FoodOntoMapV2: Food Concepts Normalization Across Food Ontologies

Gorjan Popovski[1,2(✉)], Barbara Koroušić Seljak[1], and Tome Eftimov[1]

[1] Computer Systems Department, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
{gorjan.popovski,barbara.korousic,tome.eftimov}@ijs.si
[2] Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia

Abstract. Nowadays, the existence of several available biomedical vocabularies and standards play a crucial role in understanding health information. While there is a large number of available resources in the biomedical domain, only a limited number of resources can be utilized in the food domain. There are only a few annotated corpora with food concepts, as well as a small number of rule-based food named-entity recognition systems for food concept extraction. Additionally, several food ontologies exist, each developed for a specific application scenario. To address the issue of ontology alignment, we have previously created a resource, named FoodOntoMap, that consists of food concepts extracted from recipes. The extracted concepts were annotated by using semantic tags from four different food ontologies. To make the resource more comprehensive, as well as more representative of the domain, in this paper we have extended this resource by creating a second version, appropriately named FoodOntoMapV2. This was done by including an additional four ontologies that contain food concepts. Moreover, this resource can be used for normalizing food concepts across ontologies and developing applications for understanding the relation between food systems, human health, and the environment.

Keywords: Food data normalization · Food data linking · Food ontology · Food semantics

1 Introduction

"End hunger, achieve food security and improved nutrition and promote sustainable agriculture" is one of the sustainable development goals of the United Nations set to the target date of 2030 [19]. To achieve this ambitious goal, big data and AI technologies can significantly contribute in the domain of global food and agricultural systems. A huge amount of work that has been done in biomedical predictive modelling [14,31] is enabled by the existence of diverse biomedical vocabularies and standards. Such resources play a crucial role in understanding biomedical information, as well as the sheer amount of biomedical data that is collected from numerous sources (e.g., drug, diseases, treatments, etc.).

The number of biomedical resources available to researchers is tremendous. This can pose an additional challenge when different data sets described by these resources

© Springer Nature Switzerland AG 2020
A. Fred et al. (Eds.): IC3K 2019, CCIS 1297, pp. 413–426, 2020.
https://doi.org/10.1007/978-3-030-66196-0_19

are to be fused to provide more accurate and comprehensive information than that provided by any individual data set. For example, the Unified Medical Language System (UMLS) is a system that brings and links together several biomedical vocabularies to enable interoperability between computer systems [3]. It maps these vocabularies and thus allows one to normalize the data among the various terminological systems. It additionally provides tools for natural language processing that are mainly used by systems developers in medical informatics. The UMLS system consists of over one million biomedical concepts and five million concept names, all of which stem from over 100 incorporated controlled vocabularies and classification systems. Some examples of the controlled vocabularies that are part of UMLS are CPT [2], ICD-10 [28], MeSH [9], SNOMED CT [10], and RxNorm [20].

In contrast to the biomedical domain, the food and nutrition domain is relatively low-resourced. For example, there is only one publicly available annotated corpus with food concepts, known as FoodBase [27], and there are few food named-entity recognition systems for the extraction of food and nutrient concepts [13,25]. In addition, the 0available food ontologies are developed for very specific use cases [4].

To address such shortcomings, we have created a resource that consists of food concepts extracted from recipes, named FoodOntoMap. Each food concept was assigned its corresponding semantic tags from from four food vocabularies (i.e. Hansard taxonomy and three food ontologies: FoodOn [15], OntoFood, and SNOMED CT [10]). With this, the resource provides a link between different food ontologies that can be further used to create embedding spaces for food concepts, as well as to develop applications for understanding the relation between food system, human health, and the environment [26].

In this paper, we have extended the FoodOntoMap resource by adding four additional vocabularies (i.e. ontologies: RCD [7], MeSH [9], SNMI [32], and NDDF [8]) that can be used for food data normalization. With the inclusion of these new vocabularies, the resource becomes more comprehensive, which can further assist in easier food data integration. We have appropriately named this extended resource FoodOntoMapV2.

2 Related Work

In this section, we are mainly going to discuss i) different ontologies that can be used for normalization of food concepts, and ii) food named-entity recognition methods, which are used for extracting food concepts from textual data. Additionally, we are going to provide an overview of food data normalization approaches that have already been published.

2.1 Food Semantic Resources

There are several ontologies that can be used for normalizing food concepts. Some of them are developed specifically for the food domain (i.e. FoodWiki [6], AGROVOC [5], Open Food Facts [4], Food Product Ontology [18], FOODS [30], and FoodOn [15]), while others are related to the biomedical domain, but also include links to food and environmental concepts (i.e. SNOMED CT [10], MeSH [9], SNMI [32], and NDDF [8]). A comprehensive review of these publicly available food ontologies was provided by Boulos et al. [4].

BioPortal is an open repository that consists of 816 different biomedical ontologies and services that enable access to and exploration of those ontologies [22]. Notwithstanding its definition as a repository for the biomedical domain, it contains ontologies that consist of food concepts as well.

The Hansard corpus is a collection of text and concepts created as a part of the SAMUELS project (2014–2016). One of its advantages is that it allows semantic searching of its data. More details about the semantic tags that it uses can be found in [1,29]. The concepts are organized in 37 higher level semantic groups, one of which is *Food and Drink* (i.e. AG).

Table 1 provides an overview of the most commonly used ontologies that can be explored for food data normalization, while in Table 2 their availability is presented. In our case, after manually exploring the ontologies available in BioPortal, we have concluded that the aforementioned ontologies are to be included in the final mapping methodology, as their domain is in accordance with the task and they contain a significant number of classes.

2.2 Food Information Extraction

Before the process of food data normalization can be done, we should mention that some food concepts can be presented as parts of unstructured textual data. Thus, the first step would be to perform food information extraction [25]. Information extraction (IE) is the task of automatically extracting structured information from unstructured data. In our case, the information that should be extracted are the food concepts. One way to do this is to use a named-entity recognition method (NER), whose main goal is to locate and classify the entity (concept) mentions in the unstructured data into predefined categories [21].

There is a tremendous amount of work that has been done regarding NERs for biomedical tasks, especially focusing on disease, drug, procedure entities and other similar concepts in the biomedical domain. However, the situation is completely different in the food and nutrition domain. There are only a few rule-based systems that can be used as well as some general tools that work in combination with available ontologies. A supervised machine learning (ML) model that can be used for food information extraction still does not exist, since the first annotated corpus with food entities was just recently published [27].

The UCREL Semantic Analysis System (USAS) is a framework for semantic text analysis. It can be used for named-entity recognition with 21 major classes. The class of interest to our application is "food and farming" [29]. However, one of its significant drawbacks is that it only works on a token level. To illustrate what this means, let us take the text "steamed broccoli" as an example which represents one food entity that should be extracted and annotated. In this case, the USAS tagger would extract and annotate the words "steamed" and "broccoli" as separate entities.

Moreover, there are also only two rule-based systems that can be used for extraction of food entities. One of them, titled drNER, focuses on extracting food and nutri-

Table 1. Overview of the resources that are used for the process of food data normalization.

Ontology	Description
FoodOn [15]	• Semantics for food safety, food security, agriculture and animal husbandry practices linked to food production, culinary, nutritional and chemical ingredients and processes
	• Its usage is related to research and clinical data sets in academia and government
OntoFood (OF)	• An ontology with SWRL rules regarding nutrition for diabetic patients
SNOMED CT [10]	• A standardized, multilingual vocabulary of clinical terminology that is used by physicians and other health care providers for electronic health records
	• Apart from the medical concepts that are the main focus of this ontology, it additionally contains the concept of *Food* that can be further used for food concept normalization
Read Codes, Clinical Terms Version 3 (RCD) [7]	• Available as a part of the UMLS system
MeSH [9]	• Thesaurus that is a controlled and hierarchically-organized vocabulary produced by the National Library of Medicine
	• It is used for indexing, cataloging, and searching the biomedical and health-related information
SNMI [32]	• A previous version of SNOMED CT
NDDF [8]	• Widely-known terminology regarding drugs, which combines a comprehensive set of drug elements, pricing and clinical information
	• Approved by the U.S. Food and Drug Administration (FDA)
	• It additionally consists of concepts related to herbals, nutraceutical and dietary supplements
Hansard corpus [1,29]	• A collection of concepts created as part of the SAMUELS project (2014–2016)
	• It allows semantically-based searches of its data
	• The concepts are organized in 37 higher level semantic groups, one of which is *Food and Drink* (i.e. AG)

ent concepts from evidence-based dietary recommendations [13]. Recently, this work was extended with the creation of FoodIE [25], which is another rule-based system for extracting food entities. FoodIE uses rules based on computational linguistics, which also combine and utilize the knowledge available in the Hansard corpus.

Another tool that can be used for information extraction, in general, is the NCBO Annotator [17]. It utilizes the ontologies that are part of the BioPortal [22]. The methodology leverages these ontologies to create annotations of the raw input text and returns them by adhering to semantic web standards.

Table 2. Availability of the resources that are used for the process of food data normalization.

FoodOn	https://bioportal.bioontology.org/ontologies/FOODON/?p=summary
OntoFood	https://bioportal.bioontology.org/ontologies/OF/?p=summary
SNOMED CT	https://bioportal.bioontology.org/ontologies/SNOMEDCT/?p=summary
RCD	https://bioportal.bioontology.org/ontologies/RCD/?p=summary
MeSH	https://bioportal.bioontology.org/ontologies/MESH/?p=summary
SNMI	https://bioportal.bioontology.org/ontologies/SNMI?p=summary
NDDF	https://bioportal.bioontology.org/ontologies/NDDF/?p=summary
Hansard corpus	https://www.hansard-corpus.org/

2.3 Food Data Normalization

Recently, food concept normalization poses an open research question that is highly researched by the food and nutrition science community. Food concepts that are available in unstructured data can be represented in various unstandardized ways, which simply depend on how people express themselves. It is always good practice to normalize the data in order to ease further analyses. This is a task where the same food concept, represented in different ways, should be mapped to the single corresponding food concept that exists in some food resource (e.g., taxonomy or ontology).

To propose a solution to this issue, StandFood [12] has recently been introduced. It is a semi-automatic system for classifying and describing foods according to a description and classification system. Specifically, it adheres to FoodEx2, which is proposed by the European Food Safety Agency (EFSA) [11]. It uses a combination of machine learning, methods from natural language processing, and probability theory to perform food concept normalization. Additionally, we have created a resource that consists of food concepts extracted from recipes, named FoodOntoMap. Each food concept was assigned its corresponding semantic tags from four food vocabularies (i.e. Hansard taxonomy and three food ontologies: FoodOn [15], OntoFood, and SNOMED CT [10]). With this, the resource provides a link between different food ontologies that can be further used to create embedding spaces for food concepts, as well as to develop applications for understanding the relation between food system, human health, and the environment [26]. To go one step further, we have developed a heuristic model based on lexical similarity and propose two new semantic similarity heuristics based on word embeddings [24]. Moreover, we have explored the LanguaL hierarchy [23], which is a standard used to describe foods, in order to see if different food concepts that are part of this hierarchy are linked together properly. To do this, we have trained a vector representation (i.e. embedding) for each food concept that is a part of the hierarchy and have found the most similar products for a subset of products. The results indicate that further efforts should be made to link all these standards together in order to provide a unified system for describing and standardizing food concepts.

3 Methodology

The methodology is an extension of the methodology used to build *FoodOntoMap*, presented in [26]. It is constructed by including an additional four ontologies taken from the BioPortal. With this extended pipeline the second version is created, appropriately named *FoodOntoMapV2*. The flowchart on Fig. 1 illustrates the extended methodology.

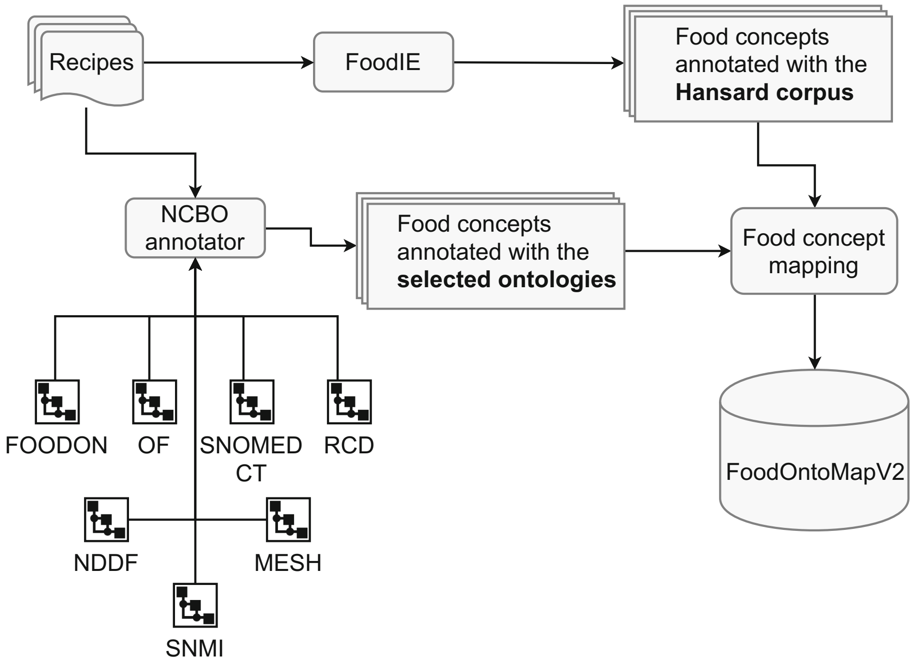

Fig. 1. Flowchart depicting the FoodOntoMapV2 methodology.

3.1 Data Collection

The data of interest is in the form of free-form text, i.e. recipe texts that are composed of i) lists of (raw or already treated) ingredients with amounts and units, and ii) descriptions of the preparation method with the specification of potential treatment methods and timing (i.e. time units and degrees). To provide a substantial amount and diversity of data, the recipes were taken from the most popular recipe sharing social media platform called AllRecipes [16]. More than 23,000 recipes were taken from the site, spanning across five different categories: “Appetizers and snacks”, “Breakfast and Lunch”, “Desserts”, “Drinks”, and “Dinners”. It is important to mention the difficulties while working with raw textual data. As it is free-form text, i.e simply natural language text, there is no predefined data representation (format) that is followed. This implicates that the sentence structures, vocabulary, and in essence the very way people express themselves vary significantly from text to text. This imposes a challenge in the pre-processing of such textual data.

3.2 Text Pre-processing

In order to provide some consistency in the recipe texts, a text pre-processing step is useful. With this step, some issues that arise due to the nature of free-form text are resolved.

In our case, the pre-processing methodology, originally a constituent of FoodIE, is done prior to running all of the NER methods. It consists of several steps:

1. Removing excess white-space characters.
2. Removing semantically irrelevant punctuation (i.e. quotation marks).
3. Substituting commonly occurring non-ASCII symbols (i.e. degree symbol(°)).
4. ASCII transliteration.
5. Standardizing fractions to decimal notation.

The pre-processing is described in more detail in [25].

3.3 Information Extraction and Annotation

NCBO Annotator with Selected Ontologies. By using the aforementioned NCBO annotator in conjunction with each ontology once, we obtain annotations for all the recipes. With this, we have seven different annotation sets, one per ontology. As each ontology is primarily constructed for various different purposes, each NCBO run provides us with unique semantic information regarding the annotated food concepts. It is important to note that not all ontologies are able to extract every mentioned food concept. Consequently, some recipe annotations are empty, which is also due to the insufficient domain coverage of the ontology in question.

Each concept extraction and annotation produced by the NCBO annotator is defined by its:

– ordinal number within the recipe text;
– *urls* - The semantic type(s) taken from the ontology;
– *text* - The textual representation of the food concept;
– *from* - The start position of the food concept in the recipe, as expressed in terms of characters from the beginning of the text;
– *to* - The end position of the food concept in the recipe, as expressed in terms of characters from the beginning of the text;
– *matchType* - The type of match that is found by the NCBO annotator.

An example of an annotated recipe can be seen in Table 3.

Table 3. An example recipe annotated by the NCBO annotator using the SNOMED CT ontology.

	SNOMED CT ID	Text	From	To	MatchType
1	226890008	MARGARINE	60	68	PREF
2	735030001	GARLIC	77	82	PREF
3	443701000124100	WORCESTERSHIRE SAUCE	90	109	PREF
4	227519005	SAUCE	105	109	PREF
5	227260004	PUMPKIN	115	121	PREF

FoodIE Method. As mentioned in Sect. 2.2, FoodIE is a rule-based system for extracting food entities. Since its rule-engine utilizes the semantic information in the Hansard corpus, each extraction is accompanied by its corresponding semantic tag from the Hansard corpus. With this, an annotation for the food concept is provided.

The FoodIE rule-based system is used to extract and annotate the food concepts in each recipe, in a similar a fashion to the usage of the NCBO annotator. This process produces one more annotation set, bringing the total to eight annotation sets containing semantic annotations for the concepts in each recipe, subject to the domain coverage of the used ontology.

Similarly to the previous method, each food concept annotation is defined by its:

- *annotation id* - The ordinal number within the recipe text;
- *offset* - The start position of the extracted food concept within the recipe, as expressed in terms of words (tokens) from the beginning of the text;
- *length* - The length of the textual representation of the food concept, as measured by its length in characters;
- *text* - The textual representation of the food concept;
- *semantic_tags* - The semantic tag(s) from the Hansard corpus which correspond to the food concept.

It is important to note that the positions of the extracted food concepts by FoodIE are measured differently compared to the NCBO annotator.

An example of an annotated recipe can be seen in Listing 1.1. This is the same convenient format used by FoodBase.

```
<document>
  <id>0recipe43</id>
  <infon key="category">Appetizers and snacks</infon>
  <infon key="full_text">
   Preheat oven to 275 degrees F (135 degrees C). Combine the margarine, salt,
   garlic salt, Worcestershire sauce and pumpkin seeds. Mix thoroughly
   and place in shallow baking dish. Bake for 1 hour, stirring occasionally.
  </infon>
  <annotation id="1">
    <location offset="15" length="9"/>
    <text>margarine</text>
    <infon key="semantic_tags">
      AG.01.f [Fat/oil];
    </infon>
  </annotation>
  <annotation id="2">
    <location offset="17" length="4"/>
```

```
    <text>salt</text>
    <infon key="semantic_tags">
      AG.01.1.01 [Salt];AG.01.w [Setting table];
    </infon>
  </annotation>
  <annotation id="3">
    <location offset="19" length="11"/>
    <text>garlic salt</text>
    <infon key="semantic_tags">
      AG.01.h.02.e [Onion/leek/garlic];AG.01.1.01 [Salt];AG.01.w [Setting table];
    </infon>
  </annotation>
  <annotation id="4">
    <location offset="22" length="20"/>
    <text>Worcestershire sauce</text>
    <infon key="semantic_tags">
      AG.01.h [Fruit and vegetables];AG.01.1.04 [Sauce/dressing];
    </infon>
  </annotation>
  <annotation id="5">
    <location offset="25" length="13"/>
    <text>pumpkin seeds</text>
    <infon key="semantic_tags">
      AG.01.h.02.f [Fruits as vegetables];
    </infon>
  </annotation>
</document>
```


Listing 1.1. The same recipe shown in Table 3, but annotated by the FoodIE method instead.

3.4 Datasets with Unique Extracted Food Concepts per Ontology

Before the food concept mapping can be done, it is useful to have a standardized representation for the food concepts found across all the different ontologies, as the semantic tags provided by each resource differ regarding their representation. After the process of extracting and annotating the recipes, all the unique food concepts extracted by each ontology can be condensed into eight simple datasets, i.e. one per ontology. Each code is represented by a single uppercase letter, followed by six digits (e.g. "A000832"). The digits are simply ordinal numbers, while each letter represents a semantic resource, namely:

A. FoodIE + Hansard corpus
B. FoodOn
C. SNOMED CT
D. OntoFood
E. Read Codes, Clinical Terms Version 3 (RCD)
F. Medical Subject Headings (MeSH)
G. Systematized Nomenclature of Medicine, International Version (SNMI)
H. National Drug Data File (NDDF)

With this, each food concept from each semantic resource has its own unique identifier by which it is represented in the final map. The number of unique food concepts per method is presented in Table 4.

Table 4. Total number of unique extracted entities per method.

Method	Number of unique extracted entities
FoodIE + Hansard corpus (A)	13111
FOODON (B)	1069
SNOMED CT (C)	583
OF (D)	111
RCD (E)	485
MESH (F)	105
SNMI (G)	42
NDDF (H)	54

3.5 Food Concept Mapping

After all the annotations have been done by using the different methods and semantic resources, the mapping process can be performed. The first thing that must be done here is to convert the positions of the food concepts into the same unit. In this case, all of the positions are converted as to be expressed in terms of words (tokens) from the beginning of the recipe text.

Now that all the positions are expressed in the same way, the mapping can be done. For each concept that is extracted and annotated by any method, we check if it is extracted and annotated by every other method. From here, three things are possible for each pair of methods:

1. The food concept is extracted and annotated by both methods.
   (a) The positions given by both methods are the same.
   - In this case, the food concepts' texts are double checked for robustness. If they are the same, a link is added in the map between the codes for both concepts. If the texts are different, an error is to be raised and the concepts are to be manually checked. However, as is expected, this did not happen when performing the mapping. Additionally, this step takes the different spelling variants of the words into account, e.g. "colour" is the same as "color".
   (b) The positions from one annotator lie within the positions from the other.
   - In this case, the food concept is only partially recognized by one of the methods. The encompassed food concept is mapped with the encompassing food concept. If several of these partial matches are present, then the semantic tags from each concept are mapped to the food concept that encompasses them. This implicates that a single food concept from one ontology can have multiple corresponding food concept mappings with another ontology. To illustrate this with an example let us consider the food concept "fruit juice". Here, one annotator might extract the concept as "fruit juice", while the other might extract two separate concepts: "fruit" and "juice". As the positions of these two concepts lie within the position of the food concept extracted by the first annotator, the semantic tags from both shorter concepts are mapped to the semantic tag of the single food concept.

2. One or both of the methods do not extract and annotate the food concept.
 - In this case, there exist no links between the food concepts, and therefore no mapping is performed.

The map consists of 1,398 instances, where each instance is a tuple of corresponding food concept codes. There are a total of eight columns, one for each semantic resource. If a food concept is not annotated by a specific semantic resource, then there is no corresponding food concept code from that ontology, and such missing data is filled with "NULL". An example entry, where all of the three possible matches occur, is presented in Table 5.

Table 5. An example of a FoodMapOntomapV2 instance which contains all types of matches.

Hansard	FOODON	SNOMEDCT	OF	RCD	MESH	SNMI	NDDF
A000630	B000066; B000022	C000018	NULL	E000018	F000007	G000002	H000003

In this example the codes represent the following food concepts:

– A000630 - "VEGETABLE OIL SPRAY"
– B000066 - "SPRAY"
– B000022 - "VEGETABLE OIL"
– C000018 - "VEGETABLE OIL"
– E000018 - "VEGETABLE OIL"
– F000007 - "VEGETABLE"
– G000002 - "VEGETABLE"
– H000003 - "VEGETABLE OIL"

Finally, the map was manually checked in order to ensure that no inconsistencies are present. The code used to perform this mapping is available at https://github.com/GorjanP/FOM_mapper_client, while the final map can be found at https://doi.org/10.5281/zenodo.3600619.

4 Discussion

Analyzing the data presented in Table 4, we can observe that the first method, which uses FoodIE and the Hansard corpus, provides the largest domain coverage of all the methods presented. This is due to the performance of the NER method FoodIE, which can extract previously unseen combinations of tokens that represent a food concept. Concretely, this means that the entities extracted and annotated by FoodIE do not have to be present as a whole in the semantic resource, as long as each token that is part of the food concept is correspondingly extracted and annotated.

Moreover, there is quite a large variation between the number of extracted food concepts among the other seven methods. As each one is built with a specific purpose in mind, it defines different points of interest that should be captured by the semantic information of the concepts it describes. However, with such a map between these different semantic resources a broader perspective is represented for each food concept that is found in it. This is especially useful regarding achieving interoperability, as well as working towards creating a definitive, universal, and standardized semantic resource for food concepts that could be used for any purpose. Such a target resource is reminiscent of the Unified Medical Language System (UMLS).

5 Conclusions

In this paper we have presented an extension of the FoodOntoMap methodology, which performs food concept normalization across different food ontologies, aptly named *FoodOntoMapV2*.

Each food ontology contains different semantic information regarding its constituent classes and concepts, as they are developed for a specific problem domain. Therefore, the process of food concept normalization across such ontologies presents an important task which aims at providing interoperability between the ontologies in the form of an ontology concept map. With this, each specific concept that is found in several ontologies is linked by this mapping.

Moreover, the same methodology can be applied in order to normalize data of any nature, provided there exist adequate semantic resources (e.g. ontologies) and a corresponding NER method.

Acknowledgements. This work was supported by the Ad Futura grant for postgraduate study; the Slovenian Research Agency Program P2-0098; and the European Union's Horizon 2020 research and innovation programme [grant agreement No 863059].

References

1. Alexander, M., Anderson, J.: The hansard corpus **1803–2003** (2012)
2. Association, A.M., et al.: CPT, 1989: Physician's Current Procedural Terminology. American Medical Association Press (1989)
3. Bodenreider, O.: The unified medical language system (UMLs): integrating biomedical terminology. Nucleic Acids Res. **32**(suppl-1), D267–D270 (2004)
4. Boulos, M.N.K., Yassine, A., Shirmohammadi, S., Namahoot, C.S., Brückner, M.: Towards an "internet of food": food ontologies for the internet of things. Future Internet **7**(4), 372–392 (2015)
5. Caracciolo, C., Stellato, A., Rajbahndari, S., Morshed, A., Johannsen, G., Jaques, Y., Keizer, J.: Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case. Int. J. Metadata Semant. Ontol. **7**(1), 65–75 (2012)
6. Çelik, D.: Foodwiki: Ontology-driven mobile safe food consumption system. Sci. World Jo. **2015** (2015)
7. Chen, Y., Lasko, T.A., Mei, Q., Denny, J.C., Xu, H.: A study of active learning methods for named entity recognition in clinical text. J. Biomed. Inform. **58**, 11–18 (2015)
8. First DataBank: National drug data file (NDDF) (2008)
9. Díaz-Galiano, M.C., García-Cumbreras, M.Á., Martín-Valdivia, M.T., Montejo-Ráez, A., Ureña-López, L.A.: Integrating MeSH ontology to improve medical information retrieval. In: Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.) CLEF 2007. LNCS, vol. 5152, pp. 601–606. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85760-0_76

10. Donnelly, K.: SNOMED-CT: The advanced terminology and coding system for eHealth. Stud. Health Technol. Inform. **121**, 279 (2006)
11. (EFSA), E.F.S.A.: The food classification and description system foodex 2 (revision 2). EFSA Supporting Publications **12**(5), 804E (2015)
12. Eftimov, T., Korošec, P., Koroušić Seljak, B.: Standfood: Standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2. Nutrients **9**(6), 542 (2017)
13. Eftimov, T., Koroušić Seljak, B., Korošec, P.: A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations. PLoS ONE **12**(6), e0179488 (2017)
14. Gligic, L., Kormilitzin, A., Goldberg, P., Nevado-Holgado, A.: Named entity recognition in electronic health records using transfer learning bootstrapped neural networks. arXiv preprint arXiv:1901.01592 (2019)
15. Griffiths, E.J., Dooley, D.M., Buttigieg, P.L., Hoehndorf, R., Brinkman, F.S., Hsiao, W.W.: Foodon: a global farm-to-fork food ontology. In: ICBO/BioCreative (2016)
16. Groves, S.: How allrecipes. com became the worlds largest food/recipe site. roi of social media (blog) (2013)
17. Jonquet, C., Shah, N., Youn, C., Callendar, C., Storey, M.A., Musen, M.: NCBO annotator: semantic annotation of biomedical data. In: International Semantic Web Conference, Poster and Demo Session, vol. 110 (2009)
18. Kolchin, M., Zamula, D.: Food product ontology: initial implementation of a vocabulary for describing food products. In: Proceeding of the 14th Conference of Open Innovations Association FRUCT, Helsinki, Finland, pp. 11–15 (2013)
19. Lartey, A.: End hunger, achieve food security and improved nutrition and promote sustainable agriculture. UN Chronicle **51**(4), 6–8 (2015)
20. Liu, S., Ma, W., Moore, R., Ganesan, V., Nelson, S.: RXNORM: prescription for electronic drug information exchange. IT Prof. **7**(5), 17–23 (2005)
21. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1), 3–26 (2007)
22. Noy, N.F., et al.: Bioportal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res. **37**(suppl-2), W170–W173 (2009)
23. Pennington, J.A., Smith, E.C., Chatfield, M.R., Hendricks, T.C.: Langual: a food-description language. Terminology. Int. J. Theoret. Appl. Issues Special. Commun. **1**(2), 277–289 (1994)
24. Popovski, G., Ispirova, G., Hadzi-Kotarova, N., Valenčič, E., Eftimov, T., Koroušić Seljak, B.: Food data integration by using heuristics based on lexical and semantic similarities. In: Proceedings of the 13th International Conference on Health Informatics (2020, in Press)
25. Popovski, G., Kochev, S., Koroušić Seljak, B., Eftimov, T.: Foodie: a rule-based named-entity recognition method for food information extraction. In: Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods, (ICPRAM 2019), pp. 915–922 (2019)
26. Popovski, G., Koroušić Seljak, B., Eftimov, T.: FoodOntoMap: linking food concepts across different food ontologies. In: Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - Volume 2: KEOD, pp. 195–202. INSTICC, SciTePress (2019). https://doi.org/10.5220/0008353201950202
27. Popovski, G., Seljak, B.K., Eftimov, T.: FoodBase corpus: a new resource of annotated food entities. Database 2019, November 2019. https://doi.org/10.1093/database/baz121
28. Quan, H., et al.: Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data. Medical care pp. 1130–1139 (2005)
29. Rayson, P., Archer, D., Piao, S., McEnery, A.: The UCREL semantic analysis system (2004)

30. Snae, C., Brückner, M.: Foods: a food-oriented ontology-driven system. In: 2nd IEEE International Conference on Digital Ecosystems and Technologies, DEST 2008 pp. 168–176. IEEE (2008)
31. Wang, Q., Zhou, Y., Ruan, T., Gao, D., Xia, Y., He, P.: Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition. J. Biomed. Inform. 103133 (2019)
32. Wang, Y.: An empirical evaluation of semantic similarity measures using the wordnet and UMLS ontologies. Ph.D. thesis, Miami University (2005)