

## 研究の背景

現在, web上で数多くの調理レシピテキストが公開されています. だれでも気軽にレシピ情報を取得することができますが, 現在の調理レシピテキストの検索は柔軟性にかける部分があります.



図1 検索結果



困ったわ...  
イーストなんて  
ないわ(´Д`)



手元にある食材のみで  
できるレシピが知りたい!



そこで...

料理に対して材料の  
**重要度**を付けよう!



## クックパッドデータセット概要

レシピのデータは株式会社クックパッドさんから頂きました.  
テーブルの関係を図2のER図に示しました.

概要

- テーブル数は12個
- レシピ関係, カテゴリ関係, 献立関係の3グループ

表1 主なテーブルのレコード数

	レシピ	材料	カテゴリ	献立
レコード数	1, 715, 595	12, 725, 006	164, 912	35, 928

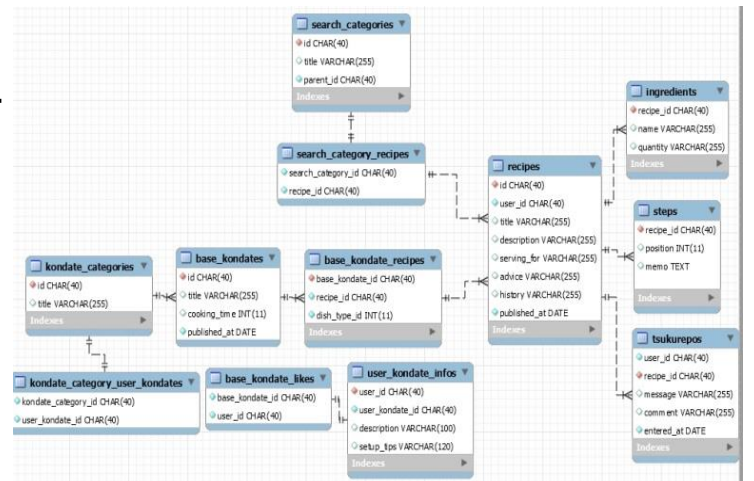


図2 データセットのER図

## Tf-idf計算法

Tf-idfとはtf値とidf値をかけあわせたもので, 文書の特徴語を示します. 今回の実験では, 以下のような計算になります.

カテゴリ内での  
食材tの出現頻度

$$tf(t, d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}}$$

$\sum_{s \in d} n_{s,d}$ : カテゴリ内での単語の総数

$n_{t,d}$ : カテゴリ内での食材tの数

食材tのカテゴリ  
間の共通度

$$idf(t) = \log \frac{N}{df(t)} + 1$$

N: 全カテゴリ数

df(t): 食材tが出現するカテゴリ数

## 実験結果

Tf-idfによる食材単語の重み付けを行いました.  
カテゴリごとの特徴語は以下ようになりました.

表1 結果

カテゴリ (tf-idf値)	食パン	スペアリブ ブ	けんちん汁
重要食材	イースト (0.2538) 強力粉 (0.2004)	スペアリブ (0.5347) コーラ (0.1528)	ごぼう (0.2841) 豆腐 (0.2054)

スペアリブの重要食材の一つであるコーラなどは料理本等では出てこないもので, 投稿型レシピサイトであるクックパッドらしい材料が抽出されました. 砂糖や塩などのよく使われる食材はtf-idf値は低くなっていました.