

卒業論文

文化依存対話のための Linked Data を用いた類

似料理検索

指導教官 村上 陽平 准教授

立命館大学 情報理工学部
先端社会デザインコース 4 回生
2600170392-6

舛田 晃一

2021 年度（春学期）卒業研究 3（CH）

令和 3 年 8 月 2 日

文化依存対話のための Linked Data を用いた類似料理検索

舛田 晃一

内容梗概

昨今新型コロナウイルスの流行によって減少してしまっているが、観光業は日本の中でも大きな主要産業である。日本に来る目的には雷門や清水寺などの歴史的建造物であったり、アニメや漫画などのサブカルチャーであったりと様々であるがその中の一つに日本の食文化も挙げられる。寿司などすでに海外で知られている食べ物もあるがほとんどの日本食は、どんな食べ物なのか詳細は知られていない。そのためその食べ物の名前だけではどのように調理されたものであるか、どのような味がするかなどの情報が想像しにくいという問題点がある。

この問題を解決する手段として観光客の出身地の文化的背景に基づく関連性 (Culturally Situated Associations, (CSA)) をもとに、出身地に合わせた類似料理を提示するというものが考えられる。このデータを提示するにあたって対話の中で観光客から出身地などの情報が必要になるため、対話エージェントを用いて対話モデルを構築し、情報を聞き出す手法をとる。この対話エージェントを作成する際、類似料理を提示する会話パターンの収集を行う必要があり、その際に DBpedia を用いて類似料理の検索を行う。しかしながら、どのような検索手法が有効であるか定まっておらず、会話パターンの収集の時に回答にバラつきが生じてしまう恐れがある。そこで本研究では類似する料理を DBpedia 上で検索する際どのような手法が有効的であるかについて検証評価を行う。具体的には DBpedia において料理のページに設定されているプロパティを収集し、その後類似される料理についてペアを設定し、ユーザーから得られる出身地の情報をもとにどのようなプロパティやカテゴリの絞り込みが類似料理の提示に有効であるかについて評価を行う。本研究で取り組む課題点は以下の二点となる。

DBpedia における有効検索手法の検討

DBpedia 内におけるプロパティは様々であり、全てのページに存在するものもあれば、同じ料理であっても一部のページにしか存在しないプロパティもある。その中で類似する料理にたどり着くためにはどのプロパティが検索に有効でどれほど広義のカテゴリに検索範囲を広げるべきか検討する必要がある。

DBpedia 内でのリソース間の文化的影響

有志によって作成されている Wikipedia をもとにしたデータベースであるため、

文化圏ごとに情報リソースが異なるケースや、カテゴリ形成がされていない場合がある。これらによってどの程度検索に影響が現れるか検証する必要がある。

一つ目の課題では DBpedia 内における料理のページにおけるプロパティのうちクエリによる検索に有効なプロパティとしてどのプロパティが挙げられるか実際の各料理の DBpedia に存在するプロパティを比較したうえで絞り込む必要がある。各料理と関連性の薄い値を持つプロパティでクエリによる検索をかけると類似料理ではない候補が挙げられる可能性が存在する。そのためクエリにかけるプロパティに関しては事前に数個に絞る。

二つ目の課題点は DBpedia の情報リソースとしての偏りである。基本的に有志によって作成される Wikipedia の各ページは話者人口による差や他言語へのリンクの有無によってページ自体は存在していても、検索する際に候補としてあがないケースが存在する。本研究では一番情報リソースの多い英語版 DBpedia をもとに類似料理の候補を挙げる際に文化圏の異なりによる影響が存在するか検証する必要がある。

本研究による貢献は以下の通りである。

DBpedia における有効検索手法の検討

現在 DBpedia にて設定されているプロパティをもとにユーザー情報に基づく条件での絞り込みと概念階層を用いた類似概念による絞り込みを行う手法の提案を行った。この検証では概念階層を用いた絞り込みに類似性がみられる候補が得られ、精度の向上には今回用いなかったプロパティまたは DBpedia 以外のデータリソースが必要であると考えられる。

DBpedia 内でのリソース間の文化的影響

今回の用いた検索パスでは同一原材料による絞り込みを行うことで検索精度が大きく落ちることが確認され、この結果から類似料理の検索する際、原材料が文化圏による影響を大きく受けることが明らかとなった。

Similar Food Search Using Linked Data for Culturally-Situated Dialogue

Koichi Masuda

Abstract

Though it has been decreased by the prevalence of the new coronavirus recently, the sightseeing industry is a large main industry even in Japan. There are various purposes for coming to Japan, such as historical buildings such as Kaminarimon and Kiyomizu-dera Temple, and subcultures such as anime and manga, among which Japanese food culture can be mentioned. There are some foods already known in foreign countries such as sushi, but details of what kind of food most Japanese foods are are unknown. Therefore, there is a problem that it is difficult to imagine information such as how the food is cooked and what kind of taste it has only by the name of the food.

As a means to solve this problem, it is considered that similar food suitable for the home country is presented on the basis of the relevance (Culturally Situated Associations, (CSA)) based on the cultural background of the home country of the tourist. In this paper, we propose a dialogue model using dialogue agents. In making this dialog agent, it is necessary to collect the conversation pattern which presents the similar food, and in that case, the similar food is searched using DBpedia. However, it is not determined what kind of retrieval technique is effective, and there is a fear that the dispersion occurs in the answer in the collection of the conversation pattern. Therefore, in this study, we examined what kind of method is effective in searching similar foods on DBpedia. Concretely, the properties set in the page of the food are collected in DBpedia, and the pair is set on the similar food afterwards, and what kind of property and category narrowing is effective for the presentation of the similar food is evaluated on the basis of the information of the hometown obtained from the user. The following two issues are addressed in this study.

Study of effective search method in DBpedia

The properties in DBpedia vary, some on all pages, and some on only some pages of the same food. It is necessary to examine which property is effective for the retrieval and how the retrieval range should be expanded to the broad category in order to reach the similar food in the inside.

Cultural impact between pages within DBpedia

Since it is a database based on Wikipedia made by volunteers, there are cases in which information resources are different in each cultural area, and cases in which categories are not formed. It is necessary to verify to what extent these will affect the retrieval.

In the first problem, it is necessary to narrow down the properties that are effective for query search among the properties on the page of food in DBpedia by comparing the properties that exist in DBpedia of each actual food. Querying for properties with values that are less relevant to each food may lead to candidates that are not similar foods. Therefore, the properties to be queried are limited to a few in advance.

The second problem is the bias of DBpedia as an information resource. There are some cases where each page of Wikipedia, which is basically created by volunteers, is not considered as a candidate when searching, even though the page itself exists depending on the difference in the number of speakers and the existence of links to other languages. In this study, it is necessary to verify whether there is the effect by the difference of cultural sphere, when the similar food candidate is mentioned on the basis of English edition DBpedia with the most information resource.

This study is expected to make it easier to propose similar foods when collecting conversational data based on CSA.

文化依存対話のための Linked Data を用いた類似料理検索

目次

第 1 章	はじめに	2
第 2 章	関連研究	4
2.1	機械翻訳の限界	4
2.2	Culturally Situated Associations	4
第 3 章	DBpedia	6
3.1	DBpedia の基本的構造	6
3.2	言語間のリソース差	7
3.3	検索プロパティの同定	8
第 4 章	文化依存の関連に基づく検索	10
4.1	SPARQL	10
4.2	検索プロセスの概要	10
4.2.1	ユーザー情報に基づく料理検索	11
4.2.2	概念階層を用いた料理検索	11
第 5 章	評価と考察	14
5.1	評価	14
5.2	考察	16
5.2.1	プロパティによる検索結果の影響	16
5.2.2	文化的背景による影響	17
5.2.3	その他の要因と展望	18
第 6 章	おわりに	22
	謝辞	24
	参考文献	25
	付録	26

第1章 はじめに

近年コロナウイルスの影響で各国間での往来が大幅に減少してしまっているが、日本において観光産業は拡大が続いており、歴史的建造物だけでなく寿司やすき焼きといった日本食に対しての注目が高まってきている。しかしながら日本食そのものに対して興味を持つ観光客は多いものの、寿司などの知名度が高い食べ物を除いて料理名を聞いただけではどのような食べ物であるか想像することは難しい。観光客の来訪が多い地域の飲食店ではメニューに料理の説明や写真などで補足を行う対応をとっていることもあるが全飲食店の総数から考えるとまだ浸透しているとは言い難い。このような疑問点を解決するにはその料理に使う食材や実際の調理法、風味などの情報を与えることで解決につなげることもできるが、これらの情報を簡潔に説明することは容易ではなく、そのうえ回答者がその料理に関する知識を有している必要がある。

これらの問題点を踏まえた解決策として DBpedia を用いた対話システムの構築が挙げられる[1]。この対話システムでは Wizard of Oz 法 (Woz 法) を用いて対話の中から類似料理を観光客の出身地の文化的背景に基づく関連性 (Culturally Situated Associations, (CSA)) をもとに必要な情報を聞き出し、DBpedia で検索をかけることにより求める。しかし、DBpedia で類似料理検索をするにあたってどのプロパティを用いて、どのようなクエリで検索をかけると有効であるか明確化されていない。

そこで本研究では DBpedia を用いて類似料理の候補を求めると得られる候補のうち、実際に類似する料理である精度はどれほどになるか、加えて検索に必要なプロパティについて分析、評価を行う。具体的には日本の食べ物から観光客の出身地域に合わせた類似する料理を、その日本食の名前と観光客の出身地、そしてその日本食のページに存在するプロパティの情報のみで求める。得られた候補のうち、実際類似する料理を候補とする割合を適合率とし、適合率を上げるために必要なプロパティやクエリについて分析を行う。本研究における課題点は以下の二点である。

DBpedia における有効検索手法の検討

DBpedia 内におけるプロパティは様々であり、全てのページに存在するものもあれば、同じ料理であっても一部のページにしか存在しないプロパティもある。その中で類似する料理にたどり着くためにはどのプロパティが検索に有効でど

れほど広義のカテゴリに検索範囲を広げるべきか検討する必要がある。

DBpedia 内でのページ間の文化的影響

有志によって作成されている Wikipedia をもとにしたデータベースであるため、文化圏ごとに情報リソースが異なるケースや、カテゴリ形成がされていない場合がある。これらによってどの程度検索に影響が現れるか検証する必要がある。

本研究では 2 章において機械翻訳の抱える問題やその問題に対する参考研究について述べる。3 章では現在の DBpedia の仕様及び検索に有効であるプロパティの同定などを行う。3 章で行った調査をもとに 4 章では本研究で行う検索手法の詳細について記述し、5 章では今回提案する検索手法の評価及び文化的背景に基づく影響について踏まえたうえで評価考察を行っていく。

第2章 関連研究

本章では本研究に至る背景について、現在異文化間でのコミュニケーションをする際に起こる問題点や参考研究について述べる。

2.1 機械翻訳の限界

近年自然言語処理技術の発達により、機械翻訳は長文であってもおおよその意味が正しく伝わるようになってきている。機械翻訳は大変便利ではあるが、いくつかの欠点も抱えている。そのうちのひとつとして自国の文化圏にないものの情報が伝わりにくいというものがある。料理においては料理名だけではその食べ物についての味、見た目、原材料、調理法などユーザーの想起を助ける情報は得られにくいという現状がある。図1のように各個人の文化圏から離れた事物に対しては機械翻訳だけでは十分な理解が得られにくい。



図 1：機械翻訳の問題例

2.2 Culturally Situated Associations

Culturally Situated Associations とは文化的背景に基づく関連性のことであり、異文化コミュニケーションをする中で相手の文化圏との認識の齟齬が生じないようにするために重要な情報である。例えば「主食である穀物」というものを想像する際、アジア圏では米、欧米では小麦と大きく異なる。そのため相手に伝わる言語でコミュニケーションをとることに加え、相手の文化圏に合わせた情報を用いることも異文化コミュニケーションにおいて重要となる。今回参考にした研究[2]では実対話データを収集する方法の一つである Wizard of Oz

法 (Woz 法) を用いた CSA に基づく類似料理の提案を行うエージェントの構築を行った。Woz 法とはシステムに扮した人間 (ウィザード) がユーザーと対話する中で、音声対話システムの設計に必要な要素の収集や知見を得るために行われる対話モデルである。この手法ではどのような人間でもデータの収集が行えるとされている[4]。参考研究では図 2 のようにエージェントの対話戦略の指示を受けて、ユーザーとシステム間での対話を繰り返しながら必要な情報を集め、必要な情報が集まれば DBpedia を用いて類似料理の検索を行いユーザーに提案するという流れである。しかしその検索方法については確立できておらず、実際 DBpedia をもとに検索した場合に得られる検索結果の精度に関しても不明なままであった。そこで本研究では参考研究の対話モデルの中でエージェントが得ることのできる情報のみで DBpedia の検索は行えるのか評価検証を行う。

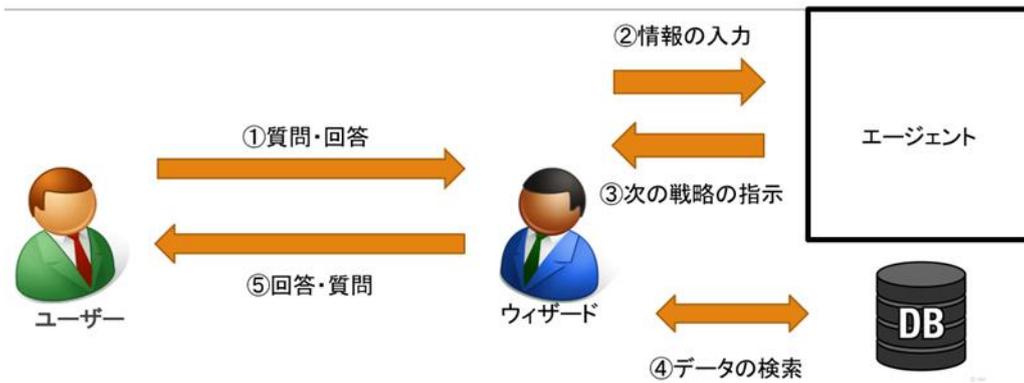


図 2 : Woz 法の対話モデル

第3章 DBpedia

本研究では DBpedia を用いて評価検証を行う。DBpedia は Wikipedia からの情報抽出を目的としたリンクトデータプロジェクトのうちのひとつである。Wikipedia はインターネット上のコミュニティのよって作成されているフリーの百科事典であり、RDF (Resource Description Framework) を用いて構造化データとして表現されている。この DBpedia では RDF データをリンクトデータの形式で公開されているだけでなく、問い合わせ言語である SPARQL を適用している。本章では検証で用いる DBpedia についての詳細と用いるデータについて述べる。

3.1 DBpedia の基本的構造

DBpedia で用いられている RDF では「主語、述語、目的語」の三要素（トリプル）で構成されている。主語はリソース、述語はプロパティ、目的語はプロパティ値を表している。例えば図 3 では「<https://dbpedia.org/page/Soba>」には”Japan” という”country” で生まれたものが書かれている」と読み取ることができる。このように各ページにおいて複数のプロパティとそれに対応するプ

The screenshot shows the DBpedia page for 'Soba'. The table below is a simplified representation of the visible content:

Property	Value
<code>dbo:abstract</code>	<ul style="list-style-type: none">Soba (そば or 蕎麦) (, Japanese pronunciation: [sobɑ] buckwheat flour, or a combination of buckwheat and wt Soba noodles are served either chilled with a dipping s; variety of settings, from "fast food" places to expensive noodle broth, to make home preparation easy. There ar noodles. Soba can nutritionally complement other grain soba. Soba contains all eight essential amino acids, inc in the Edo period. ^(en)蕎麦（そば）とは、穀物のソバの実を原料とする蕎麦麦などと区別して日本蕎麦（にほんそば）とも呼ばれ、切り（そばきり）が普及してからは、単に蕎麦と言う。作ることは少なく、大抵は小麦粉のつなぎが用いられ、味として作られる「つゆ（蕎麦汁）」は、地域によっ分かれる。蕎麦を供する場合には皿（竹籠が敷かれて蕎麦つゆを供する場合には徳利（蕎麦徳利）と猪口（おとも）もある。蕎麦は専門店のみならず、外食チェーンの状態で販売され、カップ麺としても販売されている。
<code>dbo:country</code>	<ul style="list-style-type: none"><code>dbr:Japan</code>

図 3 : DBpedia のページ例

ロパティ値が存在する。これらのプロパティは DBpedia にて設定されたものもあれば Dublin Core™ Metadata Initiative (DCMI) によって定義されたプロパティなどが存在する。

3.2 言語間のリソース差

DBpedia は多言語に対応しており、2016 年時点で 127 言語に対応をしている。しかし、DBpedia の Web サイトにおいてリンクトデータおよび SPARQL エンドポイントとして公開提供しているのは、英語版 Wikipedia を基にして作られているデータセットとなっている。そのため各言語間で情報リソース量には大きな差が開いている。DBpedia では Wikipedia の情報をもとに図 4 のようなテンプレートをもとに事前にマッピングを行っている。この作業を行うことによって DBpedia のページ作成が行われているのだが、Wikipedia の作成とは別に作業が必要になるため言語によってこのマッピング不足による情報リソース差が生じている。日本語版も例外ではなく、本家英語版と比較しても対応しているテンプレートの割合は Wikipedia 全体のテンプレートうち 5%ほどと差はないが、対応テンプレート及びプロパティの出現頻度に関しては大きく差が開いてしまっている(図 5[3])。そのため本研究では日本語版ではなく英語版の DBpedia で問い合わせを行う。

Mapping:Infobox_actor

```

{{TemplateMapping
| mapToClass = Actor
| mappings =
  {{{ PropertyMapping | templateProperty = name | ontologyProperty = foaf:name }}
  {{{ PropertyMapping | templateProperty = birth_place | ontologyProperty = birthPlace }}
}}

```

図 4 : DBpedia へのマッピング例

	日本語	英語
対応テンプレートの割合	5.52% (106 / 1,919)	5.92% (362 / 6,115)
対応プロパティの割合	2.93% (2,242 / 76,444)	3.11% (6,041 / 194,485)
対応テンプレートの出現頻度	49.1% (389,916 / 794,114)	80.92% (2,849,942 / 3,521,856)
対応プロパティの出現頻度	41.57% (3,331,271 / 8,013,323)	63.75% (29,707,828 / 46,597,066)

図 5 : 日英間のマッピング状況の比較[3]

3.3 検索プロパティの同定

前節でも触れたように DBpedia には多数のプロパティが存在している。DBpedia のデータ元となっている Wikipedia は不特定多数の人手による編集でひとつひとつ作成されているため、同ジャンルのページであっても編集者の違いによってプロパティの有無などの差が生じてしまっている。またカテゴリが異なればマッピングされるプロパティも異なる。そのため本研究を行うにあたって食べ物のカテゴリにマッピングされているプロパティのうち、どのプロパティが有効であるか同定する必要がある。

英語版 DBpedia で料理のページにおけるテンプレートを確認してみたところ 576 のプロパティが確認された。検索を行う際に検索パスに設定したプロパティが存在しない場合、候補として検索結果に列挙されないため、検索に使用するプロパティは料理のページにおいてよく設定されているプロパティが望ましい。そこで実際に 30 の料理のプロパティを集計し、576 のプロパティの中で検索に有効で出現頻度の高いプロパティは図 6 の通りとなった。下記のプロパティのほかにも全てのページにてマッピングされているそのページについての概要となる”dbo:abstract”や Wikipedia における該当ページに存在するリンク一覧である”dbo:wikiPageWikiLink”なども存在するが、リンクされているすべてのリソースがプロパティ値として設定されるため、相関関係が薄い単語まで取得してしまう恐れがあるため本研究では扱わない。この結果を受けて全品目でプロパティの存在が確認できたその料理の発祥地を表す”dbo:country”,”dbo:region”と原材料を表す”dbo:ingredient”のユーザー情報に関する 3 プロパティに加えて、概念階層を用いて類似性をもとに検索

	プロパティの存在数	補足事項
Country	30/30	
Region	22/30	場合によってはCountryより広義になることも その国全体で食べられる食べ物については定義されていないことが多い
Ingredient	26/30	ほとんどのオブジェクトが多くても2,3個程度 寿司やお茶漬けなど原材料をあまり加工せず使用しているものは一部プロパティが存在せず
Type	22/30	一部カレーなどすでに広義の食べ物には存在しない
subject	30/30	「○○の国の料理」というトピックはほぼすべてに確認

図 6 : 料理 30 品目における出現頻度上位のプロパティ

を行うために用いる” `dbo:type`” と約半数に「〇〇(国名)の〇〇(ジャンル)」の形式でプロパティ値を持つ[5]” `dc:subject`” の計 5 プロパティをもとにクエリを設定する。

第4章 文化依存の関連に基づく検索

4.1 SPARQL

本研究での問い合わせには SPARQL を用いる。SPARQL は RDF データのリソースを取り扱うための問い合わせ言語のひとつである。問い合わせの仕組みとしてはトリプルの要素のうち、問い合わせしたいものを変数として置き、RDF トリプルにマッチングする候補をデータとして抽出する。例えば図 1 の「そば」のページで原材料のプロパティ値を求めたいときは図 7 のようにクエリを設定することで求められる。

クエリ

```
select distinct ?label ?Ingredient
where {
  dbr:Soba dbo:ingredient ?o;
  rdfs:label ?label.
  ?o rdfs:label ?Ingredient.
  FILTER ( lang(?label) = "en" )
  FILTER ( lang(?Ingredient) = "en" )} LIMIT 100
```

検索結果

label	Ingredient
"Soba"@en	"Buckwheat"@en

図 7: SPARQL クエリと結果例

4.2 検索プロセスの概要

今回の評価検証では参考研究の会話シチュエーションに倣って、使用するデータリソースは DBpedia のみ、会話の中から得られる情報は旅行者が知りたい料理の「料理名」と旅行者の「出身地」の二つの情報をもとに検証を行う。

4.2.1 ユーザー情報に基づく料理検索

DBpedia には DBpedia 独自の概念である DBpediaOntology (dbo) だけでなく DCterms によって設定されたプロパティ (dct) などが存在する。本検証ではその料理がどの国で生まれたものであるかを表すプロパティ `dbo:country` を用いてユーザーの出身地域の料理に候補の絞り込みを行う。また料理の類似点として考えられる原材料についても `dbo:ingredient` プロパティを用いて、この2プロパティのプロパティ値がそれぞれ一致するリソースを類似料理の候補として扱う。

4.2.2 概念階層を用いた料理検索

また料理の類似点として概念階層を用いて共通するカテゴリ探索も行う。料理の各ページのプロパティには関連するカテゴリであることを示す `dct:subject` プロパティが設定されている。この各カテゴリにはより広義なカテゴリが `skos:broader` プロパティをもとに設定されており、例えば図 8 の例では「うどん」の関連カテゴリとして「日本の麺類」、「ピチ」の関連カテゴリとして「パスタ」がプロパティ値として設定されている。これら二つのカテゴリは「麺類」というカテゴリを広義なカテゴリとして持ち、このことから「うどん」と「ピチ」は「麺類」という共通カテゴリを上位に持つといえる。

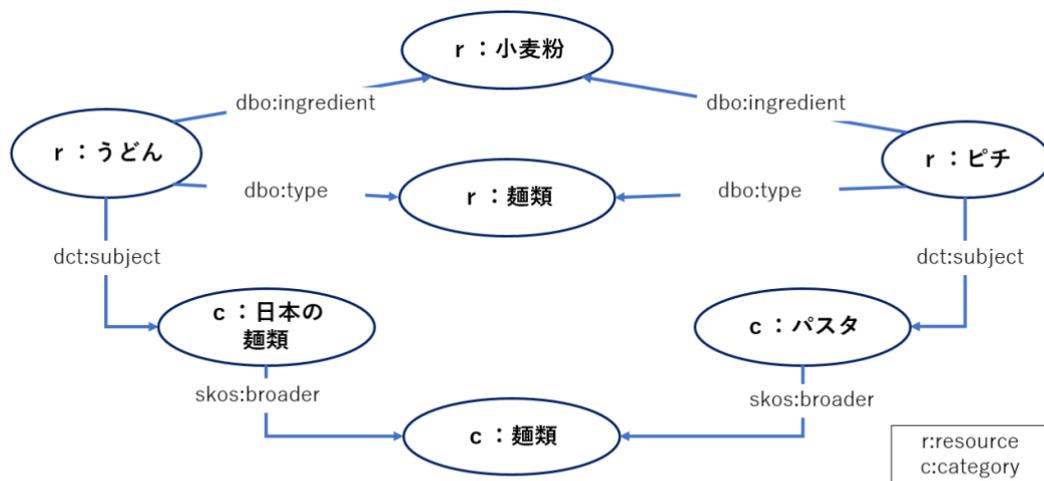


図 8 : リソースとカテゴリの関係図

また `dbo:type` プロパティはカテゴリとの関係を表すプロパティではないが、

リソース間において概念とそれに属する関係であることを示している。これらのリソース間の関係性をもとに `dct:subject`, `skos:broader` を用いた概念階層による検索 (図 9) と `dbo:type` を用いた概念階層による検索 (図 10) の二通りの検索で比較を行う。

検索の流れとしては以下の通りとなる。

1. 探索ステップを 0, 類似料理集合と類似料理候補集合を空集合に設定する。
2. 探索ステップ<3 であれば, 次に進む。そうでなければ, 検索を終了し, 類似料理集合を検索結果として返す
3. 探索ステップ=0 の場合, 検索クエリの料理が持つ `dct:subject` または `dbo:type` のプロパティ値のリソースを共通カテゴリとし, 探索ステップ=1, 2 の場合, これらのリソースに設定されている `skos:broader` または `dbo:type` のプロパティ値を共通カテゴリとする。
4. 次に探索ステップ=0, 1 であれば共通カテゴリのリソースを `skos:broader` または `dbo:type` のプロパティ値にもつリソースを得る狭義の探索を行う。加えて 3 で得られた共通カテゴリを `dct:subject` に持つあるいは共通カテゴリを `dbo:type` に持ち, それ以上狭義のリソースを持たないリソースを類似料理集合に追加する。探索ステップ=2 である場合はこのステップをスキップする。
5. 4 で得られたリソースを `dct:subject` または `dbo:type` のプロパティ値にもつリソースを類似料理候補集合に追加する。
6. 類似料理候補集合の各リソースのうち, `dbo:ingredient` のプロパティ値が, 検索クエリの料理の `dbo:ingredient` のプロパティ値と同じ, かつ, `dbo:country` のプロパティ値が検索クエリの地域情報と同じものを類似料理集合に追加する。
7. 共通カテゴリのリソースが持つ `skos:broader` または `dbo:type` のプロパティ値を共通カテゴリとし, 探索ステップをインクリメントする。
8. 類似料理候補集合を空集合に設定し, 2 に戻る。

本検証では 4.2.1 節で述べた原材料と出身国に二つのプロパティによる絞り込みに加え, 上記二通りの概念階層を用いた検索により候補を求める。

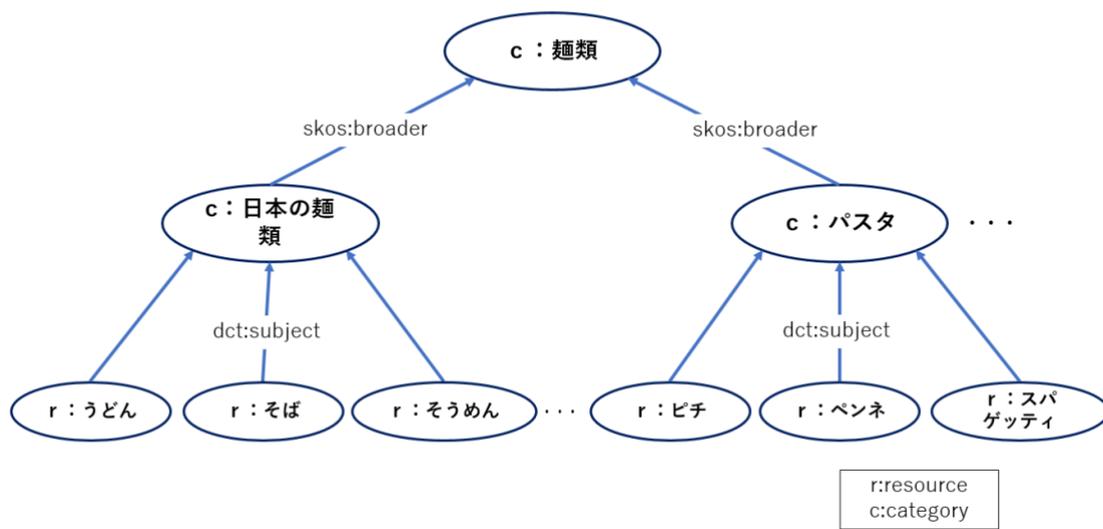


図 9 : 概念階層の例 (dct:subject,skos:broader)

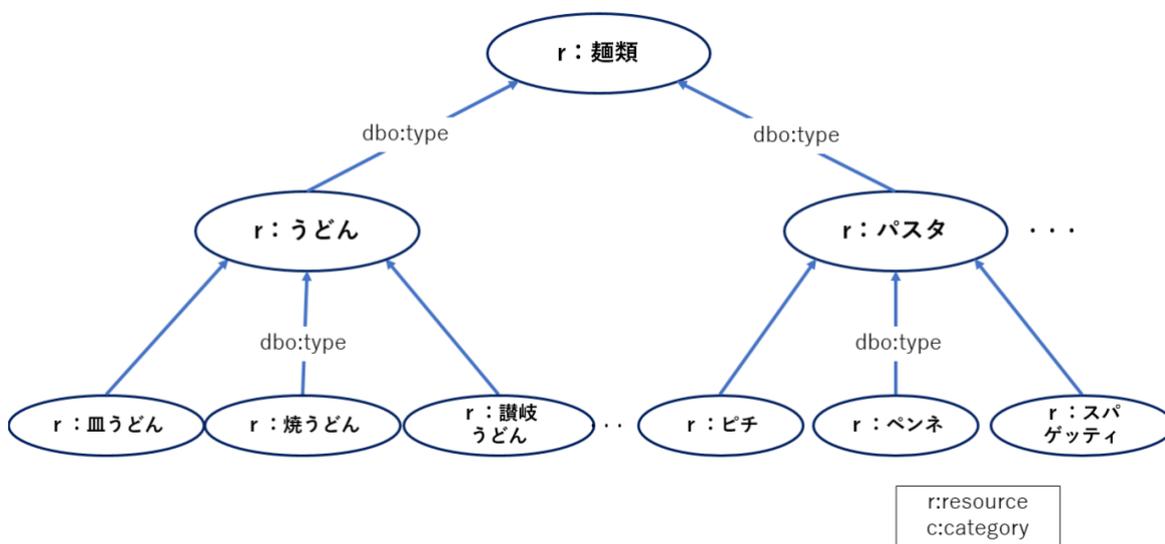


図 10 : 概念階層の例 (dbo:type)

第5章 評価と考察

本章では4章で述べた検索パスをもとに評価・考察を行う。評価の指標として得られた候補の精度である適合率を以下のように定義する。

$$\text{適合率} = \frac{\text{候補の中で類似している料理数}}{\text{類似料理の候補数}}$$

また候補として挙げられた料理が類似料理であるか判定を行うために、日本の料理に理解のある中国出身とタイ出身の協力者に協力していただき適合率の検証を行った。

5.1 評価

表 1 : dct:subject をもとにした検索結果

料理名	China	Thailand
うどん	1/8	0/3
きりたんぼ	0/7	0/3
おでん	0/21	0/6
かき氷	0/1	候補なし
冷奴	候補なし	候補なし
さつま揚げ	候補なし	候補なし
ういろう	3/22	0/6
お好み焼き	候補なし	候補なし
雑炊	0/6	0/3
けんちん汁	0/5	0/5
たくあん	候補なし	候補なし
筑前煮	0/4	0/4
たいやき	候補なし	候補なし
ヒラヤーチー	0/28	0/9
刺身	候補なし	候補なし

表 1 では `dct:subject,skos:broader` のプロパティで日本食 15 品目の検索を行った結果である。日本料理→中国料理では類似する料理が確認できたのが 15 品目中 2 品目,日本料理→タイ料理に関しては類似料理が候補として得られなかった。また類似料理が候補として得られた「うどん」,「ういろう」の適合率に関しても 12.5%,13.6%と低い値となった。

表 2 : `dbo:type` をもとにした検索結果

料理名	China	Thailand
うどん	候補なし	候補なし
きりたんぽ	候補なし	候補なし
おでん	候補なし	候補なし
かき氷	候補なし	候補なし
冷奴	候補なし	候補なし
さつま揚げ	候補なし	候補なし
ういろう	2/2	候補なし
お好み焼き	候補なし	候補なし
雑炊	候補なし	候補なし
けんちん汁	候補なし	候補なし
たくあん	候補なし	候補なし
筑前煮	候補なし	候補なし
たいやき	0/1	候補なし
ヒラヤーチー	候補なし	候補なし
刺身	候補なし	候補なし

表 2 は `dbo:type` をもとに同様に検索した結果である。ほとんどすべての項目で候補が得られないという結果になった。

5.2 考察

5.2.1 プロパティによる検索結果の影響

今回の検証では `dct:subject,skos:broader` を検索パスに用いた方が多くの候補を得ることができた。この結果に関してはプロパティの設定率が大きく影響していると考えられる。図 6 にもあるように確認したリソース全てに `dct:subject` が設定されており共通カテゴリの探索が行えていたが、`dbo:type` での検索パスはプロパティが設定されていないため探索を行うことができていない品目を複数確認した。この `dbo:type` プロパティは `wikipedia` に存在するテンプレートにマッピングされていないと `DBpedia` 上のプロパティとして設定されることがないため `dct:subject,skos:broader` の検索パスと比較した際に候補数に大きな差が生じてしまったと考えられる。一方で `dct:subject,skos:broader` の検索パスでは多くの候補こそ得られたものの適合率としてはかなり低めの値となった。一番の要因としてはプロパティ値として設定されるカテゴリによる影響があったためであると考えられる。例えば図 11 は「そば」のページに設定された `dct:subject` のプロパティ値一覧である。図 12 のように `dbc:Japanese_noodles`→`dbc:Noodles` のように候補を絞り込むのに有効な検索パスも存在するが、`dbc:Japanese_cuisine` のように「○○の国の料理」というプロパティも存在しており、このようなカテゴリが検索パスの途中で探索されることにより関連性の薄い候補が多く挙げられてしまっていた。今回の検証で複数候補を得られた「ういろう」のページにはこのようなカテゴリが設定されておらず、他の品目よりも高い適合率の結果が得られたと考えられる。

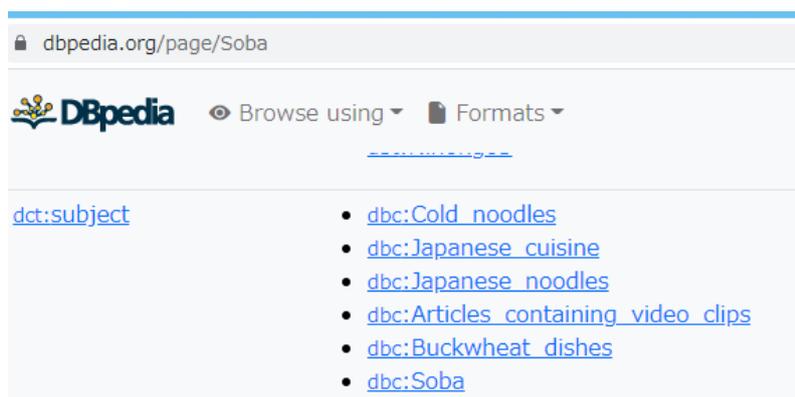


図 11 : そばの持つ `dct:subject` プロパティ値



図 12 : Japanese_noodles のプロパティ値

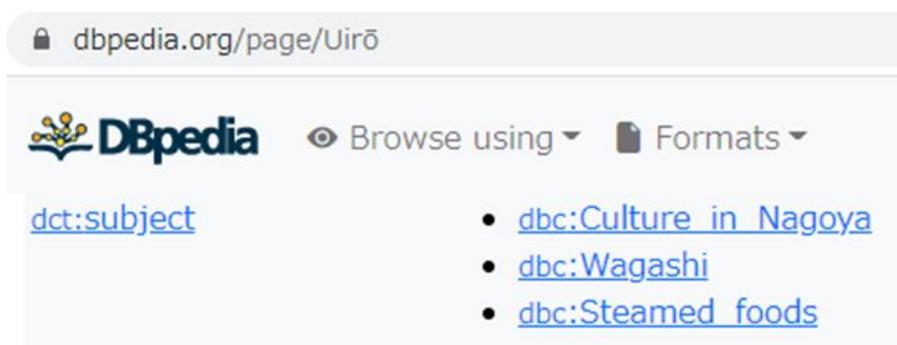


図 13 : 「ういろう」のプロパティ

5.2.2 文化的背景による影響

検索パスに用いたプロパティ以外にも影響が大きいと考えられるのが原材料を表す `dbo:ingredient` プロパティである。このプロパティもいくつかの料理では設定されておらず、設定されていない料理の共通点として原材料が特定されない点が挙げられる。今回調べた 15 品目の中で `dbo:ingredient` が設定されていなかった料理として「寿司」, 「おでん」などがある。これらの料理は多数の食材が用いられるためプロパティ値の設定が行われていないと考えられる。

一方で `dbo:ingredient` プロパティが設定されていても類似料理の候補が得られない料理も存在した。特にタイ料理では類似料理の候補が得られなかったため本検証で得られた結果をもとにタイ料理の協力者に実際に今回検証を行った 15 品目の中にタイ料理で類似している料理がないか聞き取りを行った。一部料理に関しては認知する範囲内の類似料理が存在しないものもあったが、多くの料理に関しては類似する料理の回答を得られた。それらの料理のリソースを確認し今回の検索パスで候補として挙がらなかった要因を確認すると、原材料が異なるため候補として挙がらないという事例が確認された。例えば「うどん」や「そ

うめん」の類似料理として回答を得られたタイ料理に「Khanom chin」がある。図 13 のように「うどん」や「そうめん」の原材料が小麦であるのに対して、図 14 のように「Khanom chin」の原材料は米粉となっている。今回候補を得られなかった料理全てがこの事例に該当するわけではないが、これらの検証結果により文化圏によって類似する料理であっても原材料などが異なる可能性が確認された。

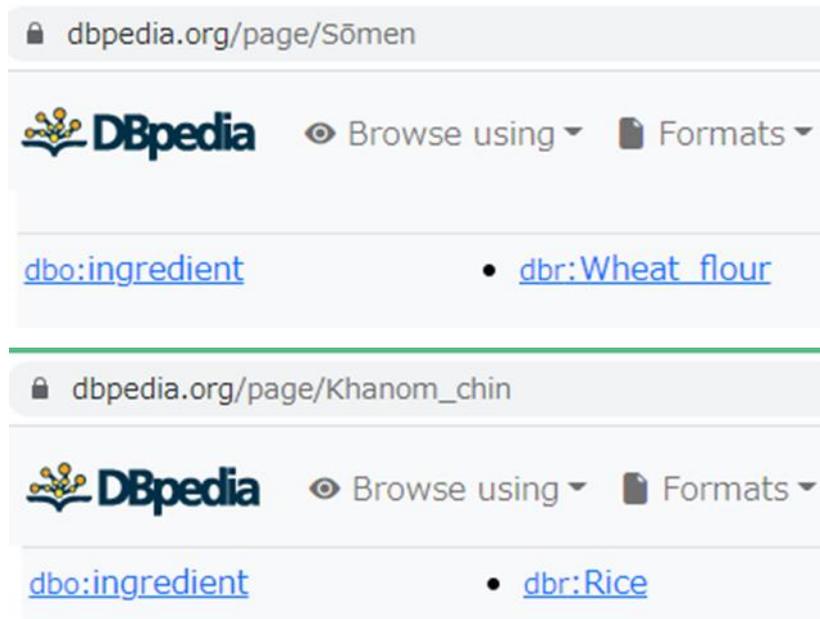


図 14 : そうめん（上）と Khanom chin（下）の原材料比較

この結果を受けて一つも候補が得られなかったタイ料理において、`dbo:type` をもちいた検索パスで原材料の絞り込みを行わずに候補の検索を行った結果表 3 の候補が得られた。表 2 同様 `dbo:type` のプロパティが設定されていない料理に関しては候補が得られなかったが、プロパティが設定されている料理に関しては候補が得られた。実際類似料理が候補として得られたのは 2 品目であったが、類似料理が得られなかった料理の候補でも風味など部分的な類似性があるものが多く確認することができた。

5.2.3 その他の要因と展望

本節ではここまで挙げた要因以外にも大きな影響を受けたと考えられる要因について述べる。

プロパティ設定の有無以外で候補が挙がらない要因として考えられるのが `dbo:country` のプロパティ値である。日本で食べられる料理は日本発祥のものも

表 3：原材料による絞り込みなしの結果

料理名	Thailand
うどん	候補なし
きりたんぽ	0/2
おでん	候補なし
かき氷	2/12
冷奴	候補なし
さつま揚げ	候補なし
ういろう	0/2
お好み焼き	候補なし
雑炊	0/2
けんちん汁	候補なし
たくあん	候補なし
筑前煮	候補なし
たいやき	1/4
ヒラヤーチー	0/1
刺身	候補なし

多い。そのため日本で食べられている食べ物の多くは dbr:Japan のプロパティ値が設定されている。しかしタイで食べられる一部の料理は東南アジア全域で食される料理なども存在する。

そのような料理は CSA をもとにした類似料理ではあるがプロパティ値にはその料理の発祥国しか設定されていないことも多く候補として挙がらない事例が存在した。このような料理には図 15 のように dbo:region のプロパティ値に「東アジア」といった国内の地域・地方ではなく、国より広域の値が設定されることも確認された。また今回の検証ではアジアに属する二カ国のみとなってしまったため他地域の料理の候補を検索する際の影響について評価を行えなかった。アジア圏では小麦の代用品として米粉が原材料に用いられることが確認されたが、このような文化差は地域ごとに異なると考えられるので、今後地域ごとに得られ

る候補にどのような差が生じるか検証が必要であると考える。

dbo:country	<ul style="list-style-type: none">• dbr:China
dbo:ingredient	<ul style="list-style-type: none">• dbr:Chinese_noodles• dbr:Broth
dbo:ingredientName	<ul style="list-style-type: none">• Wheat noodles, meat-basedbroth, vegetables or meat
dbo:region	<ul style="list-style-type: none">• dbr:East_Asia

図 15：拉麺のプロパティ

また一部の料理では類似した料理が一つのリソースに集約されて同一の料理として扱われ、その結果候補として挙がらないという事例も確認された。この事例は「たくあん」のページで確認され、タイ料理ではビートを原材料とした漬物「チャイ」が食されているが、この「チャイ」含め根菜類の漬物が「たくあん」のページに同一の食べ物として集約されており、原材料での絞り込みを行ったにもかかわらず候補として得られなかった。

本検証で提案する検索パスでは類似料理の候補を得られる事例は確認できたが全体を通じて適合率及び候補の列挙率は低い値となってしまった。現状の手法で適合率、列挙率を上げるためには以下の三点が必要となる。

- Wikipedia 上の各リソースでジャンルごとのテンプレートマッピング率の上昇
- DBpedia 全体でプロパティ設定数の向上
- 文化差による影響を考慮した検索パスの改良

本検証で共通カテゴリを求める検索パスである `dbo:type` と `dct:subject` を比較すると前者はプロパティ値の関連性が高い反面プロパティ自体の設定がされていないことがあり、後者はプロパティの設定はされているがプロパティ値が料理の特徴を表すものとなっていないことが多く確認できる。そのため Wikipedia でのテンプレートマッピングが進むことで `dpo:type` のプロパティが進むことが期待される。

検索パスに関しては現状今回使用しなかったプロパティの設定率が低いため、使用するプロパティに関しては今回用いたプロパティからおおよそ変更する点

はない可能性が高い。しかし、原材料などで文化的背景に基づく影響をもとに絞り込みを行うプロパティ値を設定することで検索精度が上がると考えられる。

またプロパティの設定がされていない、あるいは類似料理のリソースそのものが存在しないため候補として挙がらない事例も多く確認できた。そのため話者が比較的少ない言語によるリソースの確保も今後の課題になると考えられる。

第6章 おわりに

日本における重要産業となる観光業がかかえる問題点の一つに、機械翻訳だけでは伝わらない文化的背景に基づく類似料理の情報の提案がある。今回の検証では参考研究をもとに DBpedia を用いて CSA に基づく類似料理の検索手法についての提案及び検証について行った。しかし検索により類似料理の候補が得られたものもあったが、全体的には候補が得られない料理や適合率が低い料理が多数を占めることとなった。今後このような文化的背景に基づく類似料理の提案を行う際に改善すべき点、望まれる展望としてについて以下の数点が考えられる。

検索パスの改良

本研究では検索する料理と得られる候補の間の関連性として共通するカテゴリを持つことをもとに検索パスの作成を行った。今回作成した検索パスでは得られるプロパティ値全てを扱っていたが、考察でも述べた通り、`dbc:Japanese_cuisine` といった関連性が低く適合率を大きく下げる要因になりうるプロパティを検索パスから取り除くことによって適合率の上昇につながると考えられる。

DBpedia のプロパティ設定の促進

DBpedia は Wikipedia をもとにした最大級のデータベースであるが、言語により情報リソース量も異なっており、Wikipedia から DBpedia へのマッピング作業も十分には行われていないのが現状である。今回検証した料理のカテゴリにおいてもプロパティ不足が多くみられた。この作業が進むことで本研究と同様の検索手法を行う場合でも今回の検証で得られた適合率、候補の列挙率の改善が進むのではないかと期待される。また今現在確認できるプロパティでは料理の特徴を示すものが少ないため、茹でる、焼く、蒸すなどといった調理法や甘い、辛い、苦いといった風味についての情報についてプロパティ設定が進むとより効率よく類似料理の検索が行えるのではないかと考えられる。

DBpedia 以外のデータリソースの利用

また今回は参考研究をもとに DBpedia のみを用いて類似料理の候補を求めたが、詳細な調理法に関しては DBpedia のような形式ではプロパティ値として設定することが難しい。そのため DBpedia 以外で Cookpad のようなレシピなどを取り扱うページなどをリソースとして調理法の類似性をもとに検索を行うことが将来的にできるようになれば大幅に適合率を上げることも可能になると考えら

れる。

現在新型コロナウイルスの影響で観光産業は大きな打撃を受けているが、落ち着いた際にどのような文化圏の観光客でも翻訳トラブルが障害にならないようにすることが大切である。今後本研究をもとに翻訳において文化的背景に基づく影響を少しでも減らすべく、検索パスの改良、文化的背景による影響の検出などをすすめていきたい。

謝辞

本研究を行うにあたり，熱心なご指導，ご助言を賜りました村上陽平准教授に深謝申し上げます．また，ご協力いただいた被験者の **Mondheera Pituxcoosuvorn** 助教，**LI CHUANG** さん，張禹王さん，普段からお世話になっている社会知能研究室の皆様にご心より感謝申し上げます．

参考文献

- [1] Victoria Abou Khalil, Toru Ishida, Masayuki Otani, Brendan Flanagan, Hiroaki Ogata, Donghui Lin: Learning culturally situated dialogue strategies to support language learners, *Research and Practice in Technology Enhanced Learning*(2018).
- [2] Victoria Abou Khalil, Toru Ishida, Masayuki Otani, Donghui Lin: A Culturally-Situated Agent to Support Intercultural Collaboration, *Collaboration Technologies and Social Computing*, pp 130-144(2017).
- [3] 加藤文彦: DBpedia の現在 : リンクトデータ・プロジェクト, *情報管理* 第 60 巻 5 号 p. 307-315 (2017)
- [4] Klemmer, SR, Sinha, AK, Chen, J, Landay, JA, Aboobaker, N, Wang, A (2000). Suede: a Wizard of Oz prototyping tool for speech user interfaces, In the *Proceedings of the 13th annual ACM symposium on User interface software and technology* (pp. 1–10). New York: ACM
- [5] 吉岡 真治 : Wikipedia のカテゴリー階層関係の分類 を用いた日本語 Wikipedia オントロジーの分析, 2014 年度 人工知能学会全国大会 (第 28 回) 論文集 (2014), CD-ROM 2J3-4

付録

検索クエリ (dbo:type) のソースコード

```
from SPARQLWrapper import SPARQLWrapper, JSON
import sys

#エンドポイントの指定
sparql = SPARQLWrapper("http://dbpedia.org/sparql")
query22="""
#接頭辞宣言
    PREFIX dbo:<http://dbpedia.org/ontology/>
    PREFIX dbc:<http://dbpedia.org/resource/Category:>
    PREFIX dbp:<http://dbpedia.org/property/>
    PREFIX dbr:<http://dbpedia.org/resource/>
    PREFIX dct:<http://purl.org/dc/terms/>
    PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
    PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
    PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
    SELECT distinct ?label WHERE {
#検索する料理の dbo:ingredient と dbo:type のプロパティ値の取得
    dbr: "" + sys.argv[1] + "" dbo:ingredient ?Ingredient.
    dbr: "" + sys.argv[1] + "" dbo:type ?Concept.
#概念階層による検索. 本クエリでは検索する料理と類似候補が二回広義に
dbo:type のプロパティ値をたどった際に共通のプロパティ値をもつ料理を候補
とする探索となる.
    ?Concept dbo:type ?Parent.
    ?Parent dbo:type ?Grand.
    ?aParent dbo:type ?Grand.
    ?aConcept dbo:type ?aParent.
    ?candidate dbo:type ?aConcept.
#ユーザーの出身国情報と原材料の一致による絞り込み
    ?candidate dbo:country dbr: "" + sys.argv[2] + "".
```

```

?candidate dbo:ingredient ?Ingredient.
?candidate rdfs:label ?label.
?Grand rdfs:label ?Glabel.
#英語版 Wikipedia のリソースの絞り込み
  FILTER ( lang(?label) = "en" )
} LIMIT 100""
#上記クエリをもとにした SPARQL の実行
sparql.setQuery(query22)
sparql.setReturnFormat(JSON)
results = sparql.query().convert()

for result in results["results"]["bindings"]:
  print("候補:" + result["label"]["value"] + " カテゴリ:" + result["Glabel"]["value"])

```