

Mapping of ImageNet and Wikidata for Knowledge Graphs Enabled Computer Vision

Dominik Filipiak¹[\[https://orcid.org/0000-0002-4927-9992\]](https://orcid.org/0000-0002-4927-9992), Anna Fensel^{1,2}[\[https://orcid.org/0000-0002-1391-7104\]](https://orcid.org/0000-0002-1391-7104), and Agata Filipowska³[\[https://orcid.org/0000-0002-8425-1872\]](https://orcid.org/0000-0002-8425-1872)

¹Semantic Technology Institute (STI) Innsbruck, Department of Computer Science, University of Innsbruck, Austria

²Wageningen University & Research, The Netherlands

³Department of Information Systems, Poznań University of Economics and Business, Poland

Abstract. Knowledge graphs are used as a source of prior knowledge in numerous computer vision tasks. However, such an approach requires to have a mapping between ground truth data labels and the target knowledge graph. We linked the ILSVRC 2012 dataset (often simply referred to as ImageNet) labels to Wikidata entities. This enables using rich knowledge graph structure and contextual information for several computer vision tasks, traditionally benchmarked with ImageNet and its variations. For instance, in few-shot learning classification scenarios with neural networks, this mapping can be leveraged for weight initialisation, which can improve the final performance metrics value. We mapped all 1 000 ImageNet labels – 461 were already directly linked with the *exact match* property (P2888), 467 have exact match candidates, and 72 cannot be matched directly. For these 72 labels, we discuss different problem categories stemming from the inability of finding an exact match. Semantically close non-exact match candidates are presented as well. The mapping is publicly available at <https://github.com/DominikFilipiak/imagenet-to-wikidata-mapping>.

Keywords: ImageNet, Wikidata, mapping, computer vision, knowledge graphs

Introduction

Thanks to deep learning and convolutional neural networks, the field of computer vision experienced rapid development in recent years. ImageNet (ILSVRC 2012) is one of the most popular datasets used for training and benchmarking models in the classification task for computer vision. Nowadays, an intense effort can be observed in the domain of few- [17] or zero-shot learning [16] which copes with various machine learning tasks, for which training data is very scarce or even non-available. More formally, N -way K -shot learning considers a setting, in which there are N categories with K samples to learn from (typically $K \leq 20$ in few-shot learning). This is substantially harder from standard settings, as deep learning models usually rely on a large number of samples provided. One of the approaches to few-shot learning considers relying on some prior knowledge, such as the class label. This can be leveraged to improve the performance of the task. For instance, Chen et al. [4] presented Knowledge Graph Transfer Network, which uses the adjacency matrix built from knowledge graph correlations in order to create class prototypes in a few-shot learning classification. More generally, knowledge-embedded machine learning systems can use knowledge graphs as a source of information for improving performance metrics for a given task. One of these knowledge graphs is Wikidata [15], a popular collaborative knowledge graph.

Our main research goal concentrates on facilitating general-purpose knowledge graphs enabled computer vision methods, such as the aforementioned knowledge graph transfer network. In this paper, we provide a mapping between ImageNet classes and Wikidata entities, as this is the first step to achieve this goal. Our paper is inspired by and built on top of the work of Nielsen [12] – he first explored the possibility of linking ImageNet WordNet synsets with Wikidata. We also aim at providing detailed explanations for our choices and compare the results with these provided by Nielsen. Our publicly available mapping links WordNet synset used as ImageNet labels with Wikidata entities. It will be useful for the aforementioned computer vision tasks. Practical usage scenarios consider situations in which labelling data is a costly process and the considered classes can be linked to a given graph (that is, for few- or zero-shot learning tasks). However, simpler tasks, such as classification, can also use context knowledge stemming from rich knowledge graph structure (in prototype learning [18], for instance).

The remainder of this paper is structured as follows. In the next section, we briefly discuss related work. Then, in the third section, we provide detailed explanations about the mapping process, which is focused on the entities which do not have a perfect match candidate. The next section provides some analytics describing the mapping, as well as a comparison with automated matching using a NERD tool – *entity-fishing* [7]. The paper is concluded with a summary. Most importantly, the mapping is publicly available¹.

Background and Related Work

To provide a mapping between ILSVRC 2012 and Wikidata, it is necessary to define some related terms first. This requires introducing a few additional topics, such as WordNet, since some concepts (such as structure) in ILSVRC 2012 are based on the former. This section provides a comprehensive overview of these concepts. We also enlist the existing literature on the same problem of finding this specific mapping. To the best of our knowledge, there were only two attempts to achieve this goal – both are described below.

WordNet is a large lexical database of numerous (primarily) English words [11]. Nouns and verbs have a hierarchical structure and they are grouped altogether as *synsets* (sets of synonyms) in WordNet. Historically, this database paid a significant role in various pre-deep learning era artificial intelligence applications (it is still used nowadays, though). ImageNet [5] is a large image database, which inherited its hierarchical structure from WordNet. It contains 14,197,122 images and 21841 WordNet-based synsets at the time of writing, which makes it an important source of ground-truth data for computer vision. ImageNet Large Scale Visual Recognition Challenge (abbreviated as ILSVRC) [13] was an annual competition for computer vision researchers. The datasets released each year (subsets of original ImageNet) form a popular benchmark for various tasks to this day. The one released at ILSVRC 2012 is particularly popular and commonly called ImageNet² up to this date. It gained scientific attention due to the winning architecture AlexNet [9], which greatly helped to popularise deep learning. ImageNet is an extremely versatile dataset – architectures coping with it usually have been successful with different datasets as well [2]. Models trained on ImageNet are widely used for transfer learning purposes [8].

Launched in 2012, Wikidata [15] is a collaborative knowledge graph hosted by Wikimedia Foundation. It provides a convenient SPARQL endpoint. To this date, it is an active project and it is an important source of information for e.g. Wikipedia articles. Due to its popularity, size, and ubiquity, Wikidata can be considered as one of the most popular and successful knowledge graph instances along with DBpedia [1] and Freebase-powered [2] Google Knowledge graph. Given the recent interest in the ability to leverage external knowledge in computer vision tasks [4], it would be therefore beneficial to map ImageNet classes to the correspond-

¹<https://github.com/DominikFilipiak/imagenet-to-wikidata-mapping>

²From now on, we will refer to ILSVRC 2012 dataset as simply ImageNet, unless directly stated otherwise.

ing Wikidata entities. The idea itself is not new, though the full mapping was not available to this date. To the best of our knowledge, Nielsen [12] was the first to tackle this problem. He summarised the encountered problems during preparing the mapping and classified them into few categories. These categories include missing synsets on Wikidata, matching with a disambiguation page, discrepancies between ImageNet and WordNet, differences between the WordNet and the Wikidata concepts with similar names, and multiple semantically similar items in WordNet and Wikidata. Nielsen described his effort in detail, though the full mapping was not published. Independently, Edwards [6] tried to map DBpedia and Wikidata to ImageNet (in a larger sense, not ILSVRC 2012) using various pre-existing mappings and knowledge graph embeddings methods, such as TransE [3], though the results of such mapping have not been published as well. Contrary to these papers, we publish our mapping.

Mapping

This section is devoted to the mapping between the ImageNet dataset and Wikidata. First, we explain our approach in order to provide such mapping. Then, we identify and group key issues, which occurred in the mapping process. We also provide more detailed explanations for the most problematic entities.

To prepare the mapping, we follow the approach and convention presented by Nielsen. Namely, we use synset names from WordNet 3.0 (as opposed to, for example, WordNet 3.1). That is, we first check the `skos:exactMatch` (P2888) property in terms of an existing mapping between Wikidata entities and WordNet synsets. This has to be done per every ImageNet class. For example, for the ImageNet synset `n02480855` we search for P2888 equal to <http://wordnet-rdf.princeton.edu/wn30/02480855-n> using Wikidata SPARQL endpoint. Listing 1 provides a SPARQL query for this example.

```

1 SELECT *
2 WHERE {
3   ?item wdt:P2888 ?uri.
4   FILTER STRSTARTS(STR(?uri),
5     ↪ "http://wordnet-rdf.princeton.edu/wn30/02480855-n").
6 }

```

Listing 1. Matching WordNet with Wikidata entities using SPARQL.

As of November 2020, there are 461 already linked synsets out of 1000 in ImageNet using `wdt:P2888` property. For the rest, the mapping has to be provided. Unlike Edwards [6], we do not rely on automated methods, since the remaining 539 entities can be checked by hand (although we test one of them in the next section). Using manual search on Google Search, we found good `skos:exactMatch` candidates for the next 467 ImageNet classes. These matches can be considered to be added directly to Wikidata, as they directly reflect the same concept. For the vast majority of the cases, a simple heuristics was enough – one has to type the synset name in the search engine, check the first result on Wikipedia, evaluate its fitness and then use its Wikidata item link. Using this method, one can link 928 classes in total (with 467 entities matched by hand).

Sometimes, two similar concepts were yielded. Such synsets were a subject of qualitative analysis, which aimed at providing the best match. Similarly, sometimes there is no good match at all. At this stage, 72 out of 1000 classes remain unmatched. Here, we enlist our propositions for them. We encountered problems similar to Nielsen [12], though we propose a different summary of common issues. We categorised these to the following category problems:

Table 1. Mapping – hyponymy.

WordNet 3.0 synset	Wikidata Entity
n03706229 (<i>magnetic compass</i>)	Q34735 (<i>compass</i>)
n02276258 (<i>admiral</i>)	Q311218 (<i>vanessa</i>)
n03538406 (<i>horse cart</i>)	Q235356 (<i>carriage</i>)
n03976467 (<i>Polaroid camera, Polaroid Land camera</i>)	Q313695 (<i>instant camera</i>)
n03775071 (<i>mitten</i>)	Q169031 (<i>glove</i>)
n02123159 (<i>tiger cat</i>)	Q1474329 (<i>tabby cat</i>)
n03796401 (<i>moving van</i>)	Q193468 (<i>van</i>)
n04579145 (<i>whisky jug</i>)	Q2413314 (<i>jug</i>)
n09332890 (<i>lakeshore</i>)	Q468756 (<i>shore</i>)
n01685808 (<i>whiptail, whiptail lizard</i>)	Q1004429 (<i>Cnemidophorus</i>)
n03223299 (<i>doormat</i>)	Q1136834 (<i>mat</i>)
n12768682 (<i>buckeye, horse chestnut, conker</i>)	Q11009 (<i>nut</i>)
n03134739 (<i>croquet ball</i>)	Q18545 (<i>ball</i>)
n09193705 (<i>alp</i>)	Q8502 (<i>mountain</i>)
n03891251 (<i>park bench</i>)	Q204776 (<i>bench</i>)
n02276258 (<i>admiral</i>)	Q311218 (<i>vanessa</i>)

hyponymy, animals and their size, age, and sex, ambiguous synsets, and non-exact match. Each of these represents a different form of a trade-off made in order to provide the full mapping. This is not a classification in a strict sense, as some of the cases could be placed in several of the aforementioned groups.

Hyponymy. This is the situation in which the level of granularity of WordNet synsets did not match the one from Wikidata. As a consequence, some terms did not have a dedicated entity. Therefore, we performed semantic inclusion, in which we searched for a more general “parent” entity, which contained this specific case. Examples include *magnetic compass* (extended to *compass*), *mitten* (*glove*), or *whisky jug* (*jug*). The cases falling to this category are presented in Table 1.

Animals and their size, age, and sex. This set of patterns is actually a subcategory of the hyponymy, but these animal-related nouns provided several problems worth distinguishing. The first one considers a situation in which a WordNet synset describes the particular sex of a given animal. This information is often missing on Wikidata, which means that the broader semantic meaning has to be used. For example, *drake* was mapped to *duck*, whereas *ram*, *tup* to *sheep*. However, while *hen* was linked to *chicken*, for *cock*, *rooster* (n01514668) there exist an exact match (Q2216236). Another pattern considers distinguishing animals of different age and size. For instance, *lorikeet* in WordNet is defined as “any of various small lories”. As this definition is a bit imprecise, we decided to use *loriini*. In another example *eft* (juvenile newt) was linked to *newt*. Similarly, there is eastern and western *green mamba*, but WordNet defines it as “the green phase of the black mamba”. The breed of poodle has three varieties (*toy*, *miniature*, and *standard poodle*), but Wikidata does not distinguish the difference between them – all were therefore linked to *poodle* (Q38904). These mappings are summarised in Table 2.

Ambiguous synsets. This is a situation in which a set of synonyms does not necessarily consist of synonyms (at least in terms of Wikidata entities). That is, for a synset containing at least two synonyms, there is at least one possible Wikidata entity. At the same time, the broader term for a semantic extension does not necessarily exist, since these concepts can be mutually exclusive. For instance, for the synset *African chameleon*, *Chamaeleo chamaeleon* there exist two candidates on Wikidata, *Chamaeleo chamaeleon* and *Chamaelo africanus*. We choose the first one due to the WordNet definition – “a chameleon found in Africa”. Another synset,

Table 2. Mapping – animals and their size, age, and sex.

WordNet 3.0 synset	Wikidata Entity
n01847000 (<i>drake</i>)	Q3736439 (<i>duck</i>)
n01514859 (<i>hen</i>)	Q780 (<i>chicken</i>)
n01806143 (<i>peacock</i>)	Q201251 (<i>peafowl</i>)
n02412080 (<i>ram, tup</i>)	Q7368 (<i>sheep</i>)
n01820546 (<i>lorikeet</i>)	Q15274050 (<i>loriini</i>)
n01631663 (<i>eft</i>)	Q10980893 (<i>newt</i>)
n01749939 (<i>green mamba</i>)	Q194425 (<i>mamba</i>)
n02113624 (<i>toy poodle</i>)	Q38904 (<i>poodle</i>)
n02113712 (<i>miniature poodle</i>)	Q38904 (<i>poodle</i>)
n02113799 (<i>standard poodle</i>)	Q38904 (<i>poodle</i>)
n02087046 (<i>toy terrier</i>)	Q37782 (<i>English Toy Terrier</i>)

academic gown, academic robe, judge's robe contains at least two quite different notions – we have chosen *academic dress*, as this meaning seems to be dominant in the ImageNet. *Harvester, reaper* is an imprecise category in ImageNet since it offers a variety of agricultural tools, not only these suggested by the synset name. *Bonnet, poke bonnet* has a match at Wikidata's *bonnet*, though it is worth noticing that ImageNet is focused on babies wearing this specific headwear. The mapping of this category can be found in Table 3.

Non-exact match. Sometimes, however, there is no good exact match for a given synset among Wikidata entities. At the same time, the broader term might be too broad. This leads to unavoidable inaccuracies. For example, for *nipple* we have chosen its meronym, *baby bottle*. Interestingly, *nipple* exists in Polish Wikidata, though it does not have any properties, which makes it useless in further applications. Other examples involve somewhat similar meaning – *tile roof* was mapped to *roof tile*, or *steel arch bridge* to *through arch bridge*. *Plate rack* was linked to *dish drying cabinet*, though it is not entirely accurate, as the ImageNet contains pictures of things not designated to drying, but sometimes for dish representation. In other example, we map *baseball player* to *Category:baseball players*. ImageNet contains photos of different kinds of *stemware*, not only *goblet*. *Cassette* was linked to a more fine-grained synset (*audio cassette*) as the images present audio cassettes in different settings. Table 4 summarises the mappings falling into this category.

ImageNet itself is not free from errors, since it is biased towards certain skin colour, gender, or age. This is a great concern for ethical artificial intelligence scientists since models trained on ImageNet are ubiquitous. There are some ongoing efforts to fix it with a more balanced set of images, though [19]. Beyer et al. [2] enlisted numerous problems with ImageNet, such as single pair per image, restrictive annotation process, or practically duplicate classes. They proposed a set of new, more realistic labels (ReaL) and argued that models trained in such a setting achieve better performance. Even given these drawbacks, ImageNet is still ubiquitous. Naturally, the presented mapping inherits problems presented in ImageNet, such as these in which images roughly do not present what the synset name suggests. This problem was previously reported by Nielsen [12] – he described it as a discrepancy between ImageNet and WordNet. As for some examples, this might include *radiator*, which in ImageNet represents home radiator, whereas the definition on Wikidata for the same name describes a bit more broad notion (for instance, it also includes car radiators). *Monitor* is a similar example since it might be any display device, though in ImageNet it is connected mostly to a computer display. *Sunscreen, sunblock, sun blocker* represent different photos of products and their appliance on the human body, which look completely different and might be split into two distinct classes.

Table 3. Mapping – ambiguities.

	WordNet 3.0 synset		Wikidata Entity
n01694178	(<i>African chameleon, Chamaeleo chamaeleon</i>)	Q810152	(<i>Chamaeleo africanus</i>)
n02669723	(<i>academic gown, academic robe, judge's robe</i>)	Q1349227	(<i>academic dress</i>)
n02894605	(<i>breakwater, groin, groyne, mole, bulwark, seawall, jetty</i>)	Q215635	(<i>breakwater</i>)
n01755581	(<i>diamondback, diamondback rattlesnake, Crotalus adamanteus</i>)	Q744532	(<i>eastern diamondback rattlesnake</i>)
n03110669	(<i>cornet, horn, trumpet, trump</i>)	Q202027	(<i>cornet</i>)
n03899768	(<i>patio, terrace</i>)	Q737988	(<i>patio</i>)
n04258138	(<i>solar dish, solar collector, solar furnace</i>)	Q837515	(<i>solar collector</i>)
n03016953	(<i>chiffonier, commode</i>)	Q2746233	(<i>chiffonier</i>)
n02114548	(<i>white wolf, Arctic wolf, Canis lupus tundrarum</i>)	Q216441	(<i>Arctic wolf</i>)
n13133613	(<i>Ear, spike, capitulum</i>)	Q587369	(<i>Pseudanthium</i>)
n04509417	(<i>unicycle, monocycle</i>)	Q223924	(<i>unicycle</i>)
n01729322	(<i>hognose snake, puff adder, sand viper</i>)	Q5877356	(<i>hognose</i>)
n01735189	(<i>garter snake, grass snake</i>)	Q1149509	(<i>garter snake</i>)
n02017213	(<i>European gallinule, Porphyrio porphyrio</i>)	Q187902	(<i>Porphyrio porphyrio</i>)
n02013706	(<i>limpkin, aramus pictus</i>)	Q725276	(<i>limpkin</i>)
n04008634	(<i>projectile, missile</i>)	Q49393	(<i>projectile</i>)
n09399592	(<i>promontory, headland, head, foreland</i>)	Q1245089	(<i>promontory</i>)
n01644900	(<i>tailed frog, bell toad, ribbed toad, tailed toad, Ascaphus trui</i>)	Q2925426	(<i>tailed frog</i>)
n02395406	(<i>hog, pig, grunter, squealer, Sus scrofa</i>)	Q787	(<i>pig</i>)
n02443114	(<i>polecat, fitch, foulmart, foumart, Mustela putorius</i>)	Q26582	(<i>Mustela putorius</i>)
n03017168	(<i>chime, bell, gong</i>)	Q101401	(<i>bell</i>)
n02088466	(<i>bloodhound, sleuthhound</i>)	Q21098	(<i>bloodhound</i>)
n03595614	(<i>jersey, t-shirt</i>)	Q131151	(<i>t-shirt</i>)
n03065424	(<i>coil, spiral, volute, whorl, helix</i>)	Q189114	(<i>spiral</i>)
n03594945	(<i>jeep, land rover</i>)	Q946596	(<i>off-road vehicle</i>)
n01753488	(<i>horned viper, cerastes, sand viper, horned asp, Cerastes cornutus</i>)	Q1476343	(<i>Cerastes cerastes</i>)
n03496892	(<i>harvester, reaper</i>)	Q1367947	(<i>reaper</i>)
n02869837	(<i>bonnet, poke bonnet</i>)	Q1149531	(<i>bonnet</i>)

Table 4. Mapping – non-exact matches.

WordNet 3.0 synset	Wikidata Entity
n04311004 (<i>steel arch bridge</i>)	Q5592057 (<i>through arch bridge</i>)
n01737021 (<i>water snake</i>)	Q2163958 (<i>common water snake</i>)
n07714571 (<i>head cabbage</i>)	Q35051 (<i>white cabbage</i>)
n01871265 (<i>tusker</i>)	Q7378 (<i>elephant</i>)
n04344873 (<i>studio coach, day bed</i>)	Q19953097 (<i>sofa bed</i>)
n07714571 (<i>head cabbage</i>)	Q35051 (<i>(white) cabbage</i>)
n03961711 (<i>plate rack</i>)	Q1469010 (<i>dish drying cabinet</i>)
n04505470 (<i>typewriter keyboard</i>)	Q46335 (<i>typewriter</i>)
n03825788 (<i>nipple</i>)	Q797906 (<i>baby bottle</i>)
n04435653 (<i>tile roof</i>)	Q268547 (<i>roof tile</i>)
n09835506 (<i>baseball player</i>)	Q7217606 (<i>Category:baseball players</i>)
n02966687 (<i>carpenter's kit, tool kit</i>)	Q1501161 (<i>toolbox</i>)
n02860847 (<i>bobsleigh – sleigh</i>)	Q177275 (<i>bobsleigh – sport</i>)
n04493381 (<i>tub, vat</i>)	Q152095 (<i>bathtub</i>)
n03443371 (<i>goblet</i>)	Q14920412 (<i>stemware</i>)
n02978881 (<i>cassette</i>)	Q149757 (<i>audio cassette</i>)

Analytics

We also check to what extent the process can be automated, as it might be useful for larger subsets of ImageNet (in a broad sense). In this section, we present the results of such an investigation. We also provide a concise analysis of the number of direct properties, which is a crucial feature in spite of the future usage of the mapping in various computer vision settings.

Foppiano and Romary developed *entity-fishing* [7], a tool for named entity recognition and disambiguation (abbreviated as NERD). This tool can be employed in order to provide an automatic mapping between ImageNet and Wikidata. We used indexed data built from the Wikidata and Wikipedia dumps from 20.05.2020. For this experiment, each synset is split on commas. For example, a synset *barn spider, Araneus cavaticus* (n01773549) is split into two synset elements: *barn spider* and *Araneus cavaticus*. For each of these elements, the term lookup service from *entity-fishing* is called, which searches the knowledge base for given terms in order to provide match candidates. Since this service provides a list of entities ranked by its conditional probability, we choose the one with the highest value.

We start with the 461 already linked instances, which can be perceived as ground-truth data for this experiment. Among them, for 387 (84%) synset elements there was at least one correct suggestion (for example, for a previously mentioned synset *barn spider* and *Araneus cavaticus* at least one was matched to Q1306991). In particular, 286 (62%) synsets were correctly matched for all its elements (for example, for a previously mentioned synset *barn spider* and *Araneus cavaticus* were both matched to Q1306991). While these results show that NERD tools can speed up the process of linking by narrowing down the number of entities to be searched for in some cases, it does not replace manual mapping completely – especially in more complicated and ambiguous cases, which were mentioned in the previous section. Nevertheless, for the remaining 539 synsets which were manually linked, an identical NERD experiment has been performed, which resulted in similar figures. For 448 (83%) synsets, *entity-fishing* provided the same match for at least one synset element. Similarly, for 342 synsets (63%) the tool yielded the same match for all elements. Albeit these figures can be considered as relatively not low, they prove that the mapping obtained in such a way might consider some discrepancies and justify the process presented in the previous section.

Table 5. Most popular properties in the mapping (occurrences of the same properties of a given entity was counted as one).

property	label	count
P646	(Freebase ID)	932
P373	(Commons category)	927
P18	(image)	911
P8408	(KBpedia ID)	687
P910	(topic's main category)	681
P279	(subclass of)	659
P1417	(Encyclopædia Britannica Online ID)	618
P8814	(WordNet 3.1 Synset ID)	551
P31	(instance of)	524
P1014	(Art & Architecture Thesaurus ID)	482

Similarly to Nielsen, we also count the number of the direct properties available in Wikidata. This is a crucial feature since it enables to leverage knowledge graph structure. Listing 2 shows the query used for obtaining the number of properties for Q29022. The query was repeated for each mapped entity. Figure 1 depicts a histogram of direct properties for the 1000 mapped classes. This histogram presents the right-skewed distribution (normal-like after taking the natural logarithm) with the mean at 28.28 ($\sigma = 22.77$). Only one entity has zero properties (*wall clock*).

In total, there are 992 Wikidata entities used in the mapping, as some of them were used several times, like the mentioned *poodle*. These entities displayed 626 unique properties in total. The most popular ones are listed in Table 5. In the optics of computer vision, an important subset of these categories includes P373 (*Commons category*) P910 (*topic's main category*), and P279 (*subclass of*), as they directly reflect hyponymy and hypernymy with the knowledge graph. Such information can be later leveraged in the process of detecting (dis-)similar nodes in a direct way. For example, using e.g. graph path distance in SPARQL for entities sharing common ancestors considering a given property. However, SPARQL does not allow to count the number of arbitrary properties between two given entities. Using graph embedding is a potential workaround for this issue. For example, one can calculate the distances from 200-dimensional pre-trained embeddings provided by the PyTorch-BigGraph library [10]. Another possible direction considers leveraging other linked knowledge graphs, such as Freebase (*P646*), which is linked to the majority of considered instances.

```

1 SELECT (COUNT (?property) AS ?count)
2 WHERE {
3     wd:Q29022 ?property [].
4     FILTER STRSTARTS(STR(?property),
5         ↪ "http://www.wikidata.org/prop/direct/")
6 }

```

Listing 2. Counting direct properties for a single mapped entity. Based on: Nielsen [12].

Summary

In this paper, we presented a complete mapping of ILSVRC 2012 synsets and Wikidata. For 461 classes, such a mapping already existed in Wikidata with `skos:exactMatch`. For other 467

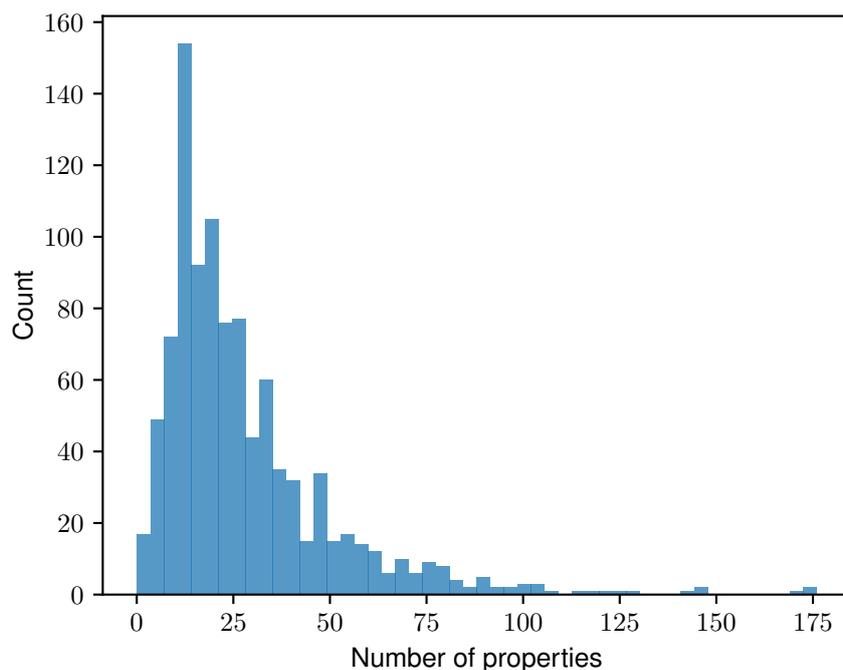


Figure 1. A histogram of direct properties.

classes, we found candidates, which match their corresponding synset. Since 72 classes do not have a direct match, we proposed a detailed justification of our choices. We also compared our mapping with the one obtained from an automated process. To the best of our knowledge, we are the first to publish the mapping ImageNet and Wikidata. The mapping is publicly available for use and validation in various computer vision scenarios.

Future work should focus on empirically testing the mapping. Our results are intended to be beneficial for general-purpose computer vision research since the graphs can be leveraged as a source of contextual information for various tasks, as our analysis showed that the vast majority of the linked entities have a certain number of direct properties. This fact can be utilised according to the given computer vision task. For example, it may be used to generate low-level entity (label) embeddings and calculate distances between them in order to create a correlation matrix used in Knowledge Graph Transfer Network [4] in the task of few-shot image classification. This architecture leverages prior knowledge regarding the semantic similarity of considered labels (called correlation matrix in the paper), which are used for creating class prototypes. These prototypes are used to help the classifier learn novel categories with only few samples available. Correlations might be calculated using simple graph path distance, as well as using more sophisticated low-dimensional knowledge graph embeddings and some distance metrics between each instance. In this case, this will result in a 1000×1000 matrix, as there are 1000 labels in ImageNet. Embeddings from pre-trained models might be used for this task (such as the aforementioned PyTorch-BigGraph embeddings).

Future work might also consider extending the mapping in a way that allows considering larger subsets of ImageNet (in a broad sense), such as ImageNet-6K [4], the dataset, which consists of 6000 ImageNet categories. Preparation of such a large mapping might require a more systematic and collaboratively-oriented approach, which can help to create, verify and reuse the results [20]. The presented approach can also be used for providing mappings with other knowledge graphs and ImageNet. Another possible application might consider further mapping to the actions, which might be particularly interesting for applications in robotics, where robots would be deciding which actions to take based on such mappings [14].

CRedit – Contributor Roles Taxonomy

Dominik Filipiak: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, writing – original draft. **Anna Fensel:** conceptualization, funding acquisition, project administration, writing – review & editing, validation, resources. **Agata Filipowska:** writing – review & editing, validation, resources.

Acknowledgements

This research was co-funded by Interreg Österreich-Bayern 2014-2020 programme project *KI-Net: Bausteine für KI-basierte Optimierungen in der industriellen Fertigung* (grant agreement: AB 292).

References

- [1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.
- [2] L. Beyer, O. J. Hénaff, A. Kolesnikov, X. Zhai, and A. v. d. Oord. Are we done with imagenet? *arXiv preprint arXiv:2006.07159*, 2020.
- [3] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26:2787–2795, 2013.
- [4] R. Chen, T. Chen, X. Hui, H. Wu, G. Li, and L. Lin. Knowledge graph transfer network for few-shot recognition. In *AAAI*, pages 10575–10582, 2020.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [6] C. Edwards. Linking knowledge graphs and images using embeddings. https://cnedwards.com/files/studyabroad_report.pdf, 2018.
- [7] L. Foppiano and L. Romary. entity-fishing: a dariah entity recognition and disambiguation service. *Journal of the Japanese Association for Digital Humanities*, 5(1):22–60, 2020.
- [8] M. Huh, P. Agrawal, and A. A. Efros. What makes imagenet good for transfer learning? *arXiv preprint arXiv:1608.08614*, 2016.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [10] A. Lerer, L. Wu, J. Shen, T. Lacroix, L. Wehrstedt, A. Bose, and A. Peysakhovich. PyTorch-BigGraph: A Large-scale Graph Embedding System. In *Proceedings of the 2nd SysML Conference*, Palo Alto, CA, USA, 2019.
- [11] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [12] F. Å. Nielsen. Linking imagenet wordnet synsets with wikidata. In *Companion Proceedings of the The Web Conference 2018*, pages 1809–1814, 2018.
- [13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

- [14] I. Stavrakantonakis, A. Fensel, and D. Fensel. Matching web entities with potential actions. In *SEMANTICS (Posters & Demos)*, pages 35–38. Citeseer, 2014.
- [15] D. Vrandečić and M. Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.
- [16] W. Wang, V. W. Zheng, H. Yu, and C. Miao. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–37, 2019.
- [17] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020.
- [18] H.-M. Yang, X.-Y. Zhang, F. Yin, and C.-L. Liu. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3474–3482, 2018.
- [19] K. Yang, K. Qinami, L. Fei-Fei, J. Deng, and O. Russakovsky. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 547–558, 2020.
- [20] A. V. Zhdanova and P. Shvaiko. Community-driven ontology matching. In *European Semantic Web Conference*, pages 34–49. Springer, 2006.