

# LexikoNet - a lexical database based on type and role hierarchies

Alexander Geyken & Norbert Schrader  
Berlin Brandenburgische Akademie der Wissenschaften,  
Jägerstr. 22/23, 10117 Berlin

## 1. Introduction

LexikoNet is a lexical ontology of German nouns, developed by Alexander Geyken and Norbert Schrader (Berlin-Brandenburg Academy of Sciences - BBAW). It is based on a concept hierarchy of more than 1,200 concept nodes that is ordered in a top-down hierarchy beginning with the concepts of CONCRETE NOUNS and ABSTRACT NOUNS. CONCRETE NOUNS are further subdivided into LIVING BEINGS, PHYSICAL OBJECTS, MATERIALS, NATURAL KINDS and SPACE. ABSTRACT NOUNS subdivide into PROPERTY, EVENT, ACTIVITY, MEASURES, DOMAINS, etc. The concept hierarchy of ARTIFACTS (underneath PHYSICAL OBJECTS) or LIVING BEINGS, for example, subdivide some levels deeper into rather specific categories such as SACRAL\_BUILDING or SPORTS\_TEAM. The hierarchy goes up to 10 levels deep. Currently, some 90,000 lexemes (corresponding to 75,000 different types) are associated to the concept hierarchy<sup>1</sup>. Lexemes in the sense of LexikoNet are dictionary senses taken from the WDG ([2]) that can be distinguished in the concept hierarchy. Lexemes are associated to concepts at the most specific level by three kinds of ISA relations (type, role and instance) as well as a meronymic and a holonymic relation.

Instances and generic terms were separated in order to enable a systematic term enrichment. For example, LexikoNet distinguishes the concepts SACRAL\_BUILDING and SACRAL\_BUILDING\_BY\_NAME. The former contains terms like 'cathedral', 'synagogue', 'mosque' etc., the latter terms named entities like 'Stephansdom', 'Paulskirche' or 'Al-Aqsa mosque'. In the same way SPORTS\_TEAM are classified as generic lexemes such as 'football team' or 'basketball team' (in German those nouns are compounds), whereas 'Arsenal London' or 'Inter Mailand' are classified as named entities, ie. SPORTS\_TEAM\_BY\_NAME.

The difference between type and role is important in many semantic fields such as animals, artifacts, plants, professions which all can be classified according to their generic resp. biologic classification or with respect to their (anthropocentric) function. For example, the above-mentioned 'mosque' is a building having a sacral role or a person can have the role as a father or a friend. In many cases, the same lexeme can have both, a type and a role. For example, the word 'donkey' in the type hierarchy is a kind of ODD-TOED UNGULATE whereas in the role hierarchy it is a kind of DOMESTIC BEAST OF BURDEN. LexikoNet organizes those lexemes in a double hierarchy. Formally speaking, this means that the hierarchy of LexikoNet corresponds to a lattice and not to a tree.

## 2. LexikoNet and WordNet

Unlike LexikoNet, WordNet [3] does not distinguish roles and types in the hypernym hierarchy. We give some arguments that this difference matters not only with respect to completeness of the lexical database but also for concrete information extraction tasks.

Consider the following examples. The word 'donkey' in WordNet (WordNet 2.1) has only the type reading even though, as shown above, it should also have the role reading:

```
S: (n) domestic ass, donkey, Equus asinus (domestic beast of burden
descended from the African wild ass; patient but stubborn)
=> S: (n) odd-toed ungulate, perissodactyl, perissodactyl mammal
(placental mammals having hooves with an odd number of toes on
each foot)
```

---

<sup>1</sup> This number is to be relativized because more than 45,000 lexemes correspond to profession nouns. However, with almost 50,000 lexemes LexikoNet remains a fairly large lexical database.

A less well known example is the word 'hurricane'. In WordNet 2.1 the following inherited hypernym hierarchy is given:

```
S: (n) hurricane (a severe tropical cyclone usually with heavy rains and winds moving a 73-136 knots (12 on the Beaufort scale))
  => S: (n) cyclone (a violent rotating windstorm)
    => S: (n) windstorm (a storm consisting of violent winds)
      => S: (n) storm, violent storm (a violent weather condition with winds 64-72 knots (11 on the Beaufort scale) and precipitation and thunder and lightning)
        => S: (n) atmospheric phenomenon (a physical phenomenon associated with the atmosphere)
```

Here, WordNet encodes the type hierarchy. However the role hierarchy is not encoded. Yet it might be very useful for information extraction tasks to know that 'hurricanes' are events that often (note here that roles are defeasible) are related to a natural catastrophe. This double classification should also be done for the term 'tsunami' which in WordNet (2.1) is only classified according to its - admittedly - predominant role\_function, but not to its type\_function as a meteorologic phenomenon:

```
S: (n) tsunami (a cataclysm resulting from a destructive sea wave caused by an earthquake or volcanic eruption) "a colossal tsunami destroyed the Minoan civilization in minutes"
  => S: (n) calamity, catastrophe, disaster, tragedy, cataclysm (an event resulting in great loss and misfortune) "the whole city was affected by the irremediable calamity"; "the earthquake was a disaster"
```

### 3. Applications and future developments

LexikoNet is currently used as a filter for lexicographic extraction tasks in the project 'Digital Dictionary of the German Language of the 20th/21st century' ([www.dwds.de](http://www.dwds.de)). It is also used to classify frequently occurring nouns, the so-called 'words of the week' in the newspaper Die ZEIT according to their semantic classes. In this application, only a very small subset of LexikoNet is used: persons, organizations, human nouns, plants, animals, events, geographical nouns, material, and food. Here, a practical use of distinguishing type and role functions of words becomes obvious. For example, 'hurricane' in a newspaper context is more likely to be used in its role function than in its type function. This has been implemented by precedence rules. The result of this classification for the last 100 issues of die ZEIT (2004 and 2005) can be found [www.dwds.de/woewo](http://www.dwds.de/woewo).

There is also a browser for LexikoNet which will be made available by the end of november 2005 under [www.dwds.de](http://www.dwds.de). Figure 1 and 2 in the appendix show the results of this browser for the word 'Esel' (donkey) according to its type and role hierarchy.

LexikoNet is currently enlarged by semi-automatic methods on the basis of a the electronic version of a huge monolingual german dictionary [www.dwds.de/wdg](http://www.dwds.de/wdg) on the basis of a semi automatic extraction of hypernyms from the definitions in the dictionary [1]

### Literature:

- [1] Geyken, Alexander & Ludwig, Rainer : *Halbautomatische Extraktion einer Hyperonymiehierarchie aus dem Wörterbuch der deutschen Gegenwartssprache*. TaCoS 2003. Gießen (D) 13.-15.6. 2003
- [2] Klappenbach, Ruth and Wolfgang Steinitz (ed) (1977). *Wörterbuch der deutschen Gegenwartssprache (WDG)*. Akademie Verlag.
- [3] Miller, George A., Christiane Fellbaum. WordNet. <http://wordnet.princeton.edu/w3wn.html>

## Appendix

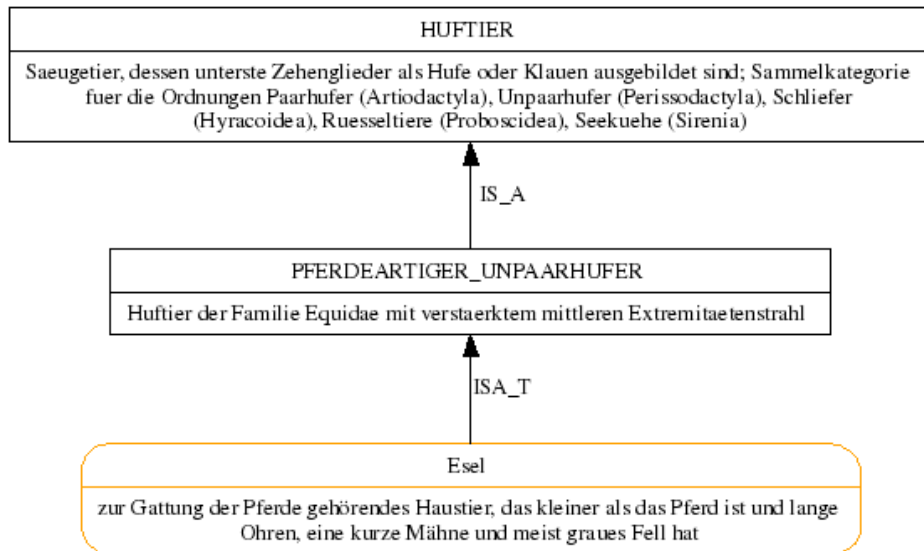


Fig. 1: Esel (donkey) is a kind of a kind of ODD-TOED UNGULATE in the type hierarchy

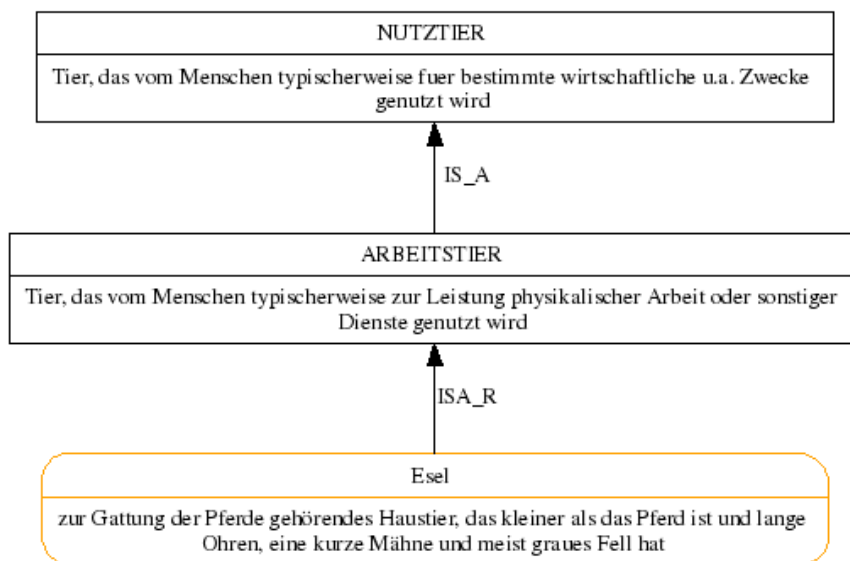


Fig. 2: Esel (donkey) is a kind of DOMESTIC BEAST OF BURDEN in role hierarchy