

料理レシピ検索の性能向上

森下 雄太† 青木 真吾†† 佐藤 隆士†††

†大阪教育大学 教育学部 教養学科 情報科学専攻

††大阪教育大学大学院 教育学研究科 総合基礎科学専攻 数理情報コース

†††大阪教育大学 情報処理センター

1 はじめに

インターネット上の情報は増加傾向にあり、ユーザは様々なサービスを楽しむことができるようになった。その中の1つに料理に関するレシピを集め、多くのユーザが共有できるようにしたレシピサイトがある。

レシピサイトには10万を超えるレシピが登録されており、その中からユーザが求める条件に適したレシピを提示する必要がある。また、ユーザの要求は多岐にわたり、幅広い対応が求められることになる。

そこで本研究ではNTCIR-11 CookingRecipeSearch (以下, RecipeSearch とする) の開催する評価型ワークショップに参加し、検索精度の向上のための自動拡張を行った。

2 タスクの概要

RecipeSearchとは料理レシピ検索に関する参加型ワークショップである。ワークショップとは複数の研究者が集まり、共通の研究対象に対して研究を行い、評価をするものである。

本研究ではアメリカのレシピ情報サイトであるYummly[2]に登録されているレシピのうち101,783件の英語レシピを検索対象とし、同じくYummlyから提供されたユーザの検索履歴の500件を検索要求として提供された。

また、500件の検索要求のそれぞれには適合と思われる正解例のレシピがそれぞれの検索要求に対し、1~3件用意されている。

それぞれの研究者は正解例のレシピを導出することができるか否かでシステムの動作を確認することができる。

また、本研究では検索精度を示すスコアとしてMAP*を採用した。

3 処理

3.1 レシピの構造

各レシピには、表1に示すように6種類のフィールドが存在する。

表1: レシピに含まれるフィールド

フィールド名	説明
RecipeID	レシピを識別するID
title	レシピのタイトル
ingredientLines	食材
preparationSteps	調理手順
attributes	レシピの属性 (カテゴリ, 地域など)
totalTimeInSeconds	調理時間 (秒)

3.2 検索要求の構造

検索要求の文章は実際にYummlyのユーザが入力した文章であるため、その様式は様々である。

料理名や風味, 季節, カテゴリ, 調理方法, 食材などが記されている。

検索要求の例を表2に示す。

表2: 検索要求の例

検索要求	備考
oregano garlic olive oil pasta	食材と料理名が記されている
carrot cake with no butter	使用してはいけない食材が記されている
rosemary beef stew crock pot	調理器具が指定されている
diet low fat liquid diet	カロリーなどの成分について記している
dessert no sugar no white flour	レシピの属性について記している
asian radish	地域が記されている

3.3 スコアリング方法

本研究では表1で示したうちのtitle, ingredientLine, preparationStep, attributesの4項目についてインデックスを作成した。

この4項目のそれぞれに対し、検索要求から検索語を作成し、検索・スコアリングを行い、その結果を併合したものをランキング結果として出力した。

†Yuta MORISHITA ††Shingo AOKI †††Takashi SATO

†Department of Arts and Sciences, Osaka Kyoiku University

††Pure and Applied Sciences Graduate School of Education, Osaka Kyoiku University

†††Information Processing Center, Osaka Kyoiku University

*Mean Average Precision の略。各検索要求に対する平均適合率 (AP:Average Point) についての平均値

3.4 検索語の作成方法

検索語の作成手順の概略図を図1に示す。検索語の作成は以下の3つのステップによって行われる。

1. 検索要求からの単語の抽出 (BASE)
2. 正解例からの単語の抽出 (ORCL)
3. 検索要求から推定される単語の抽出 (TEST)

最初に検索要求から想像される単語を抜き出し、BASEを作成する。次に、検索精度を高めるために、正解例に含まれている単語を抜き出す (ORCL)。最後に、抜き出された語群のうち、語尾変化や同義語を拡張する (TEST)。

検索要求

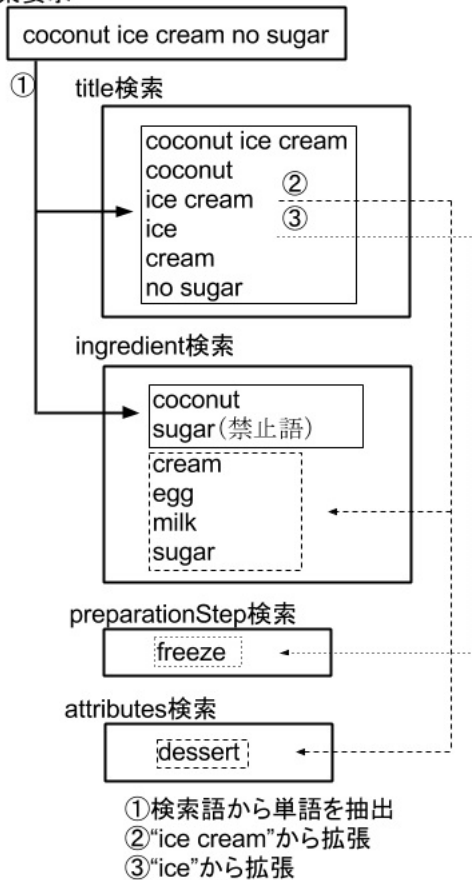


図1: 検索要求から検索語作成手順

3.5 拡張の構造

本研究では検索精度を高めるために、BASEに含まれる単語からいくつかの単語を拡張し、それを検索語として登録している。

そのために拡張用のデータベースを作成した。そのデータベースに基づき、図1のtitle検索の検索語である“ice cream”からingredient検索とattributes検索に検索語を追加している。

表3: 拡張データベースの抜粋

分類	検索語	拡張語
複数形への拡張	strawberry	strawberries
類義語への拡張	icing	frosting
題名→食材	cookie	sugar,egg,flour など
食材→食材	tropical fruit	pineapple,banana など
題名→属性	cake	dessert

3.6 評価

本研究では検索語から作成されたBASEに、正解例のサンプルから抜粋したものを加えたORCL、拡張データベースを使用したTESTの2種類の拡張手法を組み合わせることで4種類のRUN¹⁾を作成した (表4)。

表4: RUNの組み合わせ

RUNの識別名	検索語
B	BASE
BO	BASE, ORCL
BT	BASE, TEST
BOT	BASE, ORCL, TEST

表4に示したそれぞれのRUNに関して、trac_eval[3]のMAPを算出した。その結果を表5に示した。ただし、当研究室で独自に作成した正解例の一覧と、提供されたサンプルを合わせたものをQREL²⁾として使用した。表5より、検索に対する情報量が多くなるほど、検索の精度が向上しているのがわかる。

表5: 算出したMAP

RUNの識別名	MAP
B	0.6779
BO	0.6969
BT	0.7214
BOT	0.7417

4 おわりに

本研究室の手法の特徴としてレシピを4つのフィールドに分け、全文検索を可能にしたインデックスを作成し、また独自データベースによって検索語を拡張を行った点が挙げられる。これらは料理レシピの検索に特化した一つの仕組みとして、一定の効果があったことが確認できた。

参考文献

- [1] NTCIR-11 RecipeSearch, <https://sites.google.com/site/ntcir11recipesearch/>
- [2] Yummly, <http://www.yummly.com/>
- [3] Text REtrieval Conference(TREC) trec_eval, http://trec.nist.gov/trec_eval/

¹⁾一定の手法によって作成された検索結果

²⁾トピックに適合する文書の集合。RUNを評価するための基準に用いる。