Lang Resources & Evaluation (2022) 56:807–855
https://doi.org/10.1007/s10579-021-09553-5

ORIGINAL PAPER

Multiple annotation for biodiversity: developing an annotation framework among biology, linguistics and text technology

Andy Lücking[1] · Christine Driller[2] · Manuel Stoeckel[1] · Giuseppe Abrami[1] · Adrian Pachzelt[3] · Alexander Mehler[1]

Accepted: 2 July 2021 / Published online: 4 August 2021

Abstract Biodiversity information is contained in countless digitized and unprocessed scholarly texts. Although automated extraction of these data has been gaining momentum for years, there are still innumerable text sources that are poorly accessible and require a more advanced range of methods to extract relevant information. To improve the access to semantic biodiversity information, we have launched the BIOfid project (www.biofid.de) and have developed a portal to access the semantics of German language biodiversity texts, mainly from the 19th and 20th century. However, to make such a portal work, a couple of methods had to be developed or adapted first. In particular, text-technological information extraction methods were needed, which extract the required information from the texts. Such

✉ Andy Lücking
luecking@em.uni-frankfurt.de

Christine Driller
christine.driller@senckenberg.de

Manuel Stoeckel
manuel.stoeckel@stud.uni-frankfurt.de

Giuseppe Abrami
abrami@em.uni-frankfurt.de

Adrian Pachzelt
a.pachzelt@ub.uni-frankfurt.de

Alexander Mehler
mehler@em.uni-frankfurt.de

[1] TTLab, Goethe-University Frankfurt, Frankfurt, Germany

[2] Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Frankfurt, Germany

[3] University Library Johann Christian Senckenberg, Goethe-University Frankfurt, Frankfurt, Germany

methods draw on machine learning techniques, which in turn are trained by learning data. To this end, among others, we gathered the BIOfid text corpus, which is a cooperatively built resource, developed by biologists, text technologists, and linguists. A special feature of BIOfid is its multiple annotation approach, which takes into account both general and biology-specific classifications, and by this means goes beyond previous, typically taxon- or ontology-driven proper name detection. We describe the design decisions and the genuine *Annotation Hub Framework* underlying the BIOfid annotations and present agreement results. The tools used to create the annotations are introduced, and the use of the data in the semantic portal is described. Finally, some general lessons, in particular with multiple annotation projects, are drawn.

Keywords BIOfid · Biodiversity · Annotation · Semantic portal · Specialized information service · Inter-annotator agreement · Taxon · Named entity recognition

1 Introduction

Anthropocene biodiversity loss has been one of the core issues in earth and life sciences for years (Cardoso et al., 2020; Hallmann et al., 2017; Johnson et al., 2017; Seddon et al., 2016). Data on species occurrences and their adaptations to changing environmental conditions serve as an important basis for studying their distribution patterns and potential threats. FAIR data principles (Wilkinson et al., 2016) should therefore ensure a sustainable research data management to make data *findable*, *accessible*, *interoperable* and *reusable*. Even though scientists adopt these principles gradually as best practices, many studies will remain "below radar level" for various reasons. Legacy scientific literature, for example, usually falls into the latter category, because historical writings are often only available in printed form. But even if natural language texts are digitized, efficient information extraction (extracting mentions of entities and the relations between them) may not be available, since current natural language processing tools in biodiversity science have limited application range and still require testing (Thessen et al., 2012).

With regard to life sciences articles, there are also some special features that should be taken into account in semantic text analysis. A central problem is the naming of biological organisms (Akella et al., 2012; Koning et al., 2005). That naming follows an internationally accepted taxonomic nomenclature but names can also be given in vernacular forms, which often varies both regionally and temporally. There is also the practice of abbreviating scientific names (e.g., "*F. sylvatica*" instead of "*Fagus sylvatica*") or even using their full length version including the authority and the year of publication ("*Fagus sylvatica* L., 1753"). Additionally, due to numerous taxonomic revisions, a large number of synonyms and homonyms have emerged in the course of history, further affecting the (automatic) assignment of taxonomic names from texts to biological taxa.

Which species occurred when and where? This is a basic question to understand the biogeography of a species, to define its ecological preferences, and to estimate its adaptability and abundance in certain habitats.[1] However, information on species characteristics and occurrences does not necessarily follow a standard vocabulary, nor can it usually be found coherently in narrative texts. Since existing biodiversity literature, both historical and modern, incorporates a considerable amount of unstructured data with research-relevant content, there is a high demand to make these data retrospectively more FAIR (Thessen et al., 2012).

So far, bioannotation schemes and data mining tools have been prevalently developed for and applied to information extraction in biomedicine and molecular biology (e.g., Corney et al., 2004; Miyao et al., 2008). Meanwhile, data aggregators like the *Biodiversity Heritage Library*,[2] *Encyclopedia of Life*,[3] *Pangaea*,[4] or *Plazi*[5] provide various options for making biodiversity resources freely available. Methods range from digitization and encoding of taxonomic literature over data enrichment and compilation to machine readable data archiving. However, practical guidance on semantic annotation of biodiversity literature are few and far between and usually refer to English-language text corpora with a focus on taxonomy (see, e.g., Sautter et al., 2007). Beyond mere taxonomic tagging, more recent workflows also cover a much broader thematic range of biodiversity entities, but do not allow multi-label annotation, that is, (possibly) assigning more than one annotation tag to an annotation unit (Löffler et al., 2020; Nguyen et al., 2019; Thessen et al., 2018). Enhancing content retrieval and information fusion by multi-label annotation has since found its way into the biomedical domain, e.g. to detect multiple core scientific concepts at the sentence level or multi-functional genes in cancer pathways (Guan et al., 2018; Ravenscroft et al., 2016). In the first example, it was simply taken into account that a single sentence can cover different aspects of a scientific discourse, as for example the goal, the method and the result of a study. Consequently, if this sentence is only assigned to the core scientific concept “goal”, the presence of other concepts would remain unconsidered despite their relevance to the content. With regard to the word-level, mapping a single term to more than one annotation category could prove helpful to specify its meaning (pollination as reproductive process of a plant) or to resolve ambiguities (trunk as an anatomical feature of an animal or a plant). Thus, by assigning both annotation categories MORPHOLOGY and PLANT, the term trunk can be clearly identified as a plant characteristic. Multi-label annotations are also useful in the field of biological nomenclature for classifying homonyms. The genus name *Agathis* is found among insects and conifers. Adding the appropriate kingdom (PLANT respectively ANIMAL) to TAXON can resolve this ambiguity. However, biodiversity literature does not only deal with the taxonomy and morphology of plants and animals, but also those of other kingdoms (fungi, bacteria, etc.). Furthermore, biological texts can address behavioral patterns, biological processes and stages of development, which in turn also require a more distinct resolution. Instead of providing a single annotation

[1] See also the mission of the *Global Biodiversity Information Facility*, http://www.gbif.de.

[2] https://www.biodiversitylibrary.org.

[3] https://eol.org.

[4] https://www.pangaea.de.

[5] http://plazi.org

category for each specific term, (see Sect. 2.2) a higher specificity of the annotation can also be achieved by multi-labeling, while the total number of annotation categories remains limited and manageable. However, the suitability for machine learning-based classification using multi-labeling on biodiversity-related content has yet to be verified. In this article, we describe a framework in which a multi-label annotation scheme has been developed and coupled to ML fine-tuning, and finally applied to biodiversity texts within a semantic search portal.

Combining natural language processing (NLP) and machine learning (ML) techniques is the method of choice to improve automatic findability and accessibility of mentions of entities within content-rich legacy biodiversity literature. The multitude of existing methodological approaches and tools demonstrates how the underlying data and text resources influence the applied annotation schemes and information retrieval process and outcomes (cf. the different approaches to parsing of biological texts of, e.g., Krauthammer et al., 2000; Lenzi et al., 2006; Nasr & Rambow, 2004). Since the English-language literature clearly predominates in the natural sciences, there are currently no applications aiming specifically at the data mobilization of German-language full-text articles. In view of the foregoing, the German *Specialized Information Service Biodiversity Research* BIOfid (Koch et al., 2017) develops new, and (mostly) freely available,[6] routines and tools for text tagging and semantic annotation to meet the scientific community needs as well as the specifications required for biodiversity-related data with special emphasis on German-language, Central European literature of the 19th and 20th century (see www.biofid.de). This includes, in particular, a combination of named entity recognition and general ontological classification—a multiple label annotation approach which is couched in a genuine annotation framework and models texts in both a technical and a vernacular perspective.

The structure of this article is as follows: In Sect. 2, the BIOfid annotation scheme is introduced. BIOfid ’s central textual challenges in regard to the focal annotation categories TAXON, TIME, and SPACE are described. Since we are dealing with texts, we discuss some advanced linguistic distinctions and follow-up issues which have to be considered. The overall structure of the annotation process and the interaction of automatic and manual annotation—the *Annotation Hub Framework*—is summarized in Sect. 3. Section 4 introduces the annotation tools which we use to carry out the annotation work. While the tools are generic in their nature, we examine some BIOfid specific adjustments and innovations as well as the built-in agreement calculation. The section concludes with the discussion of the results of the BIOfid annotation scheme, which are fed into the semantic search web portal of BIOfid. The web portal is designed for bio-scientist users to easily access the annotated texts and extracted data and is introduced in Sect. 5. We conclude in Sect. 6.

[6] https://github.com/texttechnologylab.

2 Developing the BIOfid annotation scheme

As pointed out in Sect. 1, for the purposes of BIOfid, a general annotation scheme is required that covers taxonomic names as well as “mundane” description of (at least) organisms and their temporal and spatial relationships, and that for German-language texts. This section describes the design of an annotation scheme that aims to fulfill this purpose. The annotated data gained in this way can be used as (a part of) a training corpus for large-scale ML methods for automatic text processing (Ahmed et al., 2019), methods which can be regarded a standard in bioinformatic contexts by now (Blaschke et al., 2002). The development of an annotation scheme has to be put to test and should be regimented by annotation guidelines, since annotation is a *data generating* rather than a *data documenting* process (Consten & Loll, 2012). Accordingly, in BIOfid, annotation guidelines are collected as part of an annotation manual (Lücking et al., 2020). The main annotation classes and the rationale of their application are covered in the following subsections.

2.1 Ontological classification

2.1.1 Taxon names

One of the most conspicuous features of biological texts is the use of a certain class of proper names, namely *taxonomic names*. These are names that refer to kinds.[7] Accordingly, the first task for automatic processing of biological texts is to identify such kind-denoting proper names. Thus, there is a straightforward starting point of biological and text-technological collaboration within BIOfid, namely *Named Entity Recognition* (NER), a sub-task of information extraction (for a survey see, e.g., Nadeau & Sekine, 2007).

With regard to the term “entity”, we have to distinguish two usage traditions (Prechtl & Burkard, 2008, p. 138), a “classical” and a “logical” one. Classically, “entity” is a basic ontological notion, referring to something of independent existence—an individual. In logical semantics, however, an entity is ontologically unspecific and refers to any kind of extralinguistic object (things, concepts, propositions, events, sets, *etc.*). This is also the view of semantic data models (e.g. UML, ER model etc.) in computer science, where an entity is an *instance* of a concept. Despite that, in our annotation framework, entity and concept are two disjoint ranks; each annotation unit (that is, words) has to be specified whether it refers to something of the rank entity or concept.

Why are these digressions relevant to BIOfid and the annotation scheme developed therein? The reason is that taxonomical names are proper names, but they do not refer to an entity in the sense of an individual; rather, they can be conceived as referring to collections of individuals. Thus, from the classical perspective, taxa cannot be the referents of names, because they simply are no individuals, while

[7] To be more precise, those biological names refer to any level of the actual biological taxonomy, that is, for instance, to species, genus, family, order, class, phylum, or kingdom. Unless we want to address a certain level on this hierarchy, we simply speak of *kinds* in the following.

logical semantics is much more permissive in this respect. Of course, this has not gone unnoticed and the semantics of kind reference is a well-known fact about languages (see, e.g., Chierchia, 1998). Taxon names are therefore annotated to be of rank concept.

2.1.2 Common nouns and WordNet categories

As aforementioned, taxonomic names are not the only part of speech that is central to the questions addressed by BIOfid (cf. Sect. 1). Additionally we focus on common nouns. Let us make things more concrete with an example from the BIOfid corpus[8]:

> There has been a sleeping place of Corvidae in the outskirts of Bad Salzungen for more than two decades. The birds use a small forest with old deciduous trees near the city park for their night roost. From 1985 to 1988, once a week, the author had checked the sleeping place and noted the quantity of birds. The sleeping place is used the whole year, in summer by Jackdaws (*Corvus monedula*) and Carrion Crows (*Corvus c. corone*), in winter by Rooks (*Corvus frugilegus*), too. The maximum crowds of sleeping birds varied from 2500 to 9000 individuals. Circadium rhythm of approaching is nearly the same in every winter evening. Arrival and departure are determined by light intensity and weather. The Corvidae are very sensitive regarding disturbance at their sleeping place, but in spite of many injuries they don't change their sleeping trees.

This example highlights that biological texts do not content themselves with appellatives (e.g., *birds*) and taxonomic kind reference (e.g., *Corvus monedula*), but also contain mundane common nouns of biological impact (e.g., *outskirts*). In order to account likewise for genre-specific, scientific or vernacular names as well as for everyday descriptions, we employ a mixed classification system. "All-purpose categories" are derived from the lexical database WordNet (Fellbaum, 1998; Miller, 1995) and are used for a general ontological annotation. However, some caution is appropriate in this respect (cf. Sanfilippo et al., 2006), since WordNet includes proper name entries (e.g., "Ludwig van Beethoven") but has no `instance_of` relation at its disposal. Instead, WordNet uses lexical or sense relations throughout. This leads to a confusion between common nouns and proper names (what Oltramari et al., 2002, p. 18 call a "[c]onfusion between concepts and individuals"). In other words: WordNet is rather a lexical database or "terminological ontology" (Sowa, 2000) than an ontology *simpliciter*. However, since we distinguish "entities" from "concepts", we meet the pre-requirement for using WordNet's 26 top-level entity categories (i.e., the unique beginner synset for nouns) for ontological classification. WordNet distinguishes the following top-level categories:

[8] English summary of Klaus Schmidt (1999), "Mehrjährige Beobachtungen an einem Krähen-Dohlen-Schlafplatz in Bad Salzungen, Südwest-Thüringen". In: *Mitteilungen des Vereins Sächsischer Ornithologen* 8, Sonderheft 2, pp. 77–93.

WordNet Categories: {PERSON, HUMAN BEING}, {ANIMAL, FAUNA}, {PLANT, FLORA}, {GROUP, COLLECTION}, {SOCIETY}, {LOCATION, PLACE}, {TIME}, {COMMUNICATION}, {QUANTITY, AMOUNTS}, {EVENT, HAPPENING}, {NATURAL OBJECT}, {POSSESSION, PROPERTY}, {ATTRIBUTE, PROPERTY}, {BODY, CORPUS}, {FOOD}, {ARTIFACT}, {ACT, ACTION, ACTIVITY}, {PROCESS}, {NATURAL PHENOMENON}, {COGNITION, IDEATION}, {FEELING, EMOTION}, {MOTIVE}, {RELATION}, {SHAPE}, {STATE, CONDITION}, {SUBSTANCE}

2.1.3 Biology-specific categories

The WordNet categories are complemented by additional biology-specific categories, though. Since WordNet distinguishes only two realms of living beings: plants and animals, we have to extend its categories to include the whole variety of biological taxonomic entities. Specifically, we added the composite organism group of *lichens* as well as all missing taxonomic kingdoms, including *Archaea*, *Bacteria*, *Chromista*, *Fungi*, *Protozoa*, and *Viruses*. These are the accepted kingdoms according to one of the leading repositories on biodiversity data, the *Global Biodiversity Information Facility* (GBIF).[9] To distinguish between organism names that correspond to a taxonomic entity from those used in a more general sense, we introduced the annotation category TAXON. Based on our initial explorations of the BIOfid text corpus, we also considered it necessary to implement biology-specific annotation categories that exhibit a more refined meaning than those of related WordNet categories. In particular, for the WordNet categories to the left of the arrows, we introduced the biological category on the right of the arrow: ATTRIBUTE → MORPHOLOGY, BODY → MORPHOLOGY, LOCATION → HABITAT, PROCESS → REPRODUCTION. This enables the differentiation of more general annotations from biology-specific terms and is intended to promote the adaptation and enrichment of the ontologies underlying the semantic search in the BIOfid portal. For instance, while every habitat is a location, not every location needs to be a habitat. All categories are considered first-class citizens of the ontology, however.

Biology-specific Categories: {TAXON}, {ARCHAEA}, {BACTERIA}, {CHROMISTA}, {FUNGI}, {PROTOZOA}, {VIRUSES}, {LICHENS}, {HABITAT}, {MORPHOLOGY}, {REPRODUCTION}

These 37 annotation categories are all on the same level and constitute the basic ontological annotation grid of BIOfid. Each category comes with a description which guides its application—in case of the WordNet categories, the description is obtained from the entries in the WordNet database; descriptions of biology-specific categories are given in Appendix. However, it turned out that WordNet's beginner synset for nouns (i.e., the above-given 26 top-level categories for entities) is highly anthropocentric. For instance, *artifact* is described as "a man-made object taken as a whole".[10] This definition leaves open of how to deal with objects like a bird's nest,

[9] https://www.gbif.org/.

[10] http://wordnetweb.princeton.edu/perl/webwn?s=artifact, accessed August 31, 2020.

which seem to be an animal artifact. Following philosophical theories of action (Gould, 2007; Steward, 2009), we also conceive animals as agents. Thus, contrary to (or relaxing) the WordNet descriptions, we assume that any category that involves an agent applies to non-human agents as well.

Now, basically, any instance of any of the above-listed categories can be either referred to by means of a proper name, or described by means of predication. In the sentence *Lassie is a dog*, for example, *Lassie* is a proper name according to the classical notion: it picks out a specific individual (a dog, in this case). The common noun *dog*, as well as the corresponding technical taxon term *Canis lupus familiaris*, lacks such a discerning power. Hence, we distinguish between proper names referring to single individuals ("Lassie") from proper names and common nouns referring to other ontological classes such as sets ("dog"). For this purpose, we assign any application of an annotation label to either ENTITY (an individual referred to by a proper name) or $\overline{\text{CONCEPT}}$ (a class, or set of entities). Typographically, annotation categories that refer to a classical entity are typeset in small caps while concepts are additionally indicated by an overbar—for instance, PERS is the label for a proper name whose bearer is a human being (*Alfred Russel Wallace*), $\overline{\text{ANIMAL}}$ labels a noun that denotes a set of entities of the kingdom of animals such as *dogs*. We employ this typographic convention in the examples given throughout the paper.

2.2 Multiple classification

The ontological annotation categories outlined in the preceding section comprise both very general and more specific labels. For instance, presumably any object from the physical world can be said to be a $\overline{\text{NATURALOBJECT}}$, including animals and plants. So, for a given animal, what is the correct annotation label: $\overline{\text{NATURALOBJECT}}$ or $\overline{\text{ANIMAL}}$? Since this does not seem to be an either-or question, we decided to employ a multi-label annotation.[11] In fact, annotation units receive multiple annotation labels as a rule, not as an exception.

The following examples illustrates the multiple annotation approach of BIOfid by means of a couple of "real world" data:

- The most common multiple annotation within BIOfid probably is the annotation of taxonomic names. Each taxonomic name is marked as such (i.e., $\overline{\text{TAXON}}$) and coupled with a label indicating the biological kingdom of the taxon, such as $\overline{\text{PLANT}}$, $\overline{\text{ANIMAL}}$, or $\overline{\text{FUNGI}}$.
- The category MORPH(ology) explicitly mentions parthood.[12] Accordingly, when $\overline{\text{MORPH}}$ is used in addition to some other label, it is interpreted as "morphological part of [that other label]". For instance, a combination of $\overline{\text{MORPH}}$ and $\overline{\text{PLANT}}$

[11] In fact, such general inclusion relations as that between $\overline{\text{NATURALOBJECT}}$ and $\overline{\text{ANIMAL}}$ are "outsourced" to specific conventions, saving annotation time.

[12] "The annotation unit is about the outward appearance or inwards structure of an organism" (Lücking et al., 2020).

characterizes a part of a plant (say, its stem). That is, MORPH implements a minimal mereology.

- A garden is an artificially created location that also provides a living environment for plants and animals. Its heterogeneity is captured by the following multiple categorization: $\overline{\text{LOCATION}}$, $\overline{\text{HABITAT}}$, $\overline{\text{ARTIFACT}}$. However, since “*”garden is also a GeoNames entity (see Sect. 2.3), namely S/GDN (read: subcategory GDN in main category S), it is sufficient to use the GeoNames classification, which can be mapped onto the more elaborate multiple annotation.
- A report can be categorized as ARTIFACT, COMMUNICATION, and COGNITION, since it is man-made (ARTIFACT), conveys information (COMMUNICATION), and is the result and possibly the trigger of mental processes (COGNITION).[13]

A multiple annotation approach avoids the decision problem of choosing just one ontological label. However, it poses problems on its own, most notably, the difficulty of keeping annotations consistent. On the one hand, too permissive annotations have to be avoided. Although there is not just one “correct” ontological label in most of the cases (what is the true, single category of, say, *peduncle* or *inquiry*?), classification is by no means arbitrary. Re-using a previous example: a garden involves plants,[14] but is not a plant itself. Hence, it would go too far to label *garden* with the category $\overline{\text{PLANT}}$.

On the other hand, annotation should be as informative as possible. For instance, classifying a report merely as $\overline{\text{ARTIFACT}}$ would be correct, but not very informative. This approach would simply group together reports with other kinds of artifacts (that is, man-made objects), such as shoes, cooking spoons, or space ships. Rather, a report is also an instance of communication, and multiple annotation should reflect this. One challenge for multiple annotation projects therefore is to find the right level of granularity. Within BIOfid, this challenge is met by means of three measures:

1. Annotators discuss extracts of their annotations and highlight difficult examples at regular annotation meetings.
2. Such a meeting can result in finding annotation conventions (like the previous examples), which are compiled in the annotation manual.
3. Tool-wise, a consistent annotation is supported by a *recommendation function*, which is described in Sect. 4.1 (roughly speaking, a recommendation assigns the annotation categories chosen by the annotator for a given token to all tokens of the same lemma within a certain text span).

[13] This multilayered sortal structure of text objects is well known in lexical semantics (Pustejovsky, 1991).

[14] As mostly, if not always, one can find exceptions: a rock garden is a garden that goes without plants.

2.3 Time and space

Following the basic question posed in Sect. 1—*Which species occurred when and where?*—two categories receive special attention, namely LOC(ation) and TIME.[15] To this end, we apply GeoNames[16] locations, which subdivide WordNet's category LOC into nine major categories, while TIME is coded according to the ISO standard ISOTimeML (ISO, 2012).

GeoNames categories include geographical entities like cities, lakes, countries, or landmarks. Thus, any location is assigned to one of the following main categories, which are addressed in terms of an alphabetic character:

–A : {country, state, region,…}
–H : {stream, lake,…}
–L : {parks, area,…}
–P : {city, village,…}
–R : {road, railroad}
–S : {spot, building, farm}
–T : {mountain, hill, rock,…}
–U : {undersea}
–V : {forest, heath,…}

In addition to the nine GeoNames main classes, there are 680 sub-categories (excluding *unavailable*), which allow a very finegrained categorization.[17]

The TIME-annotation unit is about temporal entities, "the fourth coordinate that is required (along with three spatial dimensions) to specify a physical event" (WordNet[18]), including clock times ("a reading of a point in time as given by a clock", WordNet[19]). Following ISO-TimeML (ISO 2012), we distinguish DATE (referring to calendric time units), $\overline{\text{TIME}}$ (referring to daytimes, even in an unspecific way), DURATION/$\overline{\text{DURATION}}$ (referring to temporal intervals), and SET/$\overline{\text{SET}}$ (quantifying over time points or intervals, say as a result of repetition).

In addition to the above-mentioned categories, a document exhibits both a distinguished location and a distinguished date, namely the *document creation location* (DCL) and the *document creation time* (DCT), respectively (Pustejovsky, 2017a, b). DCL and DCT are used to label those locational or temporal expressions that refer to the place and time of the author writing the text. Note that DCL and DCT may be given as part of the metadata of a given text, or that they may be unknown. An example of a DCT mentioned at beginning of the main text is given in (1) (from document `3673151`).

(1) Herr cand . iur. Hepp hat *Isoetes lacustris* L. am 17. Juli dieses Jahres (1898) im Steinsee bei Grafing angetroffen.

[15] It should be noted that the BIOfid annotation focuses on the *descriptive* dimension of speech, not on its *expressive* one (Potts, 2007). That is, no contrast is made between calling a dog *dog* or *mutt* (cf. Zimmermann, 1991, p. 165). As a matter of fact, the expressive dimension is not very often referred to in the texts considered so far, as one would expect from academic writings.

[16] https://www.geonames.org/.

[17] See https://www.geonames.org/export/codes.html for a complete list.

[18] http://wordnetweb.princeton.edu/perl/webwn?s=time, the fourth dimension sense, accessed September 4, 2020.

[19] http://wordnetweb.princeton.edu/perl/webwn?s=time, the clock time sense, accessed September 4, 2020.

(*Mr. Hepp found Isoetes lacustris L. on July 17 of this year (1898) in the Steinsee near Grafing.*)

By using the demonstrative noun phrase *dieses Jahr* "*"this year the author refers to the year when he or she was actually writing the sentence. This indexical reference is resolved by the DATE given in parenthesis (*viz.* 1898). That is, 1898 can be tagged "DCT". Having applied this label, it is at disposal for resolving further indexically given temporal expressions. Later in the text we find *jetzt* 'now':

(2) Die von Schmidt bezeichnete Stelle nimmt jetzt eine kultivierte Wiese ein. (*The place designated by Schmidt is now occupied by a cultivated meadow.*)

By identifying *jetzt* 'now' with the DCT within the annotation tool (see Sect. 4.1) we receive the information that there is a cultivated meadow as of 1898.

2.4 Beyond words

In BIOfid, we pursue basically a word-based annotation.[20] However, there are a couple of phenomena that go beyond words, but nonetheless affect word annotations. We exemplarily discuss compounds, possessives, speaker's reference, and anaphora in the following.

2.4.1 Compounds

The main language of the texts investigated in BIOfid is German. Ever since Mark Twain's "The Awful German Language", German is famously known to be a compounding language. A nominal compound is a noun which consists of several other modifying components (Matthews, 1991, Sect. 5). A modifying component can be an adjective (*green tea*), a verb (*swimming pool*), or another noun (*football*)). Most nominal compounds are determinative, meaning that the modifying expression determines the head noun. From a taxonomic perspective, the head noun determines the compound's category. Hence, a compound is labeled only according to its head. For instance, *football* is labeled as an $\overline{\text{ARTIFACT}}$, and *not* (additionally) as $\overline{\text{BODY}}$.

2.4.2 Possessives

Genitive noun phrases raise the question of how to deal with possessives and relational nouns in general. Take, for instance, the following example: *the foodplant of the monophagous moorland clouded yellow* [which is *Colias palaeno*]. Here we have two nouns, the head noun *foodplant* and the modifying compound noun *moorland clouded yellow* (which itself is modified by the adjective *monophagous*). Genitives can be thought of as functions in the mathematical sense: the head noun applies to the modifying noun and returns a value. However, although the referent *type* is uniquely determined, the returned value does not need to be a specific

[20] Though in future extensions also role-based annotations will be added (cf. Sect. 6).

individual. Accordingly, the example sentence calls for a nested annotation, which, in this case, is on the level of concepts: [*foodplant of the monophaguous* [*moorland clouded yellow*]$\overline{\text{TAXON}}$,$\overline{\text{ANIMAL}}$]$\overline{\text{PLANT}}$.

2.4.3 Speaker's reference

Definite noun phrases show an interesting feature: Their usage can pick out an individual—just like a proper name does—even if this individual is unknown to the speaker. Suppose the speaker listens to a radio broadcast that announces that the jackpot was hit. Then the speaker can assert *The lottery winner must be happy*, referring to the jackpot winner, *whoever he or she is*. This usage contrasts to the noun phrase in, e.g., *Yesterday, my sister hit the jackpot*, where the genitive noun phrase *my sister* refers to a particular individual known to the speaker.[21] Within the BIOfid text corpus, there are descriptions such as *every morning, I saw the swallow leaving its bird-nest*, which are about a *specific* bird the author observed. However, there are also general statements such as *the swallow builds its bird-nest in March*, where the noun phrase receives a kind reading. We want to capture these two different usages of nouns within BIOfid. To this end, the distinction between SPECIFIC and UNSPECIFIC is introduced. As a rule of thumb, the following question guides the specific/unspecific distinction: Is the author speaking as an eyewitness? If *yes*, the annotation unit is a specific one; if *no* (e.g., if the author refers to general knowledge), the annotation unit is unspecific.

2.4.4 Anaphora

So far we have only considered nouns and noun phrases which are used by text authors as part of their real-world observations. This leads to the question of how to deal with nominal expressions that are used in other ways, most importantly anaphorically ones (nominals whose interpretation rest on their linguistic context). We have to consider two main classes in this respect, namely pronouns and anaphorically used definite noun phrases. Both kinds of expressions refer back to some preceding noun phrase in the text.[22] We can find examples of both types of nominal expression in the example extract in Sect. 2.1. In the final sentence, the plural pronoun *their* occurs, referring back to *The Corvidae*. The second sentence starts with *The birds*, referring back to *Corvidae* from the initial sentence. Hence, there are two mentions of *Corvidae* without using that name! However, a pronoun receives its interpretation from its antecedent, a computational linguistics task known as *anaphora resolution* (Mitkov, 2013). For that reason, pronouns are ignored, they do not constitute a markable in BIOfid.

In contrast to pronouns, anaphoric noun phrases exhibit a descriptive content that can be annotated. Although the noun phrase *The birds* picks up *Corvidae*, it

[21] Such issues of reference have for long been discussed in (philosophical) semantics (Donnellan, 1966; Kripke, 1977; Russell, 1905, 1910/1911).

[22] We ignore cataphoric uses here for the sake of simplicity, where an expression "refers forward". Except for direction, cataphoric and anaphoric uses work very similar.

nonetheless is about birds and can be labeled accordingly (i.e., $\overline{\text{ANIMAL}}$). Hence, anaphoric noun phrases are labeled according to their descriptive information.

As of the time writing, there are 79,813 "net" annotations (5877 of rank ENTITY and 73,936 of rank $\overline{\text{CONCEPT}}$, cf. Table 1 in Sect. 4.2). These annotations have been carried out according to an annotation hub procedure.

3 The annotation hub framework

"Annotation [...] can be a complex process potentially involving many people, stages, and tools [...]" (Finlayson & Erjavec, 2017, p. 168). Since in BIOfid, tool development and the interaction of manual and automatic (i.e., machine learned) annotations are intertwined, the annotation process involves a more complex infrastructure than acknowledged in usual annotation process models. We will sketch the annotation process in the following.

In this work, we follow MATTER (Pustejovsky & Stubbs, 2012), a conceptual framework for annotation projects. MATTER is an acronym derived from *Model*, *Annotate*, *Train* and *Test*, *Evaluate*, and *Revise*. The MATTER framework describes an annotation project cycle that focuses on machine learning (hence the "TT"). The "A" phase of MATTER includes a so-called MAMA cycle (*Model-Annotate-Model-Annotate*; Pustejovsky & Stubbs, 2012, Chap. 6).[23] Within MAMA, data annotation, guideline specification, and evaluation are re-iterated until a reasonable agreement scores are achieved. However, BIOfid starts with externally trained classifiers which are progressively refined by manual annotation.

MATTER has been extended soon, namely acknowledging additional phases called *Idea* (before *Model*), *Procure* (after *model*), and *Distribute* (final step of data distribution) (Finlayson & Erjavec, 2017). With respect to BIOfid, the *Idea* is introduced in Sect. 1, *Distribution* is covered in Sect. 5, where the semantic search portal is described which provides access to biological retrieval. In the following we describe our divergence from extended MATTER and the specific organization of the modified MAMA cycle in terms of what we call the *Annotation Hub Framework*.

Both MATTER and extended MATTER provide a revision step only after evaluation (MATTER) and eventually also during annotation (extended MATTER)—see the left panel in Fig. 1. From the practical perspective of annotation projects, in particular more complex ones, this is far too late. To make this clear: late revision would mean that after having done thousands of annotations (recall that machine learning is a goal) and getting an evaluation that is below the project standard, the annotation process has to start again in the *Model* stage (e.g., resource selection, guidelines, specification language, *etc*). All that can easily take a year and in the end it is unclear which results will be achieved.

A slightly different view is taken by the *iterative reliability testing* model of Artstein (2017), where full-scale annotation only starts after a reliable annotation schema has been established. Iterative reliability testing can straightforwardly be integrated in a slightly modified extended MATTER model, namely in terms of an

[23] We are thankful to an anonymous reviewer for pointing out the incorporation of MAMA into MATTER.

Table 1 The number of net annotations for each class, that is, the number of annotation after merging the different annotation views and resolving conflicts between overlapping annotations of the same class from different annotators

Class	Entities	Concepts	Words
Act, action, activity	21	1961	2196
Animal, fauna	4	1843	2532
Archaea	4	0	4
Artifact	305	1915	3439
Attribute, property	150	13,496	19,285
Bacteria	1	78	100
Body, corpus	7	2320	2580
Chromista	1	1	2
Cognition, ideation	28	2103	2591
Communication	84	866	1568
Event, happening	16	1499	1668
Feeling, emotion	9	811	909
Food	12	439	468
Fungi	3	414	1078
Group, collection	329	2473	3686
Habitat	0	777	859
Location, place	1655	3700	6938
Morphology	8	4659	6902
Motive	2	437	617
Natural object	57	967	1085
Natural phenomenon	22	686	752
Person, human being	2461	1112	5565
Plant, flora	146	8497	15,058
Possession, property	2	207	223
Process	6	861	940
Protozoa	0	5	5
Quantity, amount	28	4225	5940
Relation	52	3983	4978
Reproduction	5	658	671
Shape	3	1096	1238
Society	8	176	194
State, condition	17	2508	2861
Substance	25	1585	1717
Taxon	117	5808	13,006
Time	287	1768	3548
Viruses	2	2	4
Total	5877	73,936	115,207

Note that this table also includes annotations from documents that have only been annotated by a single annotator. This is why the total numbers differ between this table and Table 2

extended *Procure* stage. Since in BIOfid the annotation scheme is mainly fixed on the *Idea* and *Model* stage, and NLP routines draw on classic NER, it was opted for a dynamic combination of procuration and annotation: In parallel to the *Annotate* stage, meetings and sessions took place to build up annotation conventions (cf.

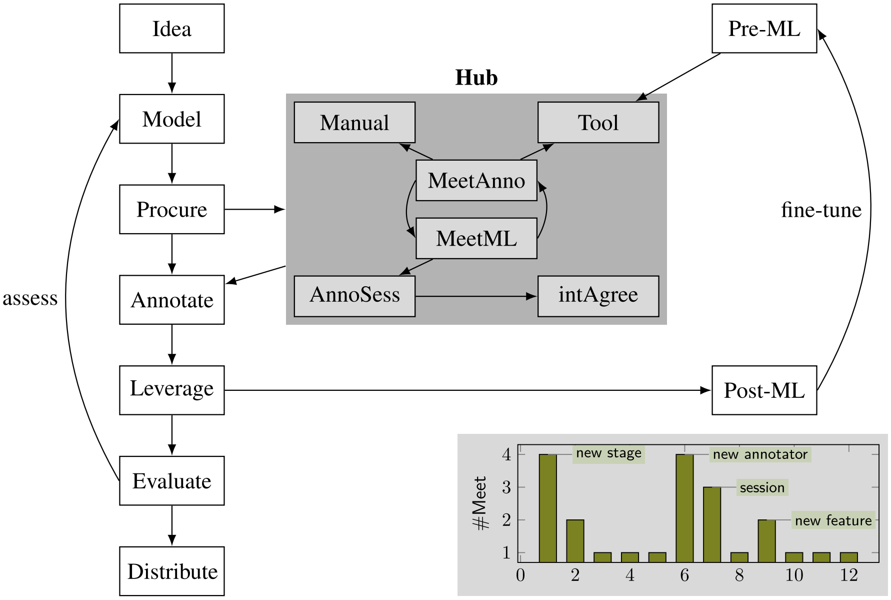


Fig. 1 Annotation Hub Framework (the inlay figure shows a tentative frequency distribution of meetings per month over one year; a new stage typically is an extension of the annotation, a new feature usually is a modification (improvement) of the annotation tool)

Sect. 2.2). Simultaneously, classifiers obtained from machine learning are used to pre-process the texts to be annotated. The classifiers are updated by the manual annotations, so that a bootstrapping cycle is initiated which should lead to an iterative improvement of annotations in both quantity and quality. This *Annotation Hub Framework* is sketched in Fig. 1. As we discuss in Sect. 4.3, we obtained mixed results from this strategy, however. In the following, we focus on manual annotation. The output of ML is fed in the publicly accessible BIOfid-portal (cf. Sect. 5).

In the *Idea* and *Model* stages some annotation scheme has to be chosen as a starting point. The starting point is not (or should not be) changeable anymore, since changes of this basic level amount to a re-start of the whole project. The starting point can be extended, though (in ISO annotation standards this is achieved by *plug-ins*, see Bunt, 2019). This approach is also followed in BIOfid. The starting point is the general *WordNet*-based classification and the biology-specific extension (cf. Sect. 2.1). The according annotation is the first layer annotation. Since temporal and location information turned out to be too underspecified in terms of broad categories *Time* and *Space*, a second annotation layer has been implemented on top of the first one. To this end, *GeoNames* and *ISOTimeML* have been chosen as more fine-grained representation formats (see Sect. 2.3). They constitute another layer of annotation, not a re-start of annotation, since the first layer annotation remains untouched, except that annotation units carrying a *Time* or *Loc* tag have additionally be classified according to the plugged-in schemes. Note, however, that every newly added annotation layer has to proceed through all the steps of an annotation process summarized in Fig. 1. This means, in particular, that the annotation guidelines have to be extended, that annotation tools are equipped with appropriate annotation functions, and that guidelines and tools have been discussed in meeting (*MeetAnno*) tested in probing annotation sessions (*AnnoSess*), which may result in intermediate agreement values (*intAgree*). This phase takes from about two weeks (for simple conventions like proper name annotation, which we elaborate shortly) to about two months (larger extensions such as the GeoNames/ISOTimeML one, which will be taken up in Sect. 4.1).

The ML part in BIOfid is constantly updated. To this end, learning is carried out on more and more external resources, mainly Wikipedia and related (i.e., hyperlinked) resources. Agreement results obtained from manual annotations are used in order to detect problematic categories. For some problematic categories specific *AnnoSess* have been devised: the task in these *CategoryConfirmation* sessions is to correct automatically generated labels on sample sentences only of specific categories in order to give feedback to ML. However, results seem to be both weak and mixed (concrete results are still to be gained, however). The ML procedure and its evaluation will be the topic of a paper on its own.

One might propose to avoid costly looping through the hub by devising a more complex annotation scheme from the outset. While this sounds like a reasonable proposal from the point of view of process optimization, it lacks ecological validity. On the one hand, it assumes that all needed or desirable extensions are known in advance, which is often not the case. On the other hand, it does not pay due attention to the interpretation effort that has to be provided by annotators, as is discussed in terms of cognitive load in Sect. 4.3. Thus, we make a plea for layered annotations.

Let us briefly make things more concrete by means of a rather straightforward example: proper name annotation. The proper name of a person is labeled as PER. That much is clear and can easily by formulated as an annotation guideline. However, proper person names are often realized as multi-tokens, for instance in case of prename–family name pairs. The reasonable thing to do now is to combine prename and family name and assign PER to the concatenated token. In order to do so, however, the annotation tool need to have the facility to create multi-tokens. Accordingly, a tool development step has to follow, during which proper name annotation at least of multi-tokens has to rest (adopting the guidelines is in this case easy, though). In further annotation sessions annotators encounter instances of proper person names that are prefixed by a title, which raises the issue in annotation meetings of how to deal with the prefix: should it be part of the multi-token, or annotated separately, say, as $\overline{\text{PERSON}}$? Since this a not a new problem, there are already guidelines for proper name annotations, such as Benikova et al. (2014), which can be relied on (prefixes are ignored, by the way). Agreeing on conventions even in such apparently minor cases is nonetheless important: the difference between the occurrence of an annotation label and the lack of an annotation label for some annotation unit simply amounts to disagreement and influences the *Evaluate* step.

The more is known of such minor complications in advance, the less scheme and tool development phases are required. This leads to the topic of expertise. In order to oversee (and foresee) potential sources requiring conventions and—equally important—to design reasonable annotation layers that prevent re-starts, acquaintance with texts, annotation procedures, and the annotation scheme landscape is helpful.

4 Implementation

In order to implement the requirements needed in the annotation process, we apply already existing tools in the field of Natural Language Processing (NLP) and digital humanities. In this context, a distinction between tools for automatic pre-processing of texts, and manual post-annotation or primary annotation is necessary. The automatic pre-processing, including the transformation of the ABBYY XML format (Adobe FineReader), tokenization, lemmatization, part-of-speech-tagging, named entity recognition, and automatic entity linking, is performed by TEXTIMAGER (Hemati et al., 2016). Furthermore, for the manual correction of possibly incorrect annotations and for the generation of training data for subsequent machine learning processes, TEXTANNOTATOR (Abrami et al., 2019; Helfrich et al., 2018) is used. At the time of beginning the BIOfid project, TEXTANNOTATOR was the only tool for UIMA-based collaborative and simultaneous multiple annotations.[24] Both tools utilize *UIMA* (Ferrucci et al., 2009), which is the de facto standard used in NLP for processing text corpora. Using UIMA, texts can be processed and ported to an XML standard (XMI), whereby a wide range of already existing tools for the individual text levels is available for pre-processing. Furthermore, both tools are integrated into a stable and flexible infrastructure (cf. Fig. 2).

The need within BIOfid to pre-process texts, implementing the current technical standards, and allowing for multiple annotation are the reasons why we do not make use of some existing bio-annotation tools. There are several XML schemas for taxonomic annotation (see Penev et al., 2011 for an overview). However, these schemas are restricted to a particular set of taxa, and do not implement a general and multiple annotation schema. In order to apply an XML schema to a digitized text, an editor such as GoldenGATE (Sautter et al., 2007) is required. While GoldenGATE allows to insert XML tags over a given text span and even provides an interface to NLP pipelines, it meanwhile lags behind the functionality, interoperability, and usability of the above-mentioned tools. Another desktop application is Phenex (Balhoff et al., 2010). It is developed for using ontologies for phenotypic annotations. Thus, it is not appropriate for the multiple annotation approach pursued within BIOfid. Both mentioned XML editors are not browser-based, which leads to a number of potential problems with regard to technical requirements, copyright restrictions (texts can remain on the server), and parallel manual processing of the text to be annotated.

[24] By now this feature may be provided by other platforms as well (e.g., Klie et al., 2018). Furthermore, since the tools are developed by project members, they can be specifically adapted to the agile annotation hub (cf. Sect. 3).

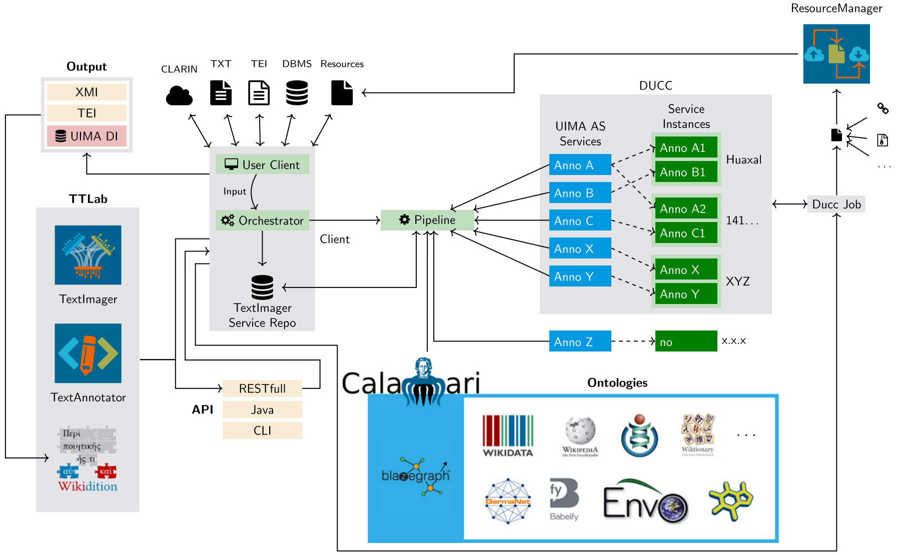


Fig. 2 The infrastructure, in which TEXTANNOTATOR and TEXTIMAGER are integrated, enables UIMA-supported processing of various resources via the *ResourceManager* (top right) as part of the *eHumanities Desktop*. A large number of different tools and languages are available for text processing (see Table 3). For large data volumes, the UIMA DUCC ("Distributed UIMA Cluster Computing") service is used, which allows processing to be upscaled on multiple servers. *Calamari* provides the ontologies required for individual pipelines. After the automatic pre-annotation, the texts can be exported in the UIMA exchange format (XMI), in TEI or can be stored in a database management system via the *UIMA-Database-Interface* (Abrami & Mehler, 2018). Afterwards, pre-processed texts can be used in *Wikidition* (Mehler et al., 2016) or further processed via TEXTANNOTATOR. All processing and annotation rights are regimented by an elaborate rights management

In short, we decided for BIOfid to use state-of-the art annotation technology and eventually adapt it to the BIOfid needs. The latter is possible since many of the tools are "homemade" anyway, which is in advantage in its own right, although tool development phases may interfere with annotation phases (cf. Sect. 3).

4.1 Tools

Both TEXTIMAGER and TEXTANNOTATOR are established tools for processing, visualizing and annotating textual corpora. However, both tools have their specific focus on annotations: TEXTIMAGER operates as a multi-server, multi-service, multi-application and multi-pipeline (where "pipeline" signifies a series of consecutive pre-processing steps) system for automatic pre-processing of textual corpora.

The tool enables distributed and process-optimized processing of texts via horizontal and vertical process distribution. Additional pipelines as well as new pre-processing software, which accept and process UIMA documents, can be flexibly and rapidly integrated into the existing infrastructure. Using this infrastructure, texts are processed through different pipelines, containing different pre-processing modules in different languages—details are given in Appendix C.

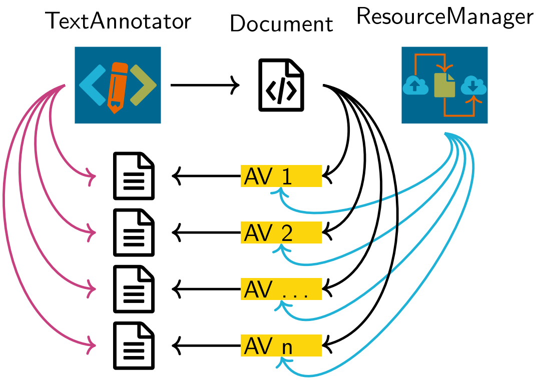


Fig. 3 Schematic diagram of the use of *Annotation Views* (AV). TEXTANNOTATOR has access to documents, which hold all their AVs that are accessible by the users. Using TEXTANNOTATOR to edit annotation texts is always bound to a tool- and user-specific AV

Documents pre-processed by TEXTIMAGER and transformed into the UIMA format can be used by TEXTANNOTATOR. TEXTANNOTATOR utilizes its own database approach for managing and using UIMA documents, which uses a MongoDB through the *UIMADatabaseInterface* (Abrami & Mehler, 2018). TEXTANNOTATOR is a browser-based annotation framework, which enables the manual annotation of documents and the management of annotation processes, based on user and group permissions. Resources can be annotated simultaneously and collaboratively in different “views” of content and subject areas with different tools. These views, *annotation views* (AV), are highly relevant for a flexible use, see Fig. 3. Using the annotation views, documents can be divided into logical layers, which can all be made accessible individually via the *ResourceManager* for individual users or groups over access permission settings (Gleim et al., 2012). This means that in annotation projects, many annotators can see only their own user annotation view, depending on their permissions, but the project manager sees all views (Abrami et al., 2020). Therefore, TEXTANNOTATOR facilitates the annotation process with independent annotators in one tool. In addition, the different AVs are also important for the later calculation of annotation agreement to select suitable documents for machine learning. Furthermore, the use of different AVs enables the collaborative and simultaneous annotation of documents.

Within BIOfid, the tool *QuickAnnotator* from the TEXTANNOTATOR suite is especially relevant. Using *QuickAnnotator*, named entity annotations and their corrections as well as word- or phrase-related classifications can be performed rapidly within a web interface, including the creation or separation of multi-tokens (Abrami et al., 2019).

QuickAnnotator is integrated into the annotation process—see the Annotation Hub Framework in Fig. 1. On the one hand, *QuickAnnotator* implements developments which are required for carrying out new annotation *stages* or special annotation *sessions* (cf. Sect. 3), on the other hand usability *features* are regularly added. For instance, extending the simple annotation of times and places by the more complex annotation of ISOTime and GeoNames (cf. Sect. 2.3) was a new annotation stage. In so-called *category confirmation* sessions, ML output for specific categories is checked. To this end, *QuickAnnotator* had to be modified so that only

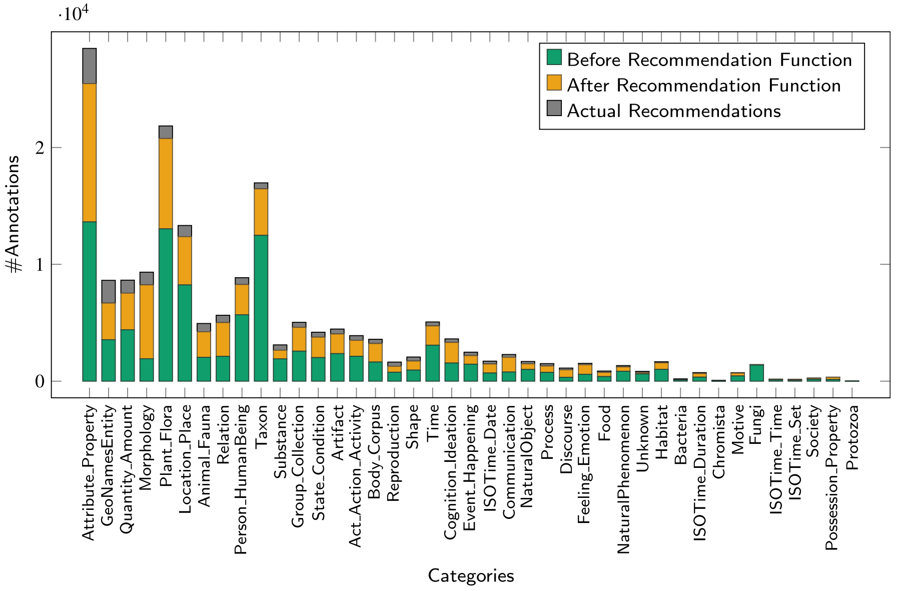
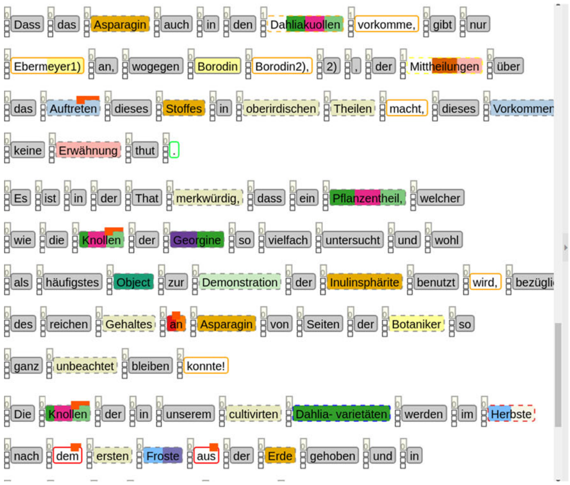


Fig. 4 The distribution of the annotated categories created by the recommendation function, decreasing from left to right (see *actual recommendations*). In total 184,409 category annotations were produced. More specifically, 97,290 annotations were produced before and 69,441 after the recommendation function has been introduced. Given that the recommendation function is a comparatively new feature, the distribution shows that recommendations are helpful because there are more annotations added based on those recommendations. Altogether 17,678 recommendations have been produced and kept

categories at question are displayed and annotations can be given as a binary decision. Preparing and testing tools for stages or sessions takes from about two weeks (category confirmation) to two months (ISOTime and GeoNames).

Annotation meetings (cf. Fig. 1) have often been the place where desirable tool features have been discussed. Going beyond basic functionality (labeling, token merge, undo/redo, comments), about half of additional features have been realized in response to suggestions for improving annotation convenience (from the latest six features, three are due to suggestions from annotators, three to tool development schedule). An example is the advanced coloring scheme illustrated in Fig. 5.

New features can also affect the annotation *process*, not only annotation convenience. To speed up the annotation of recurring words, *QuickAnnotator* includes a recommendation function, which automatically adds the annotation labels chosen by the annotator to all occurrences of the lemma of the annotation unit in question within a selectable text span (sentence-, paragraph-, text-level). All recommendations are marked as such and can be individually revoked. The number of recommendations per category generated in this way is given in Fig. 4. Note that the figure contains "gross" numbers: every annotation of each annotator has been counted irrespective of whether they agree or disagree on a given annotation unit. Multi-tokens are excluded, since they are out of the scope of the recommendation function.



Dass das Asparagin auch in den Dahliakuollen vorkomme, gibt nur
Ebermeyer1) an, wogegen Borodin Borodin2), 2) , der Mittheilungen über
das Auftreten dieses Stoffes in oberirdischen Theilen macht, dieses Vorkommen
keine Erwähnung thut .

Es ist in der That merkwürdig, dass ein Pflanzentheil, welcher
wie die Knollen der Georgine so vielfach untersucht und wohl
als häufigstes Object zur Demonstration der Inulinsphärite benutzt wird, bezügli
des reichen Gehaltes an Asparagin von Seiten der Botaniker so
ganz unbeachtet bleiben konnte!

Die Knollen der in unserem cultivirten Dahlia- varietäten werden im Herbste
nach dem ersten Froste aus der Erde gehoben und in

Options

Page 1 of 22

Normal Annotation

Rapid annotation

General

Anomaly | Diskurs | Endmarker | Historic Typo
OCR Typo | Orthographic Typo | Multiple Annotations

Please add the correction, if possible, via the comment field.

Annotation Categories

Entities | Concepts

Animal, Fauna	Archaea	Bacteria	Chromista
Fungi	Habitat	Lichen	Location, Place
Natural Object	Group, Collection	Attribute, Property	Person, Human Bein
Plant, Flora	Protozoa	Taxon	Time
Viruses	Artifact	Body, Corpus	Food
Natural Phenomenon	Substance	Society	Morphology
Possession, Property	Quantity, Amount	Relation	Reproduction
Shape	Act, Action, Activity	Cognition, Ideation	Communication
Event, Happening	Feeling, Emotion	Motive	Process

Fig. 5 The *QuickAnnotator* interface: The text is visualized token by token, which can be combined into multi-tokens. Tokens that receive an annotation are colored in the same colors as the annotation categories on the right side of the interface. Multi-label annotations are striped in various colors

Enabling annotation stages and sessions is obligatory to accomplish an annotation task. Adding new features usually is an optional add-on. Therefore, annotation meetings are also the place for deciding on which further features are actually to be realized and which may be ignored, so that tool development eventually can come to a conclusion.

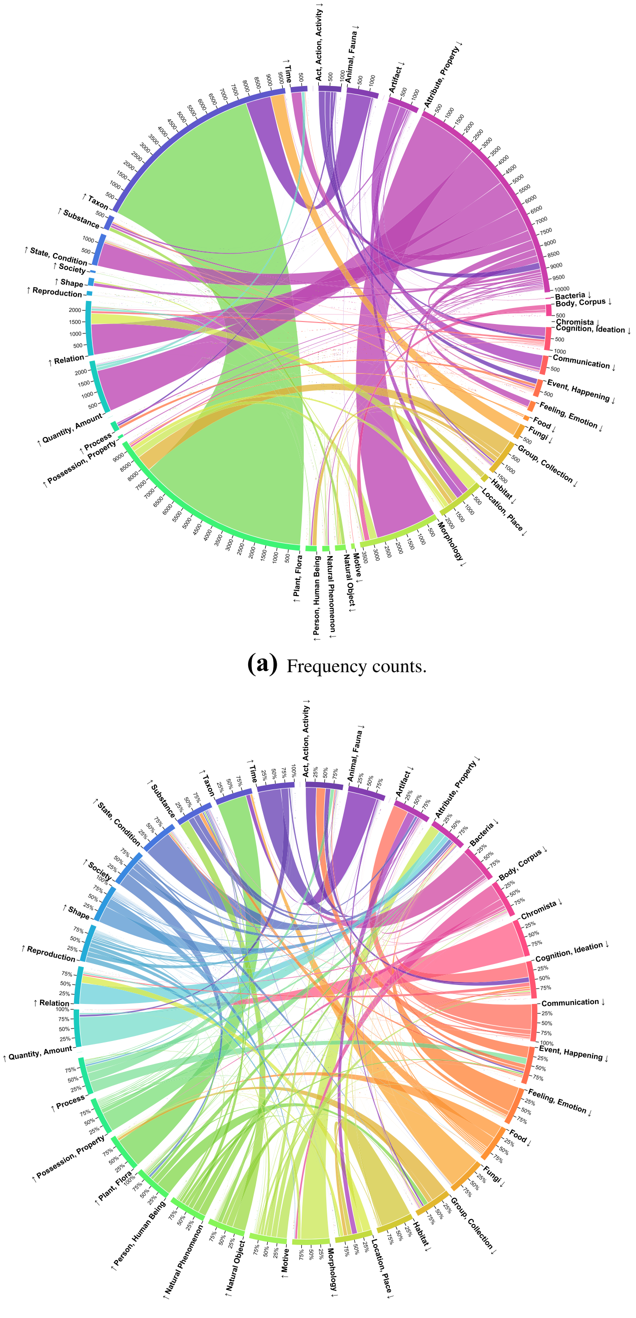
4.2 Corpus statistics and inter-annotator agreement

The texts included in the annotation process described here represent only a subset of the total BIOfid corpus, which will be semantically enriched for BIOfid users in the coming years. This subcorpus comprises articles from 15 different scientific journals on the biodiversity and ecology of organisms published between 1858 and 1914. A complete list of the individual texts is given in Appendix D. The journals are part of the (freely available) *German Botanical Journals Collection*[25] and are also partly available via the *Biodiversity Heritage Library (BHL)*.[26] The annotated subset of this collection primarily encompasses the flora (especially of botanical gardens in mountain environments) and terrestrial fauna (mammals, birds, and insects) of Central Europe. To collect the texts and metadata from BHL automatically, we created a generic harvesting tool.[27] Texts from the German Botanical Journal Collections were handled manually, since they reside in the database of a project partner. These texts contain 10,907 sentences, which in turn comprise 139,166 word form occurrences (151,783 tokens including punctuation marks).

[25] http://sammlungen.ub.uni-frankfurt.de/botanik?lang=en.

[26] https://www.biodiversitylibrary.org/browse/contributor/UBJCS#/titles.

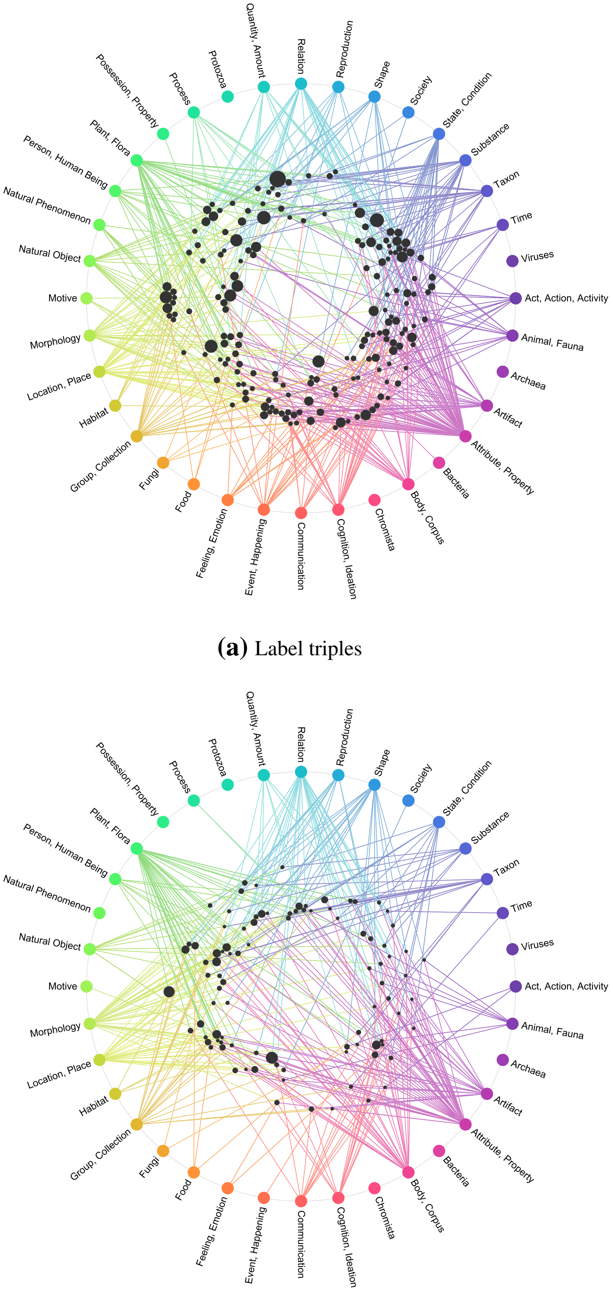
[27] https://github.com/FID-Biodiversity/LiteratureCrawler.



(a) Frequency counts.

(b) Frequency percentages.

Fig. 6 Distribution of label pairs (please enlarge on screen to see more details)



(a) Label triples

(b) Label quadruples

Fig. 7 Distribution of label triples and quadruples (please enlarge on screen to see more details)

In sum, eight annotators carried out manual annotations. During annotation, wrong labels from pre-processing have been corrected and missing labels have been added—the lion's share of annotation work. The following figures all refer to the manual annotation.

There are 79,813 "net" annotations (5877 of category ENTITY and 73,936 of category $\overline{\text{CONCEPT}}$). By "net" we mean the following: each file is annotated by at least two annotators (this is done in order to be able to assess annotation consistency). This means that for any given word there can be two (collections of) labels. These annotation labels may be partially overlapping (when annotators partially choose the same labels for the same word), or they may be (partially) disjoint (when annotators assign different labels to the word in question). Since a reduplication of annotation labels does not provide additional information for a given annotation unit, we select only unique annotation labels. For that reason, the number of annotations *per* class does not necessarily sum up to the number of annotated words. The numbers of net annotations are summarized in Table 1.

According to the multiple annotation approach, namely that multi-labeling is the rule rather than the exception, there are 61,495 multi-annotations (as of October 27, 2020). In detail: 1 label: 22,219, 2 labels: 27,601, 3 labels: 9512, 4 labels: 1674, 5 labels: 387, 6 labels: 95, 7 or more labels: 7. The majority of annotations consists of 1 to 4 labels. *Are there patterns of multi-annotations* (2 to 4 labels)? In order to obtain an answer for this question, pairs of labels (concerning the case of exactly 2 annotations) are plotted in a chord diagram by means of D3.js (www.d3js.org)—see Fig. 6. The figures are to be read as follows: The outer circle hosts the annotation labels. Their absolute frequency (Fig. 6a) and their percentage frequency (Fig. 6b) are given in the direction of the arrow attached to each label. The arc edges encode the co-occurrence frequencies of two labels. Most taxa (about 6500, or 75%), for instance, refer to the $\overline{\text{PLANT}}$ kingdom, and the majority of communications are artifacts, that is, text products rather than communication events.

In order to display combinations of more than two labels in a chord diagram, an additional layer of elements was added in terms of so-called hypernodes, that is, reifications of the "meeting points" of three or more labels as elements of hyperedges. The resulting hypergraph representations for triples and quadruples of annotated labels are given in Fig. 7. The hypernodes are displayed as black circles whose size is determined by the number of label co-occurrences. The label triples (Fig. 7a), respectively quadruples (Fig. 7b) which are involved in each hypernode are connected by label-colored edges to the hypernode. As expected, the larger the subsets of the jointly annotated labels, the smaller (in the sense of thinner and fewer edges) the graphical representations. In any event, *Plant, Flora* as well as *Relation* and *Attribute, Property* are frequently addressed categories irrespective of the size of the hyperedges: *these categories are likely to be the subject of multi-annotations regarding the texts from the field of biology considered here.* Figures 6 and 7 provide a visual overview of the most important relations. Since the project and the annotations are still under development, we refrain from enumerating concrete but intermediate figures.

From the 6938 words annotated as locations, 5566 received a GeoNames classification. There are 2816 ISOTimeML annotations (159 of type *Time*, 2013 of type *Date*, 532 of type *Duration*, and 112 of type *Set*). In addition, there are 1442 annotation units of category $\overline{\text{TIME}}$ without an ISOTimeML specification.

In particular, the multi-label annotations raises the question of annotation *consistency*. Annotation quality can be assessed with different foci of error, namely *stability* (intra-annotator agreement), *reproducibility* (inter-annotator agreement), and *accuracy* (deviation form norm) (Krippendorff, 2018). BIOfid carries out a reproducibility assessment. To this end, the inter-annotator agreement (IAA) module of TEXTANNOTATOR is used, which in turn uses the *DKPro Agreement* module which is based on the *DKPro Statistics* software library (Meyer et al., 2014). Given that the BIOfid annotation scheme supports multiple annotations, we use unitizing studies with Krippendorff's α coefficient for all agreement calculations (Krippendorff, 2018). This allows for the evaluation of annotations of an arbitrary number of annotators where multiple annotations of *different* classes from each author can cover the same span of text.

Assessing IAA furthermore takes into account a few conventions that have been agreed upon during the development and refinement of annotation practices. For instance, every organism is trivially also a NATURAL OBJECT. Since we know this in advance, having classified an annotation unit as, say, ANIMAL, it is not informative any more to assign it to NATURAL OBJECT, too. Such dependencies are collected in an *inclusion hierarchy*. The following inclusion relations are acknowledged (where "$x \prec y$" means that y is subsumed by x): BODY $\prec$ MORPH, LOC $\prec$ HABITAT,[28] COGNITION and EMOTION $\prec$ MOTIVE, ARTIFACT $\prec$ POSSESSION, BODY $\prec$ NATURAL OBJECT, MORPH $\prec$ NATURAL OBJECT, and GROUP $\prec$ SOC. Note that the category NATURAL OBJECT is completely "absorbed" by more detailed categories, it therefore has not been used as an annotation label by any annotator. The manual annotation is extended by these conventions before the unitizing agreement study is carried out. IAA is calculated on all completed annotations of all documents that have been annotated by at least two annotators. GeoNames and ISOTimeML annotations are evaluated on the most general level, that is, in terms of TIME and LOC, respectively $\overline{\text{TIME}}$ and $\overline{\text{LOC}}$. Following this approach, the resulting IAA values for each annotation category are collected in Table 2. Note that the baseline of Krippendorff's α is agreement by chance, which is $\alpha = 0$.[29] Disagreements are made use of in two heuristic respects. Firstly, they indicate categories for which it is difficult to develop a shared understanding. Secondly, multiple and even diverging annotations provide information for potential gaps in biological ontologies. To this end, manually annotated data are searched for specific categories and compared to ontology entries. By this method, in particular not yet documented vernacular names can be identified and added to the ontology.

[28] Since the distinction between entities and concepts also applies to locations, a habitat can be a spatial instance (e.g., *Black Forest*) as well as a set of multiple locations (e.g., *desert*). We are thankful to an anonymous reviewer for pointing this out.

[29] In latest test annotations sessions (cf. Sect. 3), IAA values of around 0.7 on average have been reached. The values in Table 2 obviously are worse. This can be due to having randomly chosen an exceptional "easy" session sample, or due to an increased shared understanding over time—or a mixture of both.

Table 2 Inter-annotator agreement values for all classes and their respective number of occurrences

Class	Entities		Concepts	
	Agreement	Count	Agreement	Count
Act, action, activity	– 0.0008	21	0.4292	1815
Animal, fauna	– 0.0031	20	0.8298	2052
Archaea	– 0.0002	4	0.0000	0
Artifact	0.1074	209	0.9256	4587
Attribute, property	0.1184	253	0.6109	14,473
Bacteria	0.0000	1	0.6788	101
Body, corpus	– 0.0030	15	0.4910	4358
Chromista	0.0000	1	0.0000	1
Cognition, ideation	0.1415	25	0.3888	2075
Communication	0.0247	62	0.3163	655
Event, happening	0.1919	21	0.2944	1426
Feeling, emotion	– 0.0008	12	0.2696	1099
Food	– 0.0040	11	0.2900	435
Fungi	– 0.0033	2	0.9638	823
Group, collection	0.3937	389	0.3162	2368
Habitat	0.0000	0	0.7923	1083
Location, place	0.7436	2559	0.7839	4922
Morphology	– 0.0014	5	0.4082	3250
Motive	– 0.0017	2	0.1220	369
Natural object	0.0394	71	0.7841	7485
Natural phenomenon	0.1066	32	0.3784	696
Person, human being	0.8054	2450	0.6348	1328
Plant, flora	– 0.0292	154	0.8903	11,032
Possession, property	– 0.0004	2	0.1266	178
Process	– 0.0010	4	0.2624	841
Protozoa	0.0000	0	– 0.0032	5
Quantity, amount	0.0779	164	0.5795	4403
Relation	– 0.0164	42	0.0729	2816
Reproduction	– 0.0001	3	0.4295	749
Shape	– 0.0006	2	0.2950	1088
Society	0.2774	9	0.0331	139
State, condition	– 0.0006	17	0.1767	2318
Substance	– 0.0019	24	0.4933	1942
Taxon	– 0.0041	116	0.9250	9348
Time	0.4694	409	0.6053	2189
Viruses	0.0000	2	– 0.0001	3
Overall	0.6043	7113	0.6387	92,452

The agreement was calculated from annotations of a total of seven different authors, where each document was annotated by at least two and at most four authors. The overall agreement was computed as the average of each category's agreement weighted by the number of annotations

4.3 Discussion

The IAA values are divided in two ways, according to the distinction into ENTITY and $\overline{\text{CONCEPT}}$ (cf. Sect. 2.1) and according to individual annotation categories. Since a taxonomic name (TAXON) cannot belong to category ENTITY *by definition*, the 117 occurrences documented in Table 2 are probably due to "slips of fingers" during annotation. With regard to $\overline{\text{TAXON}}$, annotators were quite consistent ($\alpha = 0.925$).[30] Apart from accidental slips, which might happen in larger-scale annotation projects (we number the two occurrences of *Chromista* to this class, too), we argue that there are three reasons for lower agreement values: (i) *cognitive load*, (ii) *lack of compositionality* (*semantic gaps*), and (iii) *under-specificity*.

4.3.1 Cognitive load

The annotation process is influenced by the cognitive load that is imposed on the annotator: How many decisions are to be made with regard to each annotation unit? Within BIOfid one might argue that annotators carry a too heavy burden. First, they have to decide on the ENTITY or $\overline{\text{CONCEPT}}$ distinction. Then they have to evaluate the textual evidence for detecting speaker's reference. Only now can they ponder the multiple annotation labels that apply to the annotation unit. The annotation should follow the guidelines, and a potential phrasal annotation has to be considered. Typographic or OCR errors have to be marked *en passant*. That is, annotators are forced to glance at the same annotation unit from various angles: they have at least to apply a mixture of word classification (multiple annotation), pragmatic interpretation (speaker reference, possibly non-literal uses) and syntactic parsing (phrasal annotation). It is obvious that this process is a demanding one which for that reason is error-prone. Projects, that want to carry out a rather complex annotation should think about structuring annotations in different layers (so that at each annotation layer only one annotation task has to be dealt with).

4.3.2 Lack of compositionality

Annotation labels are natural language words (English, in our case). Any combination of words induces functional dependencies, an implicit level of compositional structure. However, the ontological categories (Sect. 2.1) are assumed to be functionally independent, which may lead to semantic gaps. A prevalent domain of functional dependencies is mereology, which is involved, for instance, in the composite *Knollenscheibe* "*"slice of a tuber. Obviously, it is a man-made part of a plant, and it is easy and understandable to interpret the multiple annotation $\overline{\text{PLANT}}$, $\overline{\text{ARTIFACT}}$, $\overline{\text{BODY}}$ in this way. Strictly speaking, however, this is an incorrect annotation: a slice of a tuber is not a plant. But $\overline{\text{BODY}}$ (and $\overline{\text{MORPH}}$) has a

[30] Note that a taxonomic expression that is labeled by one annotator but not by the other leads to disagreement. Such a situation can happen, for instance, if the taxonomic name occurs within the header of an article in a footnote reference, where it should not be annotated, but is by one annotator mistakenly taken as a sentence of the main text.

built-in functional dependency, namely being a body part (or, with regard to the term morphology, a body part or a body quality, respectively) *of something*. Ignoring such dependencies leads to semantic gaps; filling these gaps may lead to inconsistent annotations.

4.3.3 Under-specificity

The WordNet categories are chosen since they provide a general classification scheme for common nouns. 'General' means that they have little descriptive content, or little power in constraining denotations.

Now, there are also very general common nouns that stand out due to little descriptive content. For example, consider *Verbreitung* "*"distribution. In the context of BIOfid texts, *Verbreitung* refers to the spreading of a kind within a geographic area, but is silent about any spatial or quantitative properties of the spreading. Now an "accumulation" of generality may lead to under-specification: should *distribution* be categorized as a locational attribute, a natural phenomenon, or a state/condition? Furthermore, since a distribution is always the distribution *of something*, also a functional dependency may be invoked, which leads back to compositional annotation.

Cognitive load can be managed by breaking down complex tasks into annotation stages (cf. Sect. 3). Furthermore, annotators themselves develop routines so that they carry out annotations phenomenon by phenomenon.

Lack of compositionality and under-specificity potentially aggravate each other: intuitively, the more general a category, the more functional dependencies it may be included in. This is difficult to avoid, but points at a further development: instead of sets of annotation categories, one can think of "annotation mini-grammars" which define a basic label syntax according to functional patterns. We leave this to future work.[31]

5 The BIOfid-portal—a semantic document retrieval machine

One goal of BIOfid is to make the texts, that were enriched both manually and automatically with ontological classifications, easily available for a larger audience, in particular bio-scientists. For this purpose, we implemented a semantic search web portal[32] (hereafter referred to as "BIOfid-portal") that makes use of the described pre-processed texts and tools (Sect. 2 and 4), as well as biological taxonomic ontologies (containing the hierarchical structure of kinds). The BIOfid-portal interprets the user query semantically and returns documents fitting the given query. Thanks to the ontological classification of the words and the biological

[31] It should be mentioned, however, that this direction would pose some challenges to machine learning: it is still an open question, how or whether at all neural networks model compositionality. For instance, systematicity (the ability to recombine known parts), the most relevant compositional feature for our mini-grammar interests, is not truly followed in current deep learning frameworks (Hupkes et al., 2020, p. 288).

[32] https://www.biofid.de/en/search.

taxonomy ontologies, the retrieved documents do not necessarily have to include the searched taxon name *verbatim*, but may contain a vernacular, synonymous, deprecated, or sub-species name. Consequently, when the user searches, for instance, for *Fagaceae* (the family comprising beeches and oaks), the BIOfid-portal not only retrieves documents containing this name, but also those mentioning plant species that belong to the *Fagaceae*—even though the *Fagaceae* are not specifically mentioned in the document.

The BIOfid-portal implements semantic technologies developed within the BIOfid context in multiple ways (Fig. 8). In a pre-processing step, biodiversity texts are semantically annotated (Sects. 3 and 4) to have LOCATIONs, PERSONs, and (most importantly for BIOfid) TAXONs in the text annotated and possibly linked to an ontological unique resource identifier (URI). In the pre-processing, the *UIMA* XML data, that is returned from the text processing pipeline, is restructured to a TEI[33]-like format for both human reading and HTML presentation in the BIOfid-portal. The pre-processed TEI-like texts can be indexed in the document database[34] by using a plugin that reads the annotated properties for each (multi-token) word and indexes them at the same position in the document as the word.[35] This pre-processing allows us to search for both a URI or a string in the document database and obtain the relevant documents. Hence, we can apply a graph database,[36] to search for species with specific properties (e.g. red flowers) or belonging to a systematic group (e.g. owls), and feed the resulting URIs directly to the document database.

This complexity is abstracted for the user by a search interface that translates the user query to a SPARQL query and subsequently to a list of URIs that are searched in the document database. For this purpose, a text processing pipeline, that includes taxon recognition (Ahmed et al., 2019), is used in the BIOfid-portal for *ad hoc* user query annotation and analysis of (currently only) German texts. The tokens in the user query are analyzed for their dependencies, so that species attributes (e.g. "Plants with actinomorphic flower symmetry and red flowers") can be translated properly ("actinomorphic" describing the "flower symmetry" and "red" the "flower"; with both attributes referring the "plants") to a SPARQL query by a set of rules.

Finally, the BIOfid-portal displays all relevant documents to the user. Relevant annotations are highlighted and interactive. The user also can download the documents and the respective annotated text previews for further processing on their desktop PC.

Hence, the BIOfid-portal enables bioscientists to harvest annotated texts automatically and to post-process these in their own software, e.g. searching for the reference of specific taxa and locations. Still being under development, the BIOfid-portal progressively makes more features and increased text bases available.

[33] https://tei-c.org/guidelines/p5/.

[34] Apache Solr.

[35] https://github.com/GrazingScientist/TaggedTextTokenizer.

[36] OpenLink Virtuoso.

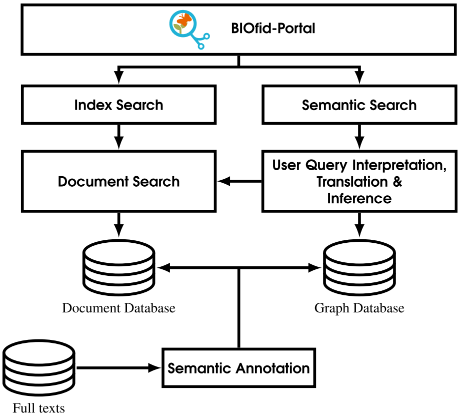


Fig. 8 Schematic data flow of a user query within the BIOfid-portal. First, full texts are semantically annotated (Sects. 3 and 4) and stored in the document database (bottom). Now, the user can query either a classical index search (left branch) or a semantic search (right branch). When selecting the semantic search, their query is semantically processed to retrieve relevant taxa URIs from biological taxonomic ontologies. These URIs are then searched in the document database for relevant documents

6 Conclusions

Within BIOfid an annotation scheme for the multiple annotation of named entities in historic biological texts has been developed. The annotation is used for the fine-tuning of machine learned classifiers and provides the link between the exploration of biodiversity literature and document retrieval via a semantic search portal. Annotation is carried out by specifically modified annotation tools, which are usable beyond the scope of BIOfid. We focused on the development of the annotation scheme and the multiple annotation approach in terms of the *Annotation Hub Framework*. We presented and discussed agreement assessments. A moral drawn is that functional dependencies of multiple annotations should not be underestimated. At least some mereological combinations should be accounted for (cf. the remarks on $\overline{\text{MORPH}}$ in Sect. 2.2), but more elaborate structures, for instance in terms of generative lexical roles (Pustejovsky 1991) or "annotation grammars", are conceivable. Additionally, complex annotation tasks should be broken down into several annotation layers to avoid an overload of annotation decisions to be made.

In future extensions, the noun-centered ontological annotation will be extended by an event-based one. This is implemented as a further annotation stage, inducing a new cycle between ML, tools and annotation in the annotation hub. An event-based annotation classifies a sentence's constituents in terms of the thematic roles they play in relation to the verb. This is a prerequisite for expanding the semantic search portal by facilities resting on relation extraction.

Acknowledgements This work is carried out within the BIOfid project. BIOfid is supported by the *Deutsche Forschungsgemeinschaft* (DFG), Grant Numbers ME 2746/5-1, MO 412/54-1, SCHN 1016/46-1. We want to thank our student helpers for their feedback during the development phase of the annotation scheme and for carrying out the lion's share of the actual annotation work.

Funding Open Access funding enabled and organized by Projekt DEAL.

Appendices

Appendix A: overview of annotation categories

WordNet Categories:	{PERSON, HUMAN BEING}, {ANIMAL, FAUNA}, {PLANT, FLORA}, {GROUP, COLLECTION}, {SOCIETY}, {LOCATION, PLACE}, {TIME}, {COMMUNICATION}, {QUANTITY, AMOUNTS}, {EVENT, HAPPENING}, {NATURAL OBJECT}, {POSSESSION, PROPERTY}, {ATTRIBUTE, PROPERTY}, {BODY, CORPUS}, {FOOD}, {ARTIFACT}, {ACT, ACTION, ACTIVITY}, {PROCESS}, {NATURAL PHENOMENON}, {COGNITION, IDEATION}, {FEELING, EMOTION}, {MOTIVE}, {RELATION}, {SHAPE}, {STATE, CONDITION}, {SUBSTANCE}
Biology-specific Categories:	{TAXON}, {ARCHAEA}, {BACTERIA}, {CHROMISTA}, {FUNGI}, {PROTOZOA}, {VIRUSES}, {LICHENS}, {HABITAT}, {MORPHOLOGY}, {REPRODUCTION}

Each annotation unit is assigned one or more annotation categories. In addition, each annotation unit is has to be specified whether it is of rank *entity* or *concept*.

Appendix B: biology-specific categories

- *Archaea*: The annotation unit is about (a) specimen(s) of archaea.

- *Bacteria*: The annotation unit is about (a) specimen(s) of bacteria.
- *Chromista*: The annotation unit is about (a) specimen(s) of chromista.
- *Fungi*: The annotation unit is about (a) specimen(s) of fungi.
- *Habitat*: The annotation unit is about the living environment of an organism. Obviously, the category habitat actually overlaps with location. However, a habitat involves more than a mere location, since it also refers to specific biotic and abiotic factors characteristic for the distribution of a species. In the context of BIOfid, habitats are therefore marked as such. Examples include proper names such as *Great Barrier Reef*, *Bayerischer Wald*, or *Bodensee* and common nouns such as *Trockenrasen*, *Auen*, or *Hochmoor*.
- *Lichens*: The annotation unit is about (a) specimen(s) of Lichens.

Table 3 Number of NLP services per language available within TEXTIMAGER

	en	de	es	fr	la	nl	pt	zh	it	da	ar	Other	∑
Tokenize	6	4	4	4	3	2	2	1	2	3	3	10	44
Lemmatization	10	4	4	2	5	1	4	0	2	0	0	11	43
POS Tagging	19	11	5	4	5	4	4	4	2	2	2	13	75
NER	15	7	4	4	0	4	1	0	1	1	0	14	51
Parsing	7	3	3	4	0	0	0	5	2	2	3	9	38
Time Rec.	1	1	1	1	0	1	1	0	1	1	1	1	10
Sentiment	3	1	0	0	0	0	0	0	0	0	0	3	7
SRL	3	2	0	0	0	0	0	0	0	0	0	2	7
Wikification	3	3	1	0	0	0	0	0	0	0	0	2	9
Coreference	4	3	0	0	0	0	0	0	0	0	0	2	9
∑	71	39	22	19	13	12	12	10	10	9	9	67	293

- *Morphology*: The annotation unit is about the outward appearance or inwards structure of an organism. Examples include words such as *Blüte*, *Rhizom*, *Femur*, or *Flügel*.
- *Protozoa*: The annotation unit is about (a) specimen(s) of protozoa.
- *Reproduction*: The annotation unit is about anything which is related to the reproduction of an organism. Examples include words such as *Gelege*, *Larve*, *Pollen*, or *Bestäubung*.
- *Taxon*: The annotation unit is about kinds. For organisms of all taxonomic ranks, their scientific or vernacular names (accepted and synonym names) will be tagged by TAXON. In addition to taxon, the kingdom (i.e. animal, plant, bacteria, fungi, archaea, chromista, fungi, protozoa, viruses) or symbiosis (lichens) to which the taxon belongs is annotated.
- *Viruses*: The annotation unit is about (a) specimen(s) of viruses.

Appendix C: list of tools and routines

Before manual annotation, all texts have been pre-processed by the following tools: *SpaCyMultiTagger* (Honnibal & Montani, 2017), *LanguageToolLemmatizer* (https://languagetool.org/dev), *MateMorphTagger* (Bohnet & Nivre, 2012), *FastTextDDC2LemmaNoPunctPOSNoFunctionwordsWithCategoriestextimagerService* (Uslu et al., 2019), *text2cwc* (Uslu, 2020), *FastTextWikipediaDisambigService* (Uslu et al., 2018), *HeidelTime* (Strötgen & Gertz, 2010), *TagMeLocalAnnotator* (Ferragina & Scaiella, 2010), *WikidataHyponyms*, *BIOfidTreeGazetteer*, *EuroWordNetTagger* (the latter tools have all been developed in the *Text Technology Lab*[37]). These tools are a subset from the tools available for the German language—see Table 3 for the number of tools that are available in general.

[37] www.texttechnologylab.org.

Appendix D: List of texts

Title	Author	Published in	Year
Die bayerischen Characeen	Giesenhagen, Karl	Berichte der Bayerischen Botanischen Gesellschaft, hrsg. v. Bayerische Botanische Gesellschaft zur Erforschung der Heimischen Flora	1892
Andreas Allescher	Schnabl, Gustav	Berichte der Bayerischen Botanischen Gesellschaft, hrsg. v. Bayerische Botanische Gesellschaft zur Erforschung der Heimischen Flora	1904
Bericht über Isoetes lacustris Linné und Marsilea quadrifolia Linné	Solereder, Hans	Berichte der Bayerischen Botanischen Gesellschaft, hrsg. v. Bayerische Botanische Gesellschaft zur Erforschung der Heimischen Flora	1899
Das spielerische Element im Leben gefangener Sperlingsvögel	Braun, Fritz	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1907
Neue Beiträge zur Flora der Kreise Danzig (Stadt, Niederung) und Putzig	Preuss, Hans	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1907
Die Moosflora von Grünhagen, Kreis Pr. Holland	Dietzow, L.	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1910
Bericht über den alpinen Garten bei der Lindauer Hütte im Gauertal	Hoock, Georg	Bericht des Vereins zum Schutze der Alpenpflanzen, hrsg. v. Verein zum Schutze der Alpenpflanzen	1913

continued

Title	Author	Published in	Year
Bericht über den Alpenpflanzengarten auf dem Schachen für das Jahr 1912	Kupper, Walter	Bericht des Vereins zum Schutze der Alpenpflanzen, hrsg. v. Verein zum Schutze der Alpenpflanzen	1913
Bericht über den Alpengarten auf dem Schachen für das Jahr 1902	Goebel, Karl von	Bericht des Vereins zum Schutze und zur Pflege der Alpenpflanzen, hrsg. v. Verein zum Schutze und zur Pflege der Alpenpflanzen	1902
Bericht über den Alpengarten bei der Lindauer Hütte im Gauertal	Hoock, Georg	Bericht des Vereins zum Schutze und zur Pflege der Alpenpflanzen, hrsg. v. Verein zum Schutze und zur Pflege der Alpenpflanzen	1907
Bericht über den alpinen Garten bei der Lindauer Hütte	Hoock, Georg	Bericht des Vereins zum Schutze und zur Pflege der Alpenpflanzen, hrsg. v. Verein zum Schutze und zur Pflege der Alpenpflanzen	1911
Ueber den derzeitigen Stand der gesetzlichen Schutzbewegung zu Gunsten der Alpenflora unter besonderer Berücksichtigung der Tätigkeit des Vereins zum Schutze und zur Pflege der Alpenpflanzen : Nachtrag III.	Schmolz, Carl	Bericht des Vereins zum Schutze und zur Pflege der Alpenpflanzen, hrsg. v. Verein zum Schutze und zur Pflege der Alpenpflanzen	1911
Bericht über den Alpenpflanzengarten auf der Neureuth	anonym	Bericht des Vereins zum Schutze und zur Pflege der Alpenpflanzen, hrsg. v. Verein zum Schutze und zur Pflege der Alpenpflanzen	1912
Ein Beitrag zur Kenntniss der fadenbildenden Bacterien	Pommer, Gustav	Mittheilungen aus dem Botanischen Institute zu Graz, hrsg. v. Botanisches Institut Graz	1886

continued

Title	Author	Published in	Year
Der Gehalt der Dahliaknollen an Asparagin und Tyrosin	Leitgeb, Hubert	Mittheilungen aus dem Botanischen Institute zu Graz, hrsg. v. Botanisches Institut Graz	1888
Mitteilungen zur Ökologie einiger sukkulenter Gewächse der Kanarischen Inseln	Burchard, Oscar	Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie	1913
Übersicht über die Lebensbedingungen und den gegenwärtigen Zustand der Pflanzendecke auf der Iberischen Halbinsel	Brandt, Max	Bericht über die Zusammenkunft der Freien Vereinigung für Pflanzengeographie und Systematische Botanik, hrsg. v. Freie Vereinigung für Pflanzengeographie und Systematische Botanik	1914
Beiträge zur Flora des Doi-Sutäp, unter vergleichender Berücksichtigung einiger anderer Höhenzüge Nord-Siams	Hosséus, Carl Curt	Bericht über die Zusammenkunft der Freien Vereinigung der Systematischen Botaniker und Pflanzengeographen, hrsg. v. Freie Vereinigung der Systematischen Botaniker und Pflanzengeographen	1908
Bericht über die botanische Exkursion von Freitag den 13. bis Sonntag den 15. September	Drude, Oscar; Schorler, Bernhard; Naumann, Arno	Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie	1908
Über die Lebensweise der Uferflora	Glück, Hugo	Bericht über die Zusammenkunft der Freien Vereinigung der Systematischen Botaniker und Pflanzengeographen, hrsg. v. Freie Vereinigung der Systematischen Botaniker und Pflanzengeographen	1909
Über die Einwanderung des arktischen Florenelementes nach Norwegen	Wille, N.	Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie	1905

continued

Title	Author	Published in	Year
Verzeichnis der bis zum Herbst 1902 in der Provinz Posen beobachteten Brombeeren/ von Professor F. Spribille	Spribille, Franz	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1903
Über die Beseitigung der Abfallstoffe mit besonderer Berücksichtigung der Posener Verhältnisse und des sogenannten biologischen Verfahrens	Wernicke, Erich	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1903
Pissodes validirostris Gyll.=strobili Redtb.	Torka, Valentin	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1904

continued

Title	Author	Published in	Year
Pilze aus der Umgegend von Alt-Boyen	Vorwerk, Kurt	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1905
Beitrag zur Flora des Kreises Wreschen	Teichert, Kurt	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1906
Botanische Ergebnisse einer Exkursion zwischen Belenczin und Tuchorze (Kr. Bomst) am 2. August 1905	Bothe, H. ; Torka, Valentin	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1906

continued

Title	Author	Published in	Year
Die Eibe	Schönke, .	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1907
Beitrag zur Pilzflora von Brudzyn im im Kreise Znin	Szulczewski, Adalbert	Zeitschrift der Naturwissenschaftlichen Abteilung der Deutschen Gesellschaft für Kunst und Wissenschaft in Posen (Naturwissenschaftlicher Verein) : zugl. Organ der Abteilung für Naturwissenschaften in Bromberg, hrsg. v. Deutsche Gesellschaft für Kunst und Wissenschaft Posen/ Naturwissenschaftliche Abteilung/Deutsche Gesellschaft für Kunst und Wissenschaft in Posen	1909
Vorläufiger Bericht über die im Auftrage des Westpreußischen Botanisch-Zoologischen Vereins in der Zeit vom 3. Juli bis 16. August 1905 ausgeführte botanische Reise	Tessendorff, Ferdinand	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1906
Die Säugetiere und Vögel Konstantinopels und seine Umgebung	Braun, Fritz	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1906
Schloss Neudeck und seine Gärten	Betten, Robert	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886

continued

Title	Author	Published in	Year
Ueberproduktion oder übermässiger Import	Krätzschmar, Hugo	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Berliner Gärtner-Markthalle	Hoffmann, Martin	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Ein Vorschlag zur vortheilhafteren Gestaltung und Besserung unserer geschäftlichen Verhältnisse	Bauer, A.	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
La Mortola : der Garten des Hrn. Thomas Hanbury	Flückiger, F. A.	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Birne: König Karl von Württemberg	Lucas, Friedrich	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Die 11. Versammlung Deutscher Pomologen und Obstzüchter in Meissen vom 29. Sept. bis 3. Okt. 1886	Wittmack, L.	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886

continued

Title	Author	Published in	Year
Mina lobota de la Lave et Lex. (Quamoclit Mina Don.)		Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Seidel's neues Roll-Haus	Hoffmann, Martin	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Die "Dell"	Schrefeld, Otto	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Maschinen, Geräthe zur Obstverwertung und Baumpflege, Obstnachbildungen etc. auf der Meissener Ausstellung/Lilium auratum	Kühn, B. L.	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Lilium auratum	Thüer, L.	Deutsche Garten-Zeitung : Wochenschrift für Gärtner u. Gartenfreunde; Organ d. Vereins zur Beförderung des Gartenbaues in den Königl. Preuss. Staaten und der Gesellschaft der Gartenfreunde Berlins	1886
Meine erzogenen parasitisch lebenden Fliegen	Brischke, C. G. A.	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1884

continued

Title	Author	Published in	Year
Zoologische Notizen IV	Anonym	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1884
Zoologische Notizen V	Anonym	Bericht des Westpreussischen Botanisch-Zoologischen Vereins, hrsg. v. Westpreußischer Botanisch-Zoologischer Verein	1885
Otto Sendtner	Ross, Hermann	Berichte der Bayerischen Botanischen Gesellschaft, hrsg. v. Bayerische Botanische Gesellschaft zur Erforschung der Heimischen Flora	1909
Kultur der Rosen im freien Lande : Anleitung zur Bereitung der Erd- und Düngerarten, welche im Rosengarten nothwendig sind	Anonym	Nestel's Rosengarten : illustr. Zeitschrift für Rosenfreunde und Rosengärtner	1866
Ueber den Rosenschnitt	Anonym	Illustrirter Rosengarten, Jg. 1875, H. 2, S. 10–11	1875
Die neueren Sommergewächse		Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Ueber gärtnerische Witterungsbeobachtungen	Fintelmann, Gustav Adolph	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Ein Besuch in dem Schauhause des Augustin'schen Gartens bei Potsdam	Anonym	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Beiträge zur Palmenzucht	Lauche, .	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Die Selaginellen der Gärten	Lauche, .	Wochenschrift für Gärtnerei und Pflanzenkunde	1858

continued

Title	Author	Published in	Year
Die Flachslilie, die Mutterpflanze des neuseeländischen Flachses : Eine Dekorationspflanze	Koch, Karl Heinrich Emil	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Die exotischen Monokotylen des Boulogner Wäldchens	Anonym	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Torenia asiatica L. als Schaupflanze	Pasewaldt, .	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Aralia spinosa L. japonica Thunb. und cachemirica Dne. : Drei Blattpflanzen des freien Landes	Anonym	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Pomologische Skizzen	Anonym	Wochenschrift für Gärtnerei und Pflanzenkunde	1858
Bildende Gartenkunst und Pflanzen-Physiognomik : Ein Vortrag	Koch, Karl Heinrich Emil	Wochenschrift für Gärtnerei und Pflanzenkunde	1859
Bildende Gartenkunst und Pflanzen-Physiognomik : Ein Vortrag	Koch, Karl Heinrich Emil	Wochenschrift für Gärtnerei und Pflanzenkunde	1859
Eine Nelken-Auswahl	Samuel, ...; Palandt, .	Wochenschrift für Gärtnerei und Pflanzenkunde	1859
Die Bromeliaceen : Aphoristische Studien	Anonym	Wochenschrift für Gärtnerei und Pflanzenkunde	1859
Nord-Kalifornien und das südliche Oregongebiet	Newberry, John S.	Wochenschrift für Gärtnerei und Pflanzenkunde	1859
Künstliche Forellenzucht im Parke der Frau-Etatsräthin Donner zu Altonaer-Neumühlen	Reimers, Th.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Ueber Keimfähigkeit und Keimkraft der Samenarten	Schinke, K.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899

continued

Title	Author	Published in	Year
Einiges über Pflanzennahrung	Schinke, C.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Die deutsche Tiefsee-Expedition		Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Unsere Sommerblumen		Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Beerenobst		Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Rentabilität der Nutzgeflügelzucht	Schinke, C.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Das praktische Besetzen der Teiche mit Abwachskarpfen	Schinke, C.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Die Kultur des Champignons	Nietner, Heinrich	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Die Blütenfarben der Wiese	Anonym	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Für unsere kleinen Naturfreunde	Günther, .	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899

continued

Title	Author	Published in	Year
Ein botanisches Paradies	Dietrich, Fr.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Etwas über Erdmischungen der Topfgewächse	Meinhardt, .	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Kultur des Cyclamen persicum	Sliwa, A.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Die Paarung der Smaragdeidechse im Terrarium	Tofohr, Otto	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Bereitung von Beerenwein	Haberlé, G.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Nützlichkeit der Regenwürmer im Feld- und Gartenbau	Schinke, C.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Der Schlangenkopffisch	Zieger, S.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Knollen-Begonien	Sliwa, A.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Die botanischen Gärten zu Breslau, Petersburg und Kew	Anonym	Wochenschrift für Gärtnerei und Pflanzenkunde	1859
Knollen-Begonien	Sliwa, A.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899

continued

Title	Author	Published in	Year
Sukkulenten	Gaerdt, Heinrich	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Die Sing- und Klettervögel der Umgegend Hamburgs	Krohn, H.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Die botanischen Gärten zu Breslau, Petersburg und Kew	Anonym	Wochenschrift für Gärtnerei und Pflanzenkunde	1859
Die Sing- und Klettervögel der Umgegend Hamburgs	Krohn, H.	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899
Vogelfütterung im Winter	Kuger, Hermann	Nerthus : ill. Wochenschr. für Tier- u. Pflanzenfreunde ; Organ für Sammler u. Freunde aller naturwiss. Zweige	1899

References

Abrami, G., & Mehler, A. (2018). A UIMA database interface for managing NLP-related text annotations. In *Proceedings of the 11th edition of the Language Resources and Evaluation Conference (LREC 2018)*, 7–12 May 2018, Miyazaki, Japan.

Abrami, G., Mehler, A., Lücking, A., Rieb, E., & Helfrich, P. (2019). TextAnnotator: A flexible framework for semantic annotations. In *Proceedings of the Fifteenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA-15)*.

Abrami, G., Mehler, A., & Stoeckel, M. (2020). TextAnnotator: A web-based annotation suite for texts. In *Proceedings of the Digital Humanities 2020 (DH 2020)*. https://doi.org/10.17613/tenm-4907, https://dh2020.adho.org/wp-content/uploads/2020/07/547_TextAnnotatorAwebbasedannotationsuitefortexts.html.

Ahmed, S., Stoeckel, M., Driller, C., Pachzelt, A., & Mehler, A. (2019). Biofid dataset: Publishing a german gold standard for named entity recognition in historical biodiversity literature. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. Association for Computational Linguistics.

Akella, L. M., Norton, C. N., & Miller, H. (2012). NetiNeti: discovery of scientific names from text using machine learning methods. *BMC Bioinformatics*, *13*, 211. https://doi.org/10.1186/1471-2105-13-211 .

Artstein, R. (2017). Inter-annotator agreement. In N. Ide & J. Pustejovsky (Eds.), *Handbook of linguistic annotation* (p. 297). Springer. https://doi.org/10.1007/978-94-024-0881-2_11.

Balhoff, J. P., Dahdul, W. M., Kothari, C. R., Lapp, H., Lundberg, J. G., Mabee, P., et al. (2010). Phenex: Ontological annotation of phenotypic diversity. *PLoS ONE*, *5*(5), e10500. https://doi.org/10.1371/journal.pone.0010500.

Benikova, D., Biemann, C., & Marc, R. (2014). NoSta-D named entity annotation for German: Guidelines and dataset. In *Proceedings of LREC 2014*.

Blaschke, C., Hirschman, L., & Valencia, A. (2002). Information extraction in molecular biology. *Briefings in Bioinformatics*, *3*(2), 154–165. https://doi.org/10.1093/bib/3.2.154.

Bohnet, B., & Nivre, J. (2012). A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics*, Jeju Island, Korea (pp. 1455–1465). https://www.aclweb.org/anthology/D12-1133.

Bunt, H. (2019). Plug-ins for content annotation of dialogue acts. In *Proceedings of the Fifteenth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-15)* (pp. 33–45).

Cardoso, P., Barton, P. S., Birkhofer, K., Chichorro, F., Deacon, C., Fartmann, T., et al. (2020). Scientists' warning to humanity on insect extinctions. *Biological Conservation*, *242*, 108426. https://doi.org/10.1016/j.biocon.2020.108426.

Chierchia, G. (1998). Reference to kinds across language. *Natural Language Semantics*, *6*(4), 339–405. https://doi.org/10.1023/A:1008324218506.

Consten, M., & Loll, A. (2012). Circularity effects in corpus studies—why annotations sometimes go round in circles. *Language Sciences*, *34*(6), 702–714. https://doi.org/10.1016/j.langsci.2012.04.010.

Corney, D. P. A., Buxton, B. F., Langdon, W. B., & Jones, D. T. (2004). BioRAT: extracting biological information from full-length papers. *Bioinformatics*, *20*(17), 3206–3213. https://doi.org/10.1093/bioinformatics/bth386.

Donnellan, K. S. (1966). Reference and definite descriptions. *The Philosophical Review*, *75*(3), 281–304.

Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Ferragina, P., & Scaiella, U. (2010). Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1625–1628).

Ferrucci, D., Lally, A., Verspoor, K., & Nyberg, E. (2009). Unstructured information management architecture (UIMA) version 1.0. OASIS Standard. https://docs.oasis-open.org/uima/v1.0/uima-v1.0.html.

Finlayson, M. A., & Erjavec, T. (2017). Overview of annotation creation: Processes and tools. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 167–191). Springer. https://doi.org/10.1007/978-94-024-0881-2_5.

Gleim, R., Mehler, A., & Ernst, A. (2012). SOA implementation of the eHumanities Desktop. In *Proceedings of the Workshop on Service-oriented Architectures (SOAs) for the Humanities: Solutions and Impacts, Digital Humanities 2012*, Hamburg, Germany.

Gould, J. L. (2007). Animal artifacts. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representaion* (pp. 249–266). Oxford University Press.

Guan, R., Wang, X., Yang, M. Q., Zhang, Y., Zhou, F., Yang, C., et al. (2018). Multi-label deep learning for gene function annotation in cancer pathways. *Scientific Reports*, *8*(1), 267. https://doi.org/10.1038/s41598-017-17842-9.

Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., et al. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE*, *12*(10), 1–21. https://doi.org/10.1371/journal.pone.0185809.

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., & Mehler, A. (2018). TreeAnnotator: Versatile visual annotation of hierarchical text relations. In *Proceedings of the 11th edition of the Language Resources and Evaluation Conference (LREC 2018)*, May 7–12, Miyazaki, Japan.

Hemati, W., Uslu, T., & Mehler, A. (2016). TextImager: A distributed UIMA-based system for NLP. In *Proceedings of the COLING 2016 System demonstrations, federated conference on computer science and information systems.*

Honnibal, M., & Montani, I. (2017). spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing, Vol. 7.

Hupkes, D., Dankers, V., Mul, M., & Bruni, E. (2020). Compositionality decomposed: How do neural networks generalise? *Journal of Artificial Intelligence Research*, *67*, 757–795. https://doi.org/10.1613/jair.1.11674.

ISO. (2012). Language resource management–semantic annotation framework (SemAF)—part 1: Time and events (SemAF-Time, ISO-TimeML). Standard ISO/IEC TR 24617-1:2012. International Organization for Standardization. https://www.iso.org/standard/37331.html.

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., et al. (2017). Biodiversity losses and conservation responses in the anthropocene. *Science*, *356*(6335), 270–275. https://doi.org/10.1126/science.aam9317.

Klie, J. C., Bugert, M., Boullosa, B., de Castilho, R. E., & Gurevych, I. (2018). The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th international conference on computational linguistics: System demonstrations* (pp. 5–9). Association for Computational Linguistics. http://tubiblio.ulb.tu-darmstadt.de/106270/

Koch, M., Kasperek, G., Hörnschemeyer, T., Mehler, A., Weiland, C., & Hausinger, A. (2017). Setup of BIOfid, a new specialised information service for biodiversity research. *Biodiversity Information Science and Standards*, *1*, e19803. https://doi.org/10.3897/tdwgproceedings.1.19803.

Koning, D., Sarkar, I. N., & Moritz, T. (2005). TaxonGrab: Extracting taxonomic names from text. *Biodiversity Informatics*, *2*, 79–82.

Krauthammer, M., Rzhetsky, A., Morozov, P., & Friedman, C. (2000). Using blast for identifying gene and protein names in journal articles. *Gene*, *259*(1), 245–252. https://doi.org/10.1016/S0378-1119(00)00431-5.

Kripke, S. A. (1977). Speaker's reference and semantic reference. *Midwest Studies in Philosophy*, *2*(1), 255–276.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.

Lenzi, L., Frabetti, F., Facchin, F., Casadei, R., Vitale, L., Canaider, S., et al. (2006). UniGene Tabulator: a full parser for the unigene format. *Bioinformatics*, *22*(20), 2570–2571. https://doi.org/10.1093/bioinformatics/btl425.

Löffler, F., Wesp, V., König-Ries, B., & Klan, F. (2020). Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? arXiv:2002.12021.

Lücking, A., Driller, C., Abrami, G., Pachzelt, A., Hemati, W., & Mehler, A. (2020). *BIOfid annotation guidelines*, version 2.8. Goethe University Frankfurt, Text Technology Laboratory; Senckenberg Nature Research Society; Frankfurt University Library.

Matthews, P. H. (1991). *Morphology*. Cambridge textbooks in linguistics (2nd ed.). Cambridge University Press.

Mehler, A., Gleim, R., vor der Brück, T., Hemati, W., Uslu, T., & Eger, S. (2016). Wikidition: Automatic lexiconization and linkification of text corpora. *Information Technology*, *58*, 70–79. https://doi.org/10.1515/itit-2015-0035.

Meyer, C. M., Mieskes, M., Stab, C., & Gurevych, I. (2014). DKPro agreement: An open-source Java library for measuring inter-rater agreement. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: System demonstrations*, Dublin City University and Association for Computational Linguistics, Dublin, Ireland (pp. 105–109). https://www.aclweb.org/anthology/C14-2023.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*, 39–41. https://doi.org/10.1145/219717.219748.

Mitkov, R. (2013). *Anaphora resolution*. Routledge.

Miyao, Y., Sagae, K., Sætre, R., Matsuzaki, T., & Tsujii, J. (2008). Evaluating contributions of natural language parsers to protein-protein interaction extraction. *Bioinformatics*, *25*(3), 394–400. https://doi.org/10.1093/bioinformatics/btn631.

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticæ Investigationes*, *30*(1), 3–26. https://doi.org/10.1075/li.30.1.03nad.

Nasr, A., & Rambow, O. (2004). Supertagging and full parsing. In *Proceedings of the 7th international workshop on tree adjoining grammar and related formalisms* (pp. 56–63).

Nguyen, Nhung T. H., Gabud, R. S., & Ananiadou, S. (2019). COPIOUS: A gold standard corpus of named entities towards extracting species occurrence from biodiversity literature. *Biodiversity Data Journal*, *7*, e29626. https://doi.org/10.3897/BDJ.7.e29626.

Oltramari, A., Gangemi, A., Guarino, N., & Masolo, C. (2002). Restructuring WordNet's top-level: The *OntoClean* approach. In *OntoLex'2 workshop, ontologies and lexical knowledge bases (LREC 2002)* (pp. 17–26).

Penev, L., Lyal, C. H., Weitzman, A., Morse, D. R., King, D., Sautter, G., et al. (2011). XML schemas and mark-up practices of taxonomic literature. *ZooKeys*, *150*, 89–116. https://doi.org/10.3897/zookeys.150.2213.

Potts, C. (2007). The expressive dimension. *Theoretical Linguistics*, *33*(2), 165–198. https://doi.org/10.1515/TL.2007.011.

Prechtl, P., & Burkard, F. P. (Eds.). (2008). *Metzler Lexikon Philosophie* (3rd ed.). J. B. Metzler'sche Verlagsbuchhandlung & Carl Ernst Poeschel GmbH.

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, *17*, 409–441.

Pustejovsky, J. (2017a). ISO-Space: Annotating static and dynamic spatial information. In N. Ide & J. Pustejovsky (Eds.), *Handbook of linguistic annotation* (pp. 989–1024). Springer. https://doi.org/10.1007/978-94-024-0881-2_37.

Pustejovsky, J. (2017b). ISO-TimeML and the annotation of temporal information. In N. Ide & J. Pustejovsky (Eds.), *Handbook of linguistic annotation* (pp. 941–968). Springer. https://doi.org/10.1007/978-94-024-0881-2_35.

Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning: A guide to corpus-building for applications*. O'Reilly Media Inc.

Ravenscroft, J., Oellrich, A., Saha, S., & Liakata, M. (2016). Multi-label annotation in scientific articles—the multi-label cancer risk assessment corpus. In N. C. C. Chair, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*. European Language Resources Association (ELRA).

Russell, B. (1905). On denoting. *Mind*, *14*(56), 479–493.

Russell, B. (1910/1911). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society, 11,* 108–128.

Sanfilippo, A., Tratz, S., Gregory, M., Chappell, A., Whitney, P., Posse, C., Paulson, P., Baddeley, B., Hohimer, R., & White, A. (2006). Automating ontological annotation with WordNet. In *Proceedings to the third international WordNet conference (GWC-06)* (pp. 22–26).

Sautter, G., Böhm, K., & Agosti, D. (2007). Semi-automated XML markup of biosystematic legacy literature with the GoldenGATE editor. *Biocomputing*. https://doi.org/10.1142/9789812772435_0037.

Seddon, N., Mace, G. M., Naeem, S., Tobias, J. A., Pigot, A. L., Cavanagh, R., et al. (2016). Biodiversity in the anthropocene: Prospects and policy. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1844), 20162094. https://doi.org/10.1098/rspb.2016.2094.

Sowa, J. F. (2000). *Knowledge representation: Logical, philosophical, and computational foundations*. Brooks/Cole.

Steward, H. (2009). Animal agency. *Inquiry*, *52*(3), 217–231. https://doi.org/10.1080/00201740902917119.

Strötgen, J., & Gertz, M. (2010). Heideltime: High quality rule-based extraction and normalization of temporal expressions. In *Proceedings of the 5th international workshop on semantic evaluation* (pp. 321–324). , Association for Computational Linguistics. http://www.aclweb.org/anthology/S10-1071.

Thessen, A. E., Cui, H., & Mozzherin, D. (2012). Applications of natural language processing in biodiversity science. *Advances in Bioinformatics*, *2012*, 391574. https://doi.org/10.1155/2012/391574.

Thessen, A. E., Preciado, J., Jain, P., Martin, J. H., Palmer, M., & Bhat, R. (2018). Automated trait extraction using ClearEarth, a natural language processing system for text mining in natural sciences. *Biodiversity Information Science and Standards*, *2*, e26080. https://doi.org/10.3897/biss.2.26080.

Uslu, T. (2020). Multi-document analysis–semantic analysis of large text corpora beyond topic modeling. PhD thesis, Goethe-University Frankfurt, Text Technology Laboratory.

Uslu, T., Mehler, A., & Baumartz, D. (2019). Computing classifier-based embeddings with the help of text2ddc. In *Proceedings of the 20th international conference on computational linguistics and intelligent text processing (CICLing 2019)*.

Uslu, T., Mehler, A., Baumartz, D., Henlein, A., & Hemati, W. (2018). fastsense: An efficient word sense disambiguation classifier. In *Proceedings of the 11th edition of the language resources and evaluation conference (LREC 2018)*, 7–12 May 2018, Miyazaki, Japan.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18.

Zimmermann, T. E. (1991). Kontextabhängigkeit. In A. von Stechow & D. Wunderlich (Eds.), *Semantik/Semantics. Ein internationales Handbuch der zeitgenössischen Forschung. An International handbook of contemporary research, no. 6 in Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK)* (pp. 156–229). de Gruyter Mouton.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.