ジャガイモの代替食材： ?
馬鈴薯の代替食材： ?

表記ゆれがあると正しく
代替食材を提示できない

・レシピ検索

ボタンク

求める検索結果

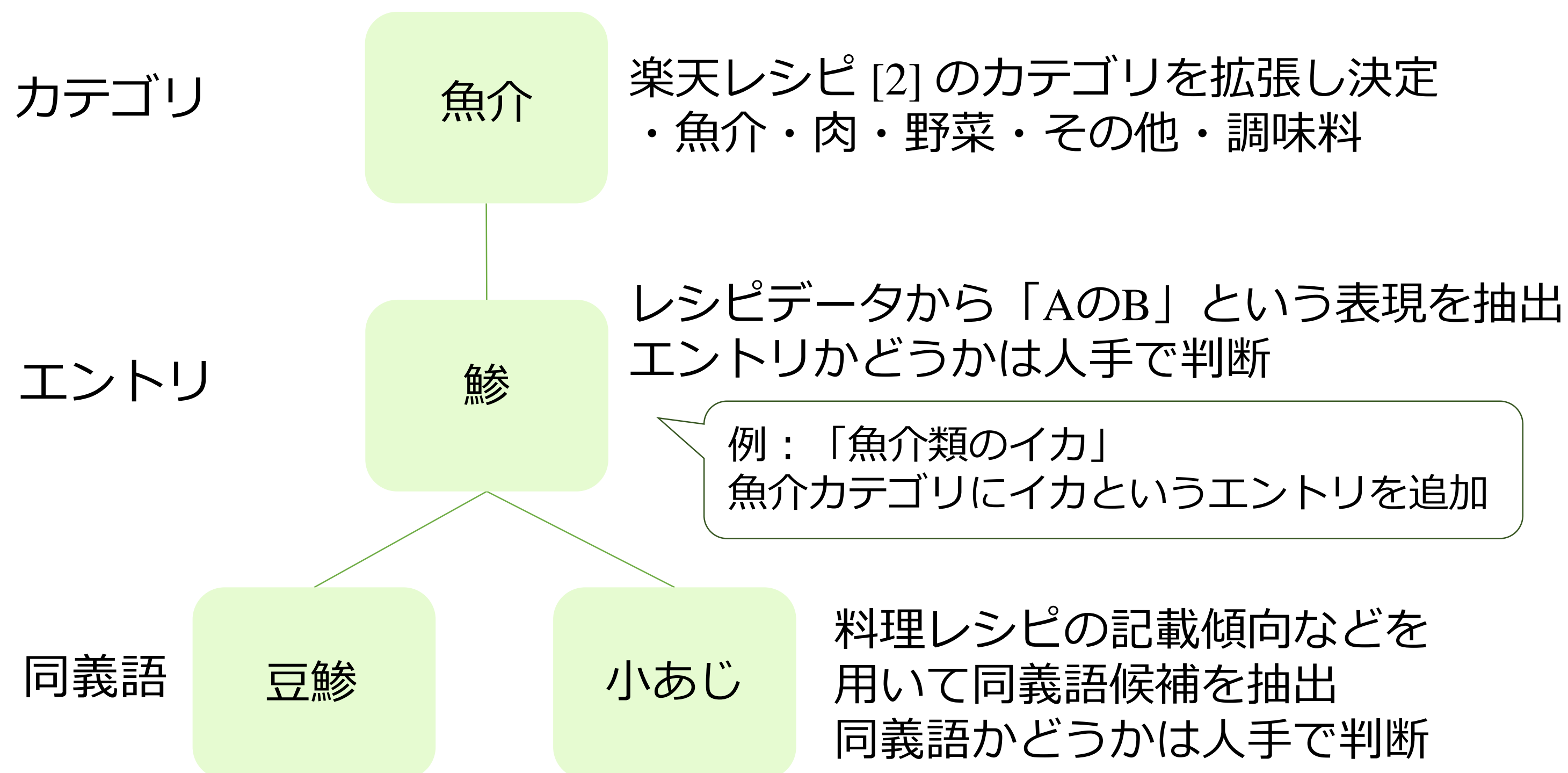
豚の角煮
豚バラ野菜炒め

表現粒度の違いがあると
正しく検索できない

実際の検索結果

ボタンクの生姜焼き
ボタンクの冷しゃぶ

料理オントロジの構築方法[1]



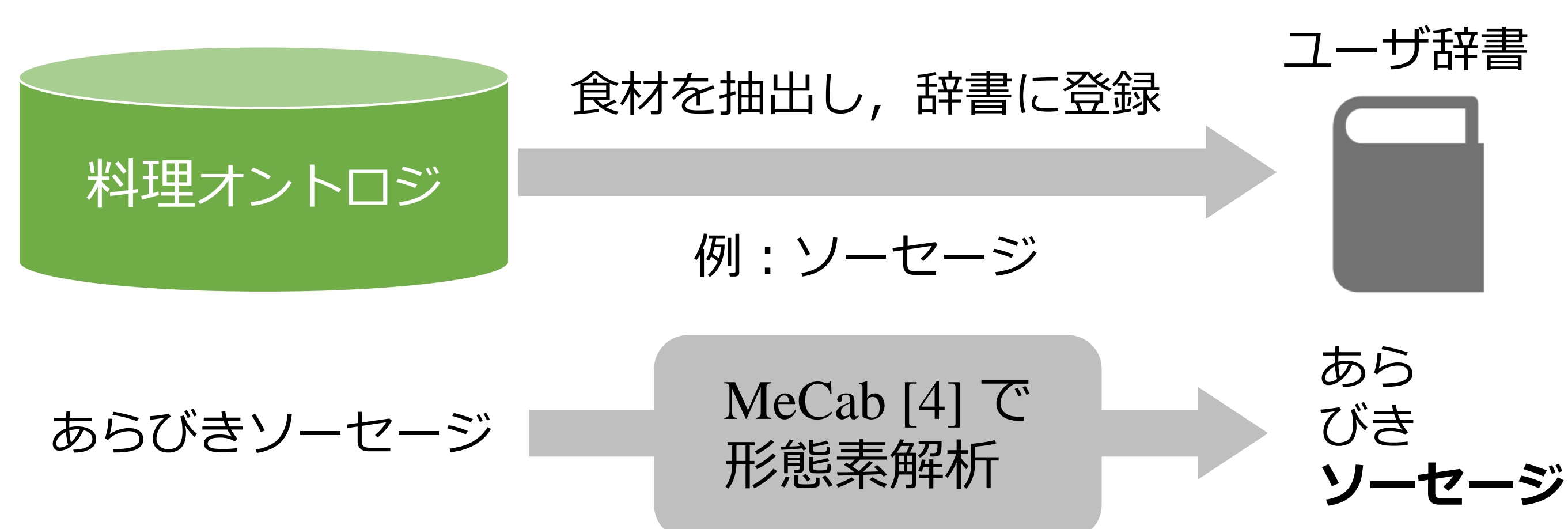
料理オントロジの評価方法

クックパッド [3] の食材データに料理オントロジの食材が
どれだけ含まれているかを料理オントロジのカバー率とする

料理オントロジのカバー率

$$= \frac{\text{クックパッド食材データに含まれる料理オントロジの食材数}}{\text{クックパッド食材データの総食材数}}$$

クックパッド食材データに含まれる料理オントロジの食材判別方法



ソーセージが料理オントロジに含まれるため
あらびきソーセージは料理オントロジがカバーできた食材とする

料理オントロジでカバーできたクックパッドデータの食材は
全体の5割であった

- 問題点
- 各カテゴリのエントリ種類数少ない
 - エントリの抽出方法が半自動的であるため、高コストかつヒューマンエラーが発生しやすい

2. 構築方法と食材の自動抽出

本研究では、日本食品標準成分表 [5] を用いてオントロジを構築する

- ✓ 各カテゴリのエントリ数が料理オントロジより多い
- ✓ 別名表記もあるため、表記ゆれにも対応できる
- ✓ 文部科学省が出している公的文書である

食材オントロジの構築方法



成分表からそのままオントロジを構築すると上位・下位概念ではない
関係ができてしまう



食材でないものの判別方法

- ・食材名の一番右には生や皮つきといった状態が多く含まれている
- ・一番右に出てくる文字列で、重複している単語は食材ではない

処理手順

いも類	さつまいも類	さつまいも	塊根	皮つき
いも類	さつまいも類	さつまいも	塊根	皮むき
いも類	さつまいも類	さつまいも	蒸し切干	
いも類	さつまいも類	むらさきいも	塊根	皮つき
いも類	さつまいも類	むらさきいも	塊根	皮むき

最下層の重複単語を削除

いも類	さつまいも類	さつまいも	塊根
いも類	さつまいも類	さつまいも	塊根
いも類	さつまいも類	さつまいも	蒸し切干
いも類	さつまいも類	むらさきいも	塊根
いも類	さつまいも類	むらさきいも	塊根

最下層の重複単語を削除

いも類	さつまいも類	さつまいも	
いも類	さつまいも類	さつまいも	
いも類	さつまいも類	さつまいも	蒸し切干
いも類	さつまいも類	むらさきいも	
いも類	さつまいも類	むらさきいも	

重複行を一つにまとめる

いも類	さつまいも類	さつまいも	
いも類	さつまいも類	さつまいも	蒸し切干
いも類	さつまいも類	むらさきいも	

各行一番右のさつまいも・蒸し切干・むらさきいもを食材とする

[5] 文部科学省「2015年度日本食品標準成分表」
http://www.mext.go.jp/a_menu/syokuhinseibun/1365419.htm, 2018年11月28日閲覧

3. 結果と考察

結果

成分表の全データ：2,191件 重複単語削除後データ：1,405件
重複単語削除後のデータにおける食材の割合：86%

考察

食材でなかったもののなかには、スイート型や塩分無調整タイプなどの
単語が含まれていた
→「～型」や「～タイプ」といった文字列一致で削除を行うことで
食材の割合を向上させる

今後の課題

- ・クックパッドの食材データ食材オントロジを適応しカバー率の計算を行う
- ・重複削除した単語の属性の自動判別を行う

[1] 土居洋子ほか(2014)「料理レシピと特許データベースからの料理オントロジの構築」

電子情報通信学会 MVE/CEA 研究会, Vol. 113, pp. 37-42

[2] 楽天レシピ https://recipe.rakuten.co.jp/, 2018年11月28日閲覧

[3] クックパッド https://cookpad.com/, 2018年11月28日閲覧

[4] MeCab http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.79.4154, 2018年11月28日閲覧