Supplementary material for: Semantic Web and Zero-Shot Learning of Large Scale Visual Classes

Appendix A: Linked Open Data
A.1 Linking process

Figure 1 presents the interlinking of the different knowledge bases used in the article. Blue ellipses represent knowledge bases. Under the name of the knowledge base figures the type and number of entities of the knowledge base mapped to Imagenet classes by our process. Blue wide arrows represent the percentage of resources matched between two knowledge bases. We can see that the main bottleneck of our linking process is between Babel synsets and DBPedia resources as only 58.0% of Babel synsets are linked to DBPedia resources.

Green rectangles correspond to the vector representations extracted from the knowledge bases. A second bottleneck concerns sense embeddings. Sense embeddings are obtained by first performing word-sense disambiguation on the English Wikipedia corpus using BabelNet. Then a word2vec model is learned on the disambiguated corpus. However, word-senses appearing less than 5 times in the whole corpus are discarded. Due to this limitation, only 18,627 (85.3%) Babel synsets find a sense embedding representation. In total, we gathered 11,609 classes for which we generated both sense embeddings and Wikipedia mappings.

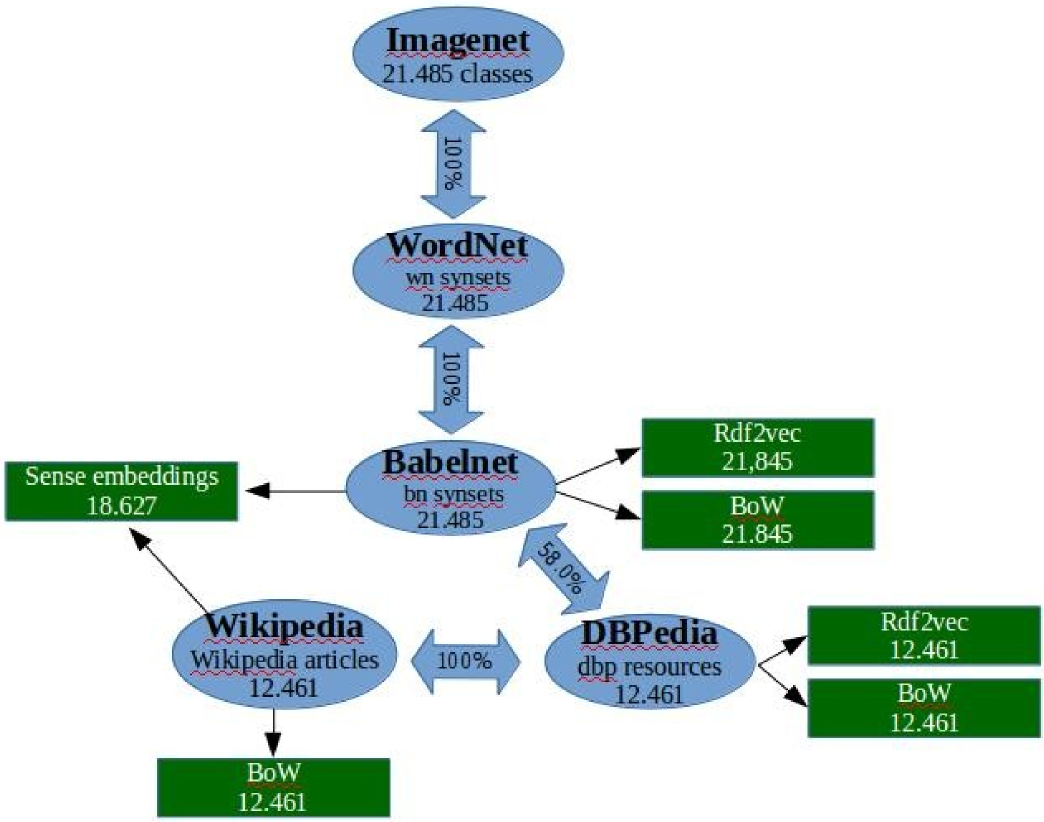

Figure 1. Schema of the linking process

Babelnet provides extensive linking of Babel synsets to DBPedia resources. It matches 4,394,879 Babel synsets to 4,396,575 DBPedia resources. However, this linking does not come in perfect one-to-one matching. Among the 21.845 Babel synsets for which Imagenet provides images, 12,461 are linked to at least one of 13,405 DBPedia resources. 185 Babel synsets are linked to more than one DBPedia resource. For these synsets, we arbitrarily select the first DBPedia resource of the list returned when querying Babelnet.

A.2 Wordnet/Wikipedia mapping evaluation

To assess the quality of the end-to-end mapping we generated between Wordnet synsets and Wikipedia articles, we manually inspect the Wikipedia articles associated to a set of 100 randomly selected Wordnet synsets. To do so, we compare the gloss of Wordnet synsets to the first introduction lines of Wikipedia articles. We subjectively assess the quality of the pair correspondence as perfect match/imperfect match and complete miss. Table 1. below illustrates examples of associations.

Quality	Wordnet synset	Wordnet gloss	Wikipedia title	Wikipedia intro
Perfect Match	n12514592	**Tagasaste, Chamaecytisus pamensis, Cytesis proliferus,** (shrub of Canary Islands having bristle-tipped oblanceolate leaves; used as cattle fodder)	*Cytisus proliferus*	***Cytisus proliferus***, **tagasaste** or **tree lucerne**, is a small spreading evergreen tree that grows 3-4m high.It is a well known fertilizer tree. It is a member of the Fabaceae (pea) family and is indigenous to the dry volcanic slopes of the Canary islands.
Perfect Match	n03304197	**Exocet** (a guided missile developed by the French government for use against ships)	Exocet	The **Exocet** (French for “flying fish”) is a French-built anti-ship missile whose various versions can be launched from surface vessels, submarines, helicopters and fixed-wing aircraft.
		...		
Imperfect Match	n08495908	**aphelion** (apoapsis in solar orbit; the point in the orbit of a planet or comet that is at the greatest distance from the sun)	Perihelion and aphelion	The **perihelion** is the point in the orbit of a celestial body where it is nearest to its orbital focus, generally a star. It is the opposite of **aphelion**, which is the point in the orbit where the celestial body is farthest from its focus
Imperfect Match	n04104500	**Roman building** (a building constructed by the ancient Romans)	Ancient Roman architecture	**Ancient Roman architecture** adopted the external language of classical Greek architecture for the purposes of the ancient Romans, but grew so different from Greek buildings as to become a new architectural style
Miss	n03238131	**dressing room** (a room in which you can change clothes)	Parts of a theater	No Intro

Miss	n02934168	**cable**, line, transmission line (a conductor for transmitting electrical or optical signals or electric power)	Cable (foreign exchange)	**Cable** (or **the cable**) is a foreign exchange term used for the GBP/USD currency pair rate (British pound vs the US dollar).
Miss	n02934168	**Fauve**, fauvist (a member of a group of French painters who followed fauvism)	Fauve (collective)	**Fauve** collective, sometimes stylized as FAUVE, is a French arts collective of music and videography established in 2010 in Paris
Miss	n10292316	**manufacturer**, producer (someone who manufactures something)	Hip hop production	**Hip hop production** is the creation of hip hop music. aspects of hip hop

Table 1. Manual inspection of Wordnet synsets/Wikipedia articles matching

Out of the 100 Wordnet synsets/Wikipedia articles pair randomly selected, we found 4 complete miss-associations and 2 slightly imperfect matches. All incorrect associations are reported in Table 1. The three last misses seem to be caused by homonyms while the reason for the first miss is less clear. The imperfect matches reveal an interesting biases: Wordnet defines semantic concepts in a formal description. Hence perihelion and aphelion are represented by two distinct resources in Wordnet. However, Wikipedia presents encyclopedic information for users to read. Hence, it is not constrained by Wordnet formalism and presents the related concepts of perihelion and aphelion in a same article. Regarding the second imperfect match, Wordnet synset refers to Roman buildings, i.e. building built following a Roman architectural style while Wikipedia article refers to the general concept of Roman architecture.

Appendix B: DBPedia datasets

To generate the DBPedia BoW feature vectors from DBPedia, we used the following DBPedia datasets: Page links, Instance types, Article Categories and Mapping based object. A brief description of these datasets is listed below. To generate RDF2vec feature representations, we only used the Page Links dataset.

1. Page links
This dataset contains internal links between DBpedia resources. The dataset was created from the internal links between articles of the English Wikipedia.

2. Instance types
Contains triples of the form dbp:resource, rdf:type, dbp:type from the mapping-based extraction.

3. Article categories
Gather links from concepts (DBPedia resources) to categories using the SKOS vocabulary.

4. Mapping based object
High-quality data extracted from Wikipedia pages Infoboxes using the mapping-based extraction (Object properties only).

Appendix C: BoW representation of graph nodes

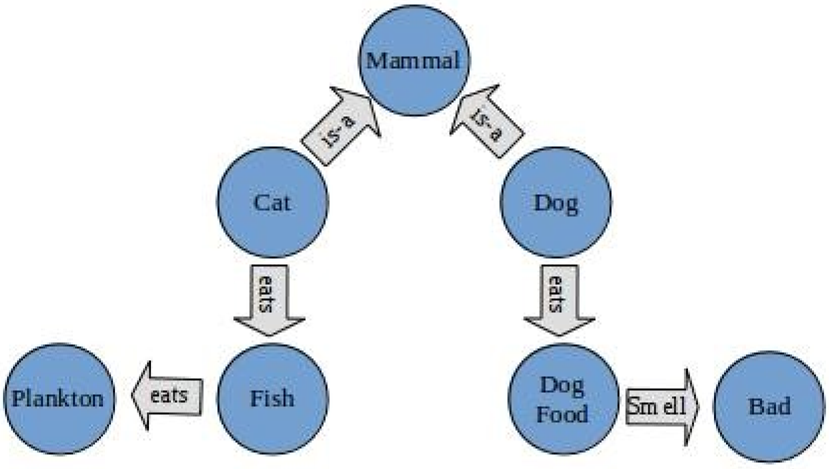

Figure 2. Example of a simple Knowledge Graph

	Outgoing relationships
Dog	(is-a, Mammal), (eats, Dog food)
Cat	(is-a, Mammal), (eats, Fish)
Vocabulary O_{kb}	(is-a, Mammal), (eats, dogfood),(eats, fish)

Table 2. Sets of outgoing relationships and vocabulary

	(is-a, Mammal)	(eats, Dog food)	(eats, Fish)
Dog	1	1	0
Cat	1	0	1

Table 3. BoW encodings of visual classes

Figure 2. represents a simple knowledge graph toy example. Like Babelnet and DBPedia, it is made of concept nodes (blue circles) and typed edges (gray arrows). In this example, we are interested in extracting semantic BoW features for the nodes Cat and Dog. Table 2 shows the set of outgoing relationship for each concept nodes and the vocabulary as the union of these two sets. The resulting BoW representation encodes these sets in a binary vector of dimension equal to the size of the vocabulary and is represented in Table 3. Both the Cat and Dog nodes share the (is-a, Mammal) feature, which establishes a correlation between the Cat and Dog vector representation.
In this toy example, we use a very simple graph and only consider a set of two concepts. The data from DBPedia and Babelnet we use is of a much larger scale. Table 4 provides few statistics to illustrate the scale of the BoW features generated.

	Toy example	Babelnet	DBPedia
Number of classes considered $n = card(S_{kb})$	2	11.609	11.609
Vocabulary size $d = card(O_{kb})$	3	473.379	348.798
Average number of relationship per class $m = \sum_{s_i \in S_{kb}} O_{s_i} / n$	2	163	83

Table 4. BoW feature statistics per knowledge base

Appendix D: Model Architecture

Figure 3. represents the model architecture presented in section 4 of the main article. It is made of two branches, the visual core model is represented in blue and the semantic branch is represented in red. The ZSL model, represented in gray combines both branches to perform classification of unseen classes.

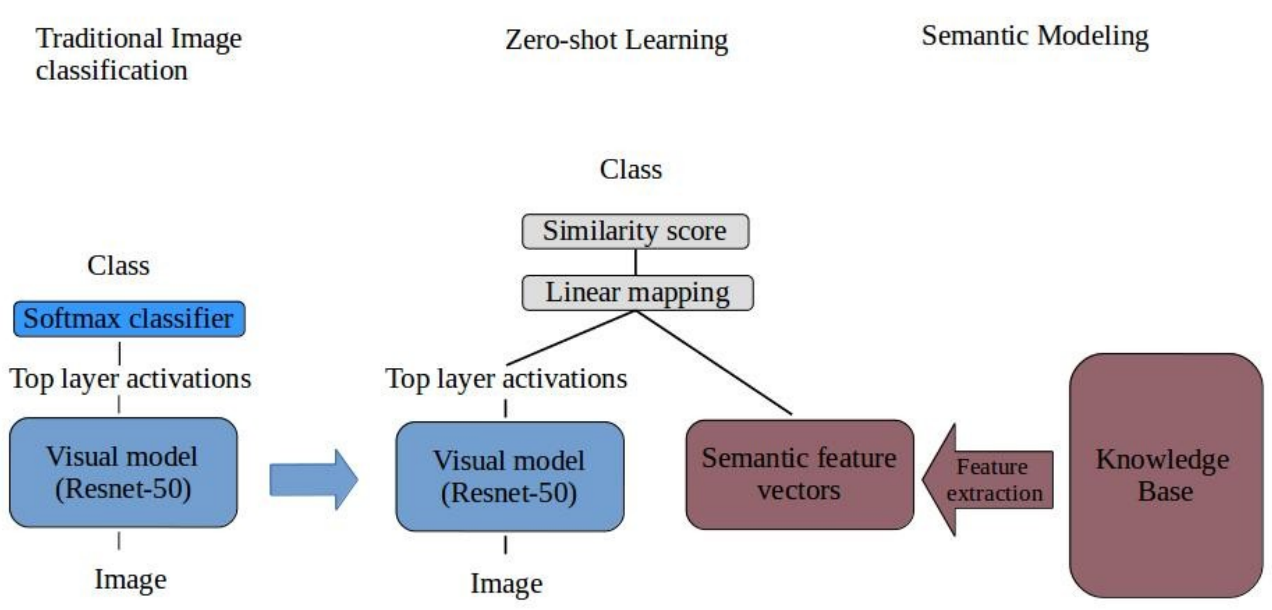

Figure 3. Architecture of our model