卒業論文

Linked Open Drug Data を用いた
外国人のための医薬品検索

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

2600180046-8

江口　美裕

2021 年度（秋学期）卒業研究 3（CH）

令和 4 年 1 月 31 日

Linked Open Drug Data を用いた外国人のための医薬品検索

江口　美裕

内容梗概

近年，グローバル化が進み外国人の長期滞在者が増加している．長期滞在者にとって，膨大な量の荷物を持っていくことは困難である．その中でも特に， 渡航先の国に持ち込める市販薬の量は，滞在期間に関わらず規定量が定められている場合がある．また，郵送による医薬品の持ち込みも，国によっては認められていない．そのため，必要になった際に渡航先のドラッグストアで自国の市販薬に対応した医薬品を探す必要がある．しかしながら，言語の分からない渡航先では，自国で使用していた医薬品を見つけにくいという問題点が挙げられる．その要因としては，製品名やパッケージが，各国のブランド独自のものであるためである．

そこで，本研究では Linked Open Drug Data（以下 LODD）を用い，各国の薬の製品名で対応する日本の製品名を検索するサービスを構築する．具体的には，私たちの身近にあるドラッグストア等で買うことのできる一般用医薬品に特化し，DrugBank，DBpedia，KEGG の 3 つの LODD を連携し，ユーザが入力した海外の製品名に近い日本の製品名を出力する．対象ユーザは，日本に長期滞在している，アメリカ，カナダ，ヨーロッパの英語圏出身者である．医薬品には，製品名と一般名という名前が付けられている．製品名は各国で異なるが，一般名は有効成分から付けられており，基本的に世界共通である．本研究では，各国と日本それぞれの製品名を結びつけるために，共通である一般名を使用し対応付ける．本手法の実現にあたり，取り組むべき課題は以下の 2 点である．

LODD の連携

本研究では，提供者の異なる 3 つの LODD を用いるため，データ形式，アクセス形式がそれぞれ異なる．このような異種のデータを連携させるためにデータを変換し，海外の一般用医薬品名から日本の一般用医薬品名に辿るパスを明らかにする必要がある．

候補の絞り込み

一般名と製品名が一対多の関係のため，海外の製品名から一般名を介して日本の製品名を辿ると，検索結果の候補となる日本の製品名が多岐にわたる．したがって，ユーザが容易に目的の医薬製品を見つけられるように，その検索結果の候補の中から，ユーザが自国で使用していた医薬製品に近い日本の製品名に絞り込みを行う必要がある．

Linked Open Drug Data (以下 LODD)は，一般に公開されている医薬品に関するデータを調査し，LODD クラウドが公開されている．本研究では，この LODD クラウドから医薬品に関する３つの LODD を使用する．DrugBank とは，薬物に関する情報を包括的に収集し，自由に閲覧できるオンラインデータベースである．DBpedia とは，Wikipedia から構造化された内容を抽出することを目的とするプロジェクトである．KEGG Medicus とは， KEGG DISEASE/DRUG データベース，日本の医療用医薬品添付文書と一般用医薬品添付文書，米国の医療用医薬品添付文書を統合して利用できるインターフェースである．前者の課題に対しては，DrugBank について取得した DrugBank XML データの処理を行った．本研究では医薬品データの中から，一般用医薬品のみに焦点を当てて，海外の製品名，DrugBank ID，国，剤型を使用する．ユーザが入力した海外の医薬品名に対し，同じ医薬品名についての DrugBank ID を抽出する．その DrugBank ID を DBpedia で検索し，KEGG ID を取得する．その後 KEGG ID を利用し，KEGG Medicus 検索ページにてスクレイピングを行ない，日本の製品名を取得する．

後者の課題に対しては，複数の一般名を含む製品名を検索する場合，それら全ての一般名を含む製品名に限定して検索を行うことで，検索候補を削減する．さらに，医薬製品の剤型や量などの指標を用いることで，自国で使用していた医薬品と使用方法が近い日本の製品名に絞り込みを行う．

提案手法の有用性を示すために，提案手法を組み込んだ外国人のための医薬品検索システムを実装し，動作確認を行った．その際，適合率を用いて評価を行う．本研究の貢献は以下の通りである．

LODD の連携

XML，RDF(SPARQL で問い合わせるデータ)，HTML という異なるフォーマットで，提供されるデータを，一般名を表す ID を介して連携させた．さらに，DBpedia を中間に配置することで，データ間で異なる ID の対応付けを可能とした．これにより，380,504 件の海外の製品名と 10,802 件の日本の製品名の紐づけを実現している．

候補の絞り込み

日本の製品名の検索結果候補から，ユーザが自国で使用している医薬製品の剤型を基準に日本の製品名を選択し絞り込みを行う手法を考案した．この手法により，平均 67%の絞り込みを達成し，ユーザによる日本の製品の選択を容易にした．さらに，この絞り込みにより適合率は 85%,再現率は 75%という結果を得た．

Drug Search for Foreigners Using Linked Open Data

Miyu Eguchi

Abstract

Because globalization is progressing, the number of long-term foreign residents has increased in recent years. Long-term residents find it difficult to bring a lot of luggage. There are some instances where the quantity of an over-the-counter (OTC) drug must be brought into the destination. Furthermore, in some countries, mailing is prohibited. As a result, when you need OTC drugs that were used in your country, we'll need to find them in a drugstore in your destination. However, there are issues that make finding what they looking for in the destination difficult. For example, each country's brand has its own product names and packaging.

In order to address this issue, we developed a system that searches for corresponding Japanese product names by product name in each country using Linked Open Drug Data (LODD). We are talking about OTC drugs, which are available at drugstores all over the place. We linked LODDs, DrugBank, DBpedia, and KEGG to generate a Japanese product name that is similar to a foreign product name entered by the user.

Product names vary by country, but generic names are derived from active ingredients and are generally applicable. Common generic names are used in this study to connect product names from each country with those from Japan. To that end, we will address the two issues listed below.

Linking LODD

We linked data provided in various formats, RDF, XML, and HTML, with various access methods, SPARQL, and REST, using IDs that indicated generic names. Furthermore, it is possible to correspond to different IDs via DBpedia. This allows us to connect 380,504 foreign product names to 10,802 Japanese product names.

Narrowing down candidates

Because of the one-to-many relationship between a generic name and a product name, there is a wide range of Japanese product names in this research results to trace foreign product names to a Japanese product name with a generic name. As a result, it is essential to easily narrow down the results to the Japanese product name that the user is looking for.

The Linked Open Drug Data (LODD) cloud researches publicly available data about the drug. We used their LODDs about the drug from this LODD cloud in this study. DrugBank Online is a comprehensive, free-to-use online database of drug and drug target information. DBpedia is a project dedicated to extracting structured content from Wikipedia. The KEGG DISEASE/DRUG database, Japanese prescription and OTC drug package inserts, and US prescription drug package inserts are all integrated and used by KEGG Medicus. To address the first problem, we process the DrugBank XML data from DrugBank.

In this study, we focused on OTC drugs from the drug data and employ foreign products such as DrugBank IDs, countries, and dosage forms. We obtained the DrugBank ID for the same user-entered name. Then, using DBpedia, look up the DrugBank ID and the KEGG ID. Then, using KEGG ID, navigate to the KEGG Medicus search page to obtain the Japanese product name. To address the latter issue, when searching for a product name that includes plural generic names, we limit the search results to product names that include generic names for all of them. Furthermore, use an indicator of the dosage form to narrow down the research.To show effectiveness, we install drug searches for foreigners and confirm the operation. In the case of, evaluate using the fit rate.

The contributions of this research are as follows.

Linking LODD

We were linked together data provided in different formats, RDF, XML, and HTML with different access methods, SPARQL and REST using IDs indicated generic name. In addition, it is possible to correspond different IDs mediating DBpedia. By doing this, it is possible to link 380,504 foreign product names with 10,802 Japanese product names.

Narrowing down candidates

Based on the dosage forms of pharmaceutical products sought by the user, we devised a method for narrowing down Japanese product names from the research result candidates. Using this method, it is possible to narrow down candidates by 67% on average. Furthermore, it increased the fit rate by 85% by narrowing down candidates.

Linked Open Drug Data を用いた外国人のための医薬品検索

目次

第1章 はじめに

近年，グローバル化が進み外国人の長期滞在者が増加している．長期滞在者にとって，膨大な量の荷物を持っていくことは困難である．その中でも特に， 渡航先の国に持ち込める市販薬の量は，滞在期間に関わらず規定量が定められている場合があり，渡航中に必要と考えられる量を超えての持ち込みは制限されている．また，郵送による医薬品の持ち込みも，国によっては認められていない．そのため，必要になった際に渡航先のドラッグストアで自国の市販薬に対応した医薬品を探す必要がある．しかしながら，言語の分からない渡航先では，自国で使用していた医薬品を見つけにくいという問題点が挙げられる．その要因としては，製品名やパッケージが，各国のブランド独自のものであるためである．

そこで，本研究ではLinked Open Drug Data（以下LODD）を用い，各国の薬の製品名で対応する日本の製品名を検索するサービスを構築する．具体的には，私たちの身近にあるドラッグストア等で買うことのできる一般用医薬品に特化し，DrugBank，DBpedia，KEGGの3つのLODDを連携し，ユーザが入力した海外の製品名に近い日本の製品名を出力する．対象ユーザは，日本に長期滞在している，アメリカ，カナダ，ヨーロッパの英語圏出身者である．医薬品には，製品名と一般名という名前が付けられている．製品名は各国で異なるが，一般名は有効成分から付けられており，基本的に世界共通である．本研究では，各国と日本それぞれの製品名を結びつけるために，共通である一般名を使用し対応付ける．本手法の実現にあたり，取り組むべき課題は以下の2点である．

LODDの連携

本研究では，提供者の異なる3つのLODDを用いるため，データ形式，アクセス形式がそれぞれ異なる．このような異種のデータを連携させるためにデータを変換し，海外の一般用医薬品名から日本の一般用医薬品名に辿るパスを明らかにする必要がある．

候補の絞り込み

一般名と製品名が一対多の関係のため，海外の製品名から一般名を介して日本の製品名を辿ると，検索結果の候補となる日本の製品名が多岐にわたる．したがって，ユーザが容易に目的の医薬製品を見つけられるように，その検索結果の候補の中から，ユーザが自国で使用していた医薬製品に近い日本の製品名に絞り込みを行う必要がある．

以下，本論文は 2 章において LODD についての関連研究を述べ，3 章では LODD 連携を用いた外国人のための医薬品検索システムの概要について述べる．4 章では，絞り込みを行う経緯を，5 章では候補の絞り込みの適切さを評価するために，適合率，再現率を用いて行った結果を述べる．

第2章 関連研究

本章では，本研究で取り扱う Linked Open Data について説明する．また，Linked Open Drug Data の関連研究[1]について記述する．

2.1 Linked Open Data

Linked Open Data は，Web 上でコンピュータ処理に適したデータを公開・共有するためのものである．以下の 4 つの原則に従うものはオープンデータとして公開されているため，誰でもデータベース間を結ぶトリプルを使ってデータベース中の関連する要素を繋ぐことが可能である．

1. あらゆる事物に URI を付与すること
2. 誰でも事物の内容が確認できるように，URI は HTTP 経由で参照できること
3. URI を参照したときは，標準の技術を使用して関係する有用な情報を利用できるようにすること
4. より多くの事物を発見できるように，他の URO へのリンクを含めること

RDF(Resource Description Framework)は，Web 技術を使用してデータを記述するためのフレームワークであり，主語(Subject)・述語(Predicate)・目的語(Object)の 3 つのリソースの組み合わせが最小単位となる．主語はリソース，述語はプロパティ，目的語はリテラルを表している．

図 1:RDF モデル

2.2 Linked Open Drug Data

Web 上には，医薬品に関する豊富な情報が掲載されている．データソースは，医薬品の化学的な結果から薬化学の結果，遺伝子発現に対する薬の影響，臨床試験における薬の結果など多岐にわたる．これらのデータは一般的に結びついていないため，洞察を得るのが容易ではない．

LODD は，一般に公開されている医薬品に関するデータを調査し，データセット

の Linked Data 表現を作成した．ここでは，医薬品の研究開発データ共有の基盤として，Linked Data の重要性が高まっていることを説明している．活動内容として，現在までに，LODD プロジェクトの参加者が作成した医薬品の研究開発に関連する 12 のオープンアクセスデータセットをリンクデータとして公開している．これらは，DrugBank，ClinicalTrials.gov，DailyMed，ChEMBL，Diseasome，TCMGeneDIT，SIDER，STITCH，Medicare formulary，最近追加された 3 つの項目 RxNorm，Unified Medical Language System と WHO Global Health Observatory である．最新の情報を得るために，元のデータセットを定期的に検索し，Linked Data の表現を更新している．

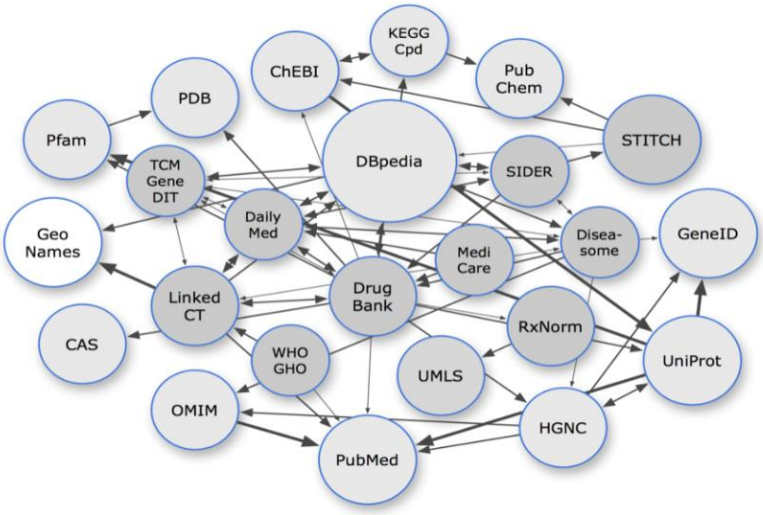

図 2: LODD クラウド

上図の LODD クラウドは，LODD データセットの一部（濃い灰色），関連するバイオメディカルデータセット（薄い灰色），関連する汎用データセット（白），およびそれらの相互接続のグラフで成り立っている．線の重さはリンクの数に対応し，矢印の方向はそのリンクを含むデータセットを示す．
このように，LODD クラウドの公開や，製薬業界が Linked Data を取り入れ始めている中で，リンクされたバイオメディカルデータセットの数はここ数年で大幅に増加しているが，エンドユーザがこれらのデータセットを探索・検索することを可能にする優れたアプリケーションはまだ著しく不足している．

第3章 外国人のための医薬品検索システム

医薬品は，大きく分かれて一般用医薬品と医療用医薬品に分類される．一般用医薬品は，医師の処方箋を必要とせずに購入でき，ドラッグストアで買える医薬品である．医療用医薬品は，医師の処方せんを必須とする医薬品である．本研究では，ユーザの入力，システムの出力それぞれ共に一般用医薬品のみとする．また，医薬品の名前は 2 種類存在する．製品名は各国で異なるが，一般名は有効成分から付けられており，基本的に世界共通である．一般名と製品名は，一対多の関係であり，本研究ではここに着目した．以下は関係性を表した図である．

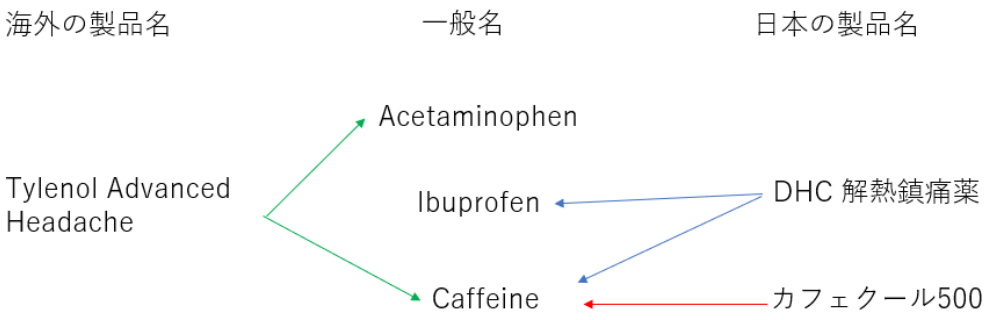

図 3:一般名と製品名の関係

一般名が一つである場合に加えて，有効成分を複数含んで成り立っている場合も存在する．このような製品の場合は，一般名が複数存在するため，DrugBank ID も複数出力される．それぞれの一般名に対する日本の製品名の検索結果を出力し，そして，共通する日本の製品名のみを抽出し，出力する．

3.1 検索プロセスの概要

３つの使用データの関連図および検索フローを以下に示す．

1. ユーザが海外の製品名を入力する．
2. 製品名で文字列一致を行い，DrugBank からこの薬の DrugBank ID を取得する．
3. 2 より得た ID と DBpedia で同じ ID を取得する．
4. DBpedia からこの薬の KEGG ID を取得する．
5. 4 より取得した ID を KEGG Medicus で検索し，日本の製品名の候補を取得する．

6. 5 の製品名の候補一覧から，ユーザの目的に沿うように，剤型を指標に絞り込む作業を行う．

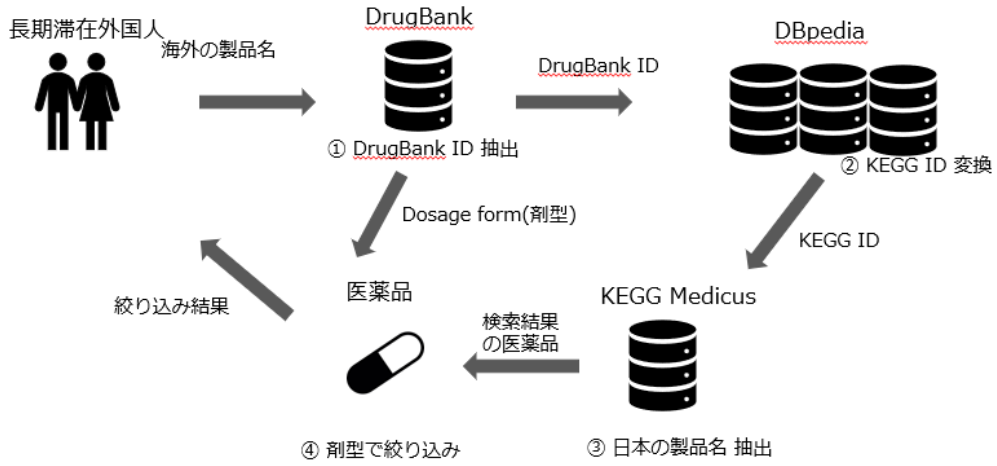

図 4:検索フロー

1 では，ユーザに海外の製品名を入力する過程において，製品名の部分一致を行っている．部分一致で合致される海外の製品名は複数存在するため，海外の製品名ごとに分けて，最終的に日本の製品名を出力する．
6 では，ユーザの目的に沿った絞り込みを行う．その過程の具体的な方法として，KEGG Medicus より抽出された日本の製品名の検索結果と，DrugBank XML データから抽出した剤型を使用する．その際，同じ海外の製品名であっても，剤型が数種類存在する場合も見受けられる．その場合には，ユーザに選択肢を設けて，剤型を選択させることで，ユーザの要望に沿った医薬品を提案できる．

3.2 LODD

本章では，LODD クラウドの中から，3 つの医薬品に関する LODD についての詳細を述べる．本研究では，DrugBank，DBpedia，KEGG の 3 つの LODD を使用する．

3.2.1 DrugBank

DrugBank[1][3]とは，薬物に関する情報を包括的に収集し，自由に閲覧できるオンラインデータベースである．

[1] https://go.drugbank.com/

表 1:DrugBank 内容

クラス	エントリー数
全エントリー	14,594件
低分子医薬品	2,717件
バイオ医薬品	1,510件
栄養補助食品	132件
実験薬	6,657以上

全部で 14,594 件の医薬品項目が存在し，各エントリーには 200 以上のデータフィールドがある．情報の半分は薬剤や化学物質データである．
本研究では，DrugBank より提供された DrugBank XML データを用いる．このデータより，DrugBank ID，name(海外の製品名)，country，dosage-form(剤型)を抽出した．データ構造の部分木を以下に示す．

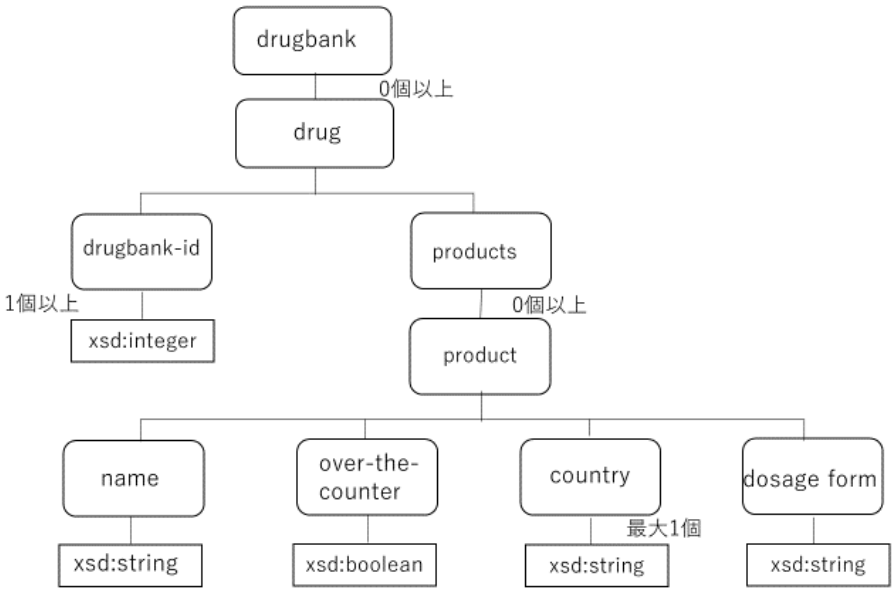

図 5:DrugBank XML データ構造の部分木

xml.etree.ElementTree モジュールを利用し，XML データの解析を行い，必要な箇所を抽出した．その結果の例は次のとおりである．

表 2:データ抽出の例

DrugBank ID	name	over-the-counter	country	dosage-form
DB00001	Refludan	FALSE	EU	Powder
DB00002	Erbitux	FALSE	Canada	Solution
DB00027	Antibiotic Cream	TRUE	Canada	Cream
DB00027	Polysporin for Stitches	TRUE	Canada	Ointment
DB00316	Tylenol	TRUE	US	Tablet, film coated
DB00316	Tylenol Cold Plus Flu Severe	TRUE	US	Tablet
DB00316	Tramacet	FALSE	Canada	Tablet

<over-the-counter>というタグは,OTC 医薬品いわゆる一般用医薬品からきている．このタグの要素は，True または False の値を持っており，<over-the-counter>の要素が True の場合は，同じ配列にある<name>の要素値が一般用医薬品の製品名を表している．一方で，<over-the-counter>の要素が False の場合は,<name>の要素値が医療用医薬品の製品名であることを示している．今回は，一般用医薬品のみに焦点を当てているため，要素が True であるものに絞り解析を行う．ユーザの入力と<name>の要素値で文字列一致したものの DrugBank ID と dosage form を出力する．ユーザの入力は，上図にあるように EU，Canada，US の英語圏のみを対象とする．

3.2.2 DBpedia

DBpedia[2][4]とは，Wikipedia から構造化された内容を抽出することを目的とするプロジェクトである．RDF を使用して情報を抽出する．DBpedia は，アクセス手段として公開 SPARQL エンドポイントを提供しているため RDF 用クエリ言語 SPARQl を使用し問い合わせを行う．

以下は本研究で使用したクエリである．

```
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?URI ?drugbank_id ?kegg_id
WHERE
{
    ?URI dbo:drugbank ?drugbank_id.
    FILTER (?drugbank_id='""" + word + """')
    ?URI dbo:kegg ?kegg_id.
    FILTER CONTAINS (?kegg_id, "D")
}
```

[2] https://www.dbpedia.org/

図 6:SPARQL クエリ例

処理の内容としては，DrugBank データから抽出した DrugBank ID がリテラルとして存在する DBpedia の医薬品ページを問い合わせする．word という Python 変数を使用し，DrugBank ID を格納している．DBpedia の医薬品ページに問い合わせ後，出力としてその医薬品にまつわる KEGG ID を取得する．FILTER を用いて絞り込む理由としては，KEGG ID が 2 種類リテラル値として存在するためである．ここでは，必要となる’D’から始まる番号を指定している．以下に，クエリの明確化のため，使用したプロパティと入力の意味を示した表を示す．

表 3:本研究で使用するプロパティ

プロパティ	リテラル
dbo:drugbank	DBXXXXX
dbo:kegg	DXXXXX

表 4:SPARQL クエリの入力とその意味

入力	入力の意味
PREFIX	後ろに示しているURIを省略して，それ以降のクエリを書くことを示す
SELECT	取得したい変数を指定する
WHERE	{}内に検索したいRDFのトリプルを記述する
FILTER	絞り込みたい場合に使用する

また，以下は RDF スキーマ，本稿のクエリの内容を可視化した例と Python 変数 word を使用する前のクエリによって問い合わせを行った結果の例である．

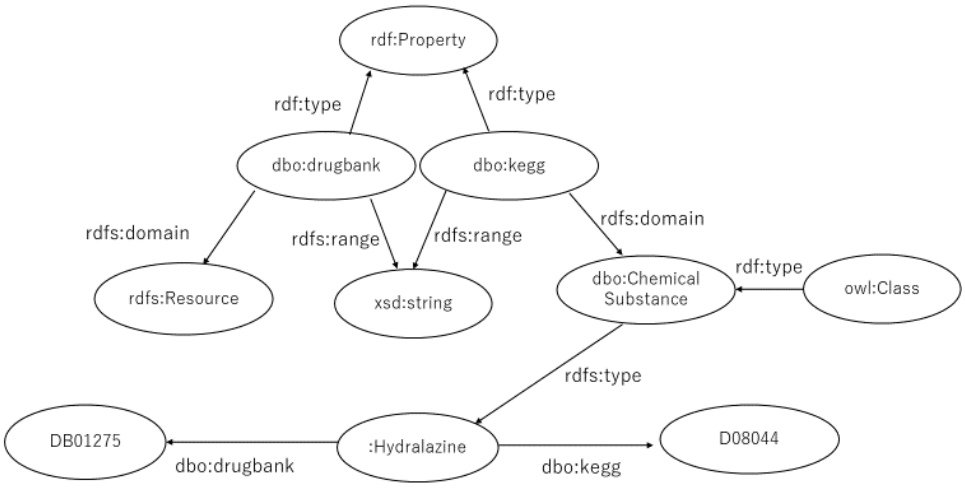

図 7:RDF スキーマ

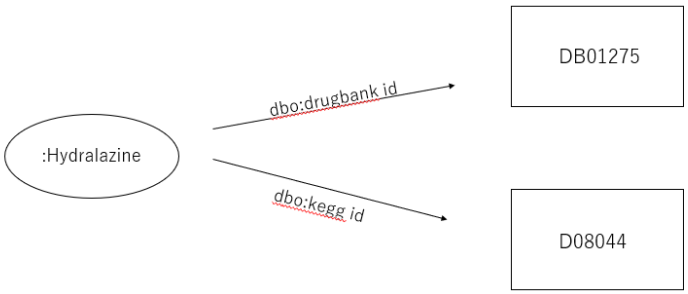

図 8:クエリの内容を可視化した例

SPARQL | HTML5 table

URI	drugbank_id	kegg_id
http://dbpedia.org/resource/Hydralazine	"DB01275"	"D08044"
http://dbpedia.org/resource/Imatinib	"DB00619"	"D08066"
http://dbpedia.org/resource/Methylprednisolone	"DB00959"	"D00407"
http://dbpedia.org/resource/Micafungin	"DB01141"	"D02465"
http://dbpedia.org/resource/Norelgestromin	"DB06713"	"D05205"
http://dbpedia.org/resource/Oseltamivir	"DB00198"	"D00900"

図 9:クエリによって問い合わせを行った結果

ここで抽出したクエリ結果を用いて，KEGG ID(D 番号)のみを抽出し，KEGG で使用する．

3.2.3 KEGG

KEGG[3][5]とは，細胞，生物，生態系などの生命システムのデータベースである．KEGG Medicus は，疾患情報を掲載している KEGG DISEASE，構造と成分の観点から医薬品情報をまとめた KEGG DRUG，疾患関連ネットワークのバリエーションをまとめた KEGG NETWORK の 3 つのデータベースから構成されているデータベースである．ここでは，日本医薬品情報センターが提供する医療用医薬品・一般用医薬品の情報も統合的に利用できる．本研究では，KEGG DRUG で一般名ごとに付けられている KEGG ID を，KEGG Medicus の検索ページの入力として，各 ID(D 番号)に対応する日本の製品名を取得する．
日本の製品名を取得するために，KEGG Medicus 検索ページで出力される Web ペ

3 https://www.genome.jp/kegg/

ージから，スクレイピングを行った．
スクレイピングを行う過程を示す．

1. DBpedia から取得した KEGG ID を item という Python 変数に格納し，ターゲットとする URL の一部分に置き換える．(以下，参照)

```
target_url =
"https://www.kegg.jp/medicus-
bin/search_drug?display=otc&thumbnail=&method=&current_submit=&search_keywo
rd={}&submit=検索".format(item)
```

2. この URL において，requests ライブラリを用いて HTML を取得する．
3. HTML からデータを取得するために，Beautiful Soap という Python の Web スクレイピングに適したライブラリを用いる．

スクレイピングを行うための HTML のデータ構造の部分木を以下に示す．

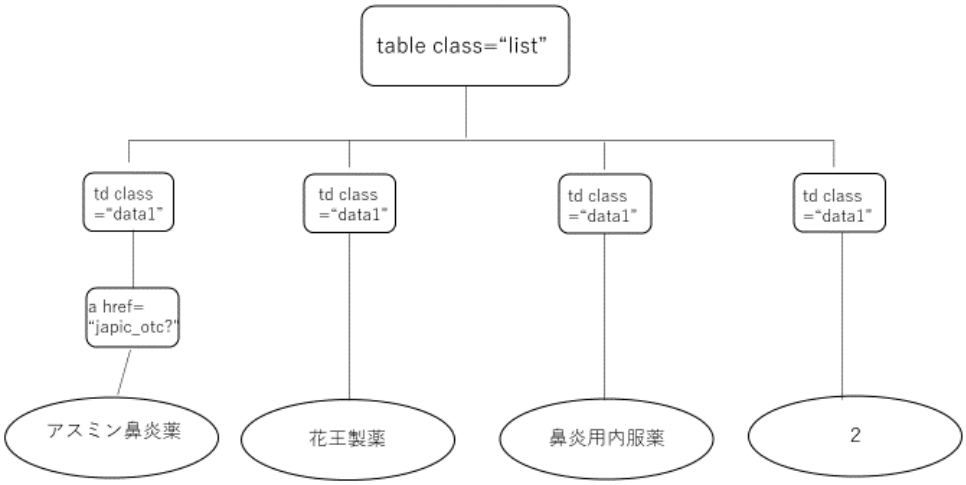

図 10:HTML のデータ構造の部分木

図 10 が示すように，<td class>の要素は 4 つに分かれており，製品名，製造提供元，小分類(医薬品分類)，リスク区分の番号である．
また，明確化のために，HTML のキャプチャを図 11 に示す．

```
<tr>

 <td class="data1"><a href="japic_otc?japic_code=J1201000189">アスミン鼻炎薬</a></td>
 <td class="data1">薬王製薬 (株) </td>
 <td class="data1">鼻炎用内服薬</td>
 <td class="data1">2</td>

</tr>

<tr>

 <td class="data1"><a href="japic_otc?japic_code=J1101000305">コンタック600ファースト</a></td>
 <td class="data1">グラクソ・スミスクライン・コンシューマー・ヘルスケア・ジャパン (株) </td>
 <td class="data1">鼻炎用内服薬</td>
 <td class="data1">2</td>

</tr>
```

図 11:HTML のデータ構造キャプチャ

今回は，KEGG ID に対応する日本の製品名の一覧を取得したいため，<a href=”japic_otc?>のタグの要素を全て出力させる．

以上，3 つの LODD を使用し，DBpedia を中間に配置することで，データ間で異なる ID の対応付けを可能とした．その結果，DrugBank と KEGG をリンクさせることが可能となり，380504 件の海外の製品名と 10802 件の日本の製品名の紐づけを実現した．

3.3 システム構成

図は本研究で提案する外国人のための医薬品検索システムの構成図である．ユーザが海外の製品名を入力すると，入力データはDrugBankが提供しているXMLデータと文字列一致する．その後出力として DrugBank ID を返す．DBpedia では，入力を DrugBank ID とし，SPARQL を用いて問い合わせを行う．その結果 KEGG ID を出力として返す．KEGG では，KEGG ID を入力として，KEGG Medicus 検索ページからスクレイピングを行い，日本の製品名を検索結果として出力する．
Web スクレイパーから取得する検索結果と XML パーサから取得する剤型のデータを用いて，絞り込み器で処理を行う．剤型で絞り込みを行った結果をユーザに

返す.

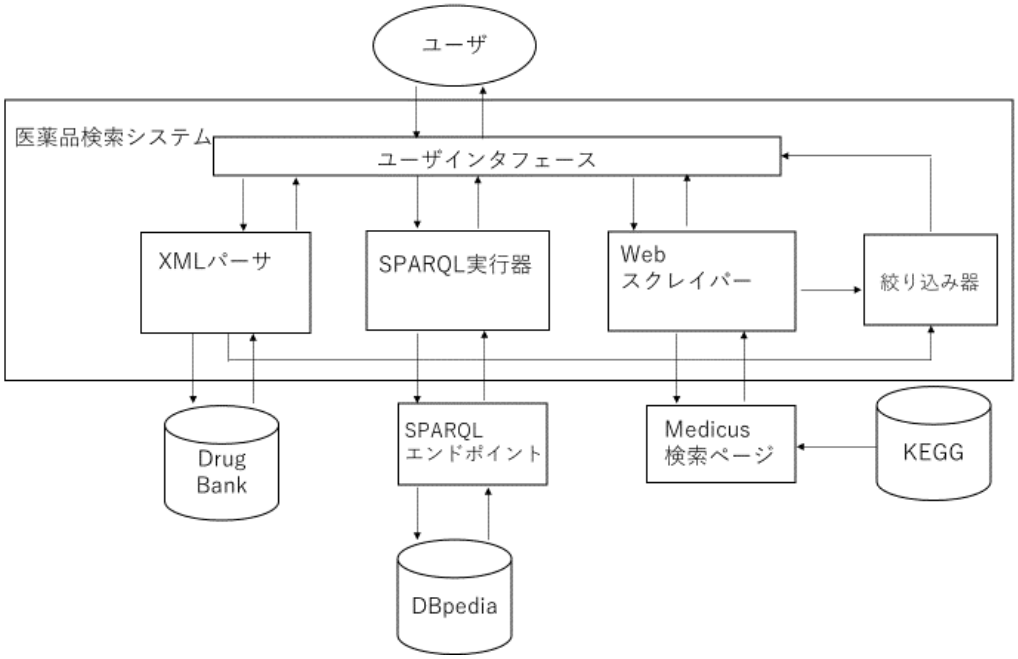

図 12:システム構成図

また，システム全体のフローチャートを下に示す.

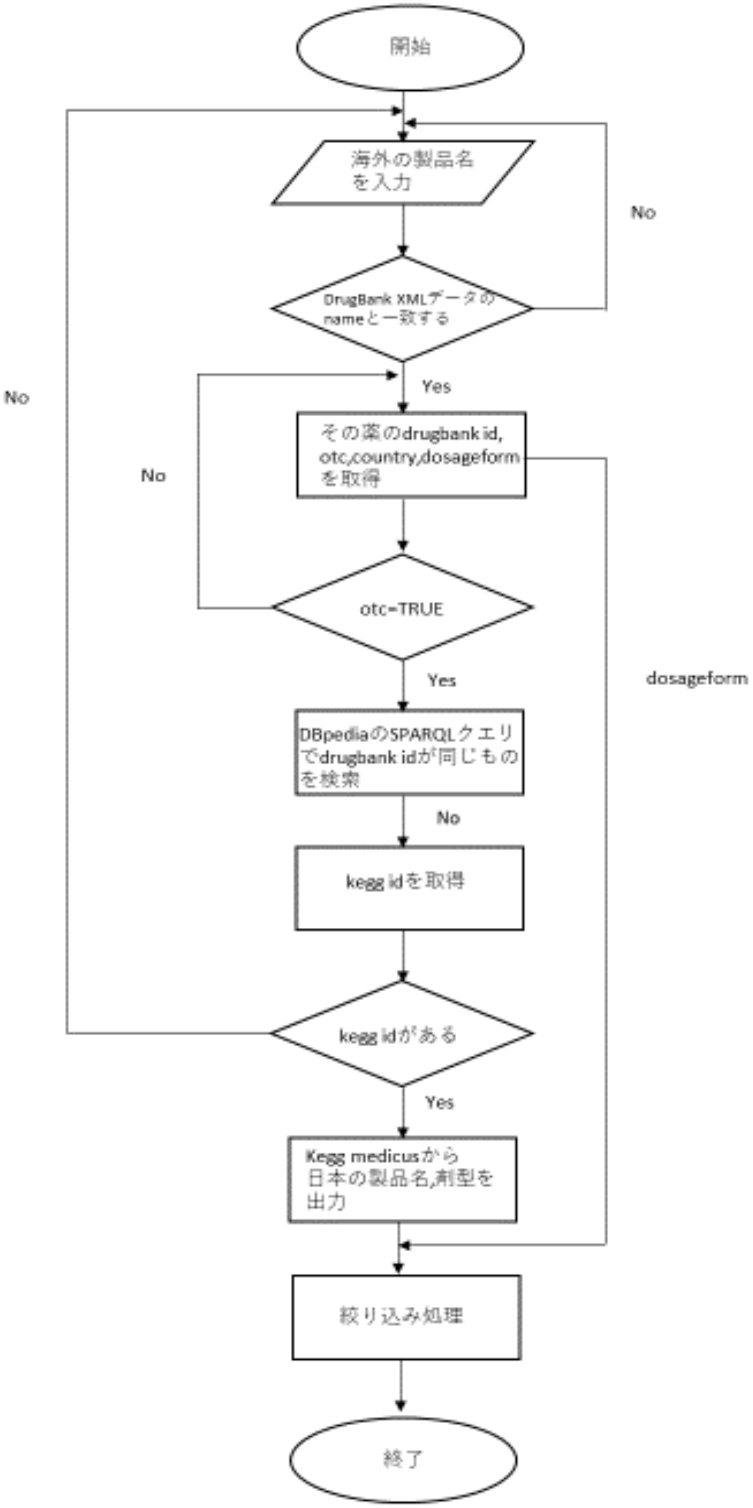

図 13:システム全体のフローチャート

第4章 検索結果の絞り込み

本章では，検索結果の日本の製品名リストから，ユーザの目的に近いものをユーザに返すために剤型を用いて絞り込みをした経緯について説明する．

4.1 剤型の対応付け

DrugBank で提供されている XML データに dosage form という剤型を表すものがある．これは，U.S Food and Drug Administration(以下，FDA とする)[4]が定めているものである．

剤型の対応付け方法としては，各 KEGG Medicus 検索ページ(図 14)から辿る医薬品情報(図 15)から，包装という項目を使用する．HTML より，包装部分と製品名を抽出し，DrugBank XML データと対応付ける．DrugBank XML データが Tablet であれば，KEGG Medicus の包装部分の X 錠の単位部分が当てはまる．これを利用し，それぞれを対応付ける．DrugBank，KEGG の 2 つの異種データを使用するため，フォーマットも言語も異なる．そこで表 5 に示すように，対応付けを行うことで，剤型による検索結果の絞り込みを行った．


KEGG MEDICUS 医薬品情報

利用法 Top

D01332 ☐成分も検索 検索 全文検索

KEGG DISEASE (0) | KEGG DRUG (1) | 医療用医薬品 (39) | 一般用医薬品 (8) | FDA Rx drug (0)

8 件中 1 ～ 8 を表示

製品名	会社名	薬効	リスク区分
アスミン鼻炎薬	薬王製薬（株）	鼻炎用内服薬	2
コンタック600ファースト	グラクソ・スミスクライン・コンシューマー・ヘルスケア・ジャパン（株）	鼻炎用内服薬	2
ザジテンAL点眼薬	グラクソ・スミスクライン・コンシューマー・ヘルスケア・ジャパン（株）	アレルギー用点眼薬	2
ザジテンAL鼻炎カプセル	グラクソ・スミスクライン・コンシューマー・ヘルスケア・ジャパン（株）	鼻炎用内服薬	2
ザジテンAL鼻炎スプレーα	グラクソ・スミスクライン・コンシューマー・ヘルスケア・ジャパン（株）	鼻炎用点鼻薬	2
ザジテンAL鼻炎スプレーαクール	グラクソ・スミスクライン・コンシューマー・ヘルスケア・ジャパン（株）	鼻炎用点鼻薬	2
ジキナAL点眼薬	（株）富士薬品	アレルギー用点眼薬	2
レガーテ点鼻薬K	白金製薬（株）	鼻炎用点鼻薬	2

8 件中 1 ～ 8 を表示

図 14:KEGG Medicus 医薬品情報ページ

4 https://www.fda.gov/

一般用医薬品：コンタック600ファースト

List Top

医薬品情報

この情報は KEGG データベースにより提供されています。
日米の医薬品添付文書はこちらから検索することができます。

製品名	コンタック600ファースト
製造販売元	佐藤薬品工業（株）
販売会社	グラクソ・スミスクライン・コンシューマー・ヘルスケア・ジャパン（株）
医薬品分類	一般用医薬品
小分類	鼻炎用内服薬　一般用医薬品分類
リスク区分	第2類医薬品　リスク区分
包装	10カプセル，20カプセル
KEGG DRUG	D01332
成分 (2カプセル中)	ケトチフェンフマル酸塩（D01332） 2.76mg ケトチフェン 2mg
添加物	D-マンニトール トウモロコシデンプン 無水ケイ酸 ステアリン酸マグネシウム ゼラチン ラウリル硫酸ナトリウム
色	白/白

図 15:KEGG Medicus 医薬品情報詳細ページ

以下は，剤型の対応付けに用いたものである．

表 5:剤型の対応付け

FDA	剤型
Liquid	液体
Capsule	カプセル
Granule	顆粒
Suppository	坐薬
Powder	パウダー
Tablet	錠剤
Syrup	シロップ
Plaster	貼付剤
Drop	点眼薬
Troche	トローチ
Ointment	軟膏
Cream	クリーム
Spray	スプレー
Gel	ジェル
Lotion	ローション
Tape	テープ
Dressing	ガーゼ
Patch	パッチ
Swab	綿棒
Injection	浣腸薬
Kit	キット
Paste	ペースト
Shampoo	シャンプー
Rinse	リンス
Lozenge	錠剤
Stick	スティック
Tincture	チンキ
Soap	石鹸
Salve	軟膏
Poultice	湿布
Plaster	絆創膏
Pastilles	トローチ
Mouthwash	含嗽薬
Liniment	塗布薬
Jelly	ゼリー
Enema	浣腸薬
Emulsion	乳液

この提案手法を通して得た出力の例を以下に示す.

表 6: 提案手法を通して得た出力の例

入力(海外の製品名)	剤型の候補	剤型の選択	検索結果(日本の製品名)	絞り込んだ候補
Ketotifen Fumarate	0 Solution/drops 1 Solution	0	コンタック600ファースト:Tablet ザジテンAL鼻炎カプセル:Capsule ザジテンAL点眼薬:Drop ジギナAL点眼薬:Drop	ザジテンAL点眼薬 ジギナ点眼薬
Pepcid AC	0 Tablet, chewable 1 Tablet, film coated 2 Tablet	2	ガスター10: Tablet ガスター10 S錠: Tablet ガスター10〈散〉: Powder ファモチジン錠M: Tablet ファモチジン錠「クニヒロ」:Tablet	ガスター10 ファモチジン錠「クニヒロ」 ファモチジン錠M ガスター10 S錠

第5章 評価

提案手法の有用性を示すために，提案手法を組み込んだ外国人のための医薬品検索システムを実装し，動作確認を行った．どの程度絞り込めたのかを検索結果数から評価し，またユーザの目的に近づいたのかを評価するために適合率，再現率を用いる．

5.1 評価

5.1.1 候補の絞り込みの適切さ

まずは，30 個のテストデータを用意し，どの程度絞り込めたのかを検索結果数から評価する．

絞り込み度 $= \frac{\text{絞り込んだ候補の数}}{\text{検索結果数}}$ として計算する．その結果を下の表 7 に示す．

今回の結果から，提案手法を用いた外国人のための医薬品検索システムは，平均67%の絞り込みを達成し，ユーザによる日本の製品の選択を容易にした．

5.1.2 適合率と再現率

次に，評価の指標として，絞り込んだ候補の精度である適合率と，再現率を用いる．ここでは，以下のように定義する．

$$\text{適合率} = \frac{\text{絞り込んだ候補のうち正しい製品の数}}{\text{絞り込んだ候補の数}}$$

$$\text{再現率} = \frac{\text{絞り込んだ候補のうち正しい製品の数}}{\text{検索結果のうち正しい製品の数}}$$

絞り込んだ候補と絞り込み結果以外が正しいかどうかを判定するために，海外の製品名，日本の製品名それぞれの用途，対象者などの，その製品にまつわる文献を参考にして評価を実施した．
その結果を表 7 に示す．

表 7:絞り込み度の適合率と再現率

海外の製品名	絞り込み度	適合率	再現率
Tylenol	7/9	7/7	7/7
Tylenol Advanced Headache	該当なし	該当なし	該当なし
Ketotifen Fumarate	2/8	2/2	2/2
Bayer	1/1	1/1	1/1
Tums	4/5	4/4	4/5
NyQuil Severe	該当なし	該当なし	該当なし
Claritin	2/2	2/2	2/2
Advil	9/14	9/9	9/14
Unisom	16/26	15/16	15/24
Vivarin	3/5	3/3	3/5
Pepcid AC	4/5	4/4	4/5
Cortizone 10	1/3	1/1	1/3
Festal Plus	1/1	1/1	1/1
Ibuprofen	9/14	9/9	9/14
Zyrtec	該当なし	該当なし	該当なし
Walgreens Clotrimazole	1/2	1/1	1/2
Clearasil	該当なし	該当なし	該当なし
Vaseline	該当なし	該当なし	該当なし
Dimenhydrinate	該当なし	該当なし	該当なし
Festal Plus	1/1	1/1	1/1
Neosporin	該当なし	該当なし	該当なし
Childrens Loratadine	2/2	0/2	0
Zylast Antiseptic	1/2	1/1	1/2
Vida Mia Povidone Iodine	22/40	16/22	16/17
Zaditor	2/8	2/2	2/2
Appedrine	2/6	0/2	0
Benadryl	該当なし	該当なし	該当なし
Alophen	12/15	12/12	12/15
Allegra-D Allergy	22/24	21/22	21/23
Chewable Multiple Vitamins for Children	2/6	0/2	0/1

第6章 考察

6.1 要因と展望

今回の検証では，30 個の海外の製品名をテストデータとして用いた．適合率は 85%，再現率は 75%を得た．結果を確認する限り，適合率が高い数値を示している医薬品が多い．剤型による絞り込みによって，成功した例と上手くいかなかった例を挙げる．Ketotifen Fumarate は，アレルギーに対する目薬である．そのため，drop(目薬)で絞り込むことで，検索結果に含まれる目以外の用途の医薬品を除外することに成功した．しかし，睡眠鎮静剤である Unisom を tablet(錠剤)で絞り込みを行うと，一つだけ湿疹，かぶれのための医薬品を含んでいた．ここから，剤型による絞り込みのみでは，うまくいかなかった例もあることが分かる．

また，結果より，該当なしや，絞り込んだ候補の中に目的とする医薬品が存在しておらず適合率が低い医薬品も存在する．該当なしという結果を生じた原因は，2 つのパターンに分かれる．第一に，絞り込み以前に検索結果が存在しない原因である．これは，リンクを辿っていく過程で，DBpedia が存在しない，DBpedia のクエリ結果に KEGG ID がない，及び KEGG ID が存在していても，一般用医薬品がないといった複数の理由から原因が生じていると考えられる．これらが意味するのは，海外で一般用医薬品として販売されているとしても，日本では処方箋を必要とする医療用医薬品として販売されているケースも存在するということである．第二に，複数の DrugBank ID から成り立っている医薬品であるという原因である．その場合，第 3 章でも述べたように，それぞれの一般名に対する日本の製品名の検索結果を出力し，そして，共通する日本の製品名のみを抽出し，出力する．しかし，共通する日本の製品名が存在しない場合は，該当なしといった結果になる．絞り込んだ候補の中に目的とする医薬品が存在しておらず適合率が低くなる原因としては，同じ有効成分を使っていてもユーザの入力とは異なる用途の日本の製品名は出力されるということにある．例えば，表 7 にある Chewable Multiple Vitamins for Children は，8 才以下の小児を対象とした医薬品であるが，検索結果として挙げられた医薬品は，小児のみならず大人用の医薬品を含んだ．そのため，適合率は低い結果になった．これは，対象年齢の情報を用いず，有効成分と剤型という指標のみを使用していることが原因である．また，適合率が 73%の Vida Mia Povidone Iodine は，本来うがい薬であるが，検索結果として挙げられたのは殺菌消毒薬の医薬品や，のどの炎症を抑える口腔咽喉薬であった．有効成分は同じでも，用途が少し違うものが存在すると

いう知見も得られた.

以上をまとめると，3 つの LODD を辿ることのメリットとして，複数のデータとデータが繋がることから，繋がる数は多岐にわたる．その一方でデメリットは，複数繋ぐ際に，どれか 1 つのデータの一部分でも欠けてしまうと，上手くリンクが繋がらないことにある．これも LODD を使用する際の一つの課題点だと考える．

6.2 提案

本研究では，DrugBank,DBpedia,KEGG の 3 つの LODD を連携し，日本の医薬品の製品名を出力した．しかしながら，前節で述べたように３つの LODD のみでは，目的とする日本の製品名を上手く取得できない場合もある．その問題を解決するために，LODD の連携に関して，各データを共通の API でサービス化することを提案する．具体的には，製品名から一般名の ID を取得するサービス，一般名 ID をターゲットの LOD の一般名に変換するサービス，及び一般名 ID から製品名を取得するサービスである．これらのサービスの 3 つの API を標準化できると，他のデータに対しても，このようなフローで検索が可能になる．

第7章 おわりに

本研究では，Linked Open Drug Data を用いた外国人のための医薬品検索システムを提案した.30 個のテストデータを用い,動作確認および検証を実施した.

本研究の貢献は以下の通りである.

LODD の連携

XML，RDF(SPARQL で問い合わせるデータ)，HTML という異なるフォーマットで，提供されるデータを，一般名を表す ID を介して連携させた．さらに，DBpedia を中間に配置することで，データ間で異なる ID の対応付けを可能とした．これにより，380,504 件の海外の製品名と 10,802 件の日本の製品名の紐づけを実現している.

候補の絞り込み

日本の製品名の検索結果候補から，ユーザが自国で使用している医薬製品の剤型を基準に日本の製品名を選択し絞り込みを行う手法を考案した．この手法により，平均 67%の絞り込みを達成し，ユーザによる日本の製品の選択を容易にした．さらに，この絞り込みにより適合率は 85%,再現率は 75%という結果を得た.

総合的に見て，適合率 8 割，再現率 7 割と半数以上の結果を達成したが，時には目的と異なる医薬品が検索結果に存在した．本研究で発見できた問題やそれに対する課題点に対しても,第 6 章で提案した標準 API を用いた方法で,他の LODD データを試すことや，絞り込みの際の指標の選択肢を増やすことでアプローチ出来るのではないかと考えた.

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指導教官の村上陽平准教授に深謝申し上げます．また普段からお世話になっている社会知能研究室の皆様にも感謝の意を表します．

参考文献

[1] Matthias Samwald, Anja Jentzsch, Christopher Bouton, Claus Stie Kallesøe, Egon Willighagen, Janos Hajagos, M Scott Marshall, Eric Prud'hommeaux, Oktie Hassanzadeh, Elgar Pichler and Susie Stephens: Linked open drug data for pharmaceutical research and development, Samwald et al. Journal of Cheminformatics (2011).

[2] Guma Lakshen, Valentina Janev, Sanja Vranes : Linked Open Drug Data:Lessons Learned, IFIP International Conference on Computer Information Systems and Industrial Management, pp 164-175 (2019).

[3] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2017 Nov 8. doi: 10.1093/nar/gkx1037.

[4] 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, 60 巻, vol. 60, no. 5, p. 307-315(2017)

[5] Minoru Kanehisa1, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi1 and Mao Tanabe : Data, information, knowledge and principle: back to metabolism in KEGG, Nucleic Acids Research, 2014, Vol. 42, Database issue, D199–D205(2013)