

# CoastTerm: a Corpus for Multidisciplinary Term Extraction in Coastal Scientific Literature

Julien DELAUNAY<sup>1,2</sup>[0009-0001-9247-5745]\*, Hanh Thi Hong  
TRAN<sup>1,3,4</sup>[0000-0002-5993-1630]\*, Carlos-Emiliano  
GONZÁLEZ-GALLARDO<sup>1</sup>[0000-0002-0787-2990], Georgeta  
BORDEA<sup>1</sup>[0000-0001-9921-8234], Mathilde DUCOS<sup>1</sup>[0009-0003-4355-1531],  
Nicolas SIDERE<sup>1</sup>[0000-0001-6719-5007], Antoine DOUCET<sup>1</sup>[0000-0001-6160-3356],  
Senja POLLAK<sup>4</sup>[0000-0002-4380-0863], and Olivier DE  
VIRON<sup>2</sup>[0000-0003-3112-9686]

<sup>1</sup> University of La Rochelle, L3i, La Rochelle, France

<sup>2</sup> University of La Rochelle, LIENSs, La Rochelle, France

<sup>3</sup> Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

<sup>4</sup> Jožef Stefan Institute, Ljubljana, Slovenia

{firstname.lastname, thi.tran, carlos.gonzalez\_gallardo}@univ-lr.fr  
senja.pollak@ijs.si

**Abstract.** The growing impact of climate change on coastal areas, particularly active but fragile regions, necessitates collaboration among diverse stakeholders and disciplines to formulate effective environmental protection policies. We introduce a novel specialized corpus comprising 2,491 sentences from 410 scientific abstracts concerning coastal areas, for the Automatic Term Extraction (ATE) and Classification (ATC) tasks. Inspired by the ARDI framework, focused on the identification of Actors, Resources, Dynamics and Interactions, we automatically extract domain terms and their distinct roles in the functioning of coastal systems by leveraging monolingual and multilingual transformer models. The evaluation demonstrates consistent results, achieving an F1 score of approximately 80% for automated term extraction and F1 of 70% for extracting terms and their labels. These findings are promising and signify an initial step towards the development of a specialized Knowledge Base dedicated to coastal areas.

**Keywords:** Automatic term extraction · ATE · Automatic term classification · ATC · terminology · coastal area · littoral.

## 1 Introduction

Coastal areas, grappling with the dual impacts of global change and human interventions, constitute a complex system wherein various dynamics continuously interact (physical, chemical, biological, societal, and others). Understanding this system necessitates examining it as an inherently anthropized environment. In

---

\* These authors contributed equally to this work

this context, many agents and mechanisms can only be understood by considering human actions, such as coastal development, activities affecting land and sea, resource management, and urbanization.

The resulting interdisciplinarity provides a very interesting use case for the automatic analysis of terminology and its use. Numerous nuanced terms from various domains like environmental science, geography, ecology, and sociology emerge in coastal literature, reflecting its interdisciplinary nature. Automatic analysis of terminology offers a systematic approach to identify and categorize these domain-specific concepts, crucial to the understanding of the dynamic nature of coastal environments. By adapting to the evolving terminological landscape, this process aids in identifying key entities within coastal systems and integrating diverse disciplinary perspectives into a coherent framework.

The ARDI framework [10] enables the identification of *Actors* influencing a territory, the *Resources* they exploit, the *Dynamics* in operation, and the *Interactions* between these agents and resources. Employing ARDI enables stakeholders to collaboratively construct a conceptual framework of the system, facilitating the development of environment protection policies and enhancing scientists’ comprehension of these territories. However, identifying key entities within such intricate systems poses considerable challenges. This difficulty is exacerbated by the numerous scientific disciplines addressing the subject matter, including but not limited to biology, oceanography, chemistry, and engineering, resulting in a huge amount of scholarly literature every year. In this context, automatic term extraction (ATE) and classification (ATC) play a crucial role in unlocking the knowledge embedded within the scientific literature.

Recently, we observed a push towards several ATE approaches with different methods, from rule-based to neural approaches. However, there is still a significant gap in the performance of ATE compared to other similar natural language processing (NLP) downstream tasks partially due to the following reasons. First, terms are inherently semantically defined to refer to domain-specific concepts [19]. Besides attempts to define the meaning of a *term*, such as a “language used in a subject field and characterized by the use of specific linguistic means of expression” (ISO 1087-1), in real-life settings, terms are not consistently defined and their definitions vary across the domains and use-cases, making it difficult to develop universal term extraction methods. Secondly, there is a lack of well-documented and transparent domain-specific corpora, even if recently some valuable efforts have been made [30].

Our main contributions are threefold and are summarized as follows:

- We propose two gold-annotated multidisciplinary datasets for ATE and ATC focusing on the domain of coastal areas<sup>5</sup>.
- Inspired by the ARDI framework [10], we propose a set of labels designed to facilitate the representation of a system independently of the study domain.
- We compare a range of ATE state-of-the-art models on our datasets and identify specific challenges inherent to our data.

<sup>5</sup> Corpus and code are available at [https://github.com/jdelaunay/coastal\\_area\\_term\\_extraction](https://github.com/jdelaunay/coastal_area_term_extraction)

## 2 Related Work

### 2.1 Term extraction datasets

Several manually annotated monolingual and multilingual domain-specific resources have been developed for term extraction systems [33], notably for scientific domains. The ACL Reference Dataset for Terminology Extraction and Classification (ACL RD-TEC) [29] serves as a benchmark for evaluating term extraction and classification in the scientific literature related to computational linguistics. It includes 300 manually annotated abstracts from articles in the ACL Anthology Reference Corpus (1978-2006), categorized into various classes. Augenstein et al. (2017) introduce a corpus for SemEval 2017 Task 10, featuring 500 double-annotated documents from the domains of computer science, materials science and physics [7]. This corpus addresses the identification and classification of keyphrases at the word level, categorizing keyphrases into *process*, *task*, and *material*.

In relation to environment studies, SPECIES-800 [27] is a corpus of 800 manually annotated abstracts for taxon mentions recognition. Constructed by randomly selecting 100 MEDLINE abstracts from various journals, it comprises 3,708 mentions of 718 unique species, referenced by 1,503 unique names. BiodivNERE [1] offers two gold standard corpora for named entity recognition (NER) and relation extraction (RE) within biodiversity studies, created from biodiversity metadata and abstracts and manually verified by experts. The corpus comprises 2,398 statements from 150 documents, with the NER corpus identifying entities such as *organism*, *environment*, *quality*, *location*, *phenomena*, and *matter*. COPIOUS [26] stands as a gold standard corpus sourced from the Biodiversity Heritage Library. Comprising more than 26K sentences from 668 documents, it classifies its 28K entities into five categories: *taxon names*, *geographical locations*, *habitats*, *temporal expressions* and *person names*.

However, to our knowledge, no annotated corpora for term extraction in the interdisciplinary study of coastal regions or related domains (such as oceans or seas) exists. Thus, CoastTerm represents a pioneering effort in this regard, paving the way for an exhaustive cross-domain and multi-disciplinary Knowledge Graph construction system for studying coastal areas.

### 2.2 Term extraction methods

Classical term extractors mainly rely on either linguistic or statistical aspects [13] or combine both [17] and applied rule-based or machine-learning methods to extract the candidate terms. More recently, the introduction of representation learning and neural networks has led to the application of various text embedding techniques for term extraction, including local-global [2], non-contextual embeddings (e.g., GloVe<sup>6</sup>, Word2Vec, skip-gram [2, 3, 21, 41]), contextual word

<sup>6</sup> <https://nlp.stanford.edu/projects/glove/>

embeddings (e.g., Flair<sup>7</sup>, BERT [21, 5, 32]), as well as their combinations (e.g., stacked Flair + BERT [5, 32]).

Neural architectures are also used as end-to-end term extraction systems, and most current systems focus on tagging-based models [15, 22]. In tagging-based mechanisms, the task was formulated as (1) sequence classification where or not the binary label of a term was assigned to each possible n-gram of a fixed length of a given sentence using different variants of BERT-based models (e.g., BERT, RoBERTa, and XLMR); or (2) token classification, where the label was assigned to each word in the given sentence following the IOB annotation format using different language models.

With the advent of transformers, several pre-trained language models have been applied as token classifiers. Above all, XLMR is now considered a benchmark for several languages [34, 37]. Cross-domain and cross-lingual learning was also applied to these benchmarks to enhance extraction performance in the absence of available annotated data [16, 22, 34–37].

In addition, there have been other experiments on ATE with the adoption of span-based methods [40] or applying generative models considering Seq2Seq models such as mBART [22] to extract candidate terms more efficiently. However, while span-based methods show their potential, the performance of generative models is still under question.

In the context of our study domains, Zhao et al. (2022) explores the extraction of knowledge from operational maritime decision-making sentences [42], Andersen et al. (2022) proposes the development of a corpus to cultivate a specialized terminology relevant to Norwegian maritime discourse [4], Mouratidis et al. (2022) performs term extraction within legal documents on maritime topics in the Greek language [25]. Similarly to our work, but in the realm of karst studies, TermFrame<sup>8</sup> serves as a KG constructed by extracting terms and triplets from English and Slovene karst corporas [28, 39]. EcoLexicon [11, 12] constitutes a KB specialized on environment, encompassing six languages (English, French, German, Modern Greek, Russian, and Spanish). Moreover, for NER tasks, TaxoNERD [23] aims to recognize taxon mentions in ecological documents, while AGRONER [38] employs unsupervised NER techniques tailored to the agriculture domain, integrating an extended BERT model with Latent Dirichlet Allocation (LDA) topic modeling.

### 3 CoastTerm corpus for term extraction

#### 3.1 Annotation process

We collected a corpus of 64,000 papers from Scopus<sup>9</sup>, spanning from 1980 to 2023, containing the terms “coastal areas” or “littoral” in their abstract or title. We initially selected randomly 600 abstracts for manual annotation, keeping a

<sup>7</sup> <https://github.com/flairNLP/flair>

<sup>8</sup> <https://termframe.ff.uni-lj.si/>

<sup>9</sup> <https://www.elsevier.com/fr-fr/solutions/scopus>

proportion of 60% regular articles and 40% surveys. Annotation efforts first enlisted the participation of two undergraduate Master’s students specialized in Earth Sciences, who were compensated for their contributions. These students were supplied with example annotations and guidelines to conduct a task of document-level joint entity and relation extraction, which involves nested ATE and ATC, coreference resolution, and document-level relation extraction. The annotation guidelines were constructed by a computer science and coastal research PhD student and a domain expert, with the assistance of an ontology expert. Only the sentences that give information about the functioning of the coastal zone were annotated, we avoided sentences that described the methodology used in the article.

We adapted the ARDI framework to extract the information related to the functioning of a described system within a scientific abstract. As a result, for ATC, we designated the following labels: “Actor” (stakeholders who consume resources and/or initiate processes) and “Resource” (goods, products, facilities, and elements, including plants and animals, utilized by stakeholders). Following the Basic Formal Ontology (BFO) [6], we replaced the “Dynamic” label in ARDI with “Process”, as defined by the Environment Ontology (ENVO) [8], referring to environmental, societal, or economic processes impacting the system and inducing changes. To enhance precision on the extracted information, we introduced the “Quality” label from the Phenotype And Trait Ontology (PATO) [14], which refers to height, concentration, or a specificity, and added a “Location” label.

Along with the two students specializing in Earth Sciences, the annotation process engaged the same domain expert and PhD student, all simultaneously annotating papers. All annotators were familiar with the annotation tool, INCEpTION [18]. Over two months, the campaign aimed to achieve dual-annotator coverage for 60% of the total annotated abstracts. Ultimately, 215 abstracts were annotated, with a mean Krippendorff’s alpha [20] of 43%, indicating a moderate agreement, but not sufficient to use directly the dataset, falling short of the minimum threshold (66%) required for direct dataset utilization, highlighting the difficulty of the manual annotation of terms. Annotations were then curated by the PhD student according to the guidelines.

We employed three domain-relevant knowledge bases (KBs) (i.e., AGROVOC<sup>10</sup>, GEMET (GEneral Multilingual Environmental Thesaurus)<sup>11</sup>, AFO (Agriculture and Forestry Ontology)<sup>12</sup>, and TAXREF-LD<sup>13</sup>) to pre-annotate pertinent terms in a secondary subset of 195 abstracts. This process significantly assisted in refining term boundaries. Annotations were then carried out by the PhD student annotator following the established guidelines, utilizing insights acquired from the initial annotation process. Subsequently, the initially fully manually annotated subset was homogenized with the KB-recommended one to produce two datasets intended for the study of coastal areas. In the context of the present

<sup>10</sup> <https://agrovoc.fao.org/browse/agrovoc/en/>

<sup>11</sup> <https://www.eionet.europa.eu/gemet/en/about/>

<sup>12</sup> <https://finto.fi/afo/en/>

<sup>13</sup> <https://github.com/frmichel/taxref-ld>

Table 1: Statistics for KB and human recommended subcorpora

	KBs recommended	Human experts recommended
Vocabulary size	7,902	6,280
Sentences	1,235	1,256
Tokens	31,147	37,983
Type-to-token ratio (TTR)	0.25	0.17
Annotated Terms	6,663	6,543
Unique Terms	3,844	4,400
Unique Terms (lemmatized)	3,539	4,110
# of “Actor”	617	602
# of “Resource”	2,236	2,052
# of “Process”	1,524	1,431
# of “Quality”	1,138	1,373
# of “Location”	1,145	1,082
# of “B”	6,663	6,543
# of “I”	4,723	6,192
# of “O”	25,415	25,710

research, both datasets were adapted manually for terminology extraction by removing relations and pronouns that indicate coreferences.

### 3.2 Dataset description

The KB-recommended corpus contains 1,235 annotated sentences, spanning across 61 different keywords, and includes 6,663 annotated terms. The human-recommended corpus contains 1,256 annotated sentences, spanning across 92 different keywords, and featuring 6,543 annotated terms. From Table 1 it can be seen that the distribution of labels is consistent across both datasets. In addition, the label “Actor” is observed to be less prevalent, whereas “Resource” emerges as the most frequently represented one. Together, these two datasets collectively span across 101 unique keywords, “*aquatic science*”, “*oceanography*”, “*ecology, evolution, behavior, and systematics*”, “*ecology*”, “*pollution*”, “*earth and planetary sciences*”, “*environmental science*”, “*management, monitoring, policy, and law*”, “*water science and technology*”, and “*geography, planning, and development*” being the most prominent. A comparison of the top 100 frequent terms shows a significant overlap of 52% between the two corpora. In total, the two datasets share 751 common terms (730 if lemmatized). We allocated 70% of the annotated sentences within each article to the training sets, 10% to the validation sets, and 20% to the test sets. An annotation example can be visualized in Figure 1.

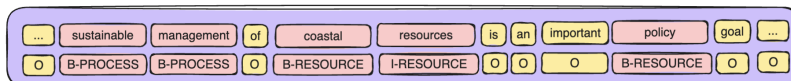


Fig. 1: A sample of the corpus annotation for term extraction and classification

## 4 Experiments

To evaluate the challenges of CoastTerm, we conducted extensive experiments using state-of-the-art ATE, extending our evaluation to ATC. Our in-depth analysis facilitates a discussion on potential future trajectories for interdisciplinary ATE and ATC research related to coastal areas.

### 4.1 Models

We considered ATE as a sequence-labeling task where the model returns a label for each token in a text sequence, using the IOB labeling mechanism [31, 34, 35]. We apply the same labeling scheme to ATC, with the addition of the term’s class following its IOB label.

We experimented with two families of language models with both base and large versions, including:

- *Monolingual pre-trained model*: We chose RoBERTa [24], a transformer-based model pre-trained on a large corpus of English data in a self-supervised fashion.
- *Multilingual pre-trained model*: We opt for XLMR [9], a transformer-based model pre-trained on 2.5TB of filtered CommonCrawl data containing 100 languages. This multilingual version of RoBERTa, achieves benchmark performance in ATE for rich-resourced languages (e.g. English) [30, 34].

### 4.2 Evaluation Metrics

To assess the performance of the ATE systems on the human-recommended and KB-recommended datasets, we juxtaposed the candidate list of unique terms extracted from the entire test set against the gold standard of the test set. We performed experiments with both lemmatized and not lemmatized terms for testing. This evaluation was conducted employing strict matching criteria, with metrics including precision (P), recall (R), and micro F1-score (F1). For the ATC task, the evaluation process remained the same but each term was accompanied by its corresponding label.

### 4.3 Results

Table 2 presents the performances of mono- and multilingual classifiers for the ATE task on both the human- and KB-recommended datasets. Regarding the KB-recommended dataset, the results demonstrate that  $\text{XLMR}_{base}$  achieves superior performance compared to  $\text{RoBERTa}_{base}$  for both lemmatized and unlemmatized text. However,  $\text{RoBERTa}_{large}$  outperforms  $\text{XLMR}_{large}$ . Conversely, on the human-recommended dataset, monolingual models exhibit better performance than multilingual ones. Additionally, it is observed that  $\text{XLMR}_{large}$  performs least effectively in this context. We also performed experiments with the combined KB- and human-recommended datasets. This configuration is referred

Table 2: Evaluation in performance of different extractors for ATE task

Models	KB			Human			Fusion		
	P	R	F1	P	R	F1	P	R	F1
<b>Not lemmatized</b>									
RoBERTa <sub>base</sub>	75.51	78.12	76.79	78.65	82.07	80.32	<b>79.08</b>	80.37	79.72
XLMR <sub>base</sub>	76.08	78.95	77.49	79.07	81.04	80.04	77.95	79.94	78.93
RoBERTa <sub>large</sub>	<b>77.23</b>	<b>79.90</b>	<b>78.54</b>	<b>80.97</b>	<b>82.90</b>	<b>81.92</b>	79.00	80.96	<b>79.97</b>
XLMR <sub>large</sub>	76.99	78.85	77.91	77.42	79.59	78.49	78.48	<b>81.45</b>	79.94
<b>Lemmatized</b>									
RoBERTa <sub>base</sub>	75.45	78.24	76.82	78.43	82.25	80.29	<b>78.80</b>	79.97	79.38
XLMR <sub>base</sub>	75.89	78.78	77.31	78.97	81.28	80.11	77.61	79.68	78.63
RoBERTa <sub>large</sub>	<b>77.32</b>	<b>79.76</b>	<b>78.52</b>	<b>80.78</b>	<b>82.79</b>	<b>81.77</b>	78.74	80.70	<b>79.71</b>
XLMR <sub>large</sub>	76.93	79.11	78.00	77.48	80.41	78.92	78.20	<b>81.21</b>	79.68

to as Fusion in Table 2 where 20% and 40% of the human-recommended dataset were allocated to the validation and test sets respectively. The KB-recommended dataset, along with the remaining 40% of the human-recommended dataset, composed the train set. Reported performances show that monolingual models outperform multilingual ones.

Table 3 displays the performance of the classifiers for the ATC task using the combined KB-recommended and human-recommended datasets. Notably, multilingual classifiers demonstrate superior performance compared to monolingual ones. Additionally, upon lemmatizing the predictions, XLMR<sub>base</sub> exhibits better performance than XLMR<sub>large</sub>.

Table 3: Evaluation in performance of different extractors for joint ATE and ATC on the fusion of KB and human datasets

Models	Not lemmatized			Lemmatized		
	P	R	F1	P	R	F1
RoBERTa <sub>base</sub>	66.36	68.93	67.62	65.96	68.74	67.32
XLMR <sub>base</sub>	66.11	<b>71.69</b>	68.79	65.95	<b>71.85</b>	<b>68.77</b>
RoBERTa <sub>large</sub>	66.60	71.05	68.75	66.05	70.85	68.37
XLMR <sub>large</sub>	<b>66.97</b>	71.16	<b>69.00</b>	<b>66.60</b>	70.96	68.71

#### 4.4 Error analysis

To assess the impact of term length on the models’ performance, we examined the proportion of correct and incorrect predictions relative to term length. Figure 2 illustrates for XLMR<sub>base</sub> that as the term length increases, the model encounters greater difficulty in accurately predicting it and might predict shorter terms. However, the vast majority of the terms consist of either one or two words which mitigates the negative impact on performance. This behavior is observed across all models.

We also report the confusion matrices for the ATC in Figure 3. The results are consistent since ambiguity might come from the fact that one entity might



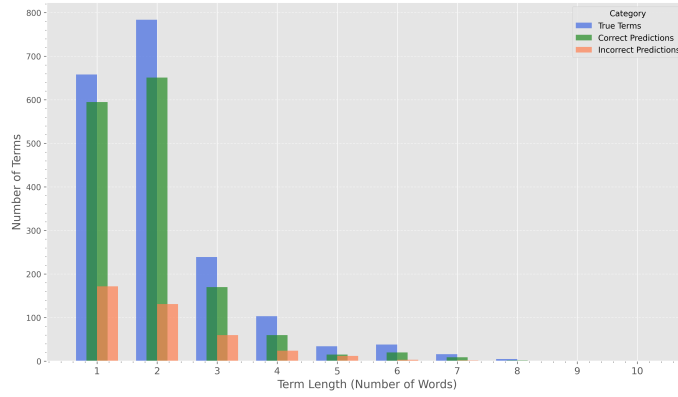


Fig. 2: Term length distribution for  $XLMR_{base}$  on the fusion of KB- and human-recommended datasets

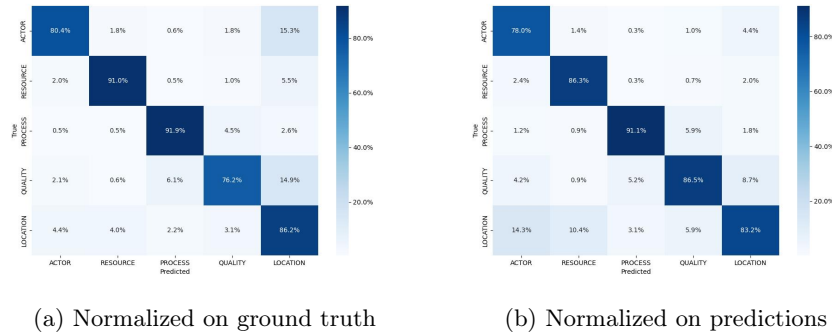


Fig. 3: Confusion matrices for  $XLMR_{base}$  on the fusion of KB- and human-recommended datasets

be an “Actor” at one point and a “Location” at another (Countries, for example); or a “Location” and a “Resource” or “Quality” depending on the context.

## 5 Conclusion and Future Works

We introduced CoastTerm, a corpus of 2,491 gold-annotated sentences for interdisciplinary automatic term extraction and classification related to the coastal area. Adapting the ARDI framework, we provided comprehensive labels applicable to various domains. Benchmarking mono- and multilingual state-of-the-art models on ATE and ATC tasks shows promising results, paving the way for interdisciplinary knowledge base construction in the coastal area domain.

## Acknowledgments

The work was supported by the TERMITRAD (2020-2019-8510010) project funded by the Nouvelle-Aquitaine Region, France, by the Slovenian Research and Innovation Agency core research program Knowledge Technologies (P2-0103) and the project Cross-lingual and Cross-domain Methods for Terminology Extraction and Alignment, a bilateral project funded by the program PROTEUS under the grant number BI-FR/23-24-PROTEUS006. We express our gratitude to Kenza HERMAN and Geraldine DUBOS for their invaluable assistance in the annotation process.

## References

1. Abdelmageed, N., Löffler, F., Feddoul, L., Algergawy, A., Samuel, S., Gaikwad, J., Kazem, A., König-Ries, B.: Biodivnere: Gold standard corpora for named entity recognition and relation extraction in the biodiversity domain. *Biodiversity Data Journal* **10** (2022)
2. Amjadian, E., Inkpen, D., Paribakht, T.S., Faez, F.: Local-global vectors to improve unigram terminology extraction. *Computerm* 2016 p. 2 (2016)
3. Amjadian, E., Inkpen, D., Paribakht, T.S., Faez, F.: Distributed specificity for automatic terminology extraction. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* **24**(1), 23–40 (2018)
4. Andersen, G.: Utilising heterogeneous language resources for term extraction in maritime domains. *Terminology* **28**(1), 1–36 (2022)
5. Andrius, U.: Automatic extraction of lithuanian cybersecurity terms using deep learning approaches. In: *Human Language Technologies–The Baltic Perspective: Proceedings of the Ninth International Conference Baltic HLT 2020*. vol. 328, p. 39. IOS Press (2020)
6. Arp, R., Smith, B., Spear, A.D.: *Building ontologies with basic formal ontology*. Mit Press (2015)
7. Augenstein, I., Das, M., Riedel, S., Vikraman, L., McCallum, A.: Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications. *arXiv preprint arXiv:1704.02853* (2017)
8. Buttigieg, P.L., Morrison, N., Smith, B., Mungall, C.J., Lewis, S.E., Consortium, E.: The environment ontology: contextualising biological and biomedical entities. *Journal of biomedical semantics* **4**, 1–9 (2013)
9. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116* (2019)
10. Etienne, M., Du Toit, D.R., Pollard, S.: Ardi: a co-construction method for participatory modeling in natural resources management. *Ecology and society* **16**(1) (2011)
11. Faber, P., León-Araúz, P., Reimerink, A.: Representing environmental knowledge in ecolexicon. In: *Languages for specific purposes in the digital era*, pp. 267–301. Springer (2013)
12. Faber, P., León-Araúz, P., Reimerink, A.: Ecolexicon: new features and challenges. *GLOBALEX* pp. 73–80 (2016)
13. Frantzi, K.T., Ananiadou, S., Tsujii, J.: The c-value/nc-value method of automatic recognition for multi-word terms. In: *International conference on theory and practice of digital libraries*. pp. 585–604. Springer (1998)

14. Gkoutos, G.V., Schofield, P.N., Hoehndorf, R.: The anatomy of phenotype ontologies: principles, properties and applications. *Briefings in Bioinformatics* **19**(5), 1008–1021 (04 2017). <https://doi.org/10.1093/bib/bbx035>, <https://doi.org/10.1093/bib/bbx035>
15. Hazem, A., Bouhandi, M., Boudin, F., Daille, B.: Termeval 2020: Taln-ls2n system for automatic term extraction. In: 6th International Workshop on Computational Terminology (COMPUTERM 2020) (2020)
16. Hazem, A., Bouhandi, M., Boudin, F., Daille, B.: Cross-lingual and cross-domain transfer learning for automatic term extraction from low resource data. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference. pp. 648–662 (2022)
17. Kessler, R., Béchet, N., Berio, G.: Extraction of terminology in the field of construction. In: 2019 First International Conference on Digital Data Processing (DDP). pp. 22–26. IEEE (2019)
18. Klie, J.C., Bugert, M., Boullosa, B., Eckart de Castilho, R., Gurevych, I.: 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. Santa Fe, New Mexico (2018), <https://www.aclweb.org/anthology/C18-2002>
19. Kockaert, H.J., Steurs, F.: Handbook of terminology, vol. 1. John Benjamins Publishing Company (2015)
20. Krippendorff, K.: Computing krippendorff’s alpha-reliability (2011)
21. Kucza, M., Niehues, J., Zenkel, T., Waibel, A., Stüker, S.: Term extraction via neural sequence labeling a comparative evaluation of strategies using recurrent neural networks. In: Interspeech. pp. 2072–2076 (2018)
22. Lang, C., Wachowiak, L., Heinisch, B., Gromann, D.: Transforming term extraction: Transformer-based approaches to multilingual term extraction across domains. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. pp. 3607–3620 (2021)
23. Le Guillarme, N., Thuiller, W.: Taxonerd: deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature. *Methods in Ecology and Evolution* **13**(3), 625–641 (2022)
24. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692 (2019)
25. Mouratidis, D., Mathe, E., Voutos, Y., Stamou, K., Keramanidis, K.L., Mylonas, P., Kanavos, A.: Domain-specific term extraction: A case study on greek maritime legal texts. In: Proceedings of the 12th Hellenic Conference on Artificial Intelligence. SETN ’22, Association for Computing Machinery, New York, NY, USA (2022). <https://doi.org/10.1145/3549737.3549751>, <https://doi.org/10.1145/3549737.3549751>
26. Nguyen, N.T., Gabud, R.S., Ananiadou, S.: Copious: A gold standard corpus of named entities towards extracting species occurrence from biodiversity literature. *Biodiversity data journal* (7) (2019)
27. Pafilis, E., Frankild, S.P., Fanini, L., Faulwetter, S., Pavloudi, C., Vasileiadou, A., Arvanitidis, C., Jensen, L.J.: The species and organisms resources for fast and accurate identification of taxonomic names in text. *PloS one* **8**(6), e65390 (2013)
28. Pollak, S., Repar, A., Martinc, M., Podpečan, V.: Karst exploration: extracting terms and definitions from karst domain corpus. *Proceedings of eLex* **2019**, 934–956 (2019)

29. QasemiZadeh, B., Schumann, A.K.: The acl rd-tec 2.0: A language resource for evaluating term extraction and entity recognition methods. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). pp. 1862–1868 (2016)
30. Rigouts Terryn, A., Hoste, V., Drouin, P., Lefever, E.: Termeval 2020: Shared task on automatic term extraction using the annotated corpora for term extraction research (acter) dataset. In: 6th International Workshop on Computational Terminology (COMPUTERM 2020). pp. 85–94. European Language Resources Association (ELRA) (2020)
31. Rigouts Terryn, A., Hoste, V., Lefever, E.: Hamlet: Hybrid adaptable machine learning approach to extract terminology. *Terminology* **27**(2), 254–293 (2021)
32. Terryn, A.R., Hoste, V., Lefever, E.: Tagging terms in text: A supervised sequential labelling approach to automatic term extraction. *Terminology: international journal of theoretical and applied issues in specialized communication* **28**(1), 157–189 (2022)
33. Tran, H.T.H., Martinc, M., Caporusso, J., Doucet, A., Pollak, S.: The recent advances in automatic term extraction: A survey. arXiv preprint arXiv:2301.06767 (2023)
34. Tran, H.T.H., Martinc, M., Doucet, A., Pollak, S.: Can cross-domain term extraction benefit from cross-lingual transfer? In: International Conference on Discovery Science. pp. 363–378. Springer (2022)
35. Tran, H.T.H., Martinc, M., Pelicon, A., Doucet, A., Pollak, S.: Ensembling transformers for cross-domain automatic term extraction. In: International Conference on Asian Digital Libraries. pp. 90–100. Springer (2022)
36. Tran, H.T.H., Martinc, M., Repar, A., Ljubešić, N., Doucet, A., Pollak, S.: Can cross-domain term extraction benefit from cross-lingual transfer and nested term labeling? *Machine Learning* pp. 1–30 (2024)
37. Tran, T.H.H., Martinc, M., Repar, A., Doucet, A., Pollak, S.: A transformer-based sequence-labeling approach to the slovenian cross-domain automatic term extraction (2022)
38. Veena, G., Kanjirangat, V., Gupta, D.: Agroner: An unsupervised agriculture named entity recognition using weighted distributional semantic model. *Expert Systems with Applications* **229**, 120440 (2023)
39. Vintar, Š., Martinc, M.: Framing karstology: From definitions to knowledge structures and automatic frame population. *Terminology* **28**(1), 129–156 (2022)
40. Wang, J., Feng, C., Liu, F., Li, X., Wang, X.: Extract then adjust: A two-stage approach for automatic term extraction. In: CCF International Conference on Natural Language Processing and Chinese Computing. pp. 236–247. Springer (2023)
41. Zhang, Z., Gao, J., Ciravegna, F.: Semre-rank: Improving automatic term extraction by incorporating semantic relatedness with personalised pagerank. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **12**(5), 1–41 (2018)
42. Zhao, X., Lian, X., Liu, P., Gao, C.: Research on knowledge extraction of maritime operational decision-making sentences. In: Proceedings of 2022 10th China Conference on Command and Control. pp. 836–845. Springer Nature Singapore, Singapore (2022)