HUKB at ChEMU 2022 Task 1: Expression-Level Information Extraction

Kojiro Machi[1], Masaharu Yoshioka[1,2,3]

[1]Graduate School of Information Science and Technology, Hokkaido University, N14 W9, Kita-ku, Sapporo-shi, Hokkaido, Japan
[2]Faculty of Information Science and Technology, Hokkaido University
[3]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University

Abstract

This paper describes our results for the three tasks at ChEMU 2022: Task 1a (named entity recognition), Task 1b (event extraction), and Task 1c (anaphora resolution). We adopted a hybrid approach using deep learning models and a small set of post-processing rules for these tasks. For Tasks 1b and 1c, we adopted a pipeline approach for relation extraction, which combined mention detection with relation classification. In addition, we proposed post-processing methods for Task 1c that considered the results of Task 1a. Our system obtained an exact match F-score of 0.9412 and a relaxed match F-score of 0.9572 for Task 1a, an exact match F-score of 0.8865 and a relaxed match F-score of 0.9027 for Task 1b, and an exact match F-score of 0.7232 and an F-score of 0.8053 for Task 1c for each test set (private score). Although our approaches tried to consider the document-level context and relationships between the tasks, limitations remained.

Keywords

Information extraction, Chemical patents, Named entity recognition, Event extraction, Anaphora resolution

1. Introduction

The automated extraction of the chemical-reaction information in patents plays an important role in collecting chemical-reaction information in reaction databases for use by synthetic chemists. Chemical patents contain important information about new chemical discoveries because any new chemical compounds are usually published via patents [1]. With the number of patents increasing rapidly, manually collecting the information written in patents not only takes time and cost but also requires expertise in the subject matter of the patents.

Since 2020, the Cheminformatics Elsevier Melbourne University (ChEMU) laboratory has identified several tasks related to information extraction from chemical patents, including expression-level information extraction [2, 3] and document-level information [4, 5]. For 2022, the ChEMU laboratory is providing five tasks for ChEMU 2022 [6].

In recent years, deep learning has been recognized as a promising approach to information extraction from chemical literature. For example, pre-trained language models such as

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

machi@eis.hokudai.ac.jp (K. Machi); yoshioka@ist.hokudai.ac.jp (M. Yoshioka)
https://www-kb.ist.hokudai.ac.jp/yoshioka/ (M. Yoshioka)
0000-0002-2096-1218 (M. Yoshioka)

BioBERT [7] and ChemBERT [8] have shown high performance in information-extraction tasks. Moreover, the best systems in previous ChEMU tasks all employed deep-learning-based approaches, together with a small amount of rule-based post-processing [9, 10]. In addition, both of these systems used a pipeline approach for relation extraction tasks [11, 12], such as the event extraction and anaphora resolution tasks.

This paper describes our results for the three tasks at ChEMU 2022: Task 1a (named entity recognition, NER), Task 1b (event extraction, EE) and Task 1c (anaphora resolution, AR). We employed hybrid approaches that used deep learning models and a small set of post-processing rules. In addition, we propose post-processing methods for Task 1c that considered the results of Task 1a.

2. Task Description

Prediction systems for the three tasks must solve two general tasks: the identification of the spans of entities and the labels (mention detection), together with the identification of relations between the spans (relation classification).

Task 1a, (i.e., NER), is a mention detection task that identifies one of the 10 entity types of labels. The set of labels contains these compounds (STARTING_MATERIAL, REAGENT_CATALYST, REACTION_PRODUCT, SOLVENT, OTHER_COMPOUND), conditions (TIME, TEMPERATURE), yields (YIELD_PERCENT, YIELD_OTHER), and a relation label (EXAMPLE_LABEL).

Task 1b, (i.e., EE), is a task that involves both mention detection and relation extraction. The mention detection task identifies events that have a relationship with entities in Task 1a and identifies relations between the events and the entities. Almost all events involve a label that is either REACTION_STEP or WORK_UP and some events involve both labels. The relation extraction task identifies relations between the events and compounds, which are annotated with ARG1, and relations between the events and conditions or yield are annotated with ARGM.

Task 1c, (i.e., AR), involves both mention detection and relation extraction tasks. The mention detection task identifies an antecedent as ENTITY and the anaphor as a label that represents their relationship. The relation extraction task identifies relations between antecedents and anaphors as a coreference relation (COREFERENCE) and four bridging relations (TRANSFORMED, REACTION_ASSOCIATED, WORK_UP, and CONTAINED).

3. Methods

We developed systems for mention detection and relation classification tasks that used ChemBERT [8], a pre-trained language model for chemistry-related documents. This approach was similar to those of the best systems adopted in previous tasks [11, 12]. Figure 1 shows our pipeline method. First, we split a snippet into sentences by using ChemDataExtractor [13]. Then, ChemBERT predicted the labels for mentions/relations. post-processing methods were adopted for mention detection in Task 1a and relation detection in Task 1c, with the aim of addressing the document-level context.

We fine-tuned ChemBERT for Task 1a. In addition, we fine-tuned multiple ChemBERTs for Tasks 1b and 1c because these tasks are more complex than Task 1a. Table 1 shows the set of

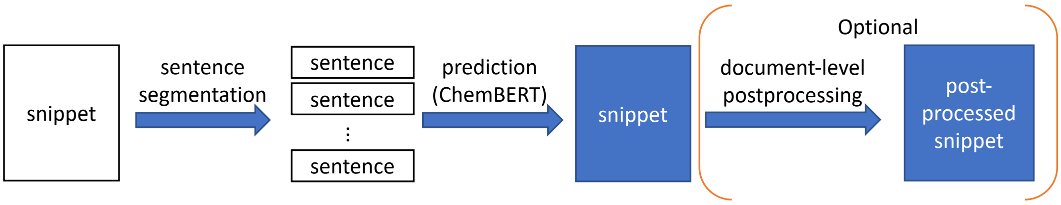

Figure 1: Overview of our pipeline method

target labels for the three tasks involving mention detection and relation classification.

Table 1
Set of target labels for the three tasks involving mention detection and relation classification

	Mention detection	Relation classification
Task 1a	Named entity	-
Task 1b	Named entity (Task 1a) Event	ARG1 ARGM
Task 1c	Candidate for coreference Candidate for bridging	Coreference Bridging

3.1. Mention Detection

For Task 1a, we trained a ChemBERT model and constructed a set of post-processing rules. First, ChemBERT was fine-tuned on the training set, excluding snippets that had overlapping mentions. A snippet was split into sentences by using ChemDataExtractor [13] and the sentences were split by a simple regex rule that used a particular tool[1] by default. Then, IOB2 labels were assigned to the tokens. We used a linear classifier to predict the label of each tokens and the input for it was the output of the first sub-token [14]. Second, two post-processing methods were applied. Because compounds in a heading section, which contains the product of the snippet (REACTION_PRODUCT) and/or the final product of the multistep reaction (OTHER_COMPOUND), cannot be distinguished without context, a post-processing method is required. The rule adopted is that if EXAMPLE_LABEL exists in a snippet, then the REACTION_PRODUCT that appears before the last EXAMPLE_LABEL is annotated with OTHER_COMPOUND. Then, if OTHER_COMPOUND appears in a heading and the same string appears as a REACTION_PRODUCT after the last source compound (STARTING_MATERIAL, REAGENT_CATALYST, or SOLVENT), then the entity is labeled as a REACTION_PRODUCT.

For Task 1b, we used the method in Task 1a for named entity recognition and trained a ChemBERT model for the detection of events in the same manner as for Task 1a.

For Task 1c, we trained one ChemBERT model for coreference and a second for bridging relations. The reason why we split the relations is because their mentions are partly different from each other. Therefore, we aimed to suppress false-positive relations caused by false-positive

[1] https://github.com/spyysalo/standoff2conll

mentions from the other ones. In these models, single-label mention detection was performed and the relation labels were given in the relation classification step. Sentences were tokenized in the same manner as for Task 1a, with B, I, O and D labels from BIOHD [15] being used for tagging because Task 1c contains discontinuous mentions. When overlapping mentions were tokenized, we used the longer entities and discarded the shorter ones. Because the number of mentions in the training set for coreference was smaller than for other datasets used for mention detection, we augmented the number of positive examples in the training set by reusing sentences that contained one or more mentions five times.

3.2. Relation Classification

For Task 1b, we trained a one ChemBERT model for ARG1 relations and a second for ARGM relations. The input to the relation classification was a sentence with the candidate pair for a relation between an event enclosed by [E1] and [/E1] tokens and a target enclosed by [E2] and [/E2] tokens. The output was a binary classification result indicating whether the pair has a relation or not. All candidate pairs in a sentence were classified and positive relations were annotated by the system. For example, if two events and three targets are included in a sentence, the number of candidate pairs would be six. This approach is similar to the Melax Tech system [11], which performed best in the ChEMU 2020 task [9]. This approach was also discussed in a general framework [16]. In the training stage, we used not only gold-standard entities but also predicted events that were generated by five systems trained on 80% of the training set, similarly to five-fold cross-validation.

For Task 1c, we trained one ChemBERT model for coreference and a second for bridging relations. The input to the relation classification was a pair of sentences representing a candidate pair for a relation between an anaphor enclosed by [E1] and [/E1] tokens and an antecedent enclosed by [E2] and [/E2]. The reason for using a pair of sentences, different from Task 1b, was that relations in Task 1c were often across sentences. If a mention was discontinuous, the first block of the mention was enclosed. The output was the label of the relation or a NO_RELATION label.

We applied two post-processing methods for Task 1c because relations that involve more than two sentences were included in the snippets. First, for coreference relations, when REACTION_PRODUCT appeared multiple times in a snippet and the sentence-level distance of the mentions was more than two, which means it cannot be found by ChemBERT, we assign a COREFERENCE relation. Second, for bridging relations, when a candidate for an antecedent did not have any anaphors, we searched for antecedent candidates for the anaphor by finding words that started with "the" and were an anaphor for another antecedent. The candidate for the anaphor that was closest to the antecedent was then selected as the anaphor. If the antecedent contained STARTING_MATERIAL, REAGENT_CATALYST or SOLVENT, the relation was annotated with REACTION_ASSOCIATED. Otherwise, the relation was annotated with a label that was the same as the already annotated anaphor after the antecedent.

In addition to the above methods, we used a post-processing tool distributed by the task organizer[2], which generates a coreference between A and C when coreferences between A and B and between B and C already exist.

[2] https://raw.githubusercontent.com/yuan-li/chemu2021/master/apply-transitive-closure.py

3.3. Experimental Settings

We used ChemBERT v3.0 [8] for the mention detection and relation classification models. ChemBERT was implemented by using AllenNLP [17] and HuggingFace Transformers [18]. We used the AdamW optimizer [19] and cross entropy loss for optimization. The models were trained on the training set for the task and evaluated on the development set and both public and private test sets. Hyperparameter values were set as follows: max sequence length=384 (covering all sequences contained in the training and development sets), batch size=16, learning rate=1e-5, and patience=7. Because the relation classification in Task 1c accepts a pair of sentences as its input, a maximum sequence length of 512 was used for this task. The validation metric for early stopping of mention detection in the development set was the F-score. For relation classification, it was the validation loss.

The performances of the various systems were evaluated with respect to both exact and relaxed matching for precision, recall, and F-score.

4. Main Results

We submitted a system with post-processing for Task 1a before the deadline and a system without post-processing after the deadline. Table 2 shows our results for Task 1a on the private set. Our system obtained an exact match F-score of 0.9412 and a relaxed match F-score of 0.9572. Table 3 shows our results for the private set in detail.

Table 2
Results for Task 1a on the private set. Here, P represents precision, R represents recall, and F represents F-score

Relation	Exact			Relaxed		
	P	R	F	P	R	F
ChemBERT (late)	0.9327	0.9349	0.9338	0.9481	0.9503	0.9492
ChemBERT + PP	0.9401	0.9422	0.9412	0.9561	0.9583	0.9572

We submitted one system for Task 1b before the deadline. We also submitted a corrected version after the deadline, having found an error related to the sequence length of the input to the system. Table 4 shows our results for Task 1b on the private set. The corrected system obtained an exact match F-score of 0.8865 and a relaxed match F-score of 0.9027. Table 5 shows our results for the private set in detail.

We submitted three systems for Task 1c before the deadline. We also submitted a corrected version after the deadline, having found an error related to the sequence length of the input to the system. Table 6 shows our results for Task 1c on the private set. The corrected system obtained an exact match F-score of 0.7232 and a relaxed match F-score of 0.8053. Table 7 shows our results on the private set in detail.

Table 3
Detailed results for Task 1a on the private set, as predicted by the post-processing version of the system

Entity	Exact			Relaxed		
	P	R	F	P	R	F
EXAMPLE_LABEL	0.9714	0.9913	0.9812	0.9714	0.9913	0.9812
OTHER_COMPOUND	0.9498	0.9486	0.9492	0.9637	0.9625	0.9631
REACTION_PRODUCT	0.9112	0.9034	0.9073	0.9433	0.9352	0.9392
REAGENT_CATALYST	0.8529	0.9050	0.8782	0.8721	0.9253	0.8979
SOLVENT	0.9284	0.9666	0.9471	0.9284	0.9666	0.9471
STARTING_MATERIAL	0.8997	0.8594	0.8791	0.9353	0.8934	0.9138
TEMPERATURE	0.9802	0.9770	0.9786	0.9901	0.9869	0.9885
TIME	0.9673	0.9741	0.9707	0.9883	0.9953	0.9918
YIELD_OTHER	0.9853	0.9711	0.9782	0.9902	0.9759	0.9830
YIELD_PERCENT	0.9750	0.9943	0.9846	0.9778	0.9972	0.9874
All	0.9401	0.9422	0.9412	0.9561	0.9583	0.9572

Table 4
Results for Task 1b on the private set

Relation	Exact			Relaxed		
	P	R	F	P	R	F
ChemBERT	0.9058	0.8685	0.8868	0.9222	0.8842	0.9028
ChemBERT Corrected (late)	0.9054	0.8684	0.8865	0.9220	0.8842	0.9027

5. Discussion

Selecting “the best model” of the various models in training was difficult. Table 8 shows the F-scores on the development and private sets. We used models that showed the best F-score on the development set for mention detection and the best validation loss for relation classification; therefore, it is not surprising that F-scores on the development sets were better than those for the private sets. In particular, the result for Task 1c (greater than 0.09) represented a large gap. An explanation for this larger gap could be that the predictions by the models were unstable because Task 1c was more difficult than the other tasks. Therefore, we must reconsider training methods when seeking a better model.

In Task 1b, all relations whose recalls were zero had fewer than five gold-standard examples (Table 5). It is quite difficult for our machine learning framework to identify such relations with a small amount of examples.

Although our system showed good results for Tasks 1a and 1b, we found errors caused by a lack of document-level information. Figure 2 shows a confusion matrix for Task 1a. Because the role of a compound depends on a reaction, it is difficult to identify the label of compounds without document-level information. Examples included errors among STARTING_MATERIAL, REAGENT_CATALYST, and SOLVENT and errors between REACTION_PRODUCT and OTHER_COMPOUND. In addition, errors between a compound for a reaction (REAGENT_CATALYST, SOLVENT, STARTING_MATERIAL) and one for a work-up

Table 5
Detailed results for Task 1b on the private set, as predicted by the corrected system. * represents relations that had fewer than five gold-standard examples

Relation	Exact			Relaxed		
	P	R	F	P	R	F
ARG1|REACTION_STEP|OTHER_COMPOUND	0.4756	0.5342	0.5032	0.5000	0.5616	0.5290
ARG1|REACTION_STEP|REACTION_PRODUCT	0.8731	0.8561	0.8645	0.9179	0.9000	0.9089
ARG1|REACTION_STEP|REAGENT_CATALYST	0.8399	0.8744	0.8568	0.8590	0.8904	0.8744
ARG1|REACTION_STEP|SOLVENT	0.8929	0.8883	0.8906	0.8596	0.8950	0.8770
ARG1|REACTION_STEP|STARTING_MATERIAL	0.8832	0.8043	0.8419	0.9036	0.8228	0.8613
ARG1|WORKUP|OTHER_COMPOUND	0.9455	0.9052	0.9249	0.9627	0.9217	0.9418
*ARG1|WORKUP|REACTION_PRODUCT	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
*ARG1|WORKUP|REAGENT_CATALYST	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
*ARG1|WORKUP|SOLVENT	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
*ARG1|WORKUP|STARTING_MATERIAL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ARGM|REACTION_STEP|TEMPERATURE	0.9262	0.8750	0.8999	0.9328	0.8811	0.9062
ARGM|REACTION_STEP|TIME	0.8976	0.9024	0.9000	0.9213	0.9261	0.9237
ARGM|REACTION_STEP|YIELD_OTHER	0.9845	0.9315	0.9573	0.9871	0.9340	0.9598
ARGM|REACTION_STEP|YIELD_PERCENT	0.9671	0.9229	0.9444	0.9701	0.9257	0.9474
ARGM|WORKUP|TEMPERATURE	0.9063	0.6541	0.7598	0.9271	0.6692	0.7773
ARGM|WORKUP|TIME	0.7895	0.4054	0.5357	0.7895	0.4054	0.5357
*ARGM|WORKUP|YIELD_OTHER	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
*ARGM|WORKUP|YIELD_PERCENT	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
All	0.9054	0.8684	0.8865	0.9220	0.8842	0.9027

Table 6
Results for Task 1c on the private set. PP_{BR} represents post-processing for bridging relations and PP_{CR} represents post-processing for coreference relations.

Relation	Exact			Relaxed		
	P	R	F	P	R	F
ChemBERT	0.7393	0.6616	0.6983	0.8222	0.7358	0.7766
ChemBERT + PP_{BR}	0.7290	0.6838	0.7057	0.8107	0.7604	0.7848
ChemBERT + PP_{BR} + PP_{CR}	0.6876	0.7307	0.7085	0.7660	0.8140	0.7893
ChemBERT + PP_{BR} + PP_{CR} Corrected (late)	0.7144	0.7322	0.7232	0.7955	0.8153	0.8053

(OTHER_COMPOUND) were found because distinguishing between them from just one sentence was sometimes difficult. Errors between REACTION_STEP and WORK_UP for Task 1b were also found to be caused by the same difficulty.

However, adopting a post-processing method for Task 1a mitigated the errors between REACTION_PRODUCT and OTHER_COMPOUND (Table 2). The post-processing methods employed in Task 1c were also useful in mitigating these errors (Table 6). However, we should note that using these methods adversely affected the precision. For example, our post-processing method for coreference generated false relations when false positive REACTION_PRODUCT existed. Therefore, we must be careful when using post-processing methods.

Table 7
Results for Task 1c on the private set, as predicted by the corrected system

Relation	Exact			Relaxed		
	P	R	F	P	R	F
COREFERENCE	0.4896	0.4882	0.4889	0.5975	0.5958	0.5967
CONTAINED	0.5054	0.6267	0.5595	0.7312	0.9067	0.8095
REACTION_ASSOCIATED	0.7368	0.7974	0.7659	0.8094	0.8760	0.8414
TRANSFORMED	0.7310	0.7576	0.7440	0.7368	0.7636	0.7500
WORK_UP	0.8219	0.8230	0.8224	0.8940	0.8952	0.8946
All	0.7144	0.7322	0.7232	0.7955	0.8153	0.8053

Table 8
Comparison of F-scores between development and private sets

Task	Exact		Relaxed	
	Development	Private	Development	Private
Task 1a	0.9548	0.9412	0.9677	0.9535
Task 1b	0.9179	0.8865	0.9294	0.9027
Task 1c	0.8168	0.7232	0.8773	0.8053

With the aim of improving our systems, we tried to consider the relationships between the tasks in some preliminary experiments. For example, we tried to construct post-processing rules for named entity mentions in Task 1a by on the results for Task 1b. However, these rules did not improve the results because it was difficult to determine which prediction (named entity or event) was correct. In addition, we tried to use the relationships not only in the forward direction (i.e., Task 1a to Task 1c), but also in the backward direction (Task 1c to Task 1a). However, improving the performance in the backward direction was also difficult because the performance of the later task was lower than in the earlier task. Despite these difficulties, the post-processing methods for Task 1c that considered the results for Task 1a did improve our system's performance. Therefore, we must conduct a more detailed analysis of the relationships between the tasks if we are to improve our systems via this approach.

Table 9 shows the results for mention detection in Task 1c on the development set. First, the detection of coreference mentions was difficult compared to identifying bridging relations. The reasons were that coreference mentions had only a small number of mentions in the training data and sometimes required inter-sentence information to extract antecedents. The significance of the data augmentation was not clear. Therefore, we must reconsider the ratios used for positive mentions.

Several errors were caused by failures in sentence splitting. For example, STARTING_MATERIAL "Ex. 18A" caused a split into two sentences by the ChemDataExtractor sentence splitter because of the "." in "Ex. 18A". A solution to this problem would involve applying a set of rules involving dependency parsing and trigger words.

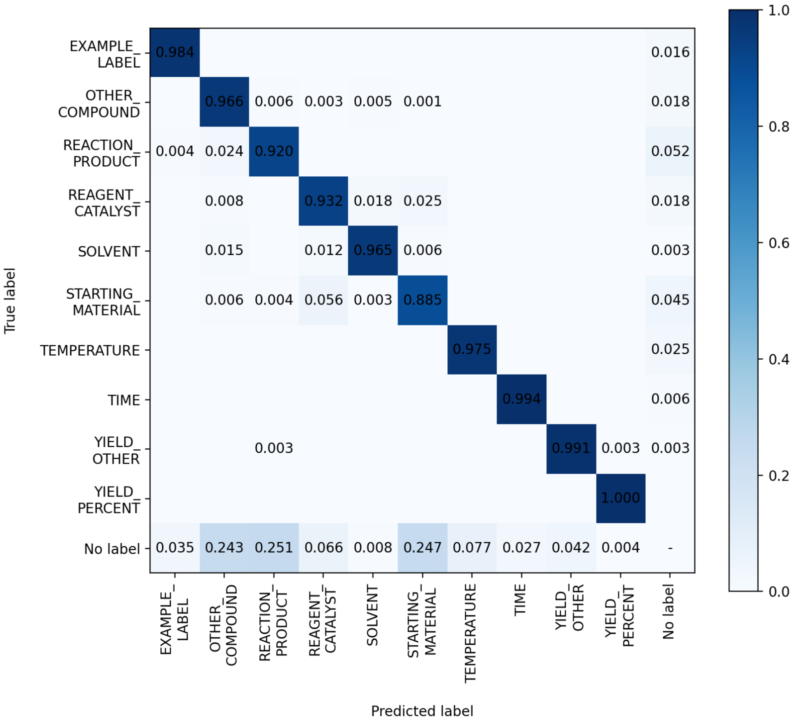

Figure 2: Normalized confusion matrix for Task 1a. Values less than 0.001 are not shown

Table 9
Results for mention detection in Task 1c on the development set

Relation	Exact			Relaxed		
	P	R	F	P	R	F
Coreference	0.8099	0.8085	0.8092	0.8746	0.8730	0.8738
Coreference with data augmentation	0.7982	0.8316	0.8145	0.8567	0.8926	0.8743
Bridging	0.9090	0.9415	0.9250	0.9542	0.9884	0.9710

6. Conclusion

This paper has reported our results for the three tasks at ChEMU 2022. We proposed hybrid methods that used ChemBERT and a small set of post-processing rules for these tasks. We employed a pipeline approach for Tasks 1b and 1c that combined mention detection and relation classification. Because we used only one or two sentences as the input to ChemBERT, this lack of document-level information suppressed the performance of the system. Although we confirmed that adopting a set of post-processing rules was effective in considering document-level information, we also confirmed that the set of rules we used was insufficient. In addition, although we tried to use relationships between the tasks to improve performance, it was difficult to construct rules that did achieve improvements. Therefore, we must conduct more detailed analyses about the relationships between the tasks.

Acknowledgments

We would like to thank the ChEMU team for providing the datasets. This work was partially supported by JSPS KAKENHI Grant Number 21K19814.

References

[1] M. Bregonje, Patents: A unique source for scientific technical information in chemistry related industry?, World Patent Information 27 (2005) 309–315. URL: https://www.sciencedirect.com/science/article/pii/S0172219005000736. doi:https://doi.org/10.1016/j.wpi.2005.05.003.

[2] J. He, D. Q. Nguyen, S. A. Akhondi, C. Druckenbrodt, C. Thorne, R. Hoessel, Z. Afzal, Z. Zhai, B. Fang, H. Yoshikawa, A. Albahem, L. Cavedon, T. Cohn, T. Baldwin, K. Verspoor, ChEMU 2020: Natural Language Processing Methods Are Effective for Information Extraction From Chemical Patents, Frontiers in Research Metrics and Analytics 6 (2021). URL: https://www.frontiersin.org/article/10.3389/frma.2021.654438. doi:10.3389/frma.2021.654438.

[3] B. Fang, C. Druckenbrodt, S. A. Akhondi, J. He, T. Baldwin, K. Verspoor, ChEMU-Ref: A Corpus for Modeling Anaphora Resolution in the Chemical Domain, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, 2021, pp. 1362–1375. URL: https://www.aclweb.org/anthology/2021.eacl-main.116.

[4] H. Yoshikawa, S. Akhondi, C. Thorne, C. Druckenbrodt, R. Hoessel, Z. Zhai, J. He, T. Baldwin, K. Verspoor, Chemical Reaction Reference Resolution in Patents (2021).

[5] Z. Zhai, C. Druckenbrodt, C. Thorne, S. A. Akhondi, D. Q. Nguyen, T. Cohn, K. Verspoor, ChemTables: a dataset for semantic classification on tables in chemical patents, Journal of Cheminformatics 13 (2021) 1–20.

[6] Y. Li, B. Fang, J. He, H. Yoshikawa, S. Akhondi, C. Druckenbrodt, C. Thorne, Z. Zhai, Z. Afzal, T. Cohn, T. Baldwin, K. Verspoor, The ChEMU 2022 Evaluation Campaign: Information Extraction in Chemical Patents, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2022, pp. 400–407.

[7] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics (2019). doi:10.1093/bioinformatics/btz682.

[8] J. Guo, A. S. Ibanez-Lopez, H. Gao, V. Quach, C. W. Coley, K. F. Jensen, R. Barzilay, Automated Chemical Reaction Extraction from Scientific Literature, Journal of Chemical Information and Modeling 62 (2022) 2035–2045. URL: https://doi.org/10.1021/acs.jcim.1c00284. doi:10.1021/acs.jcim.1c00284. arXiv:https://doi.org/10.1021/acs.jcim.1c00284, pMID: 34115937.

[9] J. He, D. Quoc Nguyen, S. A. Akhondi, C. Druckenbrodt, C. Thorne, R. Hoessel, Z. Afzal, Z. Zhai, B. Fang, H. Yoshikawa, et al., An extended overview of the CLEF 2020 ChEMU

lab: information extraction of chemical reactions from patents, in: Proceedings of CLEF (Conference and Labs of the Evaluation Forum) 2020 Working Notes, 2020.

[10] Y. Li, B. Fang, J. He, H. Yoshikawa, S. A. Akhondi, C. Druckenbrodt, C. Thorne, Z. Afzal, Z. Zhai, T. Baldwin, et al., Extended overview of ChEMU 2021: reaction reference resolution and anaphora resolution in chemical patents, CLEF (Working Notes) (2021).

[11] J. Zhang, Y. Zhang, Melaxtech: a report for clef 2020–ChEMU task of chemical reaction extraction from patent, Work Notes CLEF. Published online.[Google Scholar] (2020).

[12] R. Dutt, S. Khosla, C. P. Rosé, A pipelined approach to Anaphora Resolution in Chemical Patents., in: CLEF (Working Notes), 2021, pp. 710–719.

[13] M. C. Swain, J. M. Cole, ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature, Journal of Chemical Information and Modeling 56 (2016) 1894–1904. URL: https://doi.org/10.1021/acs.jcim.6b00207. doi:10.1021/acs.jcim.6b00207. arXiv:https://doi.org/10.1021/acs.jcim.6b00207, pMID: 27669338.

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://www.aclweb.org/anthology/N19-1423. doi:10.18653/v1/N19-1423.

[15] B. Tang, Q. Chen, X. Wang, Y. Wu, Y. Zhang, M. Jiang, J. Wang, H. Xu, Recognizing disjoint clinical concepts in clinical text using machine learning-based methods, in: AMIA annual symposium proceedings, volume 2015, American Medical Informatics Association, 2015, p. 1184.

[16] Z. Zhong, D. Chen, A Frustratingly Easy Approach for Entity and Relation Extraction, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 50–61. URL: https://aclanthology.org/2021.naacl-main.5. doi:10.18653/v1/2021.naacl-main.5.

[17] M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. F. Liu, M. Peters, M. Schmitz, L. S. Zettlemoyer, AllenNLP: A Deep Semantic Natural Language Processing Platform, 2017. arXiv:arXiv:1803.07640.

[18] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. M. Rush, Transformers: State-of-the-Art Natural Language Processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https://www.aclweb.org/anthology/2020.emnlp-demos.6.

[19] I. Loshchilov, F. Hutter, Decoupled Weight Decay Regularization, in: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019, OpenReview.net, 2019. URL: https://openreview.net/forum?id=Bkg6RiCqY7.