

IT・AI を活用した食品安全情報収集システムに関する実証  
事業

B)システム開発

(ア)アジャイル開発で生じるドキュメント

令和7年3月31日

みずほリサーチ&テクノロジーズ株式会社

情報通信研究部

# 目次

1. アジャイル開発で生じるドキュメント .....	1
1.1. クローリング・スクレイピングによる情報収集 .....	1
1.1.1. クローリング条件設定機能 .....	1
1.1.2. クローリング・スクレイピング実行機能 .....	16
1.1.3. クローリング・スクレイピング結果確認機能 .....	34
1.2. AI 等による翻訳・要約 .....	38
1.2.1. クローリング結果文章選択機能 .....	38
1.2.2. 要約機能 .....	39
1.2.3. 翻訳機能 .....	42
1.2.4. 専門用語・定訳辞書機能 .....	49
1.2.5. 記事編集機能 .....	51
1.3. 情報の保存と活用 .....	54
1.3.1. クローリング・スクレイピング結果 DB .....	54
1.3.2. 専門用語・定型辞書 DB .....	55
1.3.3. 翻訳・要約結果 DB .....	55
1.3.4. 記事 DB .....	55
1.3.5. ラベル DB .....	55
1.3.6. ラベル付与機能 .....	55
1.3.7. 記事情報参照機能 .....	68
1.3.8. クロス表表示機能 .....	71
1.3.9. 印刷機能 .....	72

## 1. アジャイル開発で生じるドキュメント

### 1.1. クローリング・スクレイピングによる情報収集

#### 1.1.1. クローリング条件設定機能

クローリング・スクレイピングを行うには、対象とする web サイトがロボット等のプログラムからのアクセスを許可している必要がある。また web サイトによっては distributed denial of service (DDos) 攻撃対策としてアクセス制限を行っていることがある。クローリング・スクレイピングを行う際には、web サイトの規約を調べ、またどのようなアクセス制限があるのかを確かめる必要がある。

##### 1.1.1.1. プログラムからの記事収集可能性調査 実施内容

食安委様より提示された 236 の記事収集サイトのリストについて、実際にプログラムからアクセスし、アクセス可能かどうかを確認した。

##### 結果

アクセス結果を「1\_プログラムからの EPA と EFSA サイトアクセス報告.xlsx」にまとめた。237 URL 中リンク切れが 20 サイト、プログラムからはアクセス出来ないサイトが 23 サイト、アクセス数制限に引っかかったサイトが 32 サイトとなった。

##### 振り返り

セキュリティ上の理由や、今後の自然言語処理技術の発展に伴い、プログラムからのアクセスを禁止するサイトは増えていくと考えられる。対象記事が多いサイトについては、特別な許可をもらうか、API の提供など、情報共有の仕組みを検討することが望ましいと思われる。

##### 1.1.1.2. プログラムからの EPA と EFSA サイトアクセス調査 実施内容

食安委様より提示された記事収集サイトから、EPA 及び EFSA について、プログラムからの取得が可能かどうかの調査を行った。

##### 結果

##### EPA

アクセス状況調査結果を表 1-1 にまとめた。

- ・ EPA の訪問 URL (7URL)に対して、アクセス間隔 1 秒で、深度 1 のクローリングのテストを 12/12~12/13 に実施した。
- ・ アクセス制限による記事取得失敗はなかった。
- ・ 2 件のエラーのうち、1 件はリンク切れ、1 件は記事サイズが AWS で使用している DB のサイズ制限(400KB)を越えてしまうため、登録ができなかったことによる。
  - (リンク切れ URL) <https://www.epa.gov/node/224289>
  - (記事サイズ制限抵触 URL)<https://iris.epa.gov/AdvancedSearch/>

表 1-1 EPA の記事取得対象 URL とプログラムによる取得状況

	リクエスト間隔 1 秒
--	-------------

URL	取得成功件数	取得失敗件数	取得失敗件数 status_code=429
<a href="https://www.epa.gov/biosolids">https://www.epa.gov/biosolids</a>	21	0	0
<a href="https://www.epa.gov/ground-water-and-drinking-water">https://www.epa.gov/ground-water-and-drinking-water</a>	44	0	0
<a href="https://www.epa.gov/iris">https://www.epa.gov/iris</a>	27	1	0
<a href="https://www.epa.gov/newsreleases/search">https://www.epa.gov/newsreleases/search</a>	154	0	0
<a href="https://www.epa.gov/pesticides/pesticide-news-stories">https://www.epa.gov/pesticides/pesticide-news-stories</a>	471	1	0
<a href="https://www.epa.gov/pfas/key-epa-actions-address-pfas">https://www.epa.gov/pfas/key-epa-actions-address-pfas</a>	43	0	0
<a href="https://www.epa.gov/pfas/press-releases-related-pfas">https://www.epa.gov/pfas/press-releases-related-pfas</a>	104	0	0
<b>合計</b>	<b>864</b>	<b>2</b>	<b>0</b>

## EFSA

アクセス状況調査結果を表 1-2 にまとめた。

- EFSA の訪問 URL (21URL) に対して、アクセス間隔 1 秒、3 秒、5 秒、10 秒での、深度 1 のクローリングのテストを 12/12~12/13 に実施した。
- <https://www.efsa.europa.eu/en/topics/topic/contaminants-feed> と <https://www.efsa.europa.eu/en/topics/topic/chemical-contaminants> は、実際にアクセスされる URL である <https://www.efsa.europa.eu/en/topics/topic/chemical-contaminants-food-feed> に結果をまとめている。
- 親 URL がアクセスエラーである場合、その配下の記事を取得することができないため、アクセス間隔により成功と失敗の総数が異なる。また、それ以外に、サイトの更新や、アクセス間隔が長くなったことにより、実行時間制限(15 分)の影響を受けている可能性がある。
- アクセス間隔 5 秒、10 秒は、全ての親 URL へのアクセスに成功している。
- アクセス間隔 10 秒は、アクセス制限による記事取得失敗はなかった。
- その他の取得失敗については、別途リストアップ(1\_EFSA\_記事取得失敗 URL リスト.xlsx)し、調査対応する。

表 1-2 EFSA の記事取得対象 URL とプログラムによる取得状況。「(429)」とは HTTP のエラーコードであり、「Too Many Requests」である。

URL	リクエスト間隔 1 秒			リクエスト間隔 3 秒			リクエスト間隔 5 秒			リクエスト間隔 10 秒		
	取得成功件数	取得失敗件数	取得失敗件数 (429)	取得成功件数	取得失敗件数	取得失敗件数 (429)	取得成功件数	取得失敗件数	取得失敗件数 (429)	取得成功件数	取得失敗件数	取得失敗件数 (429)
<a href="https://www.efsa.europa.eu/en/topics/topic/biological-hazards">https://www.efsa.europa.eu/en/topics/topic/biological-hazards</a>	9	10	8	17	2	0	17	2	0	17	2	0
<a href="https://www.efsa.europa.eu/en/topics/topic/animal-health">https://www.efsa.europa.eu/en/topics/topic/animal-health</a>	0	0	0	51	13	0	51	13	0	51	13	0
<a href="https://www.efsa.europa.eu/en/topics/topic/animal-welfare">https://www.efsa.europa.eu/en/topics/topic/animal-welfare</a>	0	0	0	36	5	0	36	5	0	36	5	0
<a href="https://www.efsa.europa.eu/en/topics/efsa-and-covid-19">https://www.efsa.europa.eu/en/topics/efsa-and-covid-19</a>	11	2	0	11	2	0	11	2	0	11	2	0
<a href="https://www.efsa.europa.eu/en/publications">https://www.efsa.europa.eu/en/publications</a>	10	40	1	10	39	0	9	38	0	8	38	0
<a href="https://www.efsa.europa.eu/en/topics/topic/cross-cutting-science">https://www.efsa.europa.eu/en/topics/topic/cross-cutting-science</a>	6	8	6	12	2	0	11	2	0	11	2	0
<a href="https://www.efsa.europa.eu/en/topics/topic/data">https://www.efsa.europa.eu/en/topics/topic/data</a>	0	1	1	17	1	0	16	1	0	16	1	0
<a href="https://www.efsa.europa.eu/en/topics/topic/emerging-risks">https://www.efsa.europa.eu/en/topics/topic/emerging-risks</a>	33	8	7	40	1	0	38	3	2	40	1	0

<a href="https://www.efsa.europa.eu/en/topics/topic/feed-additives">https://www.efsa.europa.eu/en/topics/topic/feed-additives</a>	22	2	1	23	1	0	23	1	0	23	1	0
<a href="https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging">https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging</a>	11	3	2	7	4	3	10	1	0	10	1	0
<a href="https://www.efsa.europa.eu/en/topics/topic/gmo">https://www.efsa.europa.eu/en/topics/topic/gmo</a>	0	1	1	25	4	3	27	1	0	27	1	0
<a href="https://www.efsa.europa.eu/en/topics/topic/methodology">https://www.efsa.europa.eu/en/topics/topic/methodology</a>	0	1	1	5	1	0	5	1	0	5	1	0
<a href="https://www.efsa.europa.eu/en/topics/topic/nutrition">https://www.efsa.europa.eu/en/topics/topic/nutrition</a>	0	0	0	13	1	0	12	1	0	12	1	0
<a href="https://www.efsa.europa.eu/en/topics/topic/pesticides">https://www.efsa.europa.eu/en/topics/topic/pesticides</a>	0	0	0	0	0	0	28	4	0	28	4	0
<a href="https://www.efsa.europa.eu/en/news">https://www.efsa.europa.eu/en/news</a>	27	38	12	26	37	11	36	26	0	36	26	0
<a href="https://www.efsa.europa.eu/en/calls/consultations">https://www.efsa.europa.eu/en/calls/consultations</a>	9	6	6	11	4	4	15	0	0	15	0	0
<a href="https://www.efsa.europa.eu/en/calls/data">https://www.efsa.europa.eu/en/calls/data</a>	11	1	1	11	1	1	12	0	0	12	0	0
<a href="https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/ahaw">https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/ahaw</a>	10	0	0	25	0	0	2	2	0	21	0	0
<a href="https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/ahaw#">https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/ahaw#</a>	95	5	0	71	4	0	84	3	0	55	1	0
<a href="https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/comco">https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/comco</a>	6	3	2	8	1	0	8	1	0	8	1	0
<a href="https://www.efsa.europa.eu/en/topics/topic/chemical-contaminants-food-feed">https://www.efsa.europa.eu/en/topics/topic/chemical-contaminants-food-feed</a>	20	2	1	23	1	0	23	1	0	23	1	0
合計	280	131	50	442	124	22	474	108	2	465	102	0

## 振り返り

EPA についてはアクセス制限による問題は無かったが、一つの記事の大きさが大きく、データベースに元記事情報を格納する際に問題が生じた。本番システムにおいては元記事を別ファイルに保存し、このファイル保存場所をデータベースに保存するなどの対応が必要となるだろう。

EFSA はアクセス制限が厳しく、記事収集には運の要素が入ってしまうことが分かった。アクセス制限で収集できなかった記事については、一定時間後に何度か記事収集を行うシステムが必要となるだろう。

### 1.1.1.3. クローリングによる情報収集

#### 1.1.1.3.1. サンプル URL を対象としたクローリング調査

##### 実施内容

クローリングを実行する上で下記の方針を検討した。

- ・ クローリングプログラムのメンテナンス性を考慮して、個々のページに対して作りこむのではなく、共通したロジックで各ページを収集できないかを検討する。
- ・ 指定された URL に対して、URL 内リンク先を再帰的にクロールする深度を設定する。(仕様書要求)
- ・ クラウドのマネージドサービス (AWS Lambda) 上でクローリングプログラムを実施するため、一つの指定された URL に対するプログラムの実行時間は最大 15 分とする。

モデルページとして Food and Agriculture Organization of the United Nations の Publication ページ (<https://www.fao.org/food-safety/resources/publications/en/>) (図 1-1 参照) に対してクローリングプログラムを作成し課題の抽出を行った。また食安委様から提供された URL リストに対して作成したプログラムでクローリングを実施した。

The screenshot shows the FAO Publications page. At the top, there is a blue header with the FAO logo and the text "Food and Agriculture Organization of the United Nations". To the right, it says "ENHANCED BY Google" with a search icon. Below the header, there are language options: العربية, 中文, English, Français, Русский, Español. The main content area is titled "Food safety and quality" and has a navigation bar with links: Home, Background, Food control systems, Scientific advice, Emergencies, News, and Resources. On the left, there is a sidebar with "Publications", "Tools", and "Multimedia". The main content area is titled "Publications" and features a search box with the text "Search for a publication". Below the search box, there is a "Free Text" input field and a "Categories" dropdown menu. The categories listed are: JECFA guidelines on risk assessment, Microbiological Risk Assessment (MRA) series, Microbiological risk management, Microbiological issues, and Food safety emergency/INFOSAN. A "Search" button is located below the categories. Below the search interface, there is a list of publications. The first publication is titled "99th Joint FAO/WHO Expert Committee on Food Additives (JECFA) meeting - Safety evaluation of certain food additives. Summary and conclusions, 2024". The text below the title reads: "The Ninety-ninth meeting of the Joint Food and Agriculture Organization of the United Nations (FAO)/World Health Organization (WHO) Executive Committee on Food Additives (JECFA) was held in Geneva from 11 to 20 June 2024. The purpose of the meeting was to evaluate the safety of certain food additives. The present meeting was the Ninety-ninth in a series of similar meetings. The tasks before the Committee were to (a) further elaborate principles governing the evaluation of food additives and enzymes; (b) undertake safety evaluations of certain food additives and enzymes; (c) review and prepare specifications for certain food additives and enzymes; and (d) review specifications for certain flavouring agents. 2024". Below this, there is a partial view of another publication titled "Food safety in the context of limited food availability. Risk assessment of 3-MCPD and fatty acid esters in nutrient supplements and therapeutic food. Food Safety and Quality series 25".

図 1-1 Food and Agriculture Organization of the United Nations の Publication ページ

## 結果

図 1-1 のページを 0 階層としてリンクを探索すると 1 階層目で 70 以上のリンク先があり、階層を増やすと発散してクロールが出来なくなった。そこで下記の対応を行った(図 1-2 参照)。

- ・ SNS へのリンクを無視
- ・ Header のリンクを無視
- ・ Navigation タグのリンクを無視
- ・ Footer のリンクを無視

この結果 1 階層目のリンクは 25 URL となり、この内記事と思われるリンクは 18 URL であった。2 階層目までの探索では 210 URL あり、この内記事と思われるリンクは 10 URL であった。

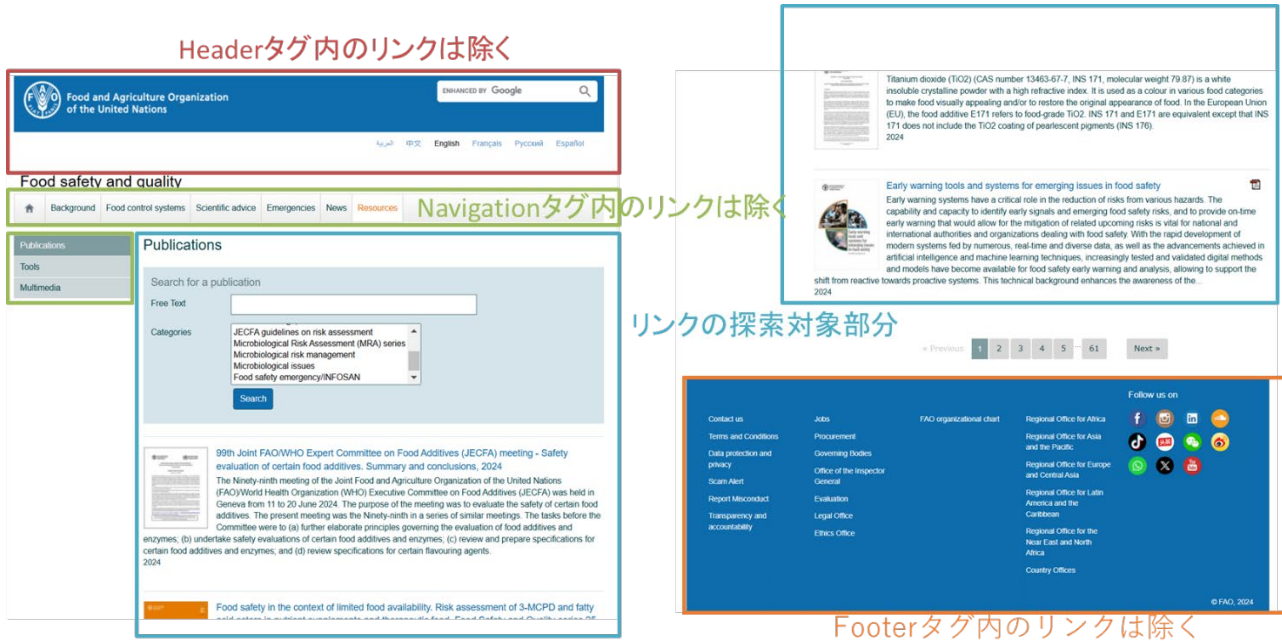


図 1-2 Food and Agriculture Organization of the United Nations の web ページの構造

### 振り返り

今回の調査で明らかになった事項を下記に示した。

- Header・footer・navigation 部分を探索してしまうとクローリングの探索が発散してしまう。
  - 現状はタグが完全一致しないと除いていない。
  - タグ名は様々なので上記の部分一致で除くように改善する必要がある。
- 階層が深くなると不要な URL が増えるため、可能であれば浅い探索が望ましい。
  - 元 URL を細かく設定するなどの対応策が考えられる。
- ページ割り(図 1-3 参照)による階層のずれは、日次でクローリングするなどして対応することを検討する。

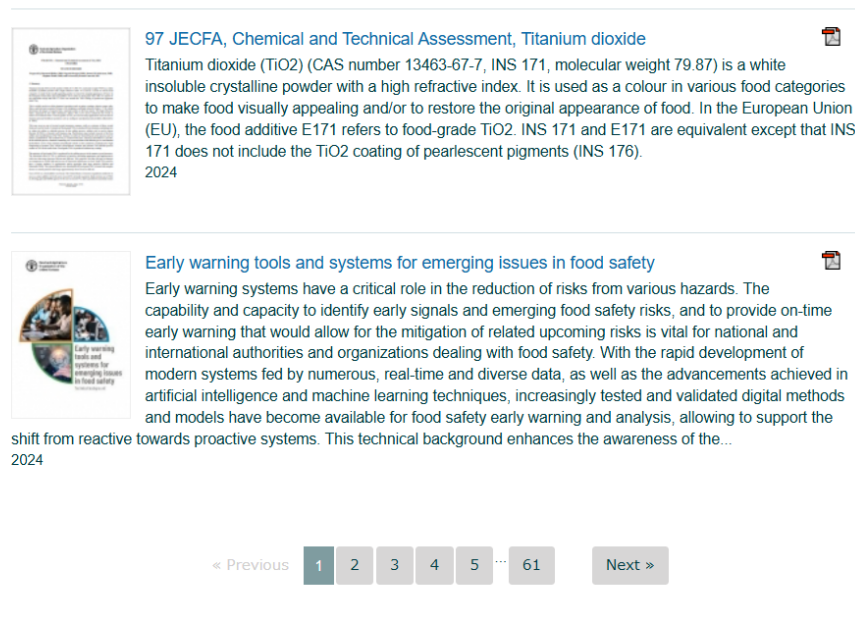


図 1-3 ページ割の例。記事一覧などの表示に使われることがある。

### 1.1.1.3.2. 記事対象 234 URL を対象としたクローリング調査

#### 実施内容

クローリングの階層が深くなると探索数が非現実的になる。そこで下記の制約を加えた。

- Header・footer・navigation 部分について、タグ名だけでなく、class 名での部分一致も含めて、探索対象外とすることで、単一ページ内のクローリング探索量を削減。
- 基本設定は 1 階層までのクローリングとして、必要に応じて 2 階層以降を設定する。
- 1 URL ごとに最大 1000 ページまでの探索とする。
- 1 URL ごとに最大 10 分間の探索とする。
- 記事一覧のページ割りについては日次で収集することでページ割りを無視する方針とする。
- URL の差を見て差分取得（収集済みのページは新たに収集しない）。
- URL が変化せず内容のみが修正されるケースについては、ハッシュ値を計算して同じならば消す。

クローリングの分割については、例えば IAEA のトップページに Latest news、Events、Publications などがあるが、実際に記事探索している「TOPICS> Water」、「TOPICS> Food and agriculture」を 1 階層までクローリングすることにより探索範囲を狭める(図 1-4)。

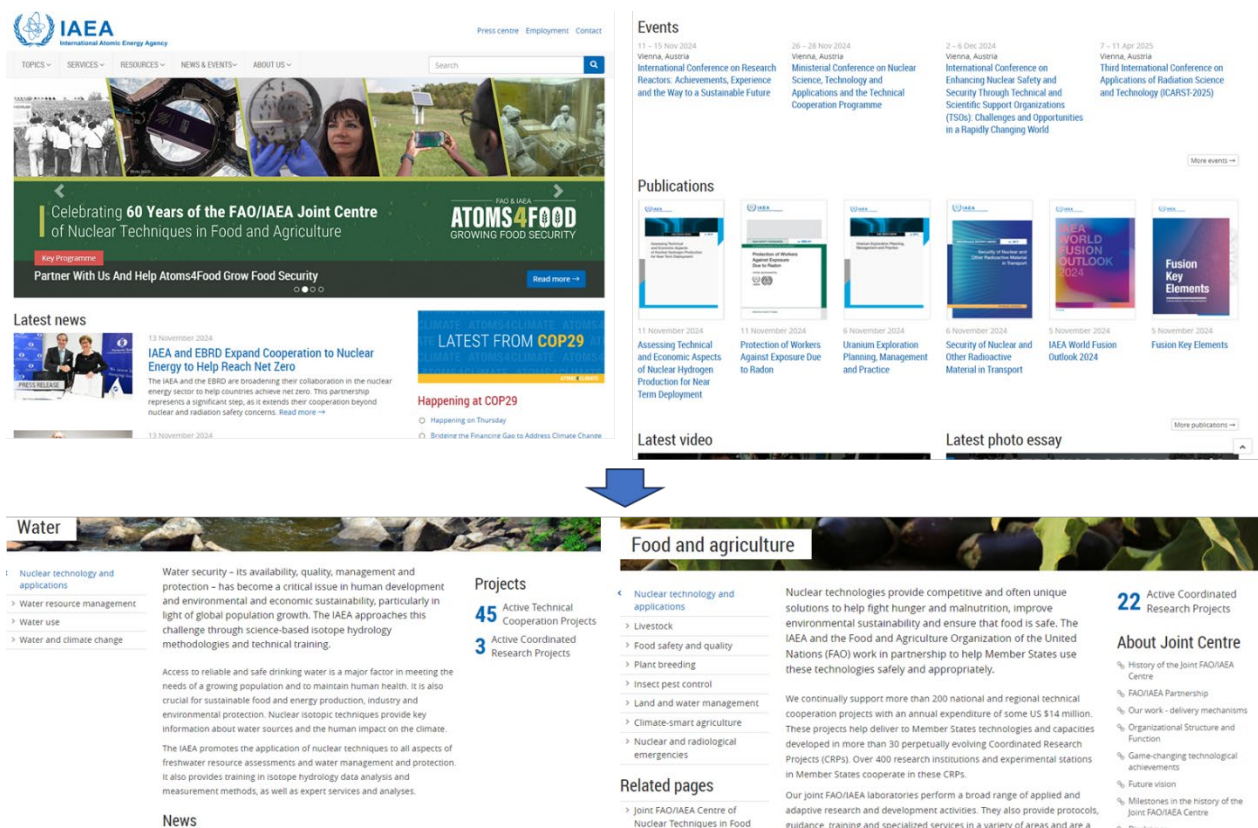


図 1-4 IAEA のトップページからクローリングを開始せず、TOPICS の下のページからクローリングを実行することにより探索 URL を減らす

#### 結果

記事対象 234 URL のクローリングにおいても実行時間が削減され、現実的な実行時間、収集ファイル数・容量となった。この結果を表 1-3 にまとめた。ただしこの内 100 URL 程度はプログラムからのアクセスが禁止されている可能性がある web サイトである。

表 1-3 食安委様が記事探索を行っている 234 URL のクローリング実行状況

URL リスト数	親ページ(0 階層目)の取得数	合計取得ページ数	実行時間(h)	合計取得容量(GB)
234	134	10,563	5.45	2.54 GB

### 振り返り

クローリングで web ページを取得することが出来ても css を含めてダウンロードできないと HTML としての表示がかなり異なる(図 1-5 参照)。

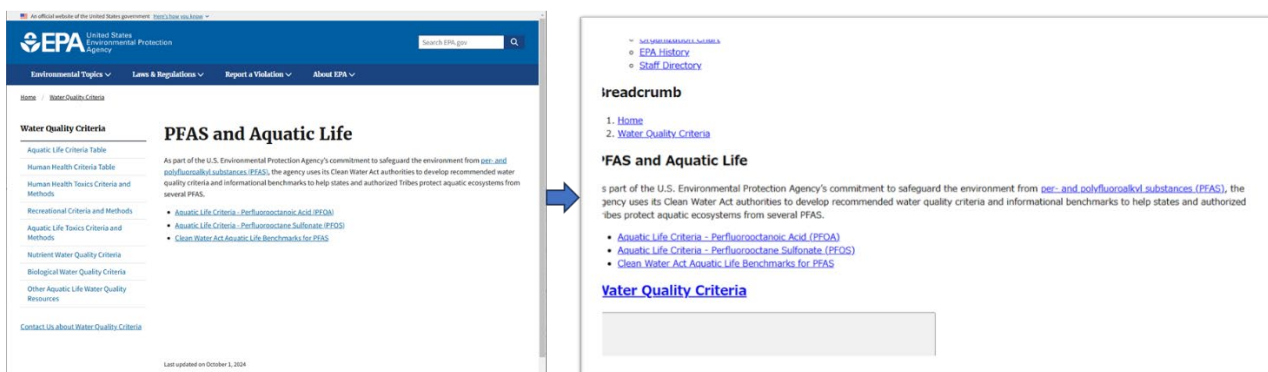


図 1-5 クローリングでダウンロードした web ページの css が無い場合の表示例

#### 1.1.1.3.3. 訪問頻度が毎日の 42 URL を対象としたクローリング調査

##### 実施内容

食安委様の方で毎日確認している 42 URL について探索深度を 1 としてクローリングを実行し、11 月 25 日～11 月 28 日の期間のクローリングの状況を確認した。

##### 結果

実際にクローリングを実行して得られた統計値を「クローリング結果の統計値.xlsx」にまとめた。結果の要約は下記である。

- ・ 2 日目～4 日目までの 3 日間の平均ページ取得数は 1 サイトあたり 55 ページ/日であった。
- ・ 2 日目～4 日目までの 3 日間の新規ページ取得件数の合計は 203 ページ、1 サイト当たり 5 ページであった。
- ・ 2 日目～4 日目までの 3 日間の更新ページ取得件数の合計は 462 ページ、1 サイト当たり 11 ページであった。
- ・ クローリングによるページのダウンロード容量は 1 サイト当たり 1 日平均 4.7MB であった。
  - 重複したページはダウンロード後に削除する。

### 振り返り

FAO のページ(<https://www.fao.org/food-safety/resources/publications/en/>)からは 11 月 26 日に更新ページを 1 件取得したが、このページからトップページ(<https://www.fao.org/home/en/>)にリンクがあり、このトップページを取得していた。トップページは毎日更新されており、このようなリンクへの対応が課題となった。

#### 1.1.1.3.4. スクレイピング対象確認

食品安全委員会事務局長からの指示により、クローリングからスクレイピングに移行することになった。本事業では EPA と EFSA の記事対象をスクレイピングすることになった。

##### 実施内容

スクレイピングで取得する記事の web ページ上の場所の確認を行い(図 1-6 参照)、スクレイピングプログラムの実装を開始した。

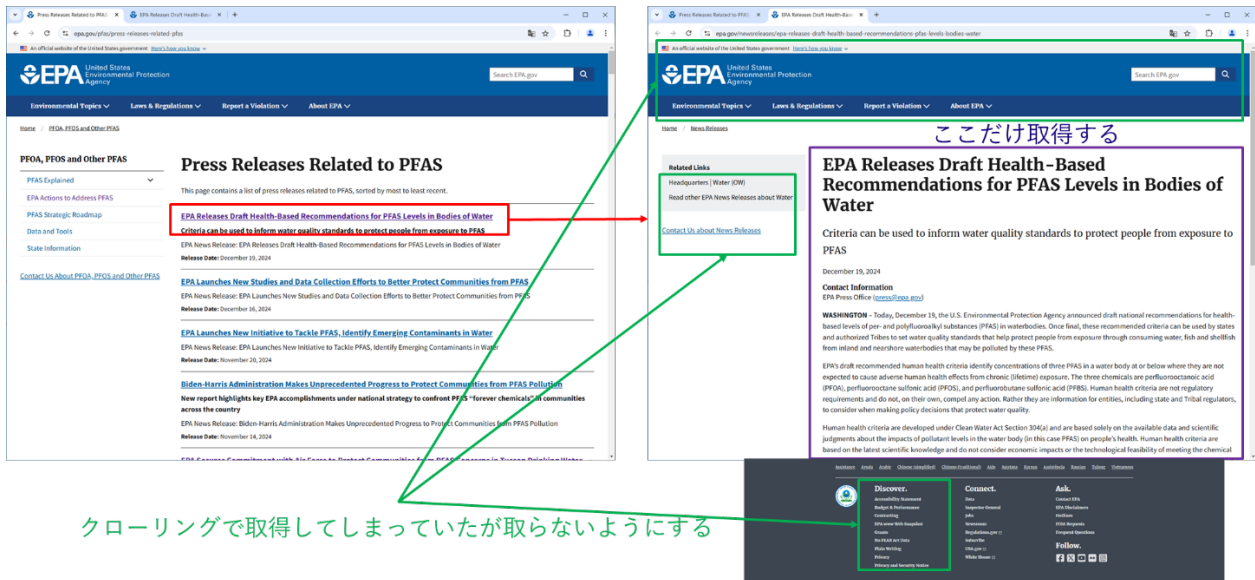


図 1-6 スクレイピング確認の例

##### 結果

スクレイピング対象記事の確認を行った資料を「1\_EPA\_EFSA スクレイピング対象確認.pptx」にまとめた。またスクレイピングプログラムの実装状況を「2\_EPA\_EFSA スクレイピング対象リスト.xlsx」にまとめた。

##### 振り返り

EFSA はアクセス制限のため、スクレイピングを行う際にアクセス間時間の調整が必要である。実際にスクレイピングプログラムを実行して、生じた課題に対応して行くことが必要となる。

#### 1.1.1.3.5. EFSA Journal アクセス確認

##### 実施内容

EFSA のスクレイピング対象から EFSA Journal の PDF ファイルを予めダウンロードして翻訳しておくことの検討を行った。

##### 結果

EFSA Journal は Wiley から提供されており、この利用規約(https://onlinelibrary.wiley.com/terms-and-conditions)にクローリングやスクレイピングが禁止であることが明記されていた(図 1-7 参照)。更に LLM による自動要約を行うことも規約に抵触すると思われる(図 1-8 参照)。ただし「学術研究目的」

であれば、別途 Wiley と交渉して API を利用したアクセスを利用して収集・要約等が可能なようである (<https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining>)。

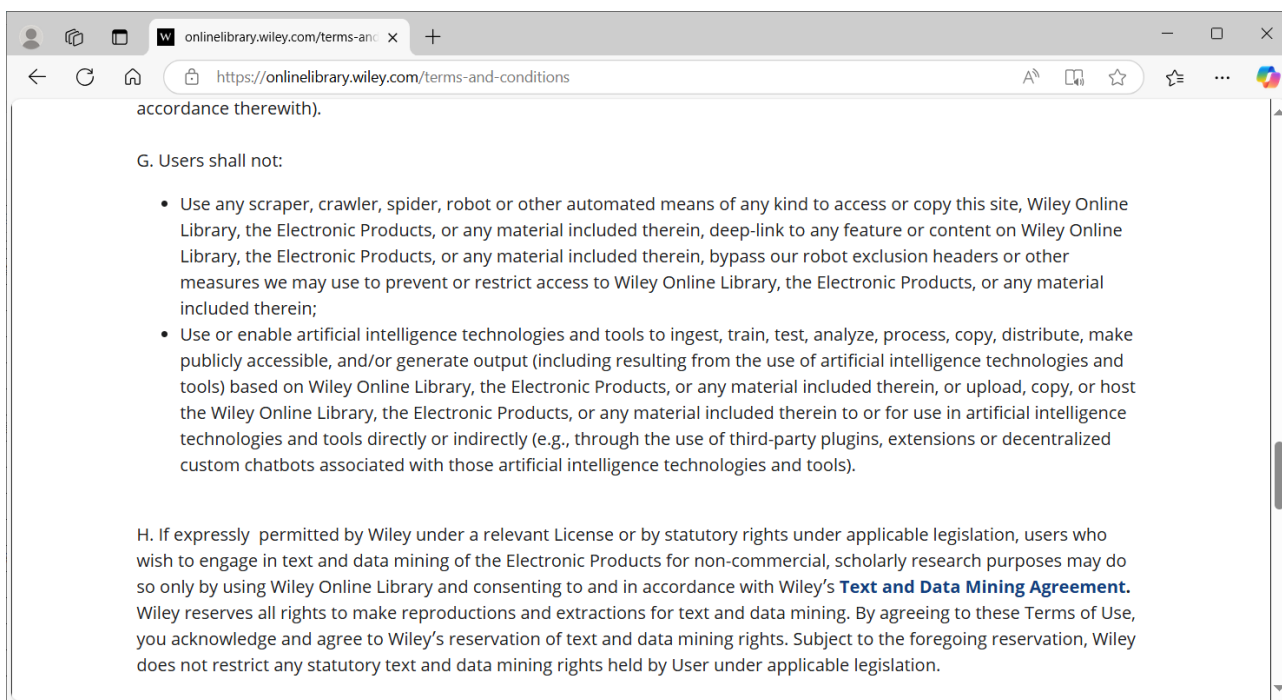


図 1-7 Wiley オンラインライブラリの利用規約の一部

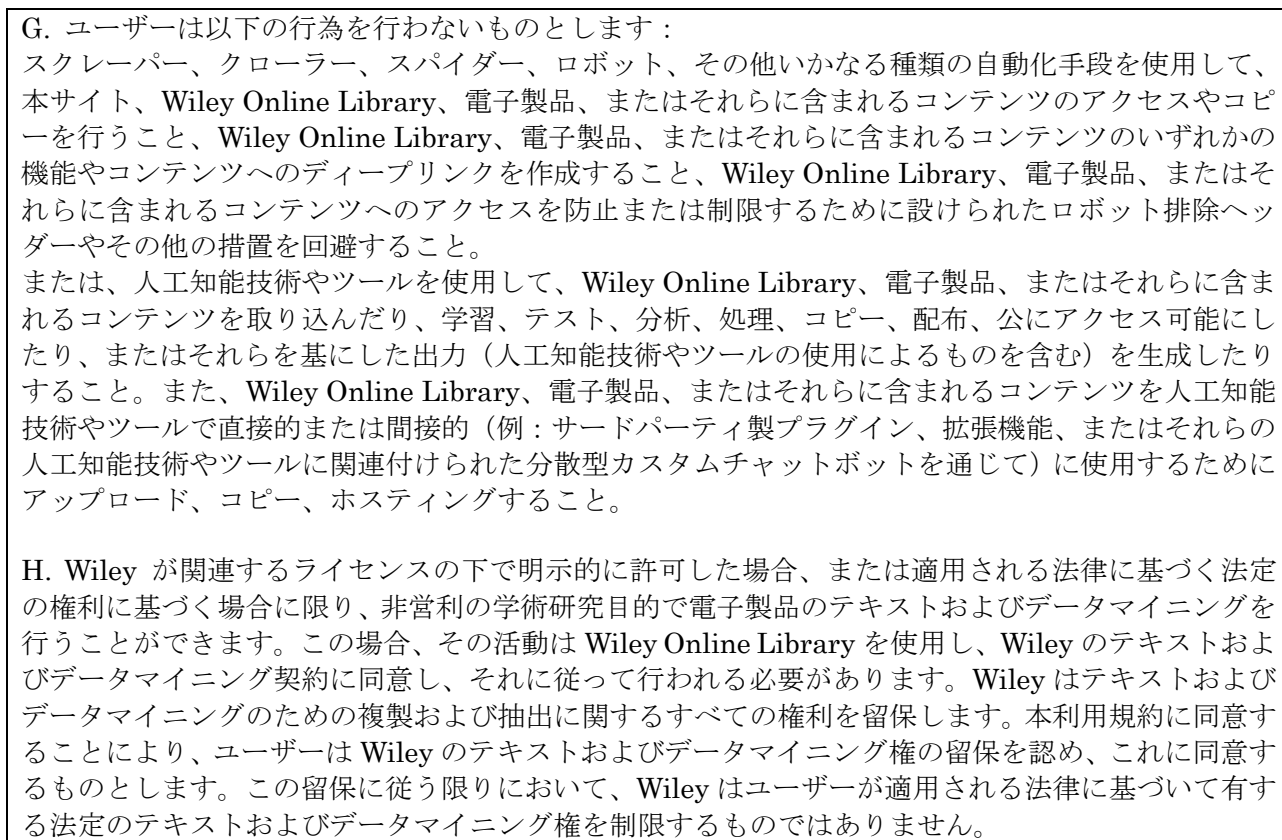


図 1-8 Wiley オンラインライブラリの利用規約の一部の日本語訳

### 振り返り

論文サイトを対象としたクローリングやスクレイピングは禁止されていくことが予想される。そうし

なければ論文要約サイトが立ち上がり、原著へのアクセスが著しく減少することが考えられる。

本システムでは EFSA の html 記事原稿に EFSA Journal の PDF ファイルのリンクを記載することにした。

#### 1.1.1.3.6. スクレイピング設定変更(その 1) 実施内容

EPA の「Press Releases Related to PFAS」(<https://www.epa.gov/pfas/press-releases-related-pfas>) は様々なタイプの記事が含まれており、長いドキュメントも多い。一方で「Release Date」情報が付与されているため(図 1-9 参照)、この情報を元に昨日と今日の記事のみにアクセスする様にスクレイピングの設定を変更した。

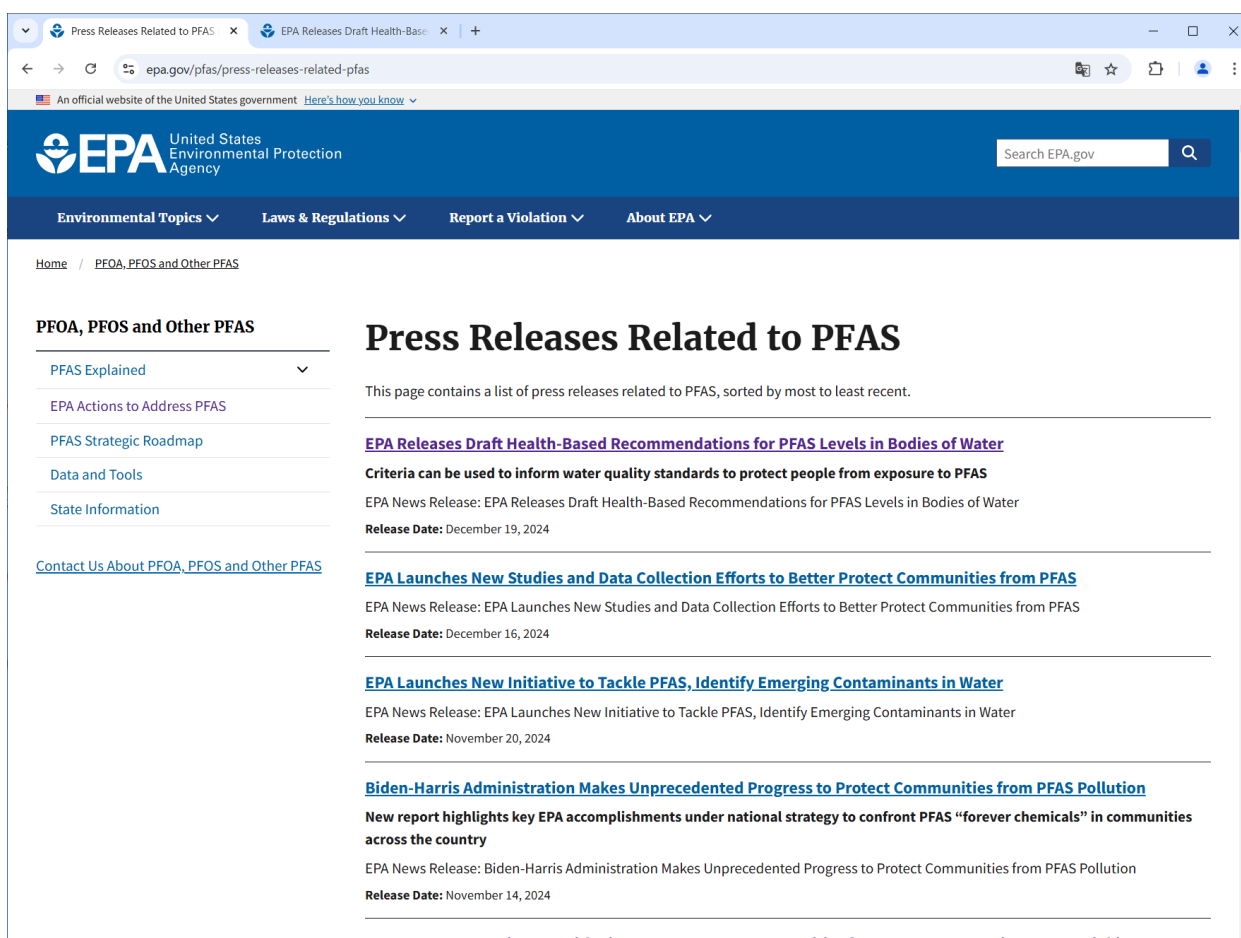


図 1-9 EPA の「Press Releases Related to PFAS」のページ

#### 結果

古い EPA の記事をスクレイピング対象から除外することが出来た。

#### 振り返り

スクレイピングを実行しつつ結果を監視して修正する処理は今後も続くと思われる。

### 1.1.1.3.7. スクレイピング設定変更(その2)

#### 課題

EPA 及び EFSA のスクレイピングにおいて、エラー表示を行い記事の取得状況をモニタリングした。この結果から

- ① EPA のサイトでは大きな PDF ファイルを一度の更新で多数公開することがあり、この場合翻訳に時間がかかり昨日と今日に公開された記事のみを取得する仕様では全ての記事を取得出来てないケースがあることがあった。
  - ② EFSA のサイトでは「429 Too Many Requests」エラーにより記事取得に失敗しているケースが多く、記事の探索が行われていない topics が発生していた。
- の課題が発生していた。

#### 実施内容

得られた課題に対応するため、記事の収集を最優先としてスクレイピングの設定を次の様に変更した。

- (ア) スクレイピング時の元記事の全文日本語翻訳処理をやめる
- (イ) PDF ファイルや html ファイルをダウンロードしてハッシュ値を比較する更新チェックをやめ、一度記事 DB に登録したターゲット URL からのダウンロードは行わない
- (ウ) 3 日前までの記事を取得対象とする
- (エ) EFSA の「Events」記事も日付探索対象(3 日前まで)とする

元記事の全文日本語翻訳は、記事作成画面から必要に応じて実行して頂くこととなる。本番システムでは夜中にデータベースを確認して全文日本語翻訳が無ければ追加する処理を加えることも検討対象になると思われるが、100 ページを超える PDF ファイルの全文翻訳が全て必要になるのかも確認が必要になると思われる。

これまでの挙動から EPA と EFSA については日付による記事の管理がきちんと行われていると判断し、ダウンロードした記事ファイルのハッシュ値の確認をやめることにした。これにより EFSA サイトへのアクセス量を減らすことが出来る。

記事公開のタイミングやアクセス多寡のエラーのため確認する日数を増やしたが、これはアクセス数を増やす変更のため、記事の全件取得についてはある程度の割切りが必要になると思われる。本番システムでは「失敗」と表示された記事を手動で追加するなどの機能が必要になると思われる。

#### 結果

下記の効果があった。

- i. 大きな PDF ファイルを同時に多数公開した場合への対応
- ii. EFSA のアクセス多寡による失敗数の軽減

#### 振り返り

今回の対策を行っても EFSA サイトは「429 Too Many Requests」エラーにより記事の確認・取得に失敗している場合が散見されている。EFSA のサイトについては今後もエラーログを確認しながら、記事取得の精度を向上する対策を検討して行きたい。

### 1.1.1.3.8. スクレイピング設定変更(その3)

#### 課題

指定された英語と日本語のペアの辞書データベースを用いた日本語翻訳を実行する。また翻訳前英語文書についても対応カ所が分かる様に表示する。

#### 実施内容

辞書データベースのペアに対して英訳及び日本語訳において対応する用語を鍵括弧「」で囲んだ。

#### 結果

スクレイピング記事において、英訳及び要約の日本語訳に専門用語辞書に登録されている文字列を鍵括弧で囲んで表示した。この例を図 1-10 に示した。「pesticide」や「risk assessment」が「農薬」や「リスク評価」に置き換わっている。

翻訳後の文章 (英語)	翻訳後の文章 (日本語)
Peer review of the 「pesticide」 「risk assessment」 of the active substance pyraclostrobin Published : 6 March 2025 Approved : 28 January 2025 Share: Share via LinkedIn Share via Bluesky Share via Facebook Wiley Online Library Full article : Read online at EFSA	翻訳後の文章がここに表示されません
要約 (英語)	要約 (日本語)
The European Food Safety Authority (EFSA) conducted a peer review of the pesticide risk assessment for pyraclostrobin, an active substance used as a fungicide and plant growth regulator. The review was based on initial assessments by Germany and Hungary, following EU regulations. It evaluated pyraclostrobin's use on cereals, maize, and potatoes, with additional maximum residue level (MRL) assessments for sweet corn and fish. The report presents reliable endpoints for regulatory risk assessment, proposed MRLs, and identifies missing information required by the regulatory framework. Concerns are noted where found. The peer review process adhered to Commission Implementing Regulation (EU) No 844/2012, as amended by Regulation (EU) No 2018/1659.	欧州食品安全機関 (EFSA) は、殺菌剤および植物成長調整剤として使用される有効成分ピラクロストロピンの「農薬」「リスク評価」のピアレビューを実施した。このレビューは、EUの規制に従い、ドイツとハンガリーによる初期評価に基づいて行われた。穀物、トウモロコシ、ジャガイモにおけるピラクロストロピンの使用を評価し、スイートコーンと魚類に対する追加の最大残留基準値 (MRL) 評価も行った。報告書は、規制上の「リスク評価」のための信頼できる「エンドポイント」、提案されたMRL、および規制の枠組みで要求される不足情報を提示している。問題点が見つかった場合はそれらが指摘されている。ピアレビューのプロセスは、規則 (EU) No 2018/1659によって改正された委員会実施規則 (EU) No 844/2012に準拠して行われた。

図 1-10 スクレイピング記事における辞書対応例。辞書に登録されている文字列は鍵括弧で囲まれる。

#### 振り返り

辞書に登録されている文字列と完全一致する場合に置換しており、文脈に依存した置換が必要な場合

には対応していない。文脈依存辞書機能のためには大規模言語モデル(LLM)の fine tuning を行うことが考えられるが、学習データが膨大に必要なこと、最近の LLM では fine tuning を行う機能が提供されていないことから現実的では無いと思われる。

### 1.1.1.4. クローリング条件 csv ファイル作成機能 実施内容

クローリング先を指定する場合、地域、国・地方、情報源などを指定し、この内部コードを付与する必要がある。これら選択して入力できるエクセルのマクロを作成し、クローリング条件 csv ファイルを出力する機能を作成した。

### 結果

作成したマクロの利用例を図 1-11 に示した。入力後 csv 出力ボタンを押すとアップロード用の csv ファイルが出力される(図 1-12)。

No	地域	AREACOD	国・地方	LOCALCC	情報源 (公的機関)	BULLETIN	深度	頻度	URL	キーワード
1	その他 (国際機関)	9	その他	250	国際連合食糧農業機関 (FAO)	295	2	日	http://www.fao.org/	
2	その他 (国際機関)	9	その他	250	国際連合食糧農業機関 (FAO)	295	2	日	http://www.fao.org/news/en/	
3	その他 (国際機関)	9	その他	250	国際連合食糧農業機関 (FAO)	295	2	日	http://www.fao.org/food-safety/resources/publications/en/	
4	欧州	5	オーストリア	123	国際原子力機関 (IAEA)	521	2	日	https://www.iaea.org/	
5	その他 (国際機関)	9	その他	250	国際がん研究機関 (IARC)	298	2	日	https://www.iarc.who.int/	
6	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	http://www.who.int/en/	
7	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/news-room/	
8	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/publications/i	
9	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/emergencies/disease-outbreak-news	
10	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/foodsafety/en/	
11	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/teams/global-influenza-programme/avian-influenza/monthly-risk-assessment-summary	
12	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/health-topics/chemical-safety#tab=tab_1	
13	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/news-room/fact-sheets/detail	
14	その他 (国際機関)	9	その他	250	国際獣疫事務局(WOAH)	297	2	日	https://www.woah.org/#/events?viewAll=true	
15	その他 (国際機関)	9	その他	250	国際獣疫事務局(WOAH)	297	2	日	https://www.woah.org/en/home/	
16	その他 (国際機関)	9	その他	250	Codex	296	2	日	http://www.fao.org/fao-who-codexalimentarius/news-and-events/en/	
17	その他 (国際機関)	9	その他	250	Codex	296	2	日	http://www.fao.org/fao-who-codexalimentarius/home/en/	
18	欧州	5	EU	111	その他	150	2	日	https://echa.europa.eu	
19	北米	3	米国	73	米国疾病管理予防センター (CDC)	104	2	日	https://www.cdc.gov/	
20	北米	3	米国	73	米国疾病管理予防センター (CDC)	104	2	日	https://www.cdc.gov/foodsafety/outbreaks/multistate-outbreaks/outbreaks-list.html	
21	北米	3	米国	73	米国疾病管理予防センター (CDC)	104	2	日	https://www.cdc.gov/prions/cwd/occurrence.html	
22	北米	3	米国	73	米国疾病管理予防センター (CDC)	104	2	日	https://www.cdc.gov/features/index.html	
23	北米	3	米国	73	米国疾病管理予防センター (CDC)	104	2	日	https://emergency.cdc.gov/han/index.asp	
24	北米	3	米国	73	米国疾病管理予防センター (CDC)	496	2	日	https://www.atsdr.cdc.gov/	
25	北米	3	米国	73	米国食品医薬品庁 (FDA)	104	2	日	https://www.cdc.gov/mmwr/index.html	
26	北米	3	米国	73	米国食品医薬品庁 (FDA)	104	2	日	https://www.fda.gov/mmdwr/index.html	
27	北米	3	米国	73	米国食品医薬品庁 (FDA)	104	2	日	https://www.fda.gov/oc/	
28	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	http://www.fda.gov/	
29	北米	3	米国	73	米国農務局(USDA)	105	2	日	https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/default.htm	
30	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/Food/NewsEvents/ConstituentUpdates/default.htm	
31	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.accessdata.fda.gov/scripts/fdccc/?set=Biocoon	
32	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/AnimalVeterinary/NewsEvents/default.htm	
33	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/oc/	
34	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/animal-veterinary/generally-recognized-safe-gras-notification-program/current-animal-food-gras-notices-inv	
35	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/food/science-research-food/food-safety-assessments	
36	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/food/chemicals-metals-pesticides-food	
37	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/food/ingredients-additives-gras-packaging-guidance-documents-regulatory-information/food-and-color-additi	
38	北米	3	米国	73	米国食品医薬品庁 (FDA)	105	2	日	https://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm272351.htm	
39	北米	3	米国	73	米国衛生研究所(NIH)	325	2	日	https://www.nih.gov/news-events	
40	北米	3	米国	73	その他	109	2	日	https://ntp.niehs.nih.gov/	
41	北米	3	米国	73	米国環境保護庁(EPA)	108	2	日	https://www.epa.gov/pesticides/pesticide-news-stories	

図 1-11 地域、国・地方、情報源をプルダウンから選択して入力を支援するクローリング条件 csv ファイル出力マクロ

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	No	地域	AREACOD	国・地方	LOCALCO	情報源 (公)	BULLETIN	深度	頻度	URL				
2	1	その他 (国)	9	その他	250	国際連合食	295	2	日	<a href="http://www.fao.org/">http://www.fao.org/</a>				
3	2	その他 (国)	9	その他	250	国際連合食	295	2	日	<a href="http://www.fao.org/news/en/">http://www.fao.org/news/en/</a>				
4	3	その他 (国)	9	その他	250	国際連合食	295	2	日	<a href="http://www.fao.org/food-safety/resources/publication">http://www.fao.org/food-safety/resources/publication</a>				
5	4	欧州	5	オーストリ	123	国際原子力	521	2	日	<a href="https://www.iaea.org/">https://www.iaea.org/</a>				
6	5	その他 (国)	9	その他	250	国際がん研	298	2	日	<a href="https://www.iarc.who.int/">https://www.iarc.who.int/</a>				
7	6	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="http://www.who.int/en/">http://www.who.int/en/</a>				
8	7	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="https://www.who.int/news-room/">https://www.who.int/news-room/</a>				
9	8	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="https://www.who.int/publications/i">https://www.who.int/publications/i</a>				
10	9	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="https://www.who.int/emergencies/disease-outbreak-r">https://www.who.int/emergencies/disease-outbreak-r</a>				
11	10	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="http://www.who.int/foodsafety/en/">http://www.who.int/foodsafety/en/</a>				
12	11	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="https://www.who.int/teams/global-influenza-program">https://www.who.int/teams/global-influenza-program</a>				
13	12	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="https://www.who.int/health-topics/chemical-safety#t">https://www.who.int/health-topics/chemical-safety#t</a>				
14	13	その他 (国)	9	その他	250	世界保健機	294	2	日	<a href="https://www.who.int/news-room/fact-sheets/detail">https://www.who.int/news-room/fact-sheets/detail</a>				
15	14	その他 (国)	9	その他	250	国際獣疫尋	297	2	日	<a href="https://wahis.woah.org/#/events?viewAll=true">https://wahis.woah.org/#/events?viewAll=true</a>				
16	15	その他 (国)	9	その他	250	国際獣疫尋	297	2	日	<a href="https://www.woah.org/en/home/">https://www.woah.org/en/home/</a>				
17	16	その他 (国)	9	その他	250	Codex	296	2	日	<a href="http://www.fao.org/fao-who-codexalimentarius/news">http://www.fao.org/fao-who-codexalimentarius/news</a>				
18	17	その他 (国)	9	その他	250	Codex	296	2	日	<a href="http://www.fao.org/fao-who-codexalimentarius/home">http://www.fao.org/fao-who-codexalimentarius/home</a>				
19	18	欧州	5	EU	111	その他	150	2	日	<a href="https://echa.europa.eu">https://echa.europa.eu</a>				

図 1-12 アップロード用のクローリング条件 csv ファイルの例

### 振り返り

Web インターフェースからの入力画面を作成する方法も考えられるが、web 画面からプルダウン選択ではコピー&ペーストなどに難がある。そこで(恐らく)使い慣れているエクセルの操作性を考慮し、エクセルのマクロで実装した。

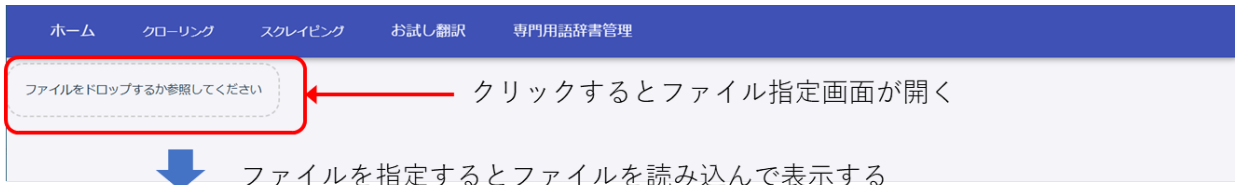
#### 1.1.1.5. アップロード画面作成

##### 実施内容

1.1.1.4 節で作成したクローリング条件 csv ファイルをアップロードする画面を作成した。

### 結果

アップロード画面とファイルを指定してアップロードした画面を図 1-13 に示した。



23 KB  
条件  
\_20241030\_151225.csv

クロール条件設定登録

No	地域	AREACODE (自動入力)	国・地方	LOCALCODE (自動入力)	情報源 (公的機関)	BULLETINCODE (自動入力)	深度	頻度	URL
1	その他 (国際機関)	9	その他	250	国際連合食糧農業機関 (FAO)	295	2	日	http://www.fao.org/
2	その他 (国際機関)	9	その他	250	国際連合食糧農業機関 (FAO)	295	2	日	http://www.fao.org/news/en/
3	その他 (国際機関)	9	その他	250	国際連合食糧農業機関 (FAO)	295	2	日	http://www.fao.org/food-safety/resources/publications/en/
4	欧州	5	オーストリア	123	国際原子力機関 (IAEA)	521	2	日	https://www.iaea.org/
5	その他 (国際機関)	9	その他	250	国際がん研究機関 (IARC)	298	2	日	https://www.iarc.who.int/
6	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	http://www.who.int/en/
7	その他 (国際機関)	9	その他	250	世界保健機関 (WHO)	294	2	日	https://www.who.int/news-room/

図 1-13 クロール条件 csv ファイルアップロード画面とファイルを指定してアップロードした画面

### 振り返り

本番システムでは記事探索対象サイトが増えた場合などで、各種コード番号の追加が発生することがあると考えられるが、この管理をどのようにするのかを検討しておく必要があるだろう。

#### 1.1.1.6. クロール条件表示

1.1.1.5 節のアップロード画面でファイルをアップロードするとクロール条件が表示される。

#### 1.1.1.7. クロール条件出力機能(csv ファイル)

1.1.1.4 節で示した通り、エクセルでクロール条件 csv ファイルを出力する仕様としたため、web インターフェースからのクロール条件出力機能は不要になった。

## 1.1.2. クローリング・スクレイピング実行機能

### 1.1.2.1. ファイル形式対応

#### 1.1.2.1.1. EPUB 形式対応

##### 実施内容

提供されたクローリング対象に EPUB 形式のデータを含むサイトがなく、フリーでダウンロード可能なサイトを検索し、「EPUB 3 Samples Project」(<https://idpf.github.io/epub3-samples/30/samples.html>)に対して検証を行うこととした。しかしながら、対象サイトの EPUB 形式へのダウンロードリンクは、リンク先の robots.txt にてロボットアクセスが禁止であったため、クローリングでデータを取得することはできなかった(他の適当なサイトを見つけることが出来なかった)。そのため EPUB 形式のデータ取得後の場合の処理を行う仕組みを構築した。EPUB 形式は電子書式等に利用されている形式であり、定められた仕様に従って記事データとして内容を抽出し、MD (markdown) 形式で登録した。参考抽出結果を「EPUB・Tableau 形式ファイルのクローリング結果.xlsx」の EPUB シートに記載した。

##### 使用ライブラリ

上記実施にあたり、表 1-4 記載の python ライブラリを使用した。

表 1-4 使用ライブラリ

ライブラリ名	利用機能
io	標準ライブラリ。バイナリデータ(PDF)の読込を行った。
PyMuPDF ( <a href="https://github.com/pymupdf/PyMuPDF">https://github.com/pymupdf/PyMuPDF</a> )	EPUB 形式から PDF 形式への変換、及び PDF 形式から MD 形式への変換を行った。

##### 結果

EPUB からテキスト部分を抽出することが可能となった。

##### 振り返り

電子書籍では写真などが含まれている場合もあり、これらの情報は含まれなくなる。電子書籍の情報を記事化するケースは多くないと思われる。

#### 1.1.2.1.2. Tableau 形式対応

##### 実施内容

提供されたクローリング対象に Tableau 形式のデータを含むサイトがなく、フリーでダウンロード可能なサイトを検索し、「The Tableau Resource」(<https://www.tableaubook.com/v9/workbooks.asp>)からダウンロードできる Tableau 形式ファイルについてクローリング方法を検討した。対象サイトのダウンロードリンクから取得できるファイルは拡張子が.twbx であり、これはワークブックファイル(.twb)と付属データの zip アーカイブである。ワークブックファイルは XML 形式で記載されており、zip アーカイブからワークブックファイルの内容を抽出し、タグを削除して記事データとして登録した。抽出結果は「EPUB・Tableau 形式ファイルのクローリング結果.xlsx」の Tableau シートに記載した。

## 使用ライブラリ

上記実施にあたり表 1-5 記載の python ライブラリを使用した。

表 1-5 使用ライブラリ

ライブラリ名	利用機能
zipfile	標準ライブラリ。バイナリデータ(zip)の読込を行った。
xml.etree.ElementTree	標準ライブラリ。XML 形式の読込とタグの削除を行った。

## 結果

Tableau 形式の内、twbx 形式の XML データに対応した。

## 振り返り

XML データのタグを除いてしまうと、何の情報か分からなくなるため XML 形式のまま記事データとして保存した方が良くかもしれない。

### 1.1.2.2. URL 及び深度指定実行機能

クロール条件 csv ファイル作成機能(1.1.1.4 節)で指定し、クロール条件 csv ファイルアップロード(1.1.1.5 節)で指定可能。

### 1.1.2.3. 不要ファイル対応

クロール実行確認表示画面(1.1.3.1 節参照)から、必要な記事を選択することで対応することになった。

### 1.1.2.4. API 利用実行機能(API 調査)

クロール先機関の Web サイトに対してクロールを行う API が存在するか調査を実施し、その結果について記載した。調査対象は、API を用いないクロールにおいて、何らかの理由によりアクセスが失敗する機関の中から、特に重要と思われる機関として選定した FDA、EFSA、USDA の 3 機関について実施した。

## 実施内容

調査対象とした FDA、EFSA、USDA の内、web スクレイピングを行う API が用意されているサイトはなかったが、FDA と USDA は特定の情報を取得する API が提供されていた(表 1-6 参照)。そこで FDA と USDA についてこれらの API の仕様を調査した。

表 1-6 FDA、EFSA、USDA の API 提供状況の概要

	FDA	EFSA	USDA
取得できる情報	<ul style="list-style-type: none"><li>● 有害事象</li><li>● 製品リコール</li></ul>	なし	製品リコール
データの更新頻度	<ul style="list-style-type: none"><li>● (有害)四半期ごと</li><li>● (リコ)毎週</li></ul>	-	逐次

API によるアクセス判定	成功	-	失敗
API を用いないクローリングによるアクセス判定	成功/失敗が半々	成功/失敗が半々	失敗
サイトポリシーによる自動クローリングの禁止の有無	無	無	無

## 結果

### FDA

FDA の提供するデータにアクセスするために openFDA API (<https://open.fda.gov/apis/>)が用意されている。openFDA は、医薬品、デバイス、食品などに関する FDA の公開データを提供する Elasticsearch ベースの API である。これらの API からリコール実施レポートや有害事象に関するデータなどを提供している。API へのすべてのクエリは、1 種類のデータに対して 1 つのエンドポイントを経由する必要がある。

表 1-7 openFDA API の基本情報

データ利用規約	<ul style="list-style-type: none"> <li>● サイトの利用規約(<a href="https://open.fda.gov/terms/">https://open.fda.gov/terms/</a>)に従う。 <ul style="list-style-type: none"> <li>➢ Creative Commons CC0 1.0 Universal</li> <li>➢ 一部のデータは第三者の著作権が適用される(明確に記載あり)</li> </ul> </li> </ul>
API 利用の制限	<ul style="list-style-type: none"> <li>● API キーを利用する。API キーは無料で取得できる。</li> <li>● API キーがない場合: IP アドレスごとに 1 分あたり 240 件のリクエスト。IP アドレスごとに 1 日あたり 1,000 件のリクエスト。</li> <li>● API キーを使用する場合: キーごとに 1 分あたり 240 件のリクエスト。キーごとに 1 日あたり 120,000 件のリクエスト。</li> </ul>

表 1-8 に openFDA API でアクセス可能なエンドポイントと取得できるデータカテゴリを記載した。本調査では、食品安全委員会事務局で巡回しているサイトの情報取得に関する Animal & Veterinary 及び Food について、カテゴリごとに具体的に取得できる情報をまとめた。

表 1-8 openFDA API のエンドポイントとカテゴリ

エンドポイント	カテゴリ
Animal & Veterinary	<ul style="list-style-type: none"> <li>● Adverse Events</li> <li>● Adverse Events</li> <li>● Product Labeling</li> <li>● NDC Directory</li> <li>● Recall Enforcement Reports</li> </ul>
Drug	<ul style="list-style-type: none"> <li>● Drugs@FDA</li> </ul>
Device	<ul style="list-style-type: none"> <li>● 510(k)</li> <li>● Classification</li> <li>● Recall Enforcement Reports</li> <li>● Adverse Events</li> <li>● Pre-market Approval</li> <li>● Recalls</li> <li>● Registrations and Listings</li> <li>● COVID-19 Serological Testing</li> <li>● Evaluations</li> </ul>
Food	<ul style="list-style-type: none"> <li>● Unique Device Identifier</li> <li>● Recall Enforcement Reports</li> </ul>

Other	<ul style="list-style-type: none"> <li>● Adverse Events</li> <li>● Historical Documents</li> <li>● NSDE</li> <li>● Substance Data Reports</li> <li>● UNII</li> </ul>
Tobacco	<ul style="list-style-type: none"> <li>● Problem Reports</li> </ul>

Animal & Veterinary API について、Adverse Events の基本情報を表 1-9 に示した。この API から取得できる情報のフィールドの説明を表 1-10 に示した。データ取得の例として、

[https://api.fda.gov/animalandveterinary/event.json?search=original\\_receive\\_date:\[20040101+TO+20161107\]&limit=1](https://api.fda.gov/animalandveterinary/event.json?search=original_receive_date:[20040101+TO+20161107]&limit=1)

にアクセスした場合の取得結果を図 1-14 に示した。

表 1-9 openFDA API Adverse Events の基本情報

概要	<ul style="list-style-type: none"> <li>● この API は、CVM が受け取った動物用医薬品や動物に使用される機器に関連する有害事象報告（AER）を、一般の人々が電子的にアクセスできるプラットフォームである openFDA.gov で公開しているもの。API に詳しい研究者や統計学者、その他の学者が、FDA が収集した大規模で貴重な公衆衛生データセットにアクセスするために利用される。</li> <li>● CVM の AER データには、動物用医薬品や動物に使用される機器に関連する有害事象や製品の欠陥についての報告が含まれている。これらには、軽度または重度の健康事象だけでなく、製品の品質問題、効果の不足、パッケージの欠陥、その他健康に関連しない問題についての苦情も含まれる。</li> <li>● 1987 年から現在までのデータにアクセス可能。四半期ごとの全データ（JSON ファイル）を一括でダウンロードすることも可能である。</li> </ul>
取得データ形式	JSON
データの更新頻度	四半期ごと
関連リンク	<ul style="list-style-type: none"> <li>● <a href="https://open.fda.gov/apis/animalandveterinary/event/">https://open.fda.gov/apis/animalandveterinary/event/</a> <ul style="list-style-type: none"> <li>➢ API による Adverse Events 取得の概要を記載</li> </ul> </li> <li>● <a href="https://www.fda.gov/animal-veterinary/product-safety-information/adverse-event-reports-animal-drugs-and-devices">https://www.fda.gov/animal-veterinary/product-safety-information/adverse-event-reports-animal-drugs-and-devices</a> <ul style="list-style-type: none"> <li>➢ 動物用医薬品および医療機器の有害事象報告について記載</li> </ul> </li> <li>● <a href="https://www.fda.gov/animal-veterinary/report-problem/how-report-animal-drug-and-device-side-effects-and-product-problems">https://www.fda.gov/animal-veterinary/report-problem/how-report-animal-drug-and-device-side-effects-and-product-problems</a> <ul style="list-style-type: none"> <li>➢ 動物用医薬品および医療機器の副作用および商品の問題を報告する方法について記載</li> </ul> </li> </ul>

表 1-10 openFDA API のフィールドの説明

フィールド名	説明
<b>animal</b>	副作用事象に関与した動物。
└── <b>age</b>	影響を受けた動物の年齢を示します。
└── └── <b>max</b>	影響を受けた動物のうち最も年長の年齢。
└── └── <b>min</b>	影響を受けた動物のうち最も若い年齢、または一匹の動物の年齢。
└── └── <b>qualifier</b>	年齢を特定する方法を示します。
└── └── <b>unit</b>	年齢の測定単位。
└── <b>breed</b>	影響を受けた動物の品種を説明します。

	—— is_crossbred	複数の品種が混ざっているか。
	—— breed_component	交雑に関与する異なる品種を説明。
	—— female_animal_physiological_status	妊娠および授乳状態。
	—— gender	動物の性別。
	—— reproductive_status	動物が去勢されているか。
	—— species	影響を受けた動物の種に関する情報。
	—— weight	影響を受けた動物の体重を示します。
	—— max	影響を受けた動物の最大体重。
	—— min	影響を受けた動物の最小体重。
	—— qualifier	体重を特定する方法を示します。
	—— unit	体重の測定単位。
	drug	副作用事象中に服用された薬。
	—— active_ingredients	獣医療用医薬品の活性成分。
	—— name	活性成分の名前。
	—— administered_by	薬を投与した個人。
	—— ae_abated_after_stopping_drug	薬の中止後に副作用が改善したか。
	—— ae_reappeared_after_resuming_drug	薬の再開後に副作用が再発したか。
	—— atc_vet_code	獣医療用医薬品の分類コード。
	—— brand_name	マーケティング承認保有者によって提示される名前
	—— dosage_form	薬の標示された用量形態。
	—— first_exposure_date	動物が薬を最初に投与された日付。
	—— frequency_of_administration	投与の頻度。
	—— unit	投与頻度の単位。
	—— value	投与頻度を示す数値。
	—— last_exposure_date	動物が最後に投与された日付。
	—— lot_expiration	薬の有効期限。
	—— lot_number	薬のロット番号。
	—— manufacturer	製造者情報。
	—— name	製造者の名前。
	—— registration_number	登録番号。
	—— manufacturing_date	薬の製造日。
	—— number_of_defective_items	不良品の数。
	—— number_of_items_returned	返品された製品の数。
	—— off_label_use	ラベル外での使用方法。
	—— previous_ae_to_drug	過去に副作用があったか。
	—— previous_exposure_to_drug	過去に曝露があったか。
	—— product_ndc	医薬品の国際製品コード。
	—— route	薬が投与された経路。
	—— used_according_to_label	ラベルに従った使用。
	duration	副作用が続いた時間。
	—— unit	持続時間の単位。
	—— value	持続時間の数値。
	health_assessment_prior_to_exposure	曝露前の健康状態。
	—— assessed_by	評価を行った者。
	—— condition	健康状態。
	number_of_animals_affected	影響を受けた動物の総数。
	number_of_animals_treated	治療された動物の数。
	onset_date	副作用の発症日。
	original_receive_date	初回情報受領日。
	outcome	副作用報告時の医療状態。
	—— medical_status	医療状態の詳細。

└── number_of_animals_affected	影響を受けた動物の総数。
primary_reporter	最も関連情報を持つ人または組織。
reaction	副作用に関連する反応。
└── accuracy	統計データの正確性。
└── number_of_animals_affected	影響を受けた動物の総数。
└── veddra_term_code	VEDDRA で使用される用語コード。
└── veddra_term_name	VEDDRA で使用される用語名。
└── veddra_version	使用される VEDDRA のバージョン。
receiver	副作用報告を受領する者。
└── city	受領都市。
└── country	受領国。
└── organization	受領組織。
└── postal_code	郵便番号。
└── state	受領州。
└── street_address	住所。
report_id	追加情報を識別するためのフィールド
secondary_reporter	関連情報を持つ他の人や組織。
serious_ae	副作用の重大性。
time_between_exposure_and_onset	初回曝露から副作用発症までの時間。
treated_for_ae	副作用に対する治療の有無。
type_of_information	報告に含まれる情報の種類。
unique_aer_id_number	独自の報告識別番号。

```
{
  "meta": {
    "disclaimer": "Do not rely on openFDA to make decisions regarding medical care. While we make every effort to ensure that data is accurate, you should assume all results are unvalidated. We may limit or otherwise restrict your access to the API in line with our Terms of Service.",
    "terms": "https://open.fda.gov/terms/",
    "license": "https://open.fda.gov/license/",
    "last_updated": "2024-10-03",
    "results": {
      "skip": 0,
      "limit": 1,
      "total": 596987
    }
  },
  "results": [
    {
      "health_assessment_prior_to_exposure": {
        "assessed_by": "Veterinarian"
      },
      "onset_date": "20040707",
      "reaction": [
        {
          "veddra_version": "7",
          "veddra_term_code": "99004",
          "veddra_term_name": "Containers, Abnormal"
        },
        {
          "veddra_version": "7",
          "veddra_term_code": "99005",
          "veddra_term_name": "Containers, Damaged"
        },
        {
          "veddra_version": "7",
          "veddra_term_code": "99022",
          "veddra_term_name": "Seal, Abnormal"
        }
      ],
      "veddra_version": "7",
    }
  ]
}
```

```

      "veddra_term_code": "99026",
      "veddra_term_name": "Tablets, Abnormal"
    }
  ],
  "receiver": {
    "organization": "Food and Drug Administration Center for Veterinary Medicine",
    "street_address": "7500 Standish Place (HFV-210) Room N403",
    "city": "Rockville",
    "state": "MD",
    "postal_code": "20855",
    "country": "USA"
  },
  "unique_aer_id_number": "-USFDACVM-2015-UNK-00620",
  "report_id": "N140971",
  "original_receive_date": "20040707",
  "primary_reporter": "Other",
  "type_of_information": "Manufacturing/Product Defect Manufacturing Defect With no Adverse Events",
  "drug": [
    {
      "lot_number": "MBD3290",
      "brand_name": "MSK",
      "dosage_form": "Tablet, chewable",
      "manufacturer": {
        "name": "MSK",
        "registration_number": "USA-USFDACVM-N140971"
      },
      "atc_vet_code": "QP54AA51",
      "active_ingredients": [
        {
          "name": "Ivermectin, Pyrantel",
          "dose": {
            "numerator": "1",
            "numerator_unit": "Unknown",
            "denominator": "1",
            "denominator_unit": "Unknown"
          }
        }
      ]
    }
  ]
}

```

図 1-14 openFDA API Adverse Events より取得したデータサンプル

openFDA API の内、Food API の Recall Enforcement Reports の基本情報を表 1-11 に示し、このフィールドの説明を表 1-12 に示した。取得データの例として図 1-15 に

[https://api.fda.gov/food/enforcement.json?search=report\\_date:\[20040101+TO+20131231\]&limit=1](https://api.fda.gov/food/enforcement.json?search=report_date:[20040101+TO+20131231]&limit=1) にアクセスした場合の取得データを示した。このデータの web 表示を図 1-16 に示した。

表 1-11 openFDA API の Food API の基本情報

概要	<ul style="list-style-type: none"> <li>● FDA に提出された製品リコールイベント情報を含むデータベースである FDA リコールエンタープライズシステム(RES)からデータを取得する。</li> <li>● RES は、FDA のリコール担当者がリコールの提出、更新、分類、終了に使用する電子データ システムである。</li> <li>● すべてのリコールがメディアや FDA のリコールに関するプレスリリースページで発表されるわけではないが、FDA が監視するすべてのリコールは、関連する危険のレベルに応じて分類され、FDA の毎週の執行レポートに含まれる。</li> </ul>
----	---

	<ul style="list-style-type: none"> <li>● このデータは 2004 年から現在までの公開可能な記録をカバーしている。データは毎週更新される。また、全データ(JSON)を一括ダウンロードすることも可能である。</li> </ul>
取得データ形式	JSON
データの更新頻度	毎週
関連リンク	<ul style="list-style-type: none"> <li>● <a href="https://open.fda.gov/apis/food/enforcement/">https://open.fda.gov/apis/food/enforcement/</a> <ul style="list-style-type: none"> <li>➢ API による Recall Enforcement Reports の概要</li> </ul> </li> <li>● <a href="https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts">https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts</a> <ul style="list-style-type: none"> <li>➢ FDA 規制対象製品の特定のリコールに関するプレスリリースやその他の公示から収集した情報を記載</li> </ul> </li> <li>● <a href="https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/enforcement-reports">https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/enforcement-reports</a> <ul style="list-style-type: none"> <li>➢ 執行報告書について記載</li> </ul> </li> <li>● <a href="https://www.accessdata.fda.gov/scripts/ires/index.cfm">https://www.accessdata.fda.gov/scripts/ires/index.cfm</a> <ul style="list-style-type: none"> <li>➢ RES のデータにアクセスできる web ページ</li> </ul> </li> </ul>

表 1-12 openFDA API の Food API のフィールド説明

フィールド名	説明文
address_1	
address_2	
city	リコール企業が所在する都市。
country	リコール企業が所在する国。
state	リコール企業が所在する米国の州
classification	回収の健康被害度を示す FDA による数値指定 (I,II,III)。
code_info	製品またはラベルに記載のロット番号や製品番号など。
distribution_pattern	初期分配地域の概要。
product_code	
product_description	回収対象製品の簡単な説明。
product_quantity	回収対象の欠陥製品の量。
product_type	回収される製品の種類。
reason_for_recall	欠陥の原因と連邦食品・医薬品・化粧品法の違反内容。
center_classification_date	
event_id	特定のリコールイベントに割り当てられた FDA の数値指定。
initial_firm_notification	公衆または顧客へのリコール通知方法。
recall_initiation_date	リコールの開始日。
recall_number	特定のリコールイベントに割り当てられた数値指定。
recalling_firm	リコールを開始した会社。
report_date	FDA がリコールの実施報告を発行した日付。
status	リコールの現状。
termination_date	
voluntary_mandated	リコールの開始者を示す説明。
meta	このセクションには免責事項とライセンス情報が含まれている。`last_updated`フィールドにはデータファイルがエクスポートされた日付が示される。
— disclaimer	openFDA データの重要な詳細とデータセットの制限。

license	openFDA 内のデータに適用されるライセンス条項のリンク。
last_updated	openFDA エンドポイントが最後に更新された日付。
<b>openfda</b>	
application_number	製品に対応する NDA、ANDA、または BLA 番号。
brand_name	製品のブランド名または商標名。
generic_name	製品の一般名。
is_original_packager	医薬品が再包装されているかどうか。
manufacturer_name	医薬品を製造する会社の名前。
nui	National Drug File Reference Terminology 内の医薬品概念に適用されるユニーク識別子。
original_packager_product_ndc	元の包装業者を識別する NDC。
package_ndc	NDC 番号のラベラー、製品、トレードパッケージサイズを識別する番号。
pharm_class_cs	医薬品の化学構造分類。
pharm_class_epc	FDA が科学的に有効で臨床的に意味があると判断した薬理学的分類。
pharm_class_pe	薬理学的分類の生理学的効果または薬力学的効果。
pharm_class_moa	薬理学的分類の作用機序。
product_ndc	NDC 番号のラベラーメーカーコードおよび製品コード。
product_type	製品の種類。
route	医薬品の投与経路。
rxcul	RxNorm Concept Unique Identifier。
spl_id	特定のバージョンの製品ラベルを識別するユニーク識別子。
spl_set_id	製品ラベルのユニーク識別子。
substance_name	医薬品の有効成分一覧。
unii	ユニーク成分識別子。
upc	ユニバーサル商品コード。

```
{
  "meta": {
    "disclaimer": "Do not rely on openFDA to make decisions regarding medical care. While we make every effort to ensure that data is accurate, you should assume all results are unvalidated. We may limit or otherwise restrict your access to the API in line with our Terms of Service.",
    "terms": "https://open.fda.gov/terms/",
    "license": "https://open.fda.gov/license/",
    "last_updated": "2024-10-23",
    "results": {
      "skip": 0,
      "limit": 1,
      "total": 4028
    }
  },
  "results": [
    {
      "country": "United States",
      "city": "Pompano Beach",
      "address_1": "2300 NW 19th St",
      "reason_for_recall": "FreshPoint South Florida is recalling sliced fresh yellow onions and salsa products processed using yellow onions sourced from Gills Onions, LLC which were recalled for Listeria monocytogenes.",
      "address_2": "",
      "product_quantity": "7 cases",
    }
  ]
}
```

```

"code_info": "Item # 302940.",
"center_classification_date": "20120808",
"distribution_pattern": "Products were distributed in South Florida.",
"state": "FL",
"product_description": "Yellow Onion. Product is labeled in part: ¥***FC Salsa Mango¥***Refrigerate
at 34 F***¥" Item # 302940 FC SALSAMANGO 2/5#. Product is fresh made - Ready To Eat.",
"report_date": "20120815",
"classification": "Class I",
"openfda": {},
"recalling_firm": "FreshPoint South Florida",
"recall_number": "F-1922-2012",
"initial_firm_notification": "Two or more of the following: Email, Fax, Letter, Press Release, Telephone,
Visit",
"product_type": "Food",
"event_id": "62750",
"termination_date": "20141008",
"more_code_info": "",
"recall_initiation_date": "20120727",
"postal_code": "33069-5227",
"voluntary_mandated": "Voluntary: Firm initiated",
"status": "Terminated"
}
]
}

```

図 1-15 openFDA API の Food API からの取得データサンプル

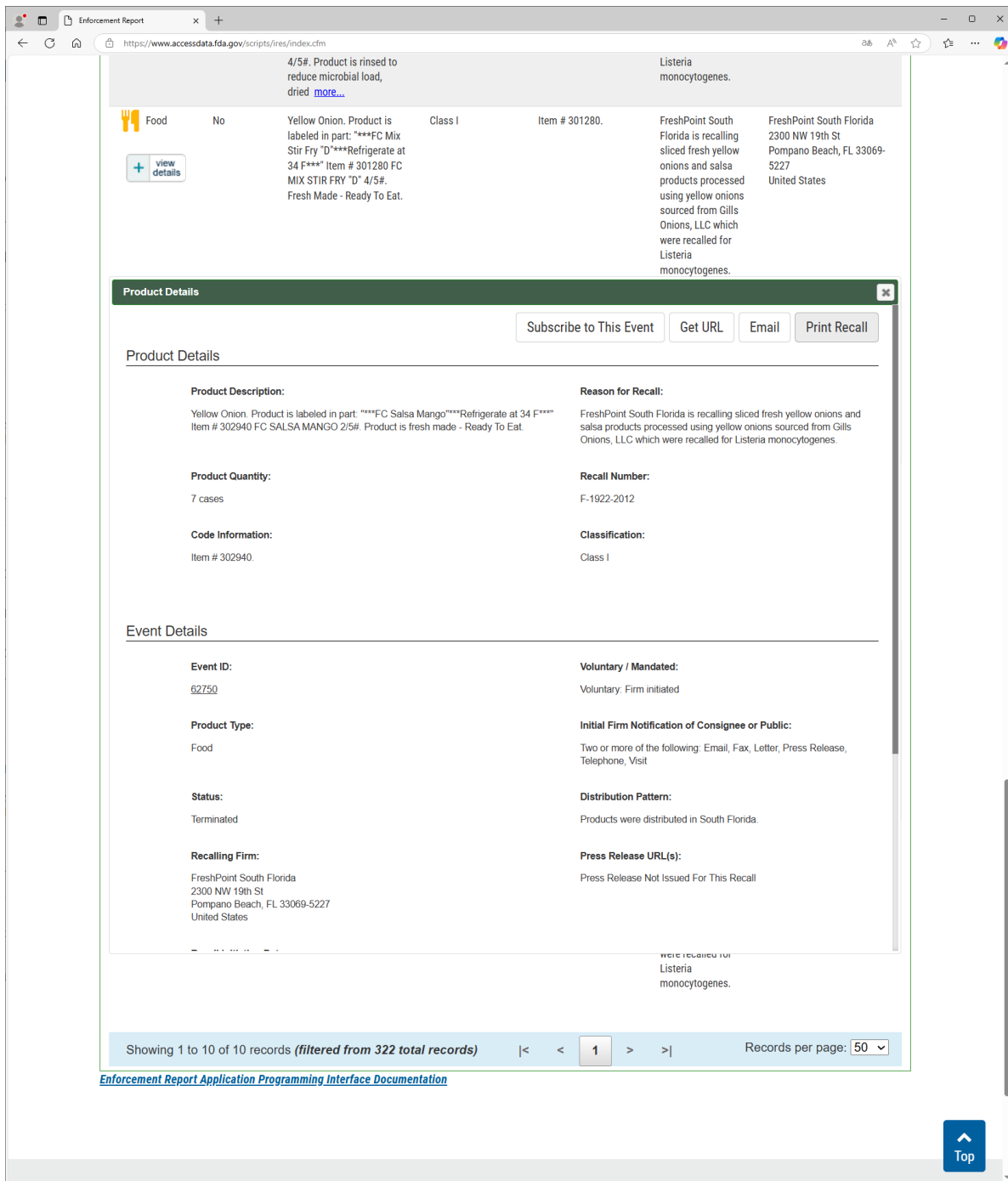


図 1-16 openFDA API の取得データサンプルに対応する RES 内データの web 表示

openFDA API の内、Food API の Adverse Events の基本情報を表 1-13 表 1-11 に示し、このフィールドの説明を表 1-14 に示した。取得データの例として図 1-17 に [https://api.fda.gov/food/event.json?search=date\\_started:\[20040101+TO+20160101\]&limit=1](https://api.fda.gov/food/event.json?search=date_started:[20040101+TO+20160101]&limit=1) にアクセスした場合の取得データを示した。このデータの web 表示を図 1-18 に示した。

表 1-13 openFDA API の Adverse Events の基本情報

概要	
	<ul style="list-style-type: none"> <li>● openFDA の食品、栄養補助食品、化粧品の有害事象 API は、食品、栄養補助食品、化粧品に関して FDA に提出された有害事象および製品苦情レポートに関する情報を含むデータベースである食品安全および応用栄養センターの有害事象報告システム (CAERS) からデータを返す。</li> </ul>

	<ul style="list-style-type: none"> <li>● このデータは 2004 年から現在までの公開可能な記録をカバーしている。データは毎週更新される。また、全データ(JSON)を一括ダウンロードすることも可能である。</li> <li>● FDA は、法律および FDA 規制で許可されている範囲で、CAERS データベースから抽出された生データ(Excel 形式)を提供する。</li> </ul>
取得データ形式	JSON
データの更新頻度	四半期ごと
関連リンク	<ul style="list-style-type: none"> <li>● <a href="https://open.fda.gov/apis/food/event/">https://open.fda.gov/apis/food/event/</a> <ul style="list-style-type: none"> <li>➢ API による Adverse Events の概要を記載</li> </ul> </li> <li>● <a href="https://www.fda.gov/food/compliance-enforcement-food/cfsan-adverse-event-reporting-system-caers">https://www.fda.gov/food/compliance-enforcement-food/cfsan-adverse-event-reporting-system-caers</a> <ul style="list-style-type: none"> <li>➢ CFSAN 有害事象報告システム (CAERS)について記載</li> </ul> </li> </ul>

表 1-14 openFDA API の Adverse Events のフィールドの説明

タグ名	説明文
<b>consumer</b>	副作用報告時の消費者関連情報。
└── <b>age</b>	副作用報告時の消費者の年齢。
└── <b>age_unit</b>	年齢が表現される単位。
└── <b>gender</b>	副作用報告時の消費者の性別。
<b>date_created</b>	報告が FDA に受理された日付。
<b>date_started</b>	副作用が始まったと考えられる日付。
<b>outcomes</b>	副作用の結果や影響。
<b>products</b>	副作用報告に関連する製品情報。
└── <b>industry_code</b>	製品の FDA 業界コード。
└── <b>industry_name</b>	製品に関連付けられた FDA 業界名。
└── <b>name_brand</b>	製品の報告されたブランド名。
└── <b>role</b>	副作用報告における製品の役割。
<b>reactions</b>	反応の MedDRA 用語。
<b>report_number</b>	報告番号。
<b>meta</b>	免責事項とライセンス情報が含まれるセクション。`last_updated`フィールドにはデータファイルがエクスポートされた日付が示される
└── <b>disclaimer</b>	openFDA データの重要な詳細とデータセットの制限。
└── <b>license</b>	openFDA 内のデータに適用されるライセンス条項のリンク。
└── <b>last_updated</b>	openFDA エンドポイントが最後に更新された日付。
└── <b>results</b>	結果に関する情報。
└── └── <b>skip</b>	結果のオフセット (ページ)。`skip`クエリパラメータによって定義される。
└── └── <b>limit</b>	この返却のレコード数。`limit`クエリパラメータによって定義される。パラメータがない場合、API は 1 つの結果を返す
└── <b>total</b>	検索条件に一致するレコードの総数。

クエリ:

[https://api.fda.gov/food/event.json?search=date\\_started:\[20040101+TO+20160101\]&limit=1](https://api.fda.gov/food/event.json?search=date_started:[20040101+TO+20160101]&limit=1)

```

{
  "meta": {
    "disclaimer": "Do not rely on openFDA to make decisions regarding medical care. While we make every effort to ensure that data is accurate, you should assume all results are unvalidated. We may limit or otherwise restrict your access to the API in line with our Terms of Service. Submission of an adverse event report does not constitute an admission that a product caused or contributed to an event. The information in these reports has not been scientifically or otherwise verified as to a cause and effect relationship and cannot be used to estimate incidence (occurrence rate) or to estimate risk.",
    "terms": "https://open.fda.gov/terms/",
    "license": "https://open.fda.gov/license/",
    "last_updated": "2024-10-29",
    "results": {
      "skip": 0,
      "limit": 1,
      "total": 48514
    }
  },
  "results": [
    {
      "report_number": "2019-CFS-000713",
      "outcomes": [
        "Death"
      ],
      "date_created": "20190125",
      "reactions": [
        "Death",
        "Injury"
      ],
      "date_started": "20100409",
      "consumer": {
        "age": "19",
        "age_unit": "year(s)",
        "gender": "Female"
      },
      "products": [
        {
          "role": "SUSPECT",
          "name_brand": "EXEMPTION 4",
          "industry_code": "53",
          "industry_name": "Cosmetics"
        },
        {
          "role": "SUSPECT",
          "name_brand": "EXEMPTION 4",
          "industry_code": "53",
          "industry_name": "Cosmetics"
        }
      ]
    }
  ]
}

```

図 1-17 openFDA API の Adverse Events の取得データサンプル

DATE_FDA_FIRST_RECEIVE	ED_REPORT	REPORT_ID	DATE_EVENT	PRODUCT_TYPE	PRODUCT	PRODUCT_CODE	DESCRIPTION	PATIENT_AGE	AGE_UNITS	SEX	CASE_MEDDRA_PREFERRED_TERMS	CASE_OUTCOME
1/25/2019	2019-CFS-000721	1/1/2009	SUSPECT	EXEMPTION 4	53	Cosmetics	60.00	year(s)	Female	Death	Death	
1/25/2019	2019-CFS-000721	1/1/2009	SUSPECT	EXEMPTION 4	53	Cosmetics	60.00	year(s)	Female	Death	Death	
1/25/2019	2019-CFS-000714	9/27/2017	SUSPECT	EXEMPTION 4	53	Cosmetics	74.00	year(s)	Female	Death, Exposure to chemical pollution, Peritoneal mesothelioma malignant	Death, Hospitalization, Other Serious or Important Medical Event	
1/25/2019	2019-CFS-000713	4/9/2010	SUSPECT	EXEMPTION 4	53	Cosmetics	19.00	year(s)	Female	Death, Injury	Death	
1/25/2019	2019-CFS-000713	4/9/2010	SUSPECT	EXEMPTION 4	53	Cosmetics	19.00	year(s)	Female	Death, Injury	Death	
1/25/2019	2019-CFS-000712	10/13/2014	SUSPECT	EXEMPTION 4	53	Cosmetics	62.00	year(s)	Female	Death, Ovarian cancer	Death, Other Serious or Important Medical Event	
1/25/2019	2019-CFS-000712	10/13/2014	SUSPECT	EXEMPTION 4	53	Cosmetics	62.00	year(s)	Female	Death, Ovarian cancer	Death, Other Serious or Important Medical Event	
1/25/2019	2019-CFS-000709	1/15/2009	SUSPECT	EXEMPTION 4	53	Cosmetics	39.00	year(s)	Female	Death, Injury	Death	
1/25/2019	2019-CFS-000709	1/15/2009	SUSPECT	EXEMPTION 4	53	Cosmetics	39.00	year(s)	Female	Death, Injury	Death	
1/25/2019	2019-CFS-000707	7/28/2003	SUSPECT	EXEMPTION 4	53	Cosmetics	47.00	year(s)	Female	Death, Injury, Ovarian cancer	Death, Other Serious or Important Medical Event	
1/25/2019	2019-CFS-000707	7/28/2003	SUSPECT	EXEMPTION 4	53	Cosmetics	47.00	year(s)	Female	Death, Injury, Ovarian cancer	Death, Other Serious or Important Medical Event	
1/25/2019	2019-CFS-000705	11/7/2014	SUSPECT	EXEMPTION 4	53	Cosmetics	43.00	year(s)	Female	Death, Injury	Death	
1/25/2019	2019-CFS-000705	11/7/2014	SUSPECT	EXEMPTION 4	53	Cosmetics	43.00	year(s)	Female	Death, Injury	Death	
1/24/2019	2020-CFS-004763		SUSPECT		23	Nuts/Edible Seed					Abdominal pain, Diarrhoea	Other Serious or Important Medical Event
1/24/2019	2019-CFS-001115	1/20/2019	SUSPECT		25	Vegetables/Vegetable Products			Male		Abdominal distension, Abdominal pain, Asthenia, Diarrhoea, Nausea, Vomiting	Visited a Health Care Provider
1/24/2019	2019-CFS-000722	1/17/2019	SUSPECT		07	Snack Food Item			Male		Food poisoning	Other Serious or Important Medical Event
1/24/2019	2019-CFS-000719	1/1/2019	SUSPECT		54	Vit/Min/Priot/Unconvy Diet(Human/Animal)	30.00	year(s)	Female		Blood glucose increased, Blood ketone body, Diabetic ketoacidosis, Diarrhoea, Fatigue, Hyperhidrosis, Nausea, Palpitations, Tremor	Other Serious or Important Medical Event

図 1-18 openFDA API の Adverse Events の取得データに対応する CRERS データ

## EFSA

EFSA には、EFSA API Developer Portal (<https://developers.efsa.europa.eu/>)が存在する(図 1-19)が、食品中に存在する物質(意図的に添加された、または自然に存在する化学物質、汚染物質、タンパク質、新規食品成分など)への慢性的な食事曝露を推定する評価ツールである DietEx Tool (<https://www.efsa.europa.eu/en/science/tools-and-resources/dietex>)への API が存在するのみである。

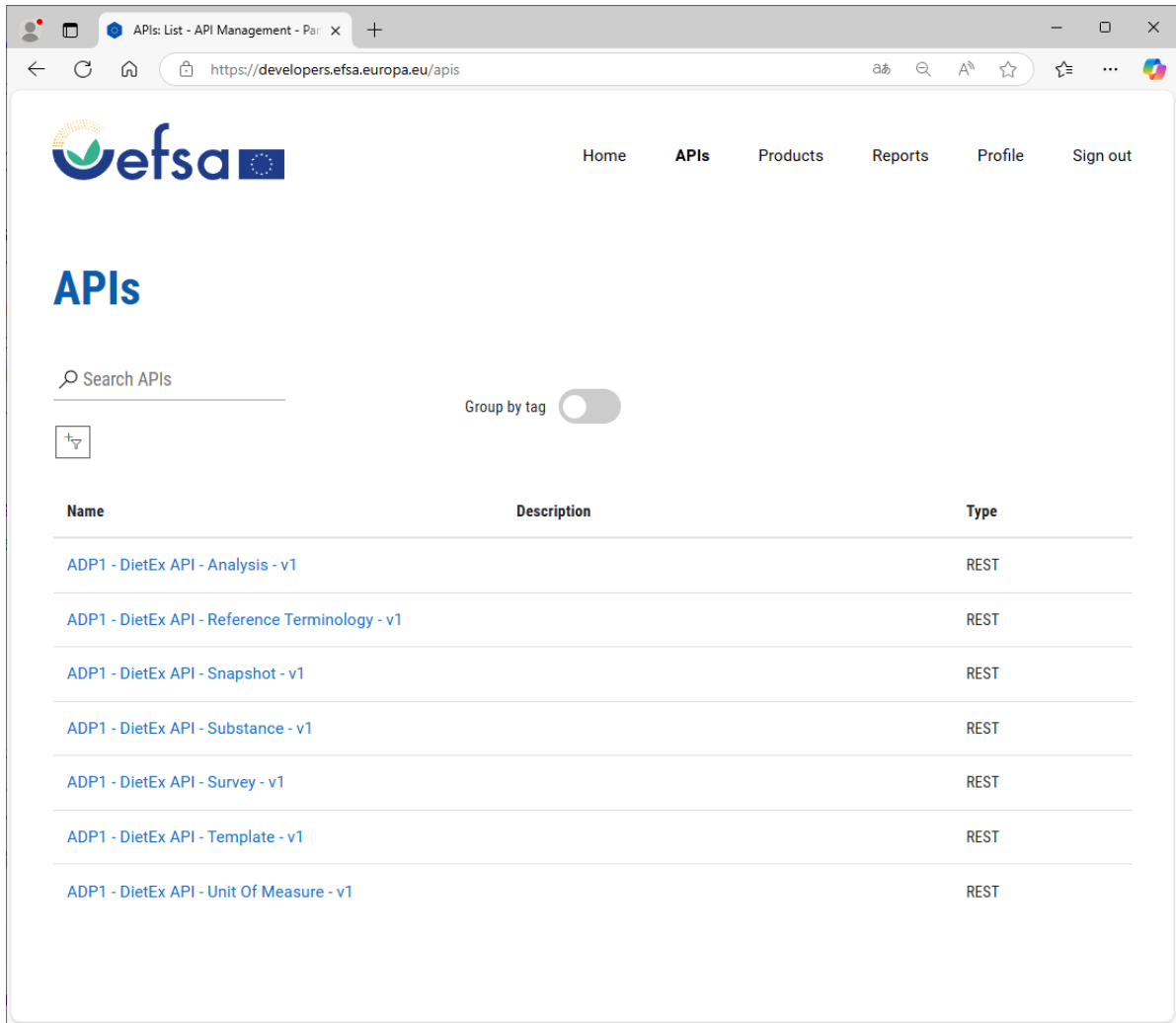


図 1-19 EFSA API Developer Portal

## USDA

USDA では、Developer Resource ページ(<https://www.usda.gov/media/digital/developer-resources>)にて各機関で公開している web サービス(API サービスを含む)を紹介している。

表 1-15 は Developer Resource ページ(<https://www.usda.gov/media/digital/developer-resources>)で紹介されている USDA の各機関で公開している web サービスへのリンクを示している。本調査では、食品安全委員会事務局で巡回しているサイトの機関であり、かつ API が提供されていることが確認できた Food Safety and Inspection Service (FSIS)の API サービスについて具体的に取得できる情報を調査した。

表 1-15 USDA の web サービス一覧

機関名	Web サービスサイトへのリンク	API サービス確認
Agricultural Marketing Service	<a href="https://www.usdalocalfoodportal.com/">https://www.usdalocalfoodportal.com/</a>	○
Agricultural Research Service	<a href="https://fdc.nal.usda.gov/api-guide.html">https://fdc.nal.usda.gov/api-guide.html</a>	○
Animal and Plant Health Inspection Service	<a href="https://vsapps.aphis.usda.gov/aims/">https://vsapps.aphis.usda.gov/aims/</a>	

Economic Research Service	<a href="https://www.ers.usda.gov/developer/">https://www.ers.usda.gov/developer/</a>	○
Farm Service Agency	記載はあるがリンク切れ。	
Food and Nutrition Service		
Food Safety and Inspection Service	<a href="https://www.fsis.usda.gov/science-data/developer-resources">https://www.fsis.usda.gov/science-data/developer-resources</a>	○
Foreign Agricultural Service		
Forest Service	<a href="https://ridb.recreation.gov/landing">https://ridb.recreation.gov/landing</a>	○
National Agricultural Statistics Service	<a href="https://quickstats.nass.usda.gov/api">https://quickstats.nass.usda.gov/api</a>	○
National Institute of Food and Agriculture		
Natural Resources Conservation Service	<ul style="list-style-type: none"> <li>● <a href="https://www.nrcs.usda.gov/programs-initiatives/sswsf-snow-survey-and-water-supply-forecasting-program/national-water-and">https://www.nrcs.usda.gov/programs-initiatives/sswsf-snow-survey-and-water-supply-forecasting-program/national-water-and</a></li> <li>● <a href="https://sdmdataaccess.nrcs.usda.gov/">https://sdmdataaccess.nrcs.usda.gov/</a></li> </ul>	
Risk Management Agency		
Rural Development		

FSIS が提供する web サービスの内、Recall API の基本情報を表 1-16 に示した。このフィールドについての説明 ([https://www.fsis.usda.gov/sites/default/files/media\\_file/documents/Recall-API-documentation.pdf](https://www.fsis.usda.gov/sites/default/files/media_file/documents/Recall-API-documentation.pdf))の一部を表 1-17 に示した。この API から取得できるデータの例を図 1-20 に示し、このサンプルデータが示す記事を図 1-21 に示した。

表 1-16 USDA FSIS が提供する Recall API の基本情報

概要	<ul style="list-style-type: none"> <li>● 食品安全検査局 (FSIS) は、肉、鶏肉、卵製品を規制しており、FSIS は、リコール対象製品のリストと製品ラベルの写真を含むリコールのニュースリリースを通じて一般に通知する。</li> <li>● FSIS は重要な公衆衛生情報の透明性と可用性を高めるために、アプリケーションプログラミングインターフェイス(API)を通じてデータを利用できるようにしています。Recall API のリコールおよび公衆衛生アラートのデータは JSON 形式で利用でき、属性ベースのクエリを提供する。</li> </ul>
API の利用規約	不明
API エンドポイント	<a href="https://www.fsis.usda.gov/fsis/api/recall/v1">https://www.fsis.usda.gov/fsis/api/recall/v1</a> (全データを確認できる)
取得データ形式	JSON
データの更新頻度	逐次
関連リンク	<ul style="list-style-type: none"> <li>● <a href="https://www.fsis.usda.gov/science-data/developer-resources/recall-api">https://www.fsis.usda.gov/science-data/developer-resources/recall-api</a> <ul style="list-style-type: none"> <li>➢ API の概要について記載</li> </ul> </li> <li>● <a href="https://www.fsis.usda.gov/sites/default/files/media_file/documents/Recall-API-documentation.pdf">https://www.fsis.usda.gov/sites/default/files/media_file/documents/Recall-API-documentation.pdf</a> <ul style="list-style-type: none"> <li>➢ API のドキュメンテーション</li> </ul> </li> <li>● <a href="mailto:fsis.webmaster@usda.gov">fsis.webmaster@usda.gov</a></li> </ul>





図 1-21 取得サンプルに対応する記事(<https://www.fsis.usda.gov/recalls-alerts/a-tu-gusto-llc-recalls-frozen-croquette-products-produced-without-benefit-inspection>)

### 振り返り

記事を収集するための API は各機関作成しておらず、API を通じた記事収集を行うことは現在は困難である。

### 1.1.3. クローリング・スクレイピング結果確認機能

#### 1.1.3.1. クローリング・スクレイピング実行確認表示機能

##### 1.1.3.1.1. クローリング結果確認画面の構築

#### 実施内容

クローリング結果を表示する画面を構築した。

#### 結果

構築したクローリング結果を表示する画面を図 1-22 に示した。

記事作成	クローリング条件URL	日時	データ形式	探索深度	ステータス	対象URLタイトル
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	0	成功	Homepage - ECHA
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	1	成功	All news - ECHA
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	1	成功	All news - ECHA
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	1	成功	All news - ECHA
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	1	成功	Maintenance messages - ECHA
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	1	成功	News archive - ECHA
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	1	成功	SCIP - ECHA
<a href="#">🔗</a>	https://www.efsa.europa.eu/en/topics/topic/pesticides	2024/12/18	html	1	成功	Loading...
<a href="#">🔗</a>	https://echa.europa.eu	2024/12/18	html	1	成功	Home - European Observatory for Nanomaterials
<a href="#">🔗</a>	https://www.efsa.europa.eu/en/topics/topic/pesticides	2024/12/18	html	1	成功	Regulation - 396/2005 - EN - EUR-Lex

図 1-22 クローリング結果画面

#### 振り返り

表示画面の URL 及びタイトルはハイパーリンクにすることにより、html ファイルおよび PDF ファイル表示機能が不要になる。

### 1.1.3.1.2. スクレイピング記事確認画面(その1)

#### 実施内容

スクレイピング実行結果は直接記事データベースに保存される。そのためスクレイピングの結果は記事一覧画面から確認することになる。このスクレイピング記事一覧画面を作成した。

#### 結果

作成したスクレイピング記事一覧画面を図 1-22 に示した。期間の絞り込みを行い、記事情報を csv ファイルに保存する画面を図 1-24 に示した。

記事表示	記事作成	スクレイピング条件URL	記事作成日時	データ形式	探索深度	ステータス	対象URLタイトル
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.epa.gov/pesticides/pesticide-news-stories	2025/01/21	html	1	成功	EPA Finalizes Settlement and Announces New Tracking Website for Endocrine Disruptor Screening Program Data Call-In Notices   US EPA
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.epa.gov/pesticides/pesticide-news-stories	2025/01/21	html	1	成功	EPA Releases Draft Biological Opinion for Carbaryl for Public Comment   US EPA
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.efsa.europa.eu/en/topics/topic/pesticides	2025/01/21	html	1	成功	Updated peer review of the pesticide risk assessment of the active substance spinosad   EFSA
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.efsa.europa.eu/en/topics/topic/cross-cutting-science	2025/01/21	html	1	成功	The EU Pollinator Hub: Operationalisation of the EU Bee Partnership Platform for Harmonised Data Collection and Sharing Among Stakeholders on Bees and Pollinators   EFSA
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/biohaz#working-groups	2025/01/20	pdf	1	成功	Microsoft Word - Minutes WG ABP 2nd meeting
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.epa.gov/pesticides/pesticide-news-stories	2025/01/20	html	1	成功	EPA Registers New Pesticide Active Ingredient Veratrine   US EPA
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/biohaz#working-groups	2025/01/20	pdf	1	成功	Minutes of the WG on Carbapenemases
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.efsa.europa.eu/en/topics/topic/biological-hazards	2025/01/20	html	1	成功	Update of the list of qualified presumption of safety (QPS) recommended microbiological agents intentionally added to food or feed as notified to EFSA 21: Suitability of taxonomic units notified to EFSA until September 2024   EFSA
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.epa.gov/pesticides/pesticide-news-stories	2025/01/20	html	1	成功	EPA Announces Interim Decisions on Chlorothalonil, Thiophanate-methyl, and Carbendazim   US EPA
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/ppr#working-groups	2025/01/20	pdf	1	成功	XXth Working Group meeting on

CSV Download

図 1-23 スクレイピング記事一覧表示画面

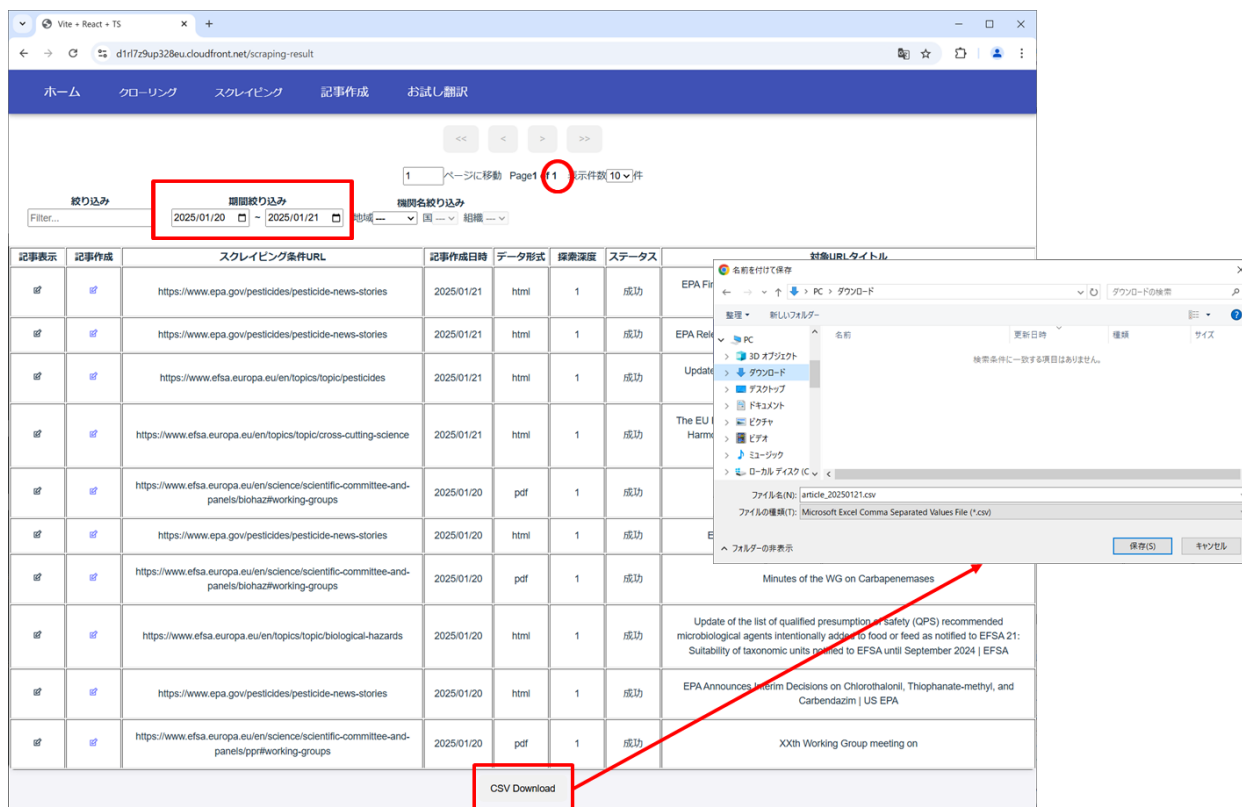


図 1-24 スクレイピング記事一覧法事画面から期間の絞り込みを行い、記事の出力を行う画面

## 振り返り

機関絞り込みの実装とスクレイピング条件 URL、対象 URL のハイパーリンク化が次のタスクである。

### 1.1.3.1.3. スクレイピング記事確認画面(その 2)

#### 実施内容

下記の変更・修正を行った。

#### (ア) URL のハイパーリンク化

- 「スクレイピング条件 URL」のハイパーリンク化と「対象 URL タイトル」への元記事 URL へのハイパーリンクの追加

#### (イ) 記事保存時間を JST に変更

- UTC 時間(協定世界時:ロンドン時間)から JST 時間(日本標準時)に変更

#### (ウ) CSV ダウンロードのタグ名変更

- 日本語タグ名で不具合が発生するため、全て英文化を実施

## 結果

URL のハイパーリンク化を行った画面を図 1-25 に示した。操作性が向上したと期待している。保存時間の変更は、画面上の表示時間が 9 時間ずれること、及び「期間絞り込み」機能でずれが生じていたことへの対応。

タグ名の対応は表 1-18 の通り。

ホーム クローリング スクレイピング お試し翻訳 専門用語辞書管理

タイトル   
 完全一致   
 AND   
   
 期間絞り込み: 2025/03/11 ~ 2025/03/18   
 地域: -- 地域を選択 --  
 記事原文   
 あいまい検索   
 OR   
 国・地域: -- 国・地域を選択 --  
 作成記事本文   
 機関: -- 機関を選択 --

**検索**

<<   <   >   >>

1 ページに移動 Page 1 of 63 表示件数 10 件

記事作成	スクレイピング条件URL	記事作成日時	データ形式	探索深度	ステータス	対象URLタイトル
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging">https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging</a>	2025/03/18			失敗	
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/calls/data">https://www.efsa.europa.eu/en/calls/data</a>	2025/03/17	html	1	成功	<a href="#">Open call for food additives analytical and use level data in food and beverages intended for human consumption   EFSA</a>
<a href="#">🔗</a>		2025/03/17			失敗	
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/calls/data">https://www.efsa.europa.eu/en/calls/data</a>	2025/03/15	html	1	成功	<a href="#">Open call for food additives analytical and use level data in food and beverages intended for human consumption   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/publications?items_per_page=30">https://www.efsa.europa.eu/en/publications?items_per_page=30</a>	2025/03/15	html	1	成功	<a href="#">Candidatus Liberibacter africanus Candidatus Liberibacter americanus Candidatus Liberibacter asiaticus Pest Report to support the ranking of EU candidate priority pests   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/calls/data">https://www.efsa.europa.eu/en/calls/data</a>	2025/03/14	html	1	成功	<a href="#">Call for Data: Length of Incubation Periods of Xylella fastidiosa Subspecies in Infected Plants   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/news?items_per_page=36">https://www.efsa.europa.eu/en/news?items_per_page=36</a>	2025/03/14	html	1	成功	<a href="#">Have your say on EFSA's performance – deadline 1 April 2025   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/publications?items_per_page=30">https://www.efsa.europa.eu/en/publications?items_per_page=30</a>	2025/03/14	html	1	成功	<a href="#">Bactrocera dorsalis Pest Report to support the ranking of EU candidate priority pests   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/publications?items_per_page=30">https://www.efsa.europa.eu/en/publications?items_per_page=30</a>	2025/03/14	html	1	成功	<a href="#">Anastrepha ludens Pest Report to support the ranking of EU candidate priority pests   EFSA</a>

図 1-25 「スクレイピング条件 URL」のハイパーリンク化と「対象 URL タイトル」への元記事 URL へのハイパーリンク化を行った画面

表 1-18 CSV 形式でダウンロードした場合のタグ名

タグ名	説明
key_words	LLM が抽出したキーワード
key_wordsJa	LLM が抽出したキーワードの日本語訳
LLMsummaryEn	LLM による英語要約
LLMsummaryJa	LLM による英語要約の日本語訳
noteText	記事データ。初期値は LLM 日本語要約
summaryEn	英語要約。初期値は LLM 英語要約
summaryJa	日本語要約。初期値は LLM 日本語要約
text_info_body	テキスト化した元記事
translateEn	テキスト化した元記事の英訳
translateJa	テキスト化した元記事の英訳の日本語訳
TITLE	LLM が抽出した元記事のタイトル
titleJa	LLM が抽出した元記事のタイトルの日本語訳

### 振り返り

AWS の標準時間が UTC のためシステムでは UTC を採用していたが、表示時間がずれるなどの致命的な欠陥となるため JST に変更した。古いスクレイピング結果を表示させる場合に時間のずれが残っ

ている。

スクレイピングを実行する時間は日本時間の 19 時としているが、EFSA の記事更新時間(イタリア時間で午前中が多そう)と重なるため日本時間の 22:00 (13:00UTC、EFSA(イタリア)14:00 時、EPA(ワシントン D.C.) 5:00 時)等に変更することが考えられる。

### 1.1.3.2. 取得ファイル一覧表示機能

クローリングによる取得ファイルは 1.1.3.1.1 節の結果確認画面で表示される。

### 1.1.3.3. タグ及び期間絞り込み表示機能(ソート機能)

1.1.3.1.2 節及び 1.1.3.1.3 節のスクレイピング記事一覧画面の機能として構築した。

### 1.1.3.4. html ファイル・PDF ファイル表示機能

1.1.3.1.3 節のスクレイピング記事一覧画面において「スクレイピング条件 URL」及び「対象 URL タイトル」のハイパーリンク化により、web ブラウザーの機能を利用して表示機能とした。

## 1.2. AI 等による翻訳・要約

### 1.2.1. クローリング結果文章選択機能

#### 1.2.1.1. 文書選択登録機能

クローリング結果表示画面の左端のボタンをクリックすることで記事編集画面になり、登録が可能である(図 1-26)。

The screenshot displays a web application interface with a search results table and a document selection screen. The search results table has a red circle around a button in the first row. A red arrow points from this button to the '原文' (Original Text) section of the document selection screen. The document selection screen shows the original text and options for translation and summarization.

記事作成	クローリング条件URL	日時	データ形式	採集深度
	https://echa.europa.eu	2024/12/18	html	0
	https://echa.europa.eu	2024/12/18	html	1
	https://echa.europa.eu	2024/12/18	html	1
	https://echa.europa.eu	2024/12/18	html	1
	https://echa.europa.eu	2024/12/18	html	1
	https://echa.europa.eu	2024/12/18	html	1
	https://www.efsa.europa.eu/en/topics/topic/pesticides	2024/12/18	html	1
	https://echa.europa.eu	2024/12/18	html	1
	https://www.efsa.europa.eu/en/topics/topic/pesticides	2024/12/18	html	1

図 1-26 クローリング結果からの記事選択機能

#### 1.2.1.2. 文書選択テキスト変換機能

クローリング及びスクレイピングの際に、PDF 文書やその他のフォーマットの文書(1.1.2.1 節参照)

は自動的にテキストに変換してデータベースに保存している。

## 1.2.2. 要約機能

### 1.2.2.1. 要約条件設定機能(語調、直訳、意識など)

1.2.5 節の記事編集機能に「要約の作成」もしくは「紹介文の作成」を選択する機能を付け、語調についても「常体(だ・である調)」及び「敬体(です・ます調)」を選択できるようにした。この時の LLM へのプロンプトを表 1-19 に纏めた。プロンプトの調整については 1.2.2.2.3 節参照。

表 1-19 要約条件および語調指定日本語訳プロンプト

英語要約プロンプト	対象の文章の詳細な要約を英語で作成してください。要約以外の説明文は不要です。
英語紹介文プロンプト	対象の文章の紹介文を英語で作成してください。紹介文以外の説明文は不要です。
常体(だ・である調) 日本語訳プロンプト	翻訳対象の文章を日本語に翻訳してください。 翻訳以外の説明文は不要で、略語はそのまま翻訳してください。 語調は「だ・である調」としてください。
敬体(です・ます調) 日本語訳プロンプト	翻訳対象の文章を日本語に翻訳してください。 翻訳以外の説明文は不要で、略語はそのまま翻訳してください。 語調は「です、ます調」としてください。

### 1.2.2.2. 要約実行機能

#### 1.2.2.2.1. EFSA 記事の翻訳、要約例の実施

##### 目的

食安委様からご提供いただいた記事データの一部を対象とし、AWS のマネージドサービスで利用可能な LLM (claude-3-5-sonnet v1) にて、翻訳、要約を実行し、挙動の確認を行った。

##### 実施内容

記事データに含まれている M\_BIBLIOGRAPHIC.xlsx ファイルの内、BULLETINCODE が 150 の記事を選択し、そこから 9 件の記事を対象とした。Excel ファイルの CONTENTS 列の記載内容から参照している URL のコンテンツをダウンロードし、翻訳・要約を実施した。

LLM で処理させる際の形式として、PDF ファイルのまま PDF からプログラムで文章だけをマークダウン形式で抽出したものの 2 パターンで試した。AWS 側の仕様により、ファイルサイズ 4.5MB の制限があるため、大きいファイルに関してはエラーとなっていた。この場合結果ファイルとしては、空の txt ファイルとして格納した。

翻訳の LLM への指示分は、

"Document の内容を翻訳してください。なるべく元の文章の形に則ってください。"

である。

要約の LLM への指示分は、

"Document の内容を要約してください。"

である。

##### 結果

対象とした文章と翻訳、要約の結果は、「5\_EFSA 記事 PDF から翻訳・要約例.zip」及び「5\_EFSA 記事 Markdown から翻訳・要約例.zip」である。

翻訳結果はファイル名の末尾に translate と記載した。今回使用した LLM はトークン数が 8,000 と

比較的少ない（精度は高い）ためか、翻訳結果が途中までとなっているものや、「ここでは、文書の内容全体を翻訳するのは量が多すぎるため、主な部分のみ要約して翻訳します:」のように、自動で全翻訳を放棄する場合があったため、入力するファイルを分割する必要があることが分かった。また

要約結果はファイル名の末尾に **summary** と記載した。翻訳と比較して、要約結果は文章量が少なくなる傾向にあるため、途中で完了しているものは見当たらず、ある程度整理された内容となっていると思われる。

### 振り返り

AWS から利用できる **claude-3-5-sonnet v1** による翻訳・要約に違和感はなく、実用可能なレベルであると思われる。ただし翻訳ではトークン数の問題から分割翻訳が必要なこと、また翻訳を指示しているにも関わらず勝手に要約する場合があるなど、プロンプトにも工夫が必要であることが分かった。

#### 1.2.2.2.2. EFSA 記事の入力フォーマットの違いによる要約の比較

##### 実施内容

EFSA の 9 件の記事について、PDF ファイルをそのまま AWS 上の Claude 3.5 Sonnet にアップロードとした場合と、PDF ファイルをマークダウン(Markdown; MD)形式と呼ばれるテキスト形式に変換した場合を比較した。

##### 結果

比較結果を「4\_EFSA 記事要約例の比較.xlsx」の AY 列に記載した。PDF ファイルのアップロードでは 4.5 MB の容量制限により要約を得られないケースがあった。要約については具体的な項目を抽出して並べるケースと、どのような検討がされているのかについてまとめるケースがある様であった。

### 振り返り

記事の内容を具体的に要約する場合と、記事の紹介の様な要約があり、これらは PDF 形式でも MD 形式でも出力されていた。これは恐らく LLM の回答揺らぎであると思われる、要約の指示文で制御できる可能性がある。「要約条件設定機能」の検討ではいくつか検討を行いたい。

#### 1.2.2.2.3. EFSA 記事の要約プロンプトの調整

##### 実施内容

PDF ファイルをマークダウン(Markdown; MD)形式と呼ばれるテキスト形式に変換し、語調

- ・「だ・である調」
- ・「です、ます調」

を指定し、

- ・要約
- ・紹介文

の作成を行った。実際に実行したプロンプトは以下の通り(try1)。

1. [記事]以下のドキュメントの内容を要約して日本語に翻訳してください。語調は「だ・である調」としてください。
2. [記事]以下のドキュメントの内容を要約して日本語に翻訳してください。語調は「です、ます調」としてください。

3. [記事]以下のドキュメントの**詳細な要約**を作成して**日本語に翻訳**してください。語調は「**だ・である調**」としてください。
4. [記事]以下のドキュメントの**詳細な要約**を作成して**日本語に翻訳**してください。語調は「**です、ます調**」としてください。
5. [記事]以下のドキュメントの**紹介文**を作成して**日本語に翻訳**してください。語調は「**だ・である調**」としてください。
6. [記事]以下のドキュメントの**紹介文**を作成して**日本語に翻訳**してください。語調は「**です、ます調**」としてください。
7. [記事]以下のドキュメントの内容を**記事の言語のまま要約**してください。
8. [記事]以下のドキュメントの**詳細な要約**を**記事の言語のまま作成**してください。
9. [記事]以下のドキュメントの**紹介文**を**記事の言語のまま作成**してください。
10. [記事]以下のドキュメントの**内容を英語で要約**してください。
11. [記事]以下のドキュメントの**詳細な要約**を**英語**で作成してください。
12. [記事]以下のドキュメントの**紹介文**を**英語**で作成してください。

## 結果

実行結果を「2\_EFSA 要約・紹介・語調検討.xlsx」にまとめた。要約対象のドキュメントが長い場合、日本語と指定しても英語で要約することがあった。また「元の言語」を指定すると、フランス語などで要約することがあった。これは名前などから誤判定をしている可能性があるのかもしれない。

「詳細な要約」と「要約」は日本語では大きな変化はなかったが、英語では顕著な差が生じた。

英語で要約してしまうことは課題であると思われる、また要約以外の文章も付与されていたことから、下記のプロンプトでも実行を行った後で、改めて日本語訳を行った(try2)。

1. [記事]以下のドキュメントの内容を**日本語で要約**してください。要約以外の説明文は不要です。語調は「**だ・である調**」としてください。
2. [記事]以下のドキュメントの内容を**日本語で要約**してください。要約以外の説明文は不要です。語調は「**です、ます調**」としてください。
3. [記事]以下のドキュメントの**詳細な要約**を日本語で作成してください。要約以外の説明文は不要です。語調は「**だ・である調**」としてください。
4. [記事]以下のドキュメントの**詳細な要約**を日本語で作成してください。要約以外の説明文は不要です。語調は「**です、ます調**」としてください。
5. [記事]以下のドキュメントの**紹介文**を日本語で作成してください。紹介文以外の説明文は不要です。語調は「**だ・である調**」としてください。
6. [記事]以下のドキュメントの**紹介文**を日本語で作成してください。紹介文以外の説明文は不要です。語調は「**です、ます調**」としてください。
7. [記事]以下のドキュメントの内容を**記事の言語で要約**してください。要約以外の説明文は不要です。
8. [記事]以下のドキュメントの**詳細な要約**を**記事の言語**で作成してください。要約以外の説明文は不要です。
9. [記事]以下のドキュメントの**紹介文**を**記事の言語**で作成してください。紹介文以外の説明文は不要です。

10. [記事]以下のドキュメントの内容を**英語で要約**してください。要約以外の説明文は不要です。
11. [記事]以下のドキュメントの**詳細な要約を英語**で作成してください。要約以外の説明文は不要です。
12. [記事]以下のドキュメントの**紹介文を英語**で作成してください。紹介文以外の説明文は不要です。

上記の実行結果に対して、

- A) 以下の記事を日本語訳してください。記事が日本語の場合はそのまま出力してください。を加えて実行した。

この場合「記事の言語で要約」ではフランス語で要約することはなかったが、日本語で要約することがあった。

### 振り返り

プロンプトの指示通りに機能していたのは下記の項目である。

- (ア) 日本語の語調の指定
- (イ) 英語での「要約」および「紹介文」の作成
- (ウ) 英語における「要約」と「詳細な要約」の違い
- (エ) 英語から日本語への翻訳

そこで記事作成機能での要約は、元記事が英語であっても英語で要約・紹介文の作成を行い、これを日本語に翻訳することで実装したい。具体的なプロンプトは下記を想定している。

- A) 以下のドキュメントの詳細な要約を英語で作成してください。要約以外の説明文は不要です。
- B) 以下のドキュメントの紹介文を英語で作成してください。紹介文以外の説明文は不要です。得られた英語の要約・紹介文を次のプロンプトで日本語訳を作成する。
- イ) 以下の記事を日本語訳してください。翻訳以外の説明文は不要です。語調は「だ・である調」としてください。
- ロ) 以下の記事を日本語訳してください。翻訳以外の説明文は不要です。語調は「です、ます調」としてください。

#### 1.2.2.3. 要約結果表示機能

1.2.5.3 節の示した通り、記事作成画面内に要約結果を表示させた。

#### 1.2.2.4. 要約保存機能

1.2.5.3 節の示した通り、記事作成画面から記事データとして要約を保存する。

### 1.2.3. 翻訳機能

#### 1.2.3.1. 翻訳条件設定機能(語調、直訳、意識など)

1.2.2.1 節と同様。

#### 1.2.3.2. 翻訳実行機能

##### 1.2.3.2.1. お試し翻訳画面作成

##### 実施内容

記事作成を支援するため、claude-3-5-sonnet v1 を使った日本語翻訳機能を作成した。

## 結果

作成した画面への遷移と画面を図 1-27 に示した。



## 振り返り

無料の google 翻訳などに比べて日本語訳の精度が高いとの使用感を頂いている。画面構成については、横並びの方が良いとの要望を受けた。

### 1.2.3.2.2. お試し翻訳画面変更とバグ対応

#### 実施内容

翻訳前の文章と翻訳語の文章を横並びに変更した。また翻訳プロンプトを修正し、翻訳のみを行うように変更した。

翻訳機能において、翻訳結果が原文の順番とずれているとのご指摘があり、これに対応。翻訳機能の高速化のため原稿を 1000 文字程度の文章に分割して翻訳していたが、翻訳結果の表示時に順番を整理する処理にミスがあった。

## 結果

変更したした画面への遷移と画面を図 1-28 に示した。不具合を修正し、正しい順番で表示されるようになった。



図 1-28 翻訳前と翻訳後を横にした画面

## 振り返り

LLM のトークン数制限と高速化のために文章を区切って翻訳を実行していたため日本語訳の順番が変わるバグを入れてしまったが、長い入力でも翻訳可能になった。

ご指摘を頂いた文章では、翻訳前の分にプロンプトエンジニアリングを行おうする記述があったが、この様な指定を有効にすることもできるので、ご検討頂きたい。

### 1.2.3.2.3. 多言語の翻訳実行調査

AWS から提供されている大規模原モデル(LLM)の内、Claude 3.5 Sonnet を利用して、アラビア語、中国語、ロシア語、英語のドキュメントの日本語翻訳を行った。

## 実施内容

アラビア語、中国語、ロシア語、英語の PDF ファイルをマークダウン形式(MD)に変換し、この日本語翻訳を行った。実際に翻訳を行ったドキュメントは「3\_翻訳サンプル.zip」の中にある PDF ファイルから変換した md ファイルである。今回の実行環境ではタイムアウトはしないが、実際の web インターフェースからの環境では AWS の API Gateway の制限によりタイムアウトするため、元のドキュメントを 2048 文字以内の文章に区切って(chunk)翻訳を行った上で結合した。この chunk 毎の実行時間とドキュメント全体の翻訳に要した時間を測定した。ただし要約と同様に英語以外の言語から直接日本語訳を指示しても英訳することが散見されたため、翻訳でも①英訳、②日本語訳の順番に行った。英語のドキュメントの場合は、そのままの文章を返す様に LLM に指示を行った。

具体的なプロンプトは下記の通りである。

- ① 以下のドキュメントを英語に翻訳してください。翻訳以外の説明文は不要で、略語はそのまま翻訳してください。英語の文章の場合はそのまま出力してください。
- ② 以下のドキュメントを日本語に翻訳してください。翻訳以外の説明文は不要で、略語はそのまま

翻訳してください。日本語の文章の場合はそのまま出力してください。

## 結果

ドキュメントのサイズと実行(英訳+日本語訳の合計時間)にかかった時間を表 1-20 に示した。英訳、日本語訳結果は「3\_翻訳サンプル.zip」の中の末尾がそれぞれ\_en.txt と \_jp.txt のファイルである。Chunk 毎の実行時間では最大でアラビア語から英語への翻訳で 516 秒かかっている場合があったが、次に時間がかかったケースは 27 秒のため、たまたまの可能性はある(LLM が混んでいた、アメリカとのネットワークに問題が発生していたなど)。中国語の場合は英訳から日本語訳に時間がかかっており、表意文字のため他言語よりも情報量が多いのかもしれない。英語からの日本語訳よりも原文からの英訳の方が通常速いが、ロシア語の場合は英訳にやや時間がかかっていた。

表 1-20 ドキュメントの大きさと翻訳実行時間

ファイル名	行数	文字数	実行時間	最大断片実行時間	2 番目実行時間
Arabic.md	4,681	37,638	65m53.755s	25.800s	22.640s
Arabic_m.md	14,107	41,078	92m56.815s	515.887s	27.163s
Chinese.md	3,512	5,249	39m30.230s	68.031s	65.564s
Chinese_s.md	1,516	6,247	17m42.558s	62.250s	60.764s
Russian.md	4,635	33,869	75m9.661s	63.910s	52.251s
Russian_s.md	224	1,043	2m3.668s	18.920s	16.923s
English.md	5,908	23,042	30m48.683s	22.212s	19.400s
English_s.md	64	222	0m27.202s	18.579s	8.266s

## 振り返り

アメリカとの通信や LLM の混雑状況など、実行時間の揺らぎが大きく API Gateway のタイムアウト時間が課題となる。中国語やロシア語からの翻訳を考えると 2048 文字から 1024 文字くらいに文字数を制限した方が良いと考えられる。

100 ページ程度のドキュメントの翻訳に 30 分～90 分程度かかっており、web インターフェース上の実行はやや難がある場合が考えられる。本番環境ではファイルをアップロードして翻訳結果のファイルをダウンロードする機能などの整備も考えられる。

### 1.2.3.2.4. 辞書機能を加えた翻訳

#### 実施内容

専門用語・定訳辞書機能の実装として、辞書ファイルに登録されている単語を置き換えて LLM に翻訳させる機能を試した。登録されている文字列を日本語で「」付きで置き換えることにより、LLM に翻訳対象外であることを知らせている。この時のプロンプトは下記である。

- ①以下のドキュメントを英語に翻訳してください。翻訳以外の説明文は不要で、略語はそのまま翻訳してください。英語の文章の場合はそのまま出力してください。
- ②以下のドキュメントを日本語に翻訳してください。ただし「」内は既に日本語に翻訳しているため、「」を含めてそのまま出力してください。翻訳以外の説明文は不要で、略語はそのまま翻訳してください。

専門用語・定訳辞書は、「食品の安全性に関する用語集.pdf」に記載の用語を用いた。この PDF ドキ

コメントでは「Risk」や「Risk Assessment」の様な形式で記載されているため、  
 (ア) 記載されている文字列に加えて、全部小文字化、先頭文字のみ大文字化をリストに加える  
 (イ) 複数形への対応として末尾に「s」付けた文字列をリストに加える(es や不規則変化には対応していない)。  
 (ウ) 英語の文字列が長い順にソートし、この順番でマッチした場合に「」付き日本語で置き換えるを行っている。最後の(ウ)は「Risk」と「Risk Assessment」の両方が辞書にある場合の対応である。この変換前後のドキュメントの例を図 1-29 及び図 1-30 に示した。

3. **Risk assessment** template and examples  
 Template  
 You can use a **risk assessment** template to help you keep a simple record of:  
 who might be harmed and how  
 what you're already doing to control the **risks**  
 what further action you need to take to control the **risks**  
 who needs to carry out the action  
 when the action is needed by  
**Risk assessment** template (Word Document Format) (.docx)  
**Risk assessment** template (Open Document Format) (.odt)  
 Example risk assessments  
 These typical examples show how other businesses have managed **risks**. You can use them as a guide to think about:  
 some of the **hazards** in your business  
 the steps you need to take to manage the **risks**  
 Do not just copy an example and put your company name to it as that would not satisfy the law and would not protect your employees. You must think about the specific **hazards** and controls your business needs.

図 1-29 英語の原文

3. 「**リスク評価**」 template and examples  
 Template  
 You can use a 「**リスク評価**」 template to help you keep a simple record of:  
 who might be harmed and how  
 what you're already doing to control the 「**リスク**」  
 what further action you need to take to control the 「**リスク**」  
 who needs to carry out the action  
 when the action is needed by  
 「**リスク評価**」 template (Word Document Format) (.docx)  
 「**リスク評価**」 template (Open Document Format) (.odt)  
 Example 「**リスク評価**」  
 These typical examples show how other businesses have managed 「**リスク**」. You can use them as a guide to think about:  
 some of the 「**ハザード**」 in your business  
 the steps you need to take to manage the 「**リスク**」  
 Do not just copy an example and put your company name to it as that would not satisfy the law and would not protect your employees. You must think about the specific 「**ハザード**」 and controls your business needs.

図 1-30 英語の原文に専門用語・定訳辞書の置き換えを行った場合

## 結果

ドキュメントのサイズと実行にかかった時間を表 1-21 に示した。専門用語・定訳辞書の置き換えを行った日本語訳のドキュメントは「3\_翻訳サンプル.zip」の中の末尾が\_jp\_withDict.txt である。表 1-20 と同じドキュメントについては実行時間を比較することができ、大きな違いは無い様である。

図 1-30 の LLM よるに翻訳を図 1-31 に示した。専門用語・定訳辞書の置き換えを行った場合の翻訳は指示文通りに機能している様である。

表 1-21 ドキュメントの大きさと辞書機能付き翻訳実行時間

ファイル名	行数	文字数	実行時間	最大断片実行時間	2 番目実行時間
Arabic.md	4,681	37,638	62m45.877s	30.254s	25.848s
Chinese_s.md	1,516	6,247	15m27.679s	53.363s	51.762s
Russian_s.md	224	1,043	1m28.585s	13.458s	11.709s
English.md	5,908	23,042	30m15.395s	28.886s	21.731s
PARC_AD6.3.md	4,413	40,369	52m49.688s	19.018s	17.458s
bambermycin.md	50	219	0m17.068s	10.016s	6.698s

### 3. 「リスク評価」テンプレートと例

テンプレート

「リスク評価」テンプレートを使用して、以下の簡単な記録を保持できます：

誰がどのように危害を受ける可能性があるか

「リスク」を制御するために既に行っていること

「リスク」を制御するためにさらに必要な対策

誰が対策を実行する必要があるか

対策がいつまでに必要か

「リスク評価」テンプレート (Word 文書形式) (.docx)

「リスク評価」テンプレート (OpenDocument 形式) (.odt)

「リスク評価」の例

これらの典型的な例は、他の企業が「リスク」をどのように管理しているかを示しています。以下について考える際のガイドとして使用できます：

あなたの事業における「ハザード」の一部

「リスク」を管理するために必要な手順

例をそのままコピーして会社名を記入するだけでは法律を満たさず、従業員を保護することもできません。あなたの事業に特有の「ハザード」と対策について考える必要があります。

図 1-31 専門用語・定訳辞書の置き換えを行ったドキュメントの日本語翻訳例

## 振り返り

専門用語・定訳辞書については単純な用語の置き換えで機能する様である。ただし大文字、小文字、複数形などの指定には検討が必要である。意図しない置き換えを防ぐためには、複数形の不規則変化は専門用語・定訳辞書に登録しておくことが望ましいだろう。

### 1.2.3.2.5. 英語括弧付与機能

#### 実施内容

英語から日本語訳した場合に辞書機能でヒットする用語を「」付きに変換する機能を作成した。

#### 結果

変換結果は下記の通り(今回は「」内は日本語に置き換わっていた)。

<p>3. 「Risk assessment」 template and examples</p> <p>Template</p> <p>You can use a 「risk assessment」 template to help you keep a simple record of:</p> <p>who might be harmed and how</p> <p>what you're already doing to control the 「risks」</p> <p>what further action you need to take to control the 「risks」</p> <p>who needs to carry out the action</p> <p>when the action is needed by</p> <p>「Risk assessment」 template (Word Document Format) (.docx)</p> <p>「Risk assessment」 template (Open Document Format) (.odt)</p> <p>Example 「risk assessments」</p> <p>These typical examples show how other businesses have managed 「risks」 . You can use them as a guide to think about:</p> <p>some of the 「hazards」 in your business</p> <p>the steps you need to take to manage the 「risks」</p> <p>Do not just copy an example and put your company name to it as that would not satisfy the law and would not protect your employees. You must think about the specific 「hazards」 and controls your business needs.</p> <p>Office-based business (PDF)</p> <p>Local shop/newsagent (PDF)</p> <p>Food preparation and service (PDF)</p> <p>Motor vehicle repair shop (PDF)</p> <p>Factory maintenance work (PDF)</p> <p>Warehouse (PDF)</p>
---

#### 振り返り

特になし。

### 1.2.3.3. 言語自動抽出機能

翻訳において、(原文が英語でも)必ず英語翻訳を実行することにより言語を自動抽出する必要がなくなった(1.2.3.2.3 節参照)。このため本機能は検討しないことになった。

### 1.2.3.4. 翻訳結果表示機能

1.2.5.3 節の示した通り、記事作成画面内に翻訳結果を表示させた。

### 1.2.3.5. 翻訳保存機能

1.2.5.3 節の示した通り、記事作成画面から記事データとして翻訳を保存する。

## 1.2.4. 専門用語・定訳辞書機能

### 1.2.4.1. 専門用語辞書の作成

#### 1.2.4.1.1. 辞書原稿作成

##### 実施内容

専門用語辞書として、食品安委員会の web page からダウンロードした「食品の安全性に関する用語集」から辞書の原稿を作成した。

##### 結果

抽出した英語、日本語のペアのリストを「専門用語・定型辞書原稿.xlsx」にまとめた。

##### 振り返り

「食品の安全性に関する用語集」は「Hot」などの単語も含まれており、必ずしも専門用語辞書として適切であるとは言えない。また複数の英語表現が一つの日本語訳になる場合がある。本事業ではこのリストを専門用語辞書の例として使用する。

#### 1.2.4.1.2. 複数形対応

##### 実施内容

複数形を作成するライブラリーInflector<sup>1</sup> (3.1.1)を組み込んで複数形を自動的に付与する辞書機能を作成した。

##### 結果

Inflector で複数形を追加した辞書を「3\_複数形対応辞書.xlsx」の「複数形対応」シートに付与した。「study」→「studies」や「analysis」→「analyses」らの追加が確認できた。尚、「全て小文字化」と「全て小文字化後に最初の文字のみ大文字化」のエントリーも付与しているので注意されたい。また、元辞書が複数形で登録されている場合、単数形は追加していない。

##### 振り返り

下記を認識しているが本番システムで対応が必要かどうかは要検討。

- 「Hazard Analysis and Critical Control Point」は「Hazard Analysis and Critical Control Points」が追加されるが「Hazard Analyses and Critical Control Points」は追加されない。
- 「IARC」などの略語も「iarc」と変換されてしまう。

#### 1.2.4.2. 登録単語対応表示機能

##### 課題

行政利用においては専門用語に定訳がある場合がある。行政用語の辞書を作成し、日本語訳において対応することが望ましい。そこで専門用語について英語と日本語のペアのリストを用意し、これを専門用語データベースに格納する機能が必要になる。

##### 実施内容

「食品の安全性に関する用語集」の用語を専門用語に見立てて辞書を作成し、これをアップロードし

---

<sup>1</sup> <https://pypi.org/project/Inflector/>

て表示し、辞書データベースに格納する機能を実装した。

## 結果

辞書データ登録・表示画面を図 1-32、図 1-33 に示した。



図 1-32 辞書データ登録・表示画面 辞書ファイルアップロード



図 1-33 辞書データ登録・表示画面 システムへの登録・削除

## 振り返り

辞書データ登録時に先頭文字の大文字、小文字、および複数形を自動生成して登録している。これらを強制的に登録するべきかどうかはご相談が必要と思われる。

### 1.2.4.3. 専門用語・定型辞書削除機能

1.2.4.2 節の機能に含まれる。

### 1.2.4.4. 専門用語・定型辞書ファイルアップロード機能

1.2.4.2 節の機能に含まれる。

### 1.2.4.5. 専門用語・定型辞書追加機能

1.2.4.2 節の機能に含まれる。

## 1.2.5. 記事編集機能

### 1.2.5.1. 記事表示機能

#### 実施内容

記事作成・表示画面を構築した。

#### 結果

構築した記事作成画面を図 1-34 に示した。記事情報を入力後、翻訳・要約を実行した画面を図 1-35 に示し、メタ情報の入力欄を図 1-36 に示した。



図 1-34 記事作成・表示画面と原文の挿入



図 1-35 記事作成画面における翻訳・要約結果実行後画面

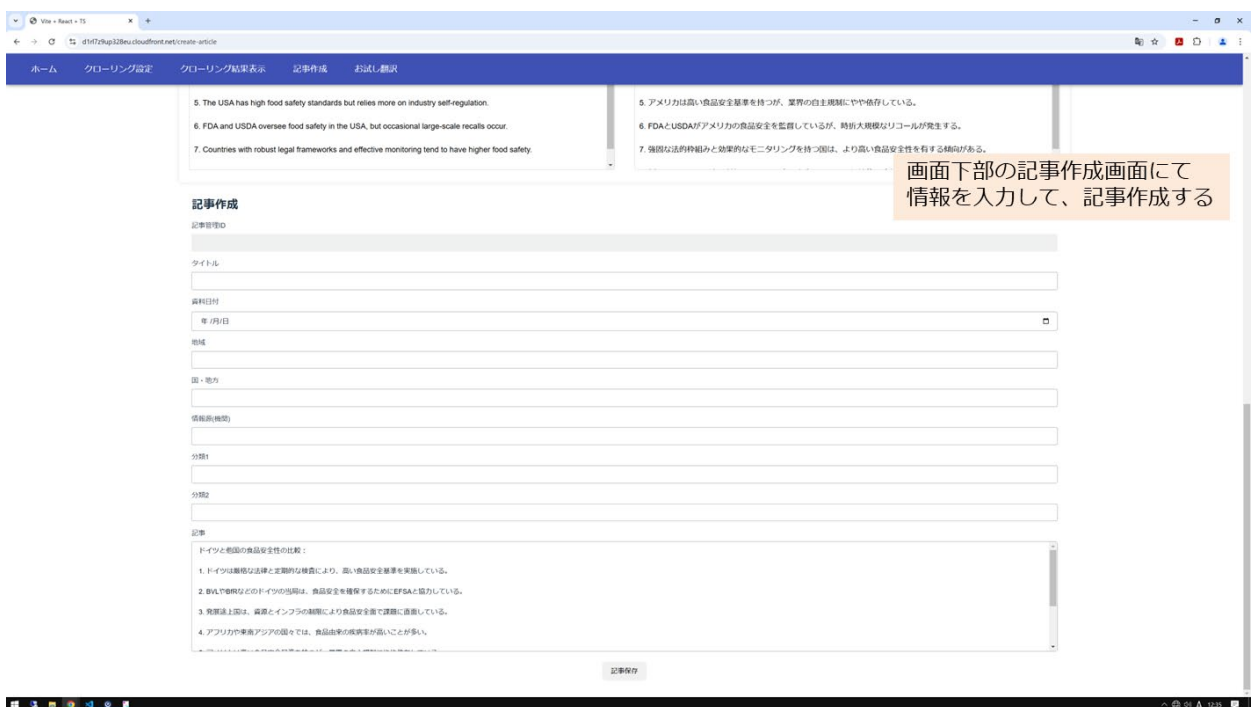


図 1-36 記事作成画面におけるメタ情報入力欄

## 振り返り

翻訳と要約を同時実行するボタンを実装したが、翻訳のみ、及び要約のみを実行するボタンを追加したい。

### 1.2.5.2. 記事表示からの印刷

本システムは web ブラウザーから利用するため、web ブラウザーの機能を使って印刷することが出来る。

### 1.2.5.3. 記事編集機能

1.2.5.1 節の記事作成・表示画面から編集することが出来る。

### 1.2.5.4. 記事保存機能

1.2.5.1 節の記事作成・表示画面から編集することが出来る。

### 1.2.5.5. 記事出力機能(csv ファイル)

1.1.3.1.3 節のスクレイピング記事一覧画面から出力することが出来る。

### 1.2.5.6. テキストからの記事追加機能

1.2.5.1 節の記事作成画面から入力する。

### 1.2.5.7. ファイルアップロード記事追加機能

本事業ではテキストをコピー&ペーストする 1.2.5.1 節の記事作成画面で代替するが、本番システムでは PDF ファイルのアップロード機能が必要になる可能性がある。この場合記事のメタ情報として情報元の URL の入力が必要なるため、URL 指定スクレイピング実行機能とするなどの協議が必要になると考えられる。

### 1.2.5.8. コメント追記機能(メタ情報:ユーザー、日時)

#### 実施内容

記事作成画面にコメントを入力・表示する機能を追加した。

#### 結果

記事作成画面にコメントを入力・表示する機能を追加した画面を図 1-37 に示した。コメントは記事作成画面の下部から「ユーザー名」と「コメント欄」に入力して「記事保存」をクリックすることで保存する。

クローリング スクレイピング お試し翻訳 専門用語辞書管理 トレンド表示

Modification of the existing maximum residue level for deltamethrin in cherries

タイトル (日本語)  
サクランボにおけるデルタメトリンの既存の最大残留基準値の修正

キーワード (英語)  
cherries, consumer risk assessment, Deltamethrin, insecticide, MRL, maximum residue level, pesticides, Bayer SAS, France, analytical methods, er

キーワード (日本語)  
サクランボ, 消費者リスク評価, デルタメトリン, 殺虫剤, MRL, 最大残留基準値, 農薬, バイエルSAS, フランス, 分析方法, 施行, 定量限界, 短期摂取

ソースURL (初期画面から作成する場合は必須)  
https://www.efsa.europa.eu/en/efsajournal/pub/9314

ソースURLタイトル  
Modification of the existing maximum residue level for deltamethrin in cherries | EFSA

日付 (必須)  
2025/03/26

地域  
欧州

国・地方  
EU

情報源(機関)  
欧州食品安全機関 (EFSA)

記事  
欧州食品安全機関 (EFSA) は、サクランボにおけるデルタメトリンの最大残留基準値 (MRL) の変更を評価した。評価の主要ポイントは以下の通りである：  
1. 要請はBayer SASによってフランスの国家機関に提出された。  
2. サクランボの新しいMRLを提案するのに十分なデータが提供された。  
3. 0.01 mg/kgの残留基準値を施行するための適切な分析方法が利用可能である。  
4. 「リスク評価」は、デルタメトリン残留物の短期および長期摂取が消費者の健康に「リスク」をもたらす可能性は低いことを示している。  
5. 以下を含む不確実性により、この評価は示唆的なものと考えられる：  
・ 特定の作物におけるトランスデルタメトリンとアルファ-R-デルタメトリンの実際の残留物に関する情報の欠如  
・ 両異性体の不完全な「毒性学的プロファイル」

コメント

日付	ユーザー	コメント
2025/3/28 20:43:12	user1	デルタメトリンは日本で認可されているのか？

コメント欄  
デルタメトリンは日本で認可されているのか？

記事保存

クローリング スクレイピング お試し翻訳 専門用語辞書管理 トレンド表示

検索対象 検索条件 検索タイプ キーワード 範囲絞り込み 種別名絞り込み

タイトル  完全一致  AND 2025/03/21 ~ 2025/03/28 地域 [ 地域を選択 ]

記事原文  あいまい検索  OR 国・地域 [ 国・地域を選択 ]

作成記事本文 機関 [ 機関を選択 ]

検索

1 ページに移動 Page 1 of 68 表示件数 10 件

リンクURL	記事作成日時	データ形式	検索深度	ステータス	対象URLタイトル
www.efsa.europa.eu/en/topics/topic/pesticides	2025/03/26	html	1	更新	Modification of the existing maximum residue level for cherries   EFSA
www.efsa.europa.eu/en/topics/topic/gmo	2025/03/26	html	1	成功	170th plenary meeting of the GMO Panel   EFSA

クローリング スクレイピング お試し翻訳 専門用語辞書管理 トレンド表示

バージョン変更  
2025-03-28 04:02:41  
2025/3/28 20:46:15

更新

翻訳・要約

EFSAの評価は、提案されたMRL変更が安全であると思われるもの、より包括的な評価のためには知識のギャップがまだ存在し、追加調査が必要であることを示している。

コメント

日付	ユーザー	コメント
2025/3/28 20:43:12	user1	デルタメトリンは日本で認可されているのか？

ユーザー名  
user1

コメント欄

記事保存

図 1-37 コメント表示・追加機能を加えた記事作成画面

## 振り返り

特になし。

### 1.2.5.9. コメント表示機能

1.2.5.8 節のコメント追記機能(メタ情報:ユーザー、日時)機能に含まれる。

## 1.3. 情報の保存と活用

### 1.3.1. クローリング・スクレイピング結果 DB

#### 1.3.1.1. クローリング結果 DB

クローリング結果を格納するデータベースを構築し、クローリング結果を格納した。

#### 1.3.1.2. スクレイピング結果 DB

##### 1.3.1.2.1. 記事データベース対応

### 課題

EPA 及び EFSA を対象としたスクレイピング実行において、

(ア) 翻訳に時間がかかり、AWS lambda の実行時間制限(15分)に抵触することがある。

(イ) 翻訳記事が大きすぎて Amazon DynamoDB の 1 レコードデータサイズ制限(400 KB)に抵触することがある。

(ウ) PDF ファイルをテキスト化したデータが Amazon DynamoDB の 1 レコードデータサイズ制限(400

KB)に抵触することがある。

(エ) HTML ファイルをテキスト化したデータが Amazon DynamoDB の 1 レコードデータサイズ制限 (400 KB)に抵触することがある。  
の課題が生じた。

### 実施内容

次の対応を行った。

- イ) LLM 翻訳に 4 分のタイムアウトを設定した。この場合翻訳結果は初期記事データに含まれなくなる。
- ロ) データベースに登録する項目にデータサイズ制限を設定し、超える部分を機械的に削った。
- ハ) 記事データは下記の優先順位でなるべく保持する様に変更
  - 1. 元記事データ
  - 2. 日本語翻訳データ
  - 3. 英語翻訳データ

### 結果

データベースに記事登録することが出来る様になったが、記事を編集すると容量制限に抵触する可能性が残る。元記事が大きな PDF・HML ファイルの場合はテキスト化した元記事 and/or 全文翻訳を DB に登録することができない。

### 振り返り

本番環境では元記事データと翻訳データは DB に登録せずに別ファイルとして保存し、このパスを DB に登録する様にするべきであろう。パブリッククラウドの場合、このファイルへのアクセス権をどの様にするのかは仕様書作成の段階で詰めておくことが望ましい。

#### 1.3.2. 専門用語・定型辞書 DB

専門用語・定型辞書データベースを構築し、1.2.4.1 節の画面から保存することが出来る。

#### 1.3.3. 翻訳・要約結果 DB

1.2.5.1 節の記事作成・表示画面から保存される記事 DB の項目として保存される。

#### 1.3.4. 記事 DB

1.2.5.1 節の記事作成・表示画面から保存する記事 DB を構築した。

#### 1.3.5. ラベル DB

1.2.5.1 節の記事作成・表示画面から保存される記事 DB の項目として保存される。

#### 1.3.6. ラベル付与機能

##### 1.3.6.1. LLM を使った情報抽出機能

M\_BIBLIOGRAPHIC.xlsx に記載の最後の 100 件の記事について LLM を使った情報抽出方法の検討を行った。

### 1.3.6.1.1. 項目抽出 実施内容

AWS から利用することができる Claude 3.5 Sonnet を利用して

食品添加物名	物質名	医薬品	農薬名	肥料名	毒物名	食品名	材料名
農作物名	工業製品名	商品名	疾患名	動物種	植物種	細菌	ウイルス

の項目について抽出できるのかの検討を行った。具体的な質問形式の例を図 1-38 に示した。

<p>[記事]以下のドキュメントから下記の[質問]に回答してください。誤字脱字、途中で切れていることがあるため、内容を予測しながら質問に回答してください。</p> <p>[質問] 食品添加物名が含まれていれば名前のみを抽出してください。含まれていなければ NA とのみ答えてください。補足情報等は回答に加えないでください。</p> <p>[記事] 欧州食品安全機関(EFSA)は 8 月 8 日、肥育用豚、肥育用のマイナー豚種及び肥育用七面鳥に使用する飼料添加物(<i>Trichoderma reesei</i> MUCL 49755 株により生産されるエンド-1,4-β-キシラナーゼ及び <i>T. reesei</i> MUCL 49754 株により生産されるエンド-1,3(4)-β-グルカナーゼからなる(AveMix(登録商標)XG 10))の認可更新の評価に関する科学的意見書(7 月 3 日採択、PDF 版 7 ページ、doi: <a href="https://doi.org/10.2903/j.efsa.2024.8951">https://doi.org/10.2903/j.efsa.2024.8951</a>)を公表した。概要は以下のとおり。 欧州委員会の要請を受け、EFSA は、肥育用豚、肥育用のマイナー豚種、肥育用七面鳥に使用する畜産飼料添加物としての <i>Trichoderma reesei</i> MUCL 49755 株により生産されるエンド-1,4-β-キシラナーゼ及び <i>T. reesei</i> MUCL 49754 株により生産されるエンド-1,3(4)-β-グルカナーゼからなる飼料添加物(AveMix(登録商標)XG 10/AveMix(登録商標)XG 10 L)の認可更新の評価に関する科学的意見書の提出を求められた。申請者は、AveMix(登録商標)XG 10 の担体原料を大豆粕から炭酸カルシウム+小麦粉又は炭酸カルシウム+セピオライトに変更することを申告した。申請者は、炭酸カルシウム+小麦粉配合の添加物 AveMix(登録商標)XG 10 及び AveMix(登録商標)XG 10 L が認可条件に適合しているエビデンスを提出した。EFSA の動物用飼料に使用される添加物及び製品又は物質に関する科学パネル(FEEDAP パネル)は、炭酸カルシウム+セピオライト配合の AveMix(登録商標)XG 10 が認可条件に適合していることを裏付けるデータが提出されていないと指摘した。同パネルは、肥育用豚、肥育用のマイナー豚種、肥育用七面鳥、消費者、環境に対して、いずれの形態の添加物も安全であると結論した。当該認可更新の枠組みにおいては、同添加物の有効性に関する評価は不要である。</p>
---

図 1-38 食品添加物名の抽出を試みた LLM への質問文の例

#### 結果

この結果を「3\_ラベル付与検討資料.xlsx」の AW~BL 列にまとめた。グレーのセルは抽出に失敗している項目であり、分類違いや検討が必要と思われる部分を赤字で示した。分類の抽出結果の精度は高くないが、全く使えないというレベルでもなく、使用法が難しいと思われる。100 件の範囲であるが、下記のような特徴が見受けられた。

1. 食品添加物名はほとんど取れない。
2. 肥料名はまず取れない。
3. 材料名には食品が入ってしまう。
4. 菌類では英語になっている場合があるが、かなり正確に取れている。
5. 今回の項目では全て NA になってしまう記事がある。

#### 振り返り

ラベル付与は記事の分類やクロス表作成に使用する予定であるが、項目ごとに正しく分類できるわけではない。決まった項目を抜き出すよりもキーワードを抽出させた方が良い可能性がある。

### 1.3.6.1.2. キーワード抽出

#### 実施内容

AWS から利用することができる Claude 3.5 Sonnet を利用して記事からキーワードの抽出を行った。質問文の例を図 1-39 に示したが、図 1-38 とは[質問]が異なるだけである。

[記事]以下のドキュメントから下記の[質問]に回答してください。誤字脱字、途中で切れていることがあるため、内容を予測しながら質問に回答してください。

[質問]  
キーワードをカンマ区切り形式で抽出してください。補足情報等は回答に加えないでください。

[記事]  
欧州食品安全機関(EFSA)は 8 月 8 日、肥育用豚、肥育用のマイナー豚種及び肥育用七面鳥に使用する飼料添加物(*Trichoderma reesei* MUCL 49755 株により生産されるエンド-1,4-β-キシラナーゼ及び *T. reesei* MUCL 49754 株により生産されるエンド-1,3(4)-β-グルカナーゼからなる(AveMix(登録商標)XG 10))の認可更新の評価に関する科学的意見書(7 月 3 日採択、PDF 版 7 ページ、doi: <https://doi.org/10.2903/j.efsa.2024.8951>)を公表した。概要は以下のとおり。 欧州委員会 の要請を受け、EFSA は、肥育用豚、肥育用のマイナー豚種、肥育用七面鳥に使用する畜産飼料添加物としての *Trichoderma reesei* MUCL 49755 株により生産されるエンド-1,4-β-キシラナーゼ及び *T. reesei* MUCL 49754 株により生産されるエンド-1,3(4)-β-グルカナーゼからなる飼料添加物(AveMix(登録商標)XG 10/AveMix(登録商標)XG 10 L)の認可更新の評価に関する科学的意見書の提出を求められた。 申請者は、AveMix(登録商標)XG 10 の担体原料を大豆粕から炭酸カルシウム+小麦粉又は炭酸カルシウム+セピオライトに変更することを申告した。申請者は、炭酸カルシウム+小麦粉配合の添加物 AveMix(登録商標)XG 10 及び AveMix(登録商標)XG 10 L が認可条件に適合しているエビデンスを提出した。EFSA の動物用飼料に使用される添加物及び製品又は物質に関する科学パネル(FEEDAP パネル)は、炭酸カルシウム+セピオライト配合の AveMix(登録商標)XG 10 が認可条件に適合していることを裏付けるデータが提出されていないと指摘した。同パネルは、肥育用豚、肥育用のマイナー豚種、肥育用七面鳥、消費者、環境に対して、いずれの形態の添加物も安全であると結論した。当該認可更新の枠組みにおいては、同添加物の有効性に関する評価は不要である。

図 1-39 キーワード抽出の質問文の例

#### 結果

この結果を「3\_ラベル付与検討資料.xlsx」の BM 列にまとめた。発信元組織や材料などが抽出されている。ただしクロス表作成では使いにくいと思われる、

- 小児科医
- 冷蔵庫
- 評価
- 食事,居住地,年齢,体重
- 返却プログラム
- 透明性

の様なキーワードも抽出していた。

#### 振り返り

項目ごとの抽出よりもキーワード抽出の方が現実的と思われる。ただし抽出キーワードをクロス表などに使うには、選択した記事リストからキーワードを抽出してこれを軸に指定できる様にするなどのデータベース側の工夫が必要となる。

### 1.3.6.1.3. EFSA chemical contaminants in food and feed のスクレイピングのからの項目抽出 実施内容

化学物質のリスク評価のドキュメントについて、構造化の抽出項目として

1. ハザードの概要
2. 毒性に関する情報
3. ばく露に関する情報
4. 他機関の評価情報
5. 評価結果
6. 評価の限界や今後の課題

の項目を頂いた。この情報に加えて

7. 対象物質
8. 対象動物

を加えて EFSA の chemical contaminants in food and feed からスクレイピングで取得した記事からの情報抽出を行った。ここで「Events」はリスク評価のドキュメントでないため「News」と「Scientific output」の記事を対象とし、期間を 2024 年 2 月から 2025 年 2 月の 12 記事を対象とした。対象記事の一覧を表 1-22 に示した。スクレイピングで取得した html ファイルをテキストに変換し、この文字列を加えて表 1-23 に示した質問文(上段)を英訳(下段)して LLM に問い合わせることにより情報抽出を行った(英語の回答を日本語に翻訳)。LLM は Claude 3.5 Sonnet v1、GPT-4o、Gemini 1.5 Pro の 3 つを試した。この結果を「3\_ChemicalContaminants.scraping.xlsx」に示した。

表 1-22 2024 年 2 月から 2025 年 2 月までの EFSA chemical contaminants in food and feed の News と Scientific output の記事一覧

date	title
20240306	Report for 2022 on the results from the monitoring of veterinary medicinal product residues and other substances in live animals and animal products
20240306	Veterinary drug residues in animals and foods remain low
20240312	Risks for animal and human health related to the presence of polychlorinated naphthalenes (PCNs) in feed and food
20240702	Risk assessment of small organoarsenic species in food
20240711	Animal dietary exposure in the risk assessment of contaminants in feed
20240715	Update of the scientific opinion on tetrabromobisphenol A (TBBPA) and its derivatives in food
20240823	EFSA Project on the use of NAMs to explore the immunotoxicity of PFAS
20241009	Genotoxicity of beauvericin
20241023	Update of the risk assessment of brominated phenols and their derivatives in food
20241209	Risk assessment of complex organoarsenic species in food
20241211	Have your say - draft risk assessment of fluoride in food and drinking water
20241217	Perchlorate in food: draft scientific opinion for public consultation

表 1-23 項目を抽出するための質問文。上段の日本語の質問を英訳して LLM に問い合わせた。

項目	質問文
対象物質	<p>下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから<b>対象物質の名称</b>のみを教えてください。余計な説明は不要で、対象物質名が見つからない場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は NA とのみ教えてください。</p> <p>The following document is a report on the risk assessment of chemical substances. Please provide only <b>the name of the target substance</b> from this document. No additional explanations are necessary. If the name of the target substance cannot be found or if it is not a report on the risk assessment of chemical substances, respond only with "NA".</p>
対象動物	<p>下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから<b>ヒトやウシなどの対象動物の名称</b>のみを教えてください。余計な説明は不要で、対象動物が見つからない場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は NA とのみ教えてください。</p> <p>The following document is a report on the risk assessment of chemical substances. Please provide only <b>the name of the target animals, such as humans or cattle</b>, from this document. No additional explanations are necessary. If the name of the target animals cannot be found or if it is not a report on the risk assessment of chemical substances, respond only with "NA".</p>
ハザードの概要	<p>下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから<b>ハザードの概要</b>を抽出してください。余計な説明は不要で、ハザードの情報が無い場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は NA とのみ教えてください。</p> <p>The following document is a report on the risk assessment of chemical substances. Please extract only <b>the summary of the hazard</b> from this document. No additional explanations are necessary. If there is no hazard information or if the document is not a report on the risk assessment of chemical substances, respond only with "NA".</p>
毒性に関する情報	<p>下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから<b>毒性に関する情報</b>を抽出してください。余計な説明は不要で、毒性に関する情報が無い場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は NA とのみ教えてください。</p> <p>The following document is a report on the risk assessment of chemical substances. Please extract only <b>the information related to toxicity</b> from this document. No additional explanations are necessary. If there is no information on toxicity or if the document is not a report on the risk assessment of chemical substances, respond only with "NA".</p>
ばく露に関する情報	<p>下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから<b>ばく露に関する情報</b>を抽出してください。余計な説明は不要で、ばく露に関する情報が無い場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は NA とのみ教えてください。</p> <p>The following document is a report on the risk assessment of chemical substances. Please extract only the information related to exposure from this document. No additional explanations are necessary. If there is no information on exposure or if the document is not a report on the risk assessment of chemical substances, respond only with "NA".</p>
他機関の評価情報	<p>下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから<b>他機関の評価情報</b>を抽出してください。余計な説明は不要で、他機関の評価情報が無い場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は NA とのみ教えてください。</p> <p>The following document is a report on the risk assessment of chemical substances. Please extract only <b>the evaluation information from other organizations</b> in this document. No additional explanations are necessary. If there is no evaluation information from other organizations or if the document is not a report on the risk assessment of chemical substances, respond only with "NA".</p>
評価結果	<p>下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから<b>評価結果</b>を抽出してください。余計な説明は不要で、評価結果の情報が無い場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は NA とのみ教えてください。</p>

	The following document is a report on the risk assessment of chemical substances. Please extract only <b>the evaluation results</b> from this document. No additional explanations are necessary. If there is no information on evaluation results or if the document is not a report on the risk assessment of chemical substances, respond only with "NA".
評価の限界や今後の課題	下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから <b>評価の限界や今後の課題</b> を抽出してください。余計な説明は不要で、評価の限界や今後の課題の情報が無い場合や下記のドキュメントが化学物質のリスク評価に関する報告で無い場合はNAとのみ答えてください。 The following document is a report on the risk assessment of chemical substances. Please extract only <b>the limitations of the assessment or future challenges</b> from this document. No additional explanations are necessary. If there is no information on the limitations or future challenges, or if the document is not a report on the risk assessment of chemical substances, respond only with "NA".
全項目	下記のドキュメントは化学物質のリスク評価に関する報告です。このドキュメントから <b>対象物質、対象動物、ハザードの概要、毒性に関する情報、ばく露に関する情報、他機関の評価情報、評価結果、評価の限界や今後の課題の情報を抽出して項目別にまとめてください</b> 。各項目の情報がなければ無理に答えず、項目の情報としてはNAとのみ答えてください。下記のドキュメントが化学物質のリスク評価に関する報告で無い場合は mismatched と答えてください。 The following document is a report on the risk assessment of chemical substances. Please extract and organize the information into the following categories: <b>target substance, target animal, hazard summary, information on toxicity, information on exposure, evaluation information from other organizations, evaluation results, and limitations of the assessment or future challenges</b> . If there is no information for a particular category, do not force an answer and respond with "NA" for that category. If the document is not related to the risk assessment of chemical substances, respond with "mismatched".

## 結果

例として 2024 年 12 月 9 日の「Risk assessment of complex organoarsenic species in food」(食品中の複雑な有機ヒ素化合物のリスク評価)を下記に示した。想定していた情報が取れているのかや、LLM による取得情報の揺らぎ等を確認し、評価方法の検討を行う必要がある。

## 元記事

元記事情報からリンク等を除いた主要部分を下記に示した。

<p><b>Keywords</b> Arsenobetaine, arsenolipids, arsenosugars, complex organoarsenic species, margin of exposure (MOE), risk assessment</p> <p><b>Abstract</b> The European Commission asked EFSA for a risk assessment on complex organoarsenic species in food. They are typically found in marine foods and comprise mainly arsenobetaine (AsB), arsenosugars and arsenolipids. For AsB, no reference point (RP) could be derived because of insufficient toxicity data. AsB did not show adverse effects in the two available repeat dose toxicity tests in rodents. It has not shown genotoxicity in in vitro assays. There is no indication of an association with adverse outcomes in human studies. The highest 95th percentile exposure for AsB was observed in 'Toddlers' with an estimate of 12.5 µg As/kg bw per day (AsB expressed as elemental arsenic). There is sufficient evidence to conclude that AsB at current dietary exposure levels does not raise a health concern. For glycerol arsenosugar (AsSugOH) a RP of 0.85 mg As/kg bw per day was derived based on the BMDL10 values for cognitive and motor function in mice. A margin of exposure (MOE) of 1000 would not raise a health concern. The highest 95th percentile estimate of exposure for AsSugOH (for adult consumers of red seaweed Nori/Laver) was 0.71 µg As/kg bw per day (AsSugOH expressed as elemental arsenic), which results in an MOE&gt;1000, not raising a health concern. Based on qualitative consideration of all identified uncertainties, it is</p>
--

regarded likely that the dietary exposures to AsB and AsSugOH do not raise a health concern. No conclusions could be drawn regarding other arsenosugars. No risk characterisation could be conducted for arsenolipids, due to the lack of data.

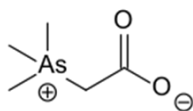
キーワード：

アルセノベタイン、アルセノリピド、アルセノ糖、複雑な有機ヒ素化合物、暴露マージン (MOE)、リスク評価

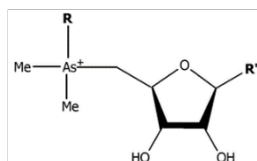
要旨

欧州委員会は EFSA に食品中の複雑な有機ヒ素化合物のリスク評価を依頼した。これらは主に海産物に含まれており、主にアルセノベタイン (AsB)、アルセノ糖、アルセノリピドで構成される。AsB については、毒性データが不十分なため参照点 (RP) を導出できなかった。AsB は、げっ歯類を用いた 2 つの利用可能な反復投与毒性試験において有害な影響を示さなかった。in vitro 試験では遺伝毒性を示していない。ヒトを対象とした研究では、有害な結果との関連性は示されていない。AsB の最も高い 95 パーセントイル暴露は「幼児」で観察され、12.5  $\mu\text{g As/kg}$  体重/日 (元素ヒ素として表現された AsB) と推定された。現在の食事暴露レベルの AsB が健康上の懸念を引き起こさないと結論付けるには十分な証拠がある。グリセロールアルセノ糖 (AsSugOH) については、マウスの認知機能と運動機能に関する BMDL10 値に基づいて、0.85 mg As/kg 体重/日の RP が導出された。暴露マージン (MOE) が 1000 以上であれば健康上の懸念は生じない。AsSugOH の最も高い 95 パーセントイル暴露推定値 (赤海藻ノリラバーの成人消費者) は 0.71  $\mu\text{g As/kg}$  体重/日 (元素ヒ素として表現された AsSugOH) であり、これは MOE > 1000 となり、健康上の懸念を引き起こさない。特定されたすべての不確実性を定性的に考慮すると、AsB と AsSugOH の食事暴露が健康上の懸念を引き起こす可能性は低いと考えられる。他のアルセノ糖については結論を出すことができなかった。アルセノリピドについては、データ不足のためリスク特性評価を行うことができなかった。

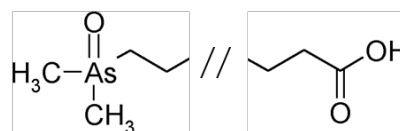
arsenobetaine (AsB)



arsenosugars



arsenolipids



Claude 3.5 Sonnet v1 による現システムでの要約は下記である。

欧州食品安全機関 (EFSA) は、主に海産物に含まれる複雑な有機ヒ素化合物 (アルセノベタイン (AsB)、ヒ素糖、ヒ素脂質) に関する食品中のリスク評価を実施した。主な調査結果は以下の通りである。

1. アルセノベタイン (AsB) :

- 毒性データが不十分なため、参照値を設定できなかった。
- げっ歯類の研究や in vitro 遺伝毒性試験で有害影響は観察されなかった。
- 最高曝露量 (95 パーセントイル) は幼児で 12.5  $\mu\text{g As/kg}$  体重/日であった。
- 現在の食事による曝露レベルは健康上の懸念を引き起こさない。

2. グリセロールヒ素糖 (AsSugOH) :

- マウスの認知機能と運動機能に基づき、参照値を 0.85 mg As/kg 体重/日と設定した。
- 曝露マージン (MOE) 1000 が安全とみなされる。
- 最高曝露量 (95 パーセントイル) は赤藻類を摂取する成人で 0.71  $\mu\text{g As/kg}$  体重/日であった。
- MOE > 1000 であり、健康上の懸念はないことを示している。

3. その他のヒ素糖とヒ素脂質 :

- 他のヒ素糖については結論を出すことができなかった。
- データ不足のため、ヒ素脂質のリスク特性評価は不可能であった。

総じて、現在の証拠と不確実性を考慮すると、AsB と AsSugOH の食事による曝露は健康上の懸念はないと考えられる。

### 対象物質

Claude 3.5 Sonnet v1 は日本語訳の時に説明文が付与されていたため英語の回答を示した。

Claude 3.5 Sonnet	complex organoarsenic species
GPT-4o	複雑な有機ヒ素化合物
Gemini 1.5 Pro	複雑な有機ヒ素種

### 対象動物

Claude 3.5 Sonnet v1 は日本語訳の時に「なし」に変換されてしまったため英語の回答を示した。

Claude 3.5 Sonnet	humans
GPT-4o	人間
Gemini 1.5 Pro	ヒト、マウス、げっ歯類

### ハザードの概要

Claude 3.5 Sonnet	なし
GPT-4o	ドキュメントから抽出された危険性の概要：,アルセノベタイン (AsB) : 毒性試験で有害な影響は観察されず、遺伝毒性もなく、人間の研究において有害な結果との関連性も認められませんでした。現在の食事による摂取量は健康上の懸念を引き起こしません。、グリセロールアルセノシュガー (AsSugOH) : 基準値 (RP) として 0.85 mg As/kg 体重/日が導き出されました。暴露マージン (MOE) が 1000 以上の場合、健康上の懸念はありません。現在の食事による摂取量は健康上の懸念を引き起こしません。、その他のアルセノシュガー : 結論を導き出すことはできませんでした。、アルセノリピッド : データ不足のためリスクの特性評価は行われていません。
Gemini 1.5 Pro	アルセノベタイン (AsB) は、げっ歯類の研究では有害な影響を示さず、in vitro での遺伝毒性も示さなかった。現在の曝露レベルでは、ヒトの研究において有害な結果は示されていない。グリセロールアルセノシュガー (AsSugOH) については、暴露マージン (MOE) > 1000 が算出され、現在の曝露レベルでは健康への懸念はないことが示された。他のアルセノシュガーについては結論を導き出すことができず、データ不足のため、アルセノリピッドのリスク特性評価は不可能であった。

### 毒性に関する情報

Claude 3.5 Sonnet	この文書には、毒性に関する以下の情報が含まれています：,アルセノベタイン (AsB) については、毒性データが不十分なため、参照ポイントを導き出すことができませんでした。AsB は、げっ歯類での 2 つの利用可能な反復投与毒性試験において有害な影響を示しませんでした。また、in vitro アッセイで遺伝毒性を示していません。、グリセロールヒ素糖 (AsSugOH) については、マウスの認知機能と運動機能に関する BMDL10 値に基づき、0.85 mg As/kg 体重/日の参照ポイントが導き出されました。、AsSugOH の曝露マージン (MOE) が 1000 以上であれば、健康上の懸念は生じないと考えられます。、他のヒ素糖の毒性に関しては、結論を出すことができませんでした。、データ不足のため、ヒ素脂質についてはリスク特性評価を行うことができませんでした。
GPT-4o	欧州委員会は、食品中の複雑な有機ヒ素化合物に関するリスク評価を EFSA (欧州食品安全機関) に依頼しました。これらは主に海産食品に含まれ、主にアルセノベタイン (AsB)、アルセノシュガー、アルセノリピッドで構成されています。AsB については、毒性データが不十分であるため、基準点 (RP) を導き出すことができませんでした。AsB は、げっ歯類を用いた 2 つの反復投与毒性試験において有害な影響を示さず、また、in vitro 試験において遺伝毒性も示しませんでした。人間の研究においても有害な結果との関連性は示されていません。AsB の 95 パーセントイルの最高曝露量は「幼児」で観察され、12.5 μg As/kg 体重/日 (AsB を元素ヒ素として表現) と推定されました。現在の食事性曝露レベル

	<p>において、AsB が健康上の懸念を引き起こすことはない結論付けるのに十分な証拠があります。、グリセロールアルセノシュガー (AsSugOH) については、マウスの認知機能および運動機能に基づく BMDL10 値に基づき、0.85 mg As/kg 体重/日という RP が導出されました。暴露マージン (MOE) が 1000 以上であれば健康上の懸念を引き起こさないとされます。AsSugOH の 95 パーセントの最高暴露量 (赤い海藻ノリ/ラバーを消費する成人) は 0.71 <math>\mu</math>g As/kg 体重/日 (AsSugOH を元素ヒ素として表現) と推定され、MOE が 1000 を超えるため、健康上の懸念を引き起こすことはありません。特定されたすべての不確実性を定性的に考慮した結果、AsB および AsSugOH の食事性暴露が健康上の懸念を引き起こさない可能性が高いと考えられます。他のアルセノシュガーについては結論を出すことができませんでした。アルセノリピッドについてはデータが不足しているため、リスク特性評価を行うことができませんでした。</p>
Gemini 1.5 Pro	<p>AsB については、毒性データが不十分なため、参照点 (RP) を導出できませんでした。AsB は、入手可能な 2 つのげっ歯類を用いた反復投与毒性試験において有害作用を示しませんでした。また、in vitro 試験では遺伝毒性を示していません。グリセロールアルセノシュガー (AsSugOH) については、マウスにおける認知機能および運動機能に対する BMDL10 値に基づき、0.85 mg As/kg 体重/日の RP が導出されました。</p>

#### ばく露に関する情報

Claude 3.5 Sonnet	<p>ドキュメントには、ばく露に関する以下の情報が含まれています：、アルセノベタイン (AsB) の 95 パーセントのばく露の最高値は、「幼児」で観察され、1 日あたり 12.5 <math>\mu</math>g As/kg 体重 (元素状ヒ素として表現された AsB) と推定されました。、グリセロールアルセノシュガー (AsSugOH) の 95 パーセントのばく露推定値の最高値は (赤海藻ノリ/ラバーを摂取する成人について)、1 日あたり 0.71 <math>\mu</math>g As/kg 体重 (元素状ヒ素として表現された AsSugOH) でした。</p>
GPT-4o	<p>AsB の 95 パーセントにおける最高の暴露量は「幼児」で観察され、その推定値は 12.5 <math>\mu</math>g As/kg 体重/日 (AsB は元素ヒ素として表現) でした。、AsSugOH の 95 パーセントにおける最高の暴露量の推定値は、(赤い海藻であるノリ/ラバーを消費する成人の場合) 0.71 <math>\mu</math>g As/kg 体重/日 (AsSugOH は元素ヒ素として表現) でした。</p>
Gemini 1.5 Pro	<p>ヒ素 (AsB) の 95 パーセントの最高曝露量は、「幼児」で観察され、推定値は 1 日あたり 12.5 <math>\mu</math>g As/kg 体重 (AsB は元素ヒ素として表される) であった。、紅藻ノリ/アマノリを摂取する成人消費者におけるヒ素糖 (AsSugOH) の 95 パーセントの曝露量の最高推定値は、1 日あたり 0.71 <math>\mu</math>g As/kg 体重 (AsSugOH は元素ヒ素として表される) であった。</p>

#### 他機関の評価情報

Claude 3.5 Sonnet	なし
GPT-4o	なし
Gemini 1.5 Pro	なし

#### 評価結果

Claude 3.5 Sonnet	なし
GPT-4o	<p>アルセノベタイン (AsB): 現在の食事摂取レベルでは、AsB が健康上の懸念を引き起こさないことを結論付けるのに十分な証拠がある。、グリセロールアルセノシュガー (AsSugOH): 参照値 (RP) として 0.85 mg As/kg 体重/日が導き出された。暴露マージン (MOE) が 1000 以上の場合、健康上の懸念を引き起こさない。AsSugOH の食事摂取は健康上の懸念を引き起こさない。、その他のアルセノシュガー: 結論を導き出すことはできなかった。、アルセノリピッド: データ不足のためリスク特性評価を行うことができなかった。</p>
Gemini 1.5 Pro	<p>AsB については、参照点 (RP) を導出できませんでした。げっ歯類を用いた 2</p>

	<p>つの反復投与毒性試験では有害作用は認められませんでした。<b>*in vitro*</b>試験では遺伝毒性は認められませんでした。ヒト試験では有害事象との関連性は示唆されませんでした。幼児における 95 パーセントイル曝露量の上限値は 1 日あたり 12.5 <math>\mu\text{g As/kg}</math> 体重でした。現在の食事曝露レベルでの AsB は健康上の懸念を引き起こしません。グリセロールアルセノシュガー (AsSugOH) については、RP は 1 日あたり 0.85 mg As/kg 体重です。MOE が 1000 以上であれば健康上の懸念を引き起こしません。紅藻類を摂取する成人における 95 パーセントイル曝露量の上限値は 1 日あたり 0.71 <math>\mu\text{g As/kg}</math> 体重であり、MOE は 1000 を超えるため、健康上の懸念を引き起こしません。AsB および AsSugOH への食事曝露は健康上の懸念を引き起こしません。他のアルセノシュガーに関する結論はありません。データ不足のため、アルセノリピドのリスク特性評価は行われていません。</p>
--	--

評価の限界や今後の課題

Claude 3.5 Sonnet	<p>文書から抽出された制限事項と今後の課題は以下の通りです: 1. アルセノベタイン(AsB)の毒性データが不十分で、参照点を導き出すことができない。 2. グリセロールアルセノ糖(AsSugOH)以外のアルセノ糖に関する結論が出せない。 3. データ不足のため、アルセノリピドのリスク特性評価を実施できない。</p>
GPT-4o	<p>・アルセノベタイン (AsB) については、基準点 (RP) を導出するための毒性データが不十分です。他のアルセノシュガーに関しては結論を出すことができませんでした。データ不足のため、アルセノリピドについてのリスク評価は実施できませんでした。</p>
Gemini 1.5 Pro	<p>他の砒糖類については結論を導き出すことができませんでした。データがないため、砒脂質のリスク特性評価を行うことができませんでした。</p>

全項目まとめて抽出

- ・ 項目別抽出では抽出していない内容を赤字で示した。
- ・ 英語で質問し、回答を日本語訳しているために項目名は LLM 間で一致していない。
- ・ GPT-4o 及び Gemini 1.5 Pro ではマークダウン形式で出力されていた。

Claude 3.5 Sonnet	<p><b>標的物質:</b> 複雑な有機ヒ素化合物、特にアルセノベタイン(AsB)、アルセノ糖 (特にグリセロールアルセノ糖 AsSugOH)、およびアルセノリピド、  <b>標的動物:</b> げっ歯類 (毒性研究用)、ヒト (曝露評価用) ,  <b>ハザードの要約:</b>AsB: げっ歯類研究で有害影響は観察されず、in vitro で遺伝毒性なし,AsSugOH: マウスで認知機能と運動機能への潜在的影響,その他のアルセノ糖とアルセノリピド: データ不足,  <b>毒性に関する情報:</b>AsB: 参照点を導出するには不十分なデータ,AsSugOH: マウスでの認知および運動効果に基づく BMDL10 から、0.85 mg As/kg 体重/日の参照点,  <b>曝露に関する情報:</b>AsB: 幼児で最高 95 パーセントイル曝露が 12.5 <math>\mu\text{g As/kg}</math> 体重/日,AsSugOH: 赤海藻を消費する成人で最高 95 パーセントイル曝露が 0.71 <math>\mu\text{g As/kg}</math> 体重/日,  <b>他の組織からの評価情報:</b> なし,  <b>評価結果:</b>AsB: 現在の食事曝露レベルでは健康上の懸念なし,AsSugOH: 健康上の懸念なし (MOE &gt; 1000) ,その他のアルセノ糖とアルセノリピド: データ不足のため結論なし,  <b>評価の限界または将来の課題:</b>AsB の参照点を導出するには毒性データが不十分,アルセノリピドのリスク特性評価のためのデータ不足,AsSugOH 以外のアルセノ糖については結論不可能,不確実性の定性的考慮により、AsB と AsSugOH の食事曝露は健康上の懸念を引き起こす可能性は低い</p>
GPT-4o	<p>### 抽出および整理された情報:  <b>**対象物質:**</b>,アルセノベタイン (AsB),グリセロールアルセノシュガー (AsSugOH),その他のアルセノシュガー,アルセノリピド,  <b>**対象動物:**</b>,齧歯類 (AsB の反復投与毒性試験に使用),マウス (AsSugOH の認知および運動機能に関する研究から参照ポイントを導出するために使用),</p>

**\*\*ハザード概要:\*\*** **\*\*アルセノベタイン (AsB):\*\*** 齧歯類での反復投与毒性試験で有害影響は観察されず。\*in vitro\* 試験で遺伝毒性なし。人間の研究でも有害な結果との関連性は認められない。 **\*\*グリセロールアルセノシュガー (AsSugOH):\*\*** マウスで特定の用量で認知および運動機能への影響が観察された。 **\*\*その他のアルセノシュガー:\*\*** データ不足のため結論は得られなかった。 **\*\*アルセノリピッド:\*\*** データ不足のためリスク評価は実施できなかった。 ,

**\*\*毒性に関する情報:\*\*** **\*\*アルセノベタイン (AsB):\*\*** 参照ポイント (RP) を導出するには毒性データが不十分。齧歯類での反復投与毒性試験や\*in vitro\* 試験で有害影響や遺伝毒性は観察されなかった。 **\*\*グリセロールアルセノシュガー (AsSugOH):\*\*** マウスの認知および運動機能に関するベンチマーク用量下限信頼限界 (BMDL10) 値に基づき、0.85 mg As/kg 体重/日という参照ポイント (RP) が導出された。 ,

**\*\*曝露に関する情報:\*\*** **\*\*アルセノベタイン (AsB):\*\*** 幼児における 95 パーセントイルの最高曝露量は 12.5  $\mu$ g As/kg 体重/日 (元素ヒ素として表記)。 **\*\*グリセロールアルセノシュガー (AsSugOH):\*\*** 赤海藻 (ノリ/ラバー) を消費する成人における 95 パーセントイルの最高曝露量は 0.71  $\mu$ g As/kg 体重/日 (元素ヒ素として表記)。 ,

**\*\*他の組織からの評価情報:\*\*** なし、

**\*\*評価結果:\*\*** **\*\*アルセノベタイン (AsB):\*\*** 現在の食事性曝露レベルでは健康上の懸念はない。 **\*\*グリセロールアルセノシュガー (AsSugOH):\*\*** 曝露マージン (MOE) が 1000 以上の場合、健康上の懸念はない。 **\*\*その他のアルセノシュガー:\*\*** 結論は得られなかった。 **\*\*アルセノリピッド:\*\*** データ不足のためリスク評価は実施できなかった。 ,

**\*\*評価の限界または将来の課題:\*\*** アルセノベタイン (AsB) の参照ポイントを導出するには毒性データが不十分。 ,アルセノリピッドのリスク評価にはデータが不足している。 ,その他のアルセノシュガーについてはデータ不足のため結論は得られなかった。

#### Gemini 1.5 Pro

**\*\*対象物質:\*\*** 食品中の複合有機ヒ素種。主にアルセノベタイン (AsB)、アルセノ糖 (特にグリセロールアルセノ糖 -AsSugOH)、およびアルセノ脂質を含む。 ,

**\*\*対象動物:\*\*** げっ歯類 (AsSugOH 毒性試験では特にマウス)、ヒト (曝露および疫学研究)。 ,

**\*\*有害性概要:\*\*** AsB はげっ歯類試験で有害作用を示さず、in vitro で遺伝毒性も示さない。AsSugOH は、マウスの認知および運動機能への影響に基づいて導出された参照点を持つ。アルセノ脂質は有害性評価のためのデータがない。 ,

**\*\*毒性情報:\*\*** AsB : げっ歯類での 2 回の反復投与毒性試験で有害作用なし、in vitro で遺伝毒性なし。AsSugOH : マウスの認知および運動機能について導出された BMDL10 値を用いて、参照点 (RP) 0.85 mg As/kg 体重/日を確立。アルセノ脂質 : データなし。 ,

**\*\*曝露情報:\*\*** AsB : 幼児における曝露量の 95 パーセントイル値は 12.5  $\mu$ g As/kg 体重/日と推定。AsSugOH : 紅藻類 (海苔) を摂取する成人における曝露量の 95 パーセントイル値は 0.71  $\mu$ g As/kg 体重/日と推定。 ,

**\*\*他機関による評価情報:\*\*** なし、

**\*\*評価結果:\*\*** AsB : 食事による曝露は健康上の懸念を引き起こさない。AsSugOH : MOE > 1000 であり、現在の曝露レベルでは健康上の懸念はない。アルセノ脂質 : データ不足のためリスク評価は不可能。全体として、不確実性を考慮すると、AsB および AsSugOH への食事曝露は健康上の懸念ではない可能性が高い。他のアルセノ糖については結論を導き出すことができなかった。 ,

**\*\*評価の限界または今後の課題:\*\*** アルセノ脂質のデータ不足によりリスク評価ができない。AsSugOH 以外の他のアルセノ糖に関するデータが必要。

項目別抽出との主な違いを下記にまとめた。

- Claude 3.5 Sonnet
  - ハザードの概要が抽出されている。
  - 評価結果が抽出されている。
- GPT-4o
  - 対象動物でヒトが抽出されていない。
  - 毒性に関する情報が、項目別ではほぼ全訳であったが要約されている。
- Gemini 1.5 Pro

- 評価結果にあった参照点(RP)の情報が消えている。

### 振り返り

項目別抽出とまとめ抽出では、同じ項目でも抽出した内容が異なることがある。また項目別抽出は長くなる傾向が見られた。実際に人が読んで分かりやすいのはスクレイピングで付与している要約であるとの意見もあった。

#### 1.3.6.1.4. EFSA chemical contaminants in food and feed の full article からの抽出

##### 実施内容

スクレイピングで取得する記事は「News」及び「Scientific Opinion」の「Abstract」であるため情報量が少なく、「食品安全関係情報データベース」に登録する記事としては全文の翻訳になることが予想される。そこで「Scientific Opinion」の9つについて EFSA Journal の「Full article」の PDF ファイルをダウンロードし、これをマークダウン形式にテキスト化した記事に対して 1.3.6.1.3 節の抽出を行った。

### 結果

この結果を「4\_ChemicalContaminants.FullArticle.xlsx」にまとめた。Claude 3.5 Sonnet v1 では1つのドキュメント、GPT-4o は3つのドキュメントがトークン数を超えたため回答を得ることが出来なかった。

### 振り返り

結果を評価するには元の PDF ファイルの内容を把握しておく必要がある。PDF ファイルは23ページから146ページであり、抽出項目の評価方法の検討が必要であろう。例えばエキスパートが抽出した正解データを100報程度用意しておき、抽出の精度評価を自動的に行える状況を構築することが考えられる。

#### 1.3.6.1.5. ハルシネーションの判定

##### 課題

記事から指定した項目の情報をデータベース化することにより、今後高度な解析作業を容易にすることが出来ると考えられる。LLM を用いて指定した項目の情報抽出を行った場合、記事に情報が無いにも関わらず別の回答をするハルシネーションが課題となっている。

##### 実施内容

近年、推論機能を持つ LLM が提供されており、ハルシネーションを判定できる可能性がある。そこで LLM を使って抽出した情報がハルシネーションかどうかを LLM に判定させることを試みた。

Claude 3.5 Sonnet v1 (anthropic.claude-3-5-sonnet-20240620-v1:0)で農薬評価書の文書断片から発がん性の有無を回答させる実験から得られたハルシネーション例(指定した農薬の発がん性の情報が無いにも関わらず、発がん性「無し」と答えた場合)を表 1-24 に示した。最後の「フロルピラウキシフェンベンジル」の場合は本文中に「本剤に発がん性は認められない」とあるが、この情報は「フェノチオカルブ」であるため NA が正解となる例である。この回答について図 1-40 に示した質問文を LLM に投げてハルシネーションの判定できるのかを試した。

表 1-24 Claude 3.5 Sonnet v1 で与えた文書断片に情報がなく「NA」と回答するべき場合に「無」と回答した時の与えた文書断片

ID	農薬名	専門家抽出	LLM 抽出
PN_170	チアジニル	有	無
<p>評価に用いた試験成績は、動物体内運命（ラット、ヤギ及びニワトリ）、植物体内運命（水稻）、作物等残留、急性毒性（ラット、ウサギ）、亜急性毒性（ラット、イヌ）、慢性毒性（イヌ）、慢性毒性/発がん性併合（ラット）、<b>発がん性（マウス）</b>、2世代繁殖（ラット）、発生毒性（ラット、ウサギ）、遺伝毒性等である。各種毒性試験結果から、チアジニル投与による影響は、主に<b>肝臓（重量増加、肝細胞肥大等）及び腎臓（尿管上皮空胞化等）に認められた。</b></p>			
PN_203	トリフロキシストロビン	無	無
<p>評価に用いた試験成績は、動物体内運命（ラット、ヤギ及びニワトリ）、植物体内運命（りんご、小麦等）、作物等残留、急性神経毒性（ラット）、亜急性毒性（ラット及びイヌ）、慢性毒性（イヌ）、慢性毒性/発がん性併合（ラット）、<b>発がん性（マウス）</b>、2世代繁殖（ラット）、発生毒性（ラット及びウサギ）、遺伝毒性、免疫毒性等である。各種毒性試験結果から、トリフロキシストロビン投与による影響は、主に<b>肝臓（肝細胞肥大等）に認められた。</b></p>			
PN_217	ピコキシストロビン	有	無
<p>各種毒性試験結果から、ピコキシストロビン投与による影響は、主に体重（増加抑制）、肝臓（肝細胞肥大：マウス）及び十二指腸（粘膜過形成及び粘液腺拡張：マウス）に認められた。神経毒性、繁殖能に対する影響、催奇形性、生体にとって問題となる遺伝毒性及び免疫毒性は認められなかった。</p>			
PN_263	フェンブコナゾール	有	無
<p>評価に用いた試験成績は、動物体内運命（ラット、ヤギ及びニワトリ）、植物体内運命（小麦、らっかせい等）、作物等残留、亜急性毒性（ラット、マウス及びイヌ）、慢性毒性（イヌ）、慢性毒性/発がん性併合（ラット）、<b>発がん性（マウス）</b>、2世代繁殖（ラット）、発生毒性（ラット及びウサギ）、遺伝毒性等である。各種毒性試験結果から、フェンブコナゾール投与による影響は、主に<b>肝臓（肝細胞肥大及び空胞化等）に認められた。催奇形性及び遺伝毒性は認められなかった。</b></p>			
PN_278	フルオピラム	有	無
<p>各種毒性試験結果から、フルオピラム投与による影響は、主に眼（ラット：角膜混濁、網膜退色等）、肝臓（重量増加、小葉中心性肝細胞肥大等）、腎臓（重量増加、慢性腎症等）及び甲状腺（ろ胞上皮細胞肥大等）に認められた。神経毒性、繁殖能に対する影響、免疫毒性及び遺伝毒性は認められなかった。（参照 1、45、68）13. 遺伝毒性試験フルオピラム原体の細菌を用いた復帰突然変異試験、チャイニーズハムスター肺由来細胞（V79）を用いた染色体異常試験及び Hprt 遺伝子座突然変異試験、マウスを用いた in vivo 小核試験が実施された。結果は表 43 に示されているとおり全て陰性であったことから、フルオピラムに遺伝毒性はないものと考えられた。（参照 1、46～50、68）51 表 43 遺伝毒性試験概要（原体）試験対象処理濃度・投与量結果 invitro 復帰突然</p>			
PN_321	フロルピラウキシフェンベンジル	無	無
<p>しかしながら、マウス骨髄での小核の誘発は低体温に起因する可能性もあり、いずれにせよ、<b>本剤に発がん性は認められない</b>ことから、これらの遺伝毒性陽性反応は発がん性と無関係であると考えられ、ADI 及び ARfD の設定は可能と考えられた。各種試験結果から、農産物中の暴露評価対象物質を<b>フェノチオカルブ（親化合物のみ）</b>と設定した。</p>			

下記の問い合わせを行った際の回答が"無"でした。この回答が正しいかどうか判断してください。

[参考]部分の情報は PDF から markdown に変換したドキュメントです。PDF からの文章の抽出は前後が入れ替わることや誤字脱字、途中で切れていることがあるため、内容を予測しながら質問に回答してください。

[質問]

{chemical}の発がん性の有無を"有"もしくは"無"で答えてください。プロンプト中に答えがなけ

れば NA とのみ答えてください。補足情報等は回答に加えないでください。

[参考]

{rag\_context}

図 1-40 ハルシネーション判定のプロンプト。{chemical}には農薬名、{rag\_context}には農薬評価書から得られた文書断片に置き換えて LLM に問い合わせる。

### 結果

図 1-40 の質問に対して、間違いと指摘した場合、及び「NA」と回答した場合を正解としたときの各 LLM の正解の回数を表 1-25 に示した。LLM は問い合わせる度に回答が異なるため 3 回同じ質問を繰り返している。全ての回答結果を「Hallucination 判定回答.xlsx」に示した。一部に回答が間違っていると指摘していながら「無」と回答している場合や回答が正しいと述べながら「NA」と答えている場合があった(いずれも不正解にカウントした)。推論機能のある Gemini 2.0 Flash Thinking や GPT-o1 ではハルシネーションと判定出来ている場合もあり、特に Gemini 2.0 Flash Thinking はかなりの頻度でハルシネーションを判定出来ていた。

表 1-25 LLM 毎にハルシネーションと同定した回数。同じ質問を 3 回行っている。

	PN_170	PN_203	PN_217	PN_263	PN_278	PN_321	合計
Gemini 2.0 Flash	0	0	0	0	0	0	0
Gemini 2.0 Flash Thinking	3	3	3	2	3	1	15
Gemini 2.0 Pro	0	3	0	0	0	0	3
GPT-4o	0	0	0	0	1	0	1
GPT-o1	1	0	3	3	3	0	10
GPT-o3-mini	0	0	1	0	3	0	4

### 振り返り

LLM の推論機能の進展により、ハルシネーションを判定できるようになってきた。複数の LLM の組み合わせや、同じ質問を何度か行った際の多数決を取るなどにより精度を高められる可能性がある。項目の抽出についても推論機能のある LLM で行うことにより、プロンプトを最適化することで高い精度で実現できる可能性があると思われる。

#### 1.3.6.2. ラベル情報出力機能(csv ファイル)

記事情報出力機能に含まれる(1.1.3.1.3 節参照)。

#### 1.3.6.3. 文章相同性によるカテゴリ付与機能

スクレイピングにおいて文書のカテゴリ情報が付与されるため、構築する必要がなくなった。

### 1.3.7. 記事情報参照機能

#### 1.3.7.1. 記事検索機能

##### 1.3.7.1.1. スクレイピング記事検索機能の実装

#### 実施内容

スクレイピング記事一覧画面に記事の絞り込み及び検索機能を追加した。追加した機能は

1. 期間絞り込み
2. 機関名絞り込み
3. タイトル検索
4. 本文検索

である。タイトルと記事本文検索では「完全一致検索」もしくは「あいまい検索」を指定することができ、キーワード検索欄にスペースで区切って検索文字列入力することにより AND 検索もしくは OR 検索を指定して実行することが出来る。

## 結果

スクレイピング記事一覧画面に検索機能を追加し(図 1-41 参照)、絞り込んだ記事データをダウンロードすることが可能になった。

記事表示	記事作成	スクレイピング条件URL	記事作成日時	データ形式	探索深度	ステータス	対象URL/タイトル
🔍	🔍	https://www.efsa.europa.eu/en/publications?item_per_page=30	2025/02/20	html	1	成功	Assessment of the feed additive consisting of carthaxanthin (Carophy® Red 10%) for breeder hens for the renewal of its authorisation (DSM Nutritional Products Ltd.)   EFSA
🔍	🔍	https://www.efsa.europa.eu/en/publications?item_per_page=30	2025/02/20	html	1	成功	Assessment of the feed additive consisting of Lactiplantibacillus plantarum CECT 4526 for all animal species for the renewal of its authorisation (Centro Sperimentale del Latte S.r.l.)   EFSA
🔍	🔍	https://www.efsa.europa.eu/en/publications?item_per_page=30	2025/02/20	html	1	成功	Assessment of the feed additive consisting of choline chloride for all animal species for the renewal of its authorisation (Andrés Peralta S.A., Tamnco B.V., Bakchem Itaka S.r.l.)   EFSA
🔍	🔍	https://www.efsa.europa.eu/en/topics/topic/cross-cutting-science	2025/02/20			失敗	https://www.efsa.europa.eu/en/events/123id/penary-meeting-scientific-committee
🔍	🔍	https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging	2025/02/20	html	1	成功	Safety evaluation of the food enzyme baculolysin from the genetically modified Bacillus amyloliquefaciens strain NPR   EFSA
🔍	🔍	https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging	2025/02/20	html	1	成功	Reporting guidance for use-level data on food additives and food flavourings - 2025   EFSA
🔍	🔍	https://www.efsa.europa.eu/en/topics/topic/pesticides	2025/02/20	html	1	成功	Evaluation of confirmatory data following the Article 12 MRL review for lerbuthylazine   EFSA
🔍	🔍	https://www.epa.gov/newsreleases/search	2025/02/20	html	1	成功	EPA Administrator Zeiden's Message to Agency Staff: 'It's Time to Return to the Office'   US EPA
🔍	🔍	https://www.epa.gov/newsreleases/search	2025/02/20	html	1	成功	EPA Announces \$147M WIFIA Loan to Increase Drinking Water Reliability in South Florida   US EPA
🔍	🔍	https://www.epa.gov/newsreleases/search	2025/02/20	html	1	成功	EPA Announces \$110-Million WIFIA Loan to Enhance Drinking Water Reliability in Drought-Prone Region of Utah   US EPA

CSV Download

図 1-41 記事一覧画面に検索機能を追加した画面

## 振り返り

本番システムでは、元記事、英訳記事、日本語訳記事、英語要約、日本語要約、LLM 抽出キーワードについても検索機能を追加することが検討されると思われる。この時、長文の元記事を別ファイルとして保存している場合には、全文検索システムを別に導入などの検討が必要になるだろう。

### 1.3.7.1.2. スクレイピング記事検索機能の追加

#### 実施内容

検索対象をタイトル、記事本文、作成記事本文に変更した。

#### 結果

検索対象をタイトル、記事本文、作成記事本文に変更し、この画面図 1-42 に示した。

記事作成	スクレイピング条件URL	記事作成日時	データ形式	探索深度	ステータス	対象URLタイトル
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging">https://www.efsa.europa.eu/en/topics/topic/food-ingredients-and-packaging</a>	2025/03/18			失敗	
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/calls/data">https://www.efsa.europa.eu/en/calls/data</a>	2025/03/17	html	1	成功	<a href="#">Open call for food additives analytical and use level data in food and beverages intended for human consumption   EFSA</a>
<a href="#">🔗</a>		2025/03/17			失敗	
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/calls/data">https://www.efsa.europa.eu/en/calls/data</a>	2025/03/15	html	1	成功	<a href="#">Open call for food additives analytical and use level data in food and beverages intended for human consumption   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/publications?items_per_page=30">https://www.efsa.europa.eu/en/publications?items_per_page=30</a>	2025/03/15	html	1	成功	<a href="#">Candidatus Liberibacter africanus Candidatus Liberibacter americanus Candidatus Liberibacter asiaticus Pest Report to support the ranking of EU candidate priority pests   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/calls/data">https://www.efsa.europa.eu/en/calls/data</a>	2025/03/14	html	1	成功	<a href="#">Call for Data: Length of Incubation Periods of Xylella fastidiosa Subspecies in Infected Plants   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/news?items_per_page=36">https://www.efsa.europa.eu/en/news?items_per_page=36</a>	2025/03/14	html	1	成功	<a href="#">Have your say on EFSA's performance – deadline 1 April 2025   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/publications?items_per_page=30">https://www.efsa.europa.eu/en/publications?items_per_page=30</a>	2025/03/14	html	1	成功	<a href="#">Bactrocera dorsalis Pest Report to support the ranking of EU candidate priority pests   EFSA</a>
<a href="#">🔗</a>	<a href="https://www.efsa.europa.eu/en/publications?items_per_page=30">https://www.efsa.europa.eu/en/publications?items_per_page=30</a>	2025/03/14	html	1	成功	<a href="#">Anastrepha ludens Pest Report to support the ranking of EU candidate priority pests   EFSA</a>

図 1-42 検索対象をタイトル、記事本文、作成記事本文に変更した画面

#### 振り返り

本番環境ではキーワードも検索対象に加えることが考えられる。

### 1.3.7.2. 検索結果表示機能

検索を実行し、検索結果が表示された画面を図 1-43 に示した。

ホーム クローリング スクレイピング お試し翻訳 専門用語辞書管理

**検索対象**  
 タイトル  
 記事原文  
 作成記事本文

**検索条件**  
 完全一致  
 あいまい検索

**検索タイプ**  
 AND  
 OR

**キーワード**

**期間絞り込み**  
 ~

**機関名絞り込み**  
 地域:   
 国・地域:   
 機関:

検索結果: 1件  
     
 ページに移動 Page 1 of 1 表示件数 10件

記事作成	スクレイピング条件URL	記事作成日時	データ形式	探索深度	ステータス	対象URLタイトル
	<a href="https://www.efsa.europa.eu/en/publications?items_per_page=30">https://www.efsa.europa.eu/en/publications?items_per_page=30</a>	2025/03/14	html	1	成功	Bactrocera dorsalis Pest Report to support the ranking of EU candidate priority pests   EFSA

[CSV Download](#)

図 1-43 スクレイピング記事一覧から検索を実行し、検索結果が表示された画面

### 1.3.8. クロス表表示機能

スクレイピング対象が EPA と EFSA のみであり、また期間が 2025 年 1 月～3 月のため、有効なクロス表を作成することが困難であることから、キーワード頻度表示機能を作成することになった。

#### 1.3.8.1. クロス表条件設定機能

スクレイピング対象が EPA と EFSA のみであり、また期間が 2025 年 1 月～3 月のため全期間を対象とすることにした。本番環境では期間を選択可能とし、必要に応じて機関のリストを指定できる機能を作成することが考えられる。

#### 1.3.8.2. クロス表結果表示機能

キーワード頻度表示を行った画面を図 1-44 に示した。

## トレンド表示

キーワード	回数
EFSA	228
EPA	134
risk assessment	95
European Commission	70
European Food Safety Authority	53
food safety	51
environmental protection	46
public health	40
dietary exposure	36
PFAS	36
plant health	34
food enzyme	29
pesticide	29
pesticides	27
scientific opinion	25
pest report	24
EU candidate priority pests	24
working group	23
quarantine pests	23
pest prioritisation	23

図 1-44 キーワード頻度表示機能画面

## 1.3.9. 印刷機能

本システムは web アプリケーションとして構築しているため、web ブラウザーの機能として印刷が可能である。