PcMSP: A Dataset for Scientific Action Graphs Extraction from Polycrystalline Materials Synthesis Procedure Text

Xianjun Yang[1], **Ya Zhuo**[2], **Julia Zuo**[2], **Xinlu Zhang**[1], **Stephen Wilson**[2], **Linda Petzold**[1]
[1]Department of Computer Science [2]Department of Materials Science and Engineering
University of California, Santa Barbara
{xianjunyang, yzhuo, jlzuo, xinluzhang, stephendwilson, petzold}@ucsb.edu

Abstract

Scientific action graphs extraction from materials synthesis procedures is important for reproducible research, machine automation, and material prediction. But the lack of annotated data has hindered progress in this field. We demonstrate an effort to annotate **P**oly**c**rystalline **M**aterials **S**ynthesis **P**rocedures (**PcMSP**) from 305 open access scientific articles for the construction of synthesis action graphs. This is a new dataset for material science information extraction that simultaneously contains the synthesis sentences extracted from the experimental paragraphs, as well as the entity mentions and intra-sentence relations. A two-step human annotation and inter-annotator agreement study guarantee the high quality of the PcMSP corpus. We introduce four natural language processing tasks: sentence classification, named entity recognition, relation classification, and joint extraction of entities and relations. Comprehensive experiments validate the effectiveness of several state-of-the-art models for these challenges while leaving large space for improvement. We also perform the error analysis and point out some unique challenges that require further investigation. We will release our annotation scheme, the corpus, and codes to the research community to alleviate the scarcity of labeled data in this domain[1].

1 Introduction

Synthesis procedural texts are written in instructional languages (Grishman, 2001; Grishman and Kittredge, 2014) to represent the step-by-step reactions, but also contain the distinct features in specific domains, such as the domain notations, writing styles, and journal requirements. The synthesis procedures of materials science articles include valuable information for new materials prediction (Raccuglia et al., 2016), laboratory automation (Coley et al., 2019) and knowledge graph construction (Mrdjenovich et al., 2020). However, available datasets are extremely limited, despite the notable work by (Mysore et al., 2017, 2019; Friedrich et al., 2020; O'Gorman et al., 2021).

Synthesis Paragraph

Polycrystalline[Descriptor] sample of composition *Sr2CoO4*[Material_target] was *synthesized*[operation] under *high pressure*[Property_pressure] at *high temperature*[Property_temperature]. Starting materials of *SrO2*[Material_recipe] and *Co*[Material_recipe] were *well*[Descriptor] *mixed*[operation] in a *molar ratio*[Descriptor] of *SrO2*[Material_recipe] : *Co*[Material_recipe]=*2 : 1*[Value]. The *mixture*[Material-intermedium] was *sealed*[operation] into *a*[Value] *gold*[Descriptor] *capsule*[Device]. ... The crystal structure of the polycrystalline sample was identified by the powder X-ray diffraction (XRD, Rigaku Smart- lab3), using Cu-Kα radiation (λ=1.54184Å). ...

Table 1: An example of a synthesis paragraph from our dataset with index srep27712 (Li et al., 2016).

The goal of information extraction from procedures is to construct the action graphs, which refer to all the steps in a synthesis making up a Directed Acyclic Graph (DAG) (Mysore et al., 2019; Kulkarni et al., 2018) (as can be seen from one example in Figure 1). This can be further breakdown into three tasks: sentence classification, named entity recognition (NER), and relation extraction (RE). Previous research (Mysore et al., 2017, 2019) either annotates the whole synthesis paragraph in the general inorganic domain, ignoring the non-synthesis sentences and subdomain discrepancy or only focuses on entity mentions (Friedrich et al., 2020; O'Gorman et al., 2021).

To fill this gap, we focus on one important category of polycrystalline materials and simultane-

[1]https://github.com/Xianjun-Yang/PcMSP

Findings of the Association for Computational Linguistics: EMNLP 2022, pages 6033–6046
December 7-11, 2022

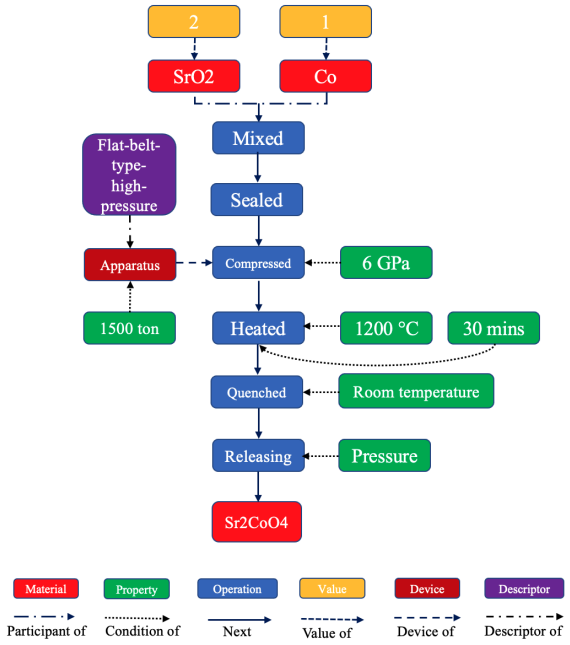

Figure 1: A synthesis action graph constructed from Table 1.

ously include all three tasks. The annotation guidelines are designed by materials experts after comprehensive discussion, and the new dataset is subsequently labeled with a two-round annotation.

The key contributions of this paper include:

- We contribute a new large-scale dataset, as well as an annotation scheme with high quality for information extraction in materials science.
- We conduct comprehensive experiments on four tasks, sentence classification, named entity recognition, relation extraction, and joint extraction to provide baselines.
- We perform error analysis and point out unique challenges and potential use of this dataset for future research.

2 Related Work

Scientific information extraction

With the fast-growing volume of scholarly publications, it is highly demanding to extract structured information from large-scale scientific literature in many domains (Augenstein et al., 2017; Luan et al., 2018; Jiang et al., 2019; Beltagy et al., 2019; Buscaldi et al., 2019), like biomedical domain (Shah et al., 2003; Lai et al., 2021; Zhang et al., 2021; Lewis et al., 2020; Kulkarni et al., 2018) and chemistry domain (Rocktäschel et al., 2012; He et al., 2020). In the field of materials science, there have been few attempts in this direction, leaving many unexplored challenges for research (Hong et al., 2021). Recent research mainly focuses on knowledge base construction (Jiang et al., 2019; Luan et al., 2018), new materials discovery (Isayev, 2019), and automation of lab procedures (Vaucher et al., 2020; Tamari et al., 2021; Steiner et al., 2019). (Beltagy et al., 2019) trained a Bidirectional Encoder Representations from Transformers model (SciBERT) on 1.14M scientific papers from Semantic Scholar for scientific information extraction.

Materials procedures information extraction

In the area of annotation of materials synthesis procedures, (Mysore et al., 2019) annotate 230 general materials synthesis paragraphs for NER and RE tasks. Similar work is also undertaken by (Friedrich et al., 2020), in which 45 open access scholarly articles are labeled for experiment-describing sentence classification, NER, and slot filling tasks. However, in contrast to our works, their annotation scheme focuses on the full text rather than the experimental section. (Kuniyoshi et al., 2020) annotate the synthesis process of all-solid-state batteries from the scientific literature, but their corpus is not publicly available. (Walker et al., 2021) release MatBERT trained on 50 million materials science paragraphs to explore the impact of domain-specific pre-training on NER task. Also of interest, (O'Gorman et al., 2021) recently create the largest corpus for entity mentions extraction in both general domain and subdomain from material synthesis text, but the relations between entities are still missing.

Named entity recognition and relation extraction

Many neural network-based models have been proposed for named entity recognition, for example, (Huang et al., 2015; Lample et al., 2016; Panchendrarajan and Amaresan, 2018). The core idea uses one encoding layer (e.g. Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997), BERT) for representation and one additional conditional random fields (CRF (Lafferty et al., 2001)) layer for sequence labeling. Then relations are predicted based on either gold entities or predicted entities, and PURE (Zhong and Chen, 2021) designs two separate encoders for joint extraction of

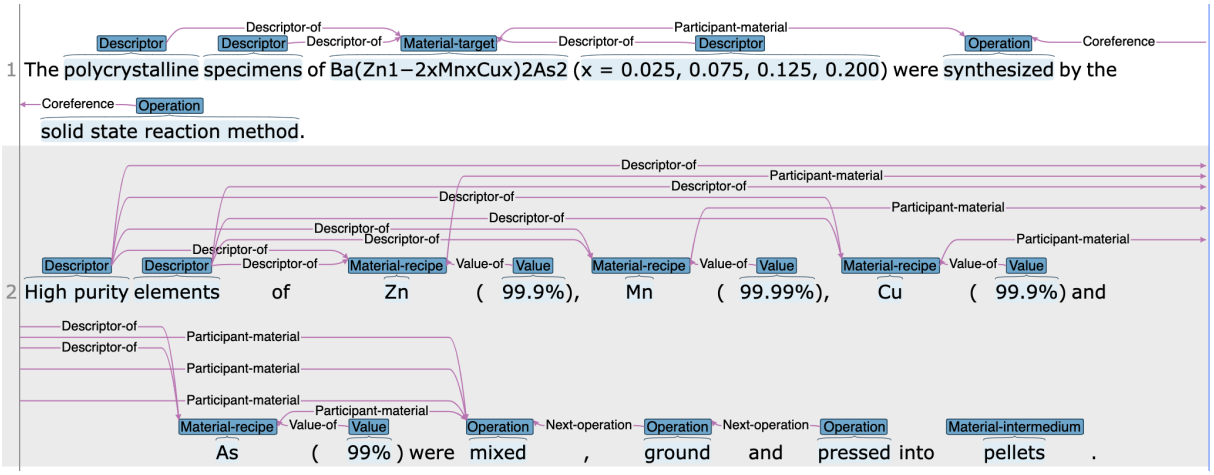

Figure 2: An annotated PcMSP example on the INCEpTION platform, taken from srep15507 (Man et al., 2015).

entities and relations. We adopt their model for our tasks due to its super performance.

3 The Selection of Our Dataset

Here we talk about the importance of our selection and how is it different from other materials procedural text corpora.

Why do we choose inorganic polycrystalline materials? There are a number of sub-categories within solid-state inorganic materials. For example, materials can be divided based on function and properties, such as the battery or thermoelectric materials. Synthesis within both categories largely falls within the broader category of solid-state synthesis and even then, there is a high degree of overlap with other function categories, such as quantum and magnetic materials. More importantly, **those materials are usually in the form of polycrystalline**. Other subcategories relate to form factors, for instance, single-crystalline synthesis often starts with a **polycrystalline** synthesis and therefore has a high degree of overlap with solid-state synthesis.

Inorganic polycrystal compounds span combinations of the entire periodic table and different chemical bonding schemes, such that their synthesis typically takes place under extreme conditions, such as high temperature and pressure. Reaction pathways are therefore difficult to characterize without specialized equipment and are not well established for any given material. In particular, solid-state reactions, which are the main techniques to synthesize inorganic polycrystalline materials, are particularly similar to a "black box", where materials scientists can only make educated guesses to the procedure or stability of a new reaction. This presents a prime opportunity (Mysore et al., 2017, 2019) for compiling published inorganic synthesis data in order to demystify the black box of solid-state inorganic materials synthesis and create datasets for future text mining endeavors. While there have been efforts within general solid-state materials (Mysore et al., 2017, 2019; O'Gorman et al., 2021) and battery materials subcategory (Friedrich et al., 2020), this work aims to extend the subcategory of inorganic solid-state synthesis methods in order to address the frequent overlap and "borrowing" of materials between subdisciplines of materials science.

Why do we discard characterization sentences? Inorganic reactions typically involve relatively few reactions from a set of precursors and there are very few purification pathways for solid materials compared to organic materials or liquids. Therefore, characterizations of solid-state inorganic reactions are seldom reported in literature unless they proceed to complete purity within standard measurement fidelity. This is in contrast to organic materials where there are a number of important characterization metrics in a compound, such as molecular weight in polymers or reaction yield. Therefore, these standard characterization measurements do not add valuable information for a researcher attempting to recreate the reported synthesis method and we decide to discard these characterization sentences.

Why do we annotate sentence, entity, and relation simultaneously? A full action graph consists of both entities and relations extracted from experimental-describing sentences. However, most previous research either ignores the annotation of sentence or relation information, making them in-

complete for action graph construction. To fill this gap, we aim to annotate all pertinent information jointly.

4 Description of the Annotation

4.1 Selection of synthesis procedures for annotation

We begin by harvesting the polycrystalline materials synthesis-related open access publications from the main journal publishers by searching keywords (e.g. 'polycrystalline+synthesis'). The journals that we used include Physical Review Journals[2], Nature journals[3], Science journals[4], Journal of the American Chemical Society[5], Advanced Materials[6], Journal of Physics Condensed Matter[7], Chemistry of Materials[8] and ArXiv[9]. After the collection of 305 publications, each portable document format (PDF) document is converted into a plain text file by pdfminer[10]. The experimental paragraphs usually appear in the experimental section within an article and are selected by one materials expert. To improve the data quality, the selected paragraphs are double-checked by another annotator to ensure their correctness. And some missing sentences caused by the conversion process are also added. Finally, the collected paragraphs are prepared for the next step of annotations.

4.2 Sentence annotation

Based on the selected paragraphs from the aforementioned step, each document is annotated on the semantic annotation platform INCEpTION (Klie et al., 2018), and the sentence segmentation is carried out automatically[11]. Each line represents all tokens of one sentence, and the annotation is done on the token level. In practice, only the synthesis-related sentences are annotated for NER and RC. The resulting unlabeled sentences automatically obtain non-synthesis labels. This process resulted in 1497 synthesis-related sentences and 971 non-related sentences. It is worthwhile to point out that several selected paragraphs also contain single crystal synthesis (this occurs < 1%), but we do not take those as synthesis-related sentences so as to focus purely on polycrystalline synthesis. In general, most non-synthesis sentences are relevant to the characterization of materials, description of devices, etc. While synthesis sentences typically describe the synthesis actions conducted in the experiments. For example, in Table 1, the first two sentences are synthesis-related while the remaining sentences are not.

4.3 Entity type annotation

We defined 13 entity types to include the most useful entity mentions, which are decided by the materials experts. Each span of continuous words is labeled as a certain kind of entity type. There are five general categories of labels, namely **Material**: Material-target, Material-recipe, Material-intermedium and Material-others, **Property**: Property-time, Property-temperature, Property-rate and Property-pressure, **Operation**, **Item**: Value, Brand, Device and **Descriptor**. Every general coarse-grained category can further be divided into one or several fine-grained types. The full definitions of these labels can be found in the following.

Material-target: final material (or products) of the material synthesis process, usually refers to only one target in a typical procedural paragraph, but can appear as multiple target materials (this occurs less than 1%).

Material-recipe: raw material used to synthesize the final product, can be fundamental elements(like Si), compounds(like $SrO2$), or precursors of other polycrystalline materials.

Material-intermedium: an intermediate material produced during the synthesis process that is subsequently used as participants in the following reactions.

Material-others: materials that are not compositionally related to the final material or used as solvents (like $water$) to provide reaction conditions.

Operation: an individual action performed by the experimenters, which is often represented by verbs or a particular overall synthesis method, like $Solid-state-reaction$.

Property-time: a time condition associated with an operation, which is usually composed of numerical values and time units.

[2]https://journals.aps.org/
[3]https://www.nature.com/
[4]https://www.science.org/journals
[5]https://pubs.acs.org/journal/jacsat
[6]https://onlinelibrary.wiley.com/journal/15214095
[7]https://iopscience.iop.org/journal/0953-8984
[8]https://pubs.acs.org/journal/cmatex
[9]https://arxiv.org/
[10]https://pdfminersix.readthedocs.io/en/latest/
[11]InCeption uses Java's built-in sentence segmentation algorithm with US locale.

Dataset	Domain	Procedure only	Documents	Sentence type	Sentences	Entity type	Entities	Relation type	Relations
MSPT	General	✗	230	✗	2112	21	20849	16	18402
SOFC-Exp	Subdomain	✗	45	2	853	16	5095	✗	✗
SC-CoMIcs	Subdomain	✗	1000	✗	6639	7	42337	✗	✗
MS-MENTIONS	General	-	595	✗	7980	14	44295	✗	✗
Our PcMSP	Subdomain	✓	305	2	2468	13	14592	8	13968

Table 2: Corpus statistics of our PcMSP and previous datasets for materials science. ✗denotes that no such information is contained in the corresponding corpus. - denotes that the corpus has not been released yet.

Property-temperature: a temperature condition associated with an operation, which is usually composed of numerical values and temperature units.
Property-rate: a rate condition associated with an operation, which is usually composed of numerical values and rate units. The rates can be rotation speed, cooling, or heating rates, etc.
Property-pressure: a pressure condition associated with an operation, which is not only in the form of value and units but also can be a certain condition like vacuum, helium, or air.
Value: numerical values and their corresponding units. In addition, we include specifications like "around", "over", “more than” or “between” in the annotation span (e.g., “around 250 g,” and “over 20 mol”). We do not include time, temperature, pressure, or rate in this category, as they are already included in properties.
Device: mentions of the type of device used in the corresponding operation, which can contain the device name and serial number.
Brand: the brand name or source laboratory associated with the equipment or material.
Descriptor: description of an operation or a material or a value that does not apply to properties but is necessarily included for clear descriptions.

4.4 Relation type annotation

The previous two steps provide us with the labeled entity mentions within each sentence. We then connect each entity pair by a relation type when there is a believed necessary connection, according to the definition of agreement study. The full descriptions of relation labels are listed in the following.
Participant-material: materials that are involved in one operation process, and we also mark the target material and its synthesis action as this label.
Device-of-operation: a device used in an operation.
Condition-of: indicates the conditions of an operation (such as the temperature, time, and pressure) for performing an operation.
Value-of: expresses the relationships between participated material and their weight, mass, volume, or purity, and also represents the relationship between the device and its serial number.
Next-operation: represents the order of an operation sequence that one operation that happens following the previous operation. Note that we assume the linear sequence of synthesis operations happens sentence by sentence, which is true for most cases.
Brand-of: expresses the relationships between a raw material or device and its manufacturer name or source laboratory.
Descriptor-of: the descriptor for the material, device, or operation that can not be covered by other labels.
Coreference: represents the same material or operation in the same sentence.

Besides, according to the largest Document-level relation extraction dataset (Yao et al., 2019), around 40% of relations exist across multiple sentences. But cross-sentence relation is out of our scope for current work and we leave it for future investigation.

5 Inter-annotator Agreement Study

We perform a two-round agreement study to ensure that our corpus has a high quality of annotation. Before undertaking the formal annotation, all four annotators participate in a discussion of the formulation rules and discuss the necessary entity and relation types. In the warm-up exercise, all annotators annotate the same documents individually and then compare and discuss the results together to achieve better agreement on annotation. After the agreements are formulated, in the first-round annotation four annotators are randomly assigned different documents to work on. It takes around twenty to thirty minutes to annotate one document on average for all annotators. When all of the annotations are finished, two of the four annotators select several typical examples for analysis and

Round	Sen.	En1.	En2.	Re1.	Re2.
First-round	80.13	56.41	92.8	48.51	90.2
Second-round	85.06	69.81	93.44	53.63	91.03

Table 3: Two-round inter-annotator agreement study measured by Fleiss' Kappa.

eventually set more rules for annotating the most debatable parts. In the second round of annotation, two lead annotators individually re-annotate half of the documents, guaranteeing that there are no significant differences or mistakes. It takes around 500 hours for our material expert team in total to create this corpus to guarantee high quality.

We use Fleiss' Kappa to measure the agreement scores between our four annotators. The result is shown in Table 3, with substantially high agreement scores. We can see obvious improvements in all aspects from the first to second round annotation, demonstrating the effectiveness of our annotation pipeline. We use five metrics to measure the agreement score: Sen. refers to sentence agreement, En1. means span boundaries and type are both correct, En2. means matched type on same spans, Re1. represents complete relation triple with correct entities and Re2. stands for correct relation type on same entities. More details are discussed in Appendix D.

6 Statistics of Corpus and Problem Formulation

In this section, we describe the statistics of this new dataset, the comparison with precious corpora, and formulated tasks.

6.1 PcMSP corpus

We outline the main material science corpus in Table 2, including Materials Science Procedural Text (MSPT) (Mysore et al., 2019), SC-CoMIcs (Yamaguchi et al., 2020), SOFC-exp (Friedrich et al., 2020) and MS-MENTIONS (O'Gorman et al., 2021), as well as our PcMSP corpus. Among those corpora, MSPT focuses on general solid-state compounds and is most similar to ours. But MSPT contains annotation for all sentences in synthesis procedural paragraphs, even though many of those sentences are actually describing material characteristic methods rather than synthesis procedures. On the other hand, the SC-CoMIcs and MS-MENTIONS only contain entity mentions, without any sentence or relation labels. In addition, the SOFC-exp corpus focuses on the whole articles rather than the procedural text and does not contain full annotation of entity-to-entity relations. The provided relations in the original SOFC-exp dataset are constructed by only linking slot fillers to the syntactically closest EXPERIMENT mention.

Item	Train	Validation	Test
Synthesis procedures	243	31	31
Sentences	1972	275	221
Avg. sentence length	27.24	26.22	27.21
Avg. sentences/Doc	8.12	8.87	7.13
Entities	11585	1507	1516
Entity types	13	13	13
Relations	11176	1376	1435
Relation types	8	8	8
Tokens	53720	7210	6014

Table 4: Statistics of our annotated dataset.

Our new PcMSP dataset simultaneously contains the sentence, entity, and relation annotation from 305 polycrystalline synthesis-related open access publications. Among the 2468 sentences extracted from the synthesis paragraphs, 1497 sentences are identified as the synthesis description involved in an experiment. A total of 14608 entity mentions with 13 entity types and 13987 relations with 8 relation types are labeled by materials experts. We further show more corpus statistics for the training, validation, and test set in Table 4. We provide the train/validation/test split for potential use in the future.

6.2 Task definition

The PcMSP corpus labels every sentence with entity mentions and relations among entity pairs. Formally, given a sentence of n words $s = \{w_1, ..., w_n\}$ with the labeled sentence type, entity set $\mathcal{E}$ and relation set $\mathcal{R}$, four information extraction tasks are introduced:

1) SC: classification of the sentence as an experimental procedure sentence or irrelevant sentences, 2) NER: recognition of all named entities mentions in $\mathcal{E}$, 3) RE: identification of the entity pair relations in $\mathcal{R}$ and 4) Joint: joint extraction of all entities and relations.

7 Results and Analysis

We present the main experimental results in this section, and more modeling details are included in Appendix B. PURE refers to the advanced joint extraction model by (Zhong and Chen, 2021). For all the experiments, we use the *bert-base-uncased* (Devlin

	Dev	Test		
Model	F1	P	R	F1(%)
BERT-base	87.84	89.43	85.92	87.20
SciBERT	88.38	89.84	88.12	88.85
MatBERT	89.44	**91.71**	89.13	90.16
Human evaluation	-	90.74	**90.62**	**90.62**

Table 5: Experiment-describing sentence classification results in terms of F1 score on the test set. Scores are reported on macro average.

	Dev	Test		
Model	F1	P	R	F1(%)
BERT + PURE	77.06	79.23	77.24	78.23
MatBERT + PURE	76.98	**79.56**	**79.36**	**79.46**
SciBERT + PcMSP	79.46	77.32	78.91	78.84
+ MS-Mentions	91.55	-	-	91.47
+ MSPT	82.8	-	-	78.15
+ SOFC-Exp	73	-	-	78.57
Human evaluation	-	**90.05**	**89.26**	**89.46**

Table 6: Named entity recognition results in terms of F1 score on the PcMSP test set.

et al., 2019), *scibert-scivocab-uncased* (Beltagy et al., 2019), and *matbert-base-uncased* (Walker et al., 2021) as encoders. Generally, BERT with domain-specific pretraining considerably improves the performance.

7.1 Sentence classification

We summarize the results for the experiment-describing sentence detection in Table 5. For this binary classification task, we fine-tune the BERT, SciBERT, and MatBERT (Walker et al., 2021) models, resulting in an F1 score of 87.20, 88.85, and 90.16%, respectively. The best result is achieved by MatBERT, demonstrating the usefulness of domain-specific pretraining. The close-human performance of sentence classification stems from the obvious difference in expression between synthesis-describing sentences and others. Generally, synthesis-describing sentences contain 1) the material's chemical formulas, 2) the operations (usually certain verbs), and 3) experimental conditions. In contrast, other sentences often describe the characterization approaches which are totally different. In conclusion, synthesis sentence detection is the foundation for other downstream tasks and the high detection accuracy guarantees the success of our workflow for other downstream tasks.

Entity Label	Number	P	R	F1
$Brand$	21	66.67	80.00	72.73
$Descriptor$	324	61.34	74.30	67.20
$Device$	79	66.67	79.37	72.46
$Material-intermedium$	96	55.68	50.52	52.97
$Material-others$	27	1.00	16.67	28.57
$Material-recipe$	150	70.66	75.16	72.84
$Material-target$	65	67.74	68.85	68.29
$Operation$	329	82.30	84.51	83.39
$Property-pressure$	41	62.22	70.00	65.88
$Property-rate$	15	92.31	92.31	92.31
$Property-temperature$	77	76.74	79.52	78.11
$Property-time$	72	83.08	85.71	84.38
$Value$	187	76.63	87.58	81.74
Overall	1483	**77.32**	**78.91**	**78.84**
Human evaluation	-	**90.05**	**89.26**	**89.46**

Table 7: NER per label performance on the PcMSP test set by SciBERT.

7.2 Named entity recognition

In Table 6, we present the NER results obtained from different models. Based on the synthesis procedure sentences detected earlier, we train the models only on the experiment-describing sentences, ignoring irrelevant sentences. The SciBERT model is trained with one CRF layer for sequence labeling and the MatBERT is stacked with one additional forward layer for span-based tagging. The MatBERT model with PURE achieves the best F1 result of 79.46%, although a large gap of 10 points still exists compared with the human agreement score. When looking at all the label performance from Table 7, recognizing the labels such as $Property-rate$, $Property-time$ and $Operation$ achieves good scores of 92.31%, 84.38%, and 83.39%, respectively. On the contrary, the recognition is still difficult for labels like $Material-others$, $Material-interdium$, etc. One possible reason might be those mentions require cross-sentence reasoning, while the current model is only trained on single sentences. We also report SciBERT results on other previously mentioned materials procedural datasets and the overall sentence-level results are very consistent. Thus, a promising direction for improving the results is to include paragraph-level context or use cross-domain transfer learning and we leave this for future work.

7.3 Relation classification

In this section, the modeling is performed on gold entities to investigate individual modeling capability. The relation classification results are provided in Table 8. For entity pairs without any relation,

Relation Label	Number	P	R	F1
$Brand-of$	25	85.19	92.00	88.46
$Condition-of$	212	90.73	87.74	89.21
$Coreference$	140	72.79	70.71	71.74
$Descriptor-of$	349	83.92	88.25	86.03
$Device-of-operation$	87	86.59	81.61	84.02
$Next-operation$	109	84.62	90.83	87.61
$Participant-material$	296	80.74	80.74	80.74
$Value-of$	217	87.67	88.48	88.07
NA	7102	97.62	97.42	97.52
Overall	8534	**85.54**	**86.42**	**85.93**
Human evaluation	-	**96.82**	**97.69**	**97.37**

Table 8: RE per label performance on the PcMSP test set.

a 'NA' label is given for modeling. Here, the human agreement score is calculated by treating one annotation as gold and another one as predictions. Among all of the relation modeling results in Table 8, we can see that the F1 score is almost always above 80%, demonstrating promising prediction results on all label levels. In particular, the $Condition-of$ and $Brand-of$ relation predictions achieve a high F1 score of 89.21% and 88.46%, respectively. But $Coreference$ prediction is more difficult, achieving only 71.74 points. Overall, the RE modeling achieves comparable results to those of human annotators, although leaving more than 10% points for improvement. Similarly, we believe cross-sentence information can further improve the results and leave it for further investigation.

7.4 Joint entity and relation extraction

Previous sections consider entity and relation extraction separately, but the practical scenario involves joint extraction of entities and relations. Here we use the super performing joint extraction PURE (Zhong and Chen, 2021) model to evaluate the joint extraction performance. The PURE model first produces all the possible entities and then uses these predicted entities for relation extraction. Following their work, the evaluation is conducted on three metrics: (1) **Ent**: a predicted entity is correct only if the predicted span boundaries and entity type are both correct. (2) **Rel**: a predicted relation type is correct given the correct boundaries of two spans. (3) **Rel+**: in addition to the boundaries requirements, the predicted entity must conserve the correct type.

As can be seen from Table 9, the joint model demonstrates a 79.46% F1 score in terms of the entity prediction. As for the relation prediction, a much lower F1 score is observed for both Rel and Rel+, with 66.69% and 62.53% respectively. This is not unexpected since the RE relies on the previous entity prediction result and the error inevitably propagates. Compared with previous individual extraction, the joint extraction achieves lower results and leaves a large margin for improvement. Considering the goal of action graphs extraction from procedures is the joint extraction of all entities and relations, we encourage more research towards better modeling. Also of notice, the current joint evaluation is on a single sentence, while more realistic end-to-end extraction is conducted on the whole paragraph. And cross-sentence relations will also preserve in such a scenario, but this is out of the scope of this work.

Joint	P	R	F1 (%)
Ent	79.56	79.35	79.46
Rel	67.55	65.85	66.69
Rel+	63.33	61.74	62.53

Table 9: Joint entity and relation extraction results on test set.

8 Conclusion

In summary, we contribute a new dataset PcMSP collected from 305 open access scholarly publications for action graphs construction from material synthesis procedures. The two-round human expert's annotations guarantee the high quality of the dataset, which is evident by the agreement study. Based on this new dataset, we perform sentence classification, named entity recognition, and relation extraction tasks. We also experiment with the joint extraction of entities and relations. Several good-performing neural models are utilized to provide competitive baselines, although leaving a big gap compared with the human upper bound. To alleviate the data scarcity of this domain, we will make our dataset publicly available.

Some future directions would be to investigate incorporating cross-sentence context, improving the joint extraction results, performing paragraph-level end-to-end extraction, as well as using our PcMSP to investigate domain adaptation. For example, pre-training with distant supervision in the materials domain might also help improve the results. Considering the high labeling cost, how to efficiently transfer knowledge into other domains to reduce human annotations is also of great importance.

Limitations

Even though we try our best to guarantee high annotation quality, inaccurate labels may still exist. We are not responsible for any products derived from our dataset. Also, the real-world end-to-end actions graphs construction involves the whole pipeline and will inevitably face the error propagation problem.

Ethics Statement

We notice that our data source comes from open access publications and we make our dataset publicly available, but further use might also fall into potential limitations required by certain journals. Besides, in our annotation process, all the annotators are paid as research assistants following the campus policy.

Acknowledgements

We thank the anonymous ARR and EMNLP reviewers for their insightful comments related to this paper. We thank the insightful suggestions from Lei Li for an early version of the manuscript. We gratefully acknowledge support from the UC Santa Barbara NSF Quantum Foundry funded via the Q-AMASEi program under NSF award DMR-1906325. Any opinion or conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Isabelle Augenstein, Mrinal Das, Sebastian Riedel, Lakshmi Vikraman, and Andrew McCallum. 2017. Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 546–555.

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. Scibert: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620.

Gorachand Biswal and KL Mohanta. 2021. A recent review on iron-based superconductor. *Materials Today: Proceedings*, 35:207–215.

Davide Buscaldi, Danilo Dessì, Enrico Motta, Francesco Osborne, and Diego Reforgiato Recupero. 2019. Mining scholarly data for fine-grained knowledge graph construction. In *CEUR Workshop Proceedings*, volume 2377, pages 21–30.

Connor W Coley, Dale A Thomas, Justin AM Lummiss, Jonathan N Jaworski, Christopher P Breen, Victor Schultz, Travis Hart, Joshua S Fishman, Luke Rogers, Hanyu Gao, et al. 2019. A robotic platform for flow synthesis of organic compounds informed by ai planning. *Science*, 365(6453).

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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)*, pages 4171–4186.

Annemarie Friedrich, Heike Adel, Federico Tomazic, Johannes Hingerl, Renou Benteau, Anika Marusczyk, and Lukas Lange. 2020. The sofc-exp corpus and neural approaches to information extraction in the materials science domain. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1255–1268.

Ralph Grishman. 2001. Adaptive information extraction and sublanguage analysis. In *Proc. of IJCAI 2001*, pages 1–4.

Ralph Grishman and Richard Kittredge. 2014. *Analyzing language in restricted domains: sublanguage description and processing*. Psychology Press.

Jiayuan He, Dat Quoc Nguyen, Saber A Akhondi, Christian Druckenbrodt, Camilo Thorne, Ralph Hoessel, Zubair Afzal, Zenan Zhai, Biaoyan Fang, Hiyori Yoshikawa, et al. 2020. Overview of chemu 2020: named entity recognition and event extraction of chemical reactions from patents. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 237–254. Springer.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Zhi Hong, Logan Ward, Kyle Chard, Ben Blaiszik, and Ian Foster. 2021. Challenges and advances in information extraction from scientific literature: a review. *JOM*, pages 1–18.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Alaa AF Husain, Wan Zuha W Hasan, Suhaidi Shafie, Mohd N Hamidon, and Shyam Sudhir Pandey. 2018. A review of transparent solar photovoltaic technologies. *Renewable and sustainable energy reviews*, 94:779–791.

Olexandr Isayev. 2019. Text mining facilitates materials discovery.

Tianwen Jiang, Tong Zhao, Bing Qin, Ting Liu, Nitesh V Chawla, and Meng Jiang. 2019. The role of" condition" a novel scientific knowledge graph

representation and construction model. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1634–1642.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics.

Chaitanya Kulkarni, Wei Xu, Alan Ritter, and Raghu Machiraju. 2018. An annotated corpus for machine reading of instructions in wet lab protocols. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 97–106.

Fusataka Kuniyoshi, Kohei Makino, Jun Ozawa, and Makoto Miwa. 2020. Annotating and extracting synthesis process of all-solid-state batteries from scientific literature. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1941–1950.

John D. Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*.

Tuan Lai, Heng Ji, ChengXiang Zhai, and Quan Hung Tran. 2021. Joint biomedical entity and relation extraction with knowledge-enhanced collective inference. *arXiv preprint arXiv:2105.13456*.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270.

Patrick Lewis, Myle Ott, Jingfei Du, and Veselin Stoyanov. 2020. Pretrained language models for biomedical and clinical tasks: Understanding and extending the state-of-the-art. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*, pages 146–157.

Qiuhang Li, Xueping Yuan, Lei Xing, and Mingxiang Xu. 2016. Magnetization and magneto-transport staircaselike behavior in layered perovskite sr2coo4 at low temperature. *Scientific Reports*, 6(1):1–6.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3219–3232.

Huiyuan Man, Shengli Guo, Yu Sui, Yang Guo, Bin Chen, Hangdong Wang, Cui Ding, and FL Ning. 2015. Ba (zn1- 2xmnxcux) 2as2: A bulk form diluted ferromagnetic semiconductor with mn and cu codoping at zn sites. *Scientific reports*, 5(1):1–9.

David Mrdjenovich, Matthew K Horton, Joseph H Montoya, Christian M Legaspi, Shyam Dwaraknath, Vahe Tshitoyan, Anubhav Jain, and Kristin A Persson. 2020. propnet: A knowledge graph for materials science. *Matter*, 2(2):464–480.

Sheshera Mysore, Zachary Jensen, Edward Kim, Kevin Huang, Haw-Shiuan Chang, Emma Strubell, Jeffrey Flanigan, Andrew McCallum, and Elsa Olivetti. 2019. The materials science procedural text corpus: Annotating materials synthesis procedures with shallow semantic structures. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 56–64.

Sheshera Mysore, Edward Kim, Emma Strubell, Ao Liu, Haw-Shiuan Chang, Srikrishna Kompella, Kevin Huang, Andrew McCallum, and Elsa Olivetti. 2017. Automatically extracting action graphs from materials science synthesis procedures. *arXiv preprint arXiv:1711.06872*.

Tim O'Gorman, Zach Jensen, Sheshera Mysore, Kevin Huang, Rubayyat Mahbub, Elsa Olivetti, and Andrew McCallum. 2021. Ms-mentions: Consistently annotating entity mentions in materials science procedural text. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1337–1352.

Rrubaa Panchendrarajan and Aravindh Amaresan. 2018. Bidirectional lstm-crf for named entity recognition. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*.

Hong-Jie Peng, Jia-Qi Huang, Xin-Bing Cheng, and Qiang Zhang. 2017. Review on high-loading and high-energy lithium–sulfur batteries. *Advanced Energy Materials*, 7(24):1700260.

Huabei Peng, Jie Chen, Yongning Wang, and Yuhua Wen. 2018. Key factors achieving large recovery strains in polycrystalline fe–mn–si-based shape memory alloys: A review. *Advanced Engineering Materials*, 20(3):1700741.

Paul Raccuglia, Katherine C Elbert, Philip DF Adler, Casey Falk, Malia B Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A Friedler, Joshua Schrier, and Alexander J Norquist. 2016. Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601):73–76.

Tim Rocktäschel, Michael Weidlich, and Ulf Leser. 2012. Chemspot: a hybrid system for chemical named entity recognition. *Bioinformatics*, 28(12):1633–1640.

Parantu K Shah, Carolina Perez-Iratxeta, Peer Bork, and Miguel A Andrade. 2003. Information extraction from full text scientific articles: where are the keywords? *BMC bioinformatics*, 4(1):1–9.

Sebastian Steiner, Jakob Wolf, Stefan Glatzel, Anna Andreou, Jarosław M Granda, Graham Keenan, Trevor Hinkley, Gerardo Aragon-Camarasa, Philip J Kitson, Davide Angelone, et al. 2019. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science*, 363(6423).

Ronen Tamari, Fan Bai, Alan Ritter, and Gabriel Stanovsky. 2021. Process-level representation of scientific protocols with interactive annotation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2190–2202.

Huihsin Tseng, Pi-Chuan Chang, Galen Andrew, Dan Jurafsky, and Christopher D Manning. 2005. A conditional random field word segmenter for sighan bakeoff 2005. In *Proceedings of the fourth SIGHAN workshop on Chinese language Processing*.

Alain C Vaucher, Federico Zipoli, Joppe Geluykens, Vishnu H Nair, Philippe Schwaller, and Teodoro Laino. 2020. Automated extraction of chemical synthesis actions from experimental procedures. *Nature communications*, 11(1):1–11.

Nicholas Walker, Amalie Trewartha, Haoyan Huo, Sanghoon Lee, Kevin Cruse, John Dagdelen, Alexander Dunn, Kristin Persson, Gerbrand Ceder, and Anubhav Jain. 2021. The impact of domain-specific pre-training on named entity recognition tasks in materials science. *Available at SSRN 3950755*.

Kyosuke Yamaguchi, Ryoji Asahi, and Yutaka Sasaki. 2020. Sc-comics: A superconductivity corpus for materials informatics. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6753–6760.

Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 2019. Docred: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 764–777.

Zixuan Zhang, Nikolaus Parulian, Heng Ji, Ahmed Elsayed, Skatje Myers, and Martha Palmer. 2021. Fine-grained information extraction from biomedical literature based on knowledge-enriched abstract meaning representation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6261–6270.

Zexuan Zhong and Danqi Chen. 2021. 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*, pages 50–61.

A Background on Polycrystalline Materials

Polycrystalline materials are solids composed of small randomly oriented crystallites, also called grains, with the size varying from a few nanometers to several millimeters. Most of the inorganic solid materials available in macroscopic quantities are in fact polycrystals, including common metals, ceramics, and rocks. They provide versatility in numerous applications such as superconductors, batteries, photovoltaic cells, and shape memory alloys (Husain et al., 2018; Peng et al., 2018, 2017; Biswal and Mohanta, 2021).

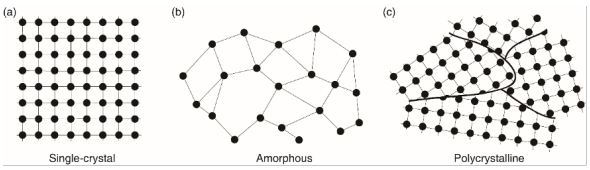

Figure 3: Material classification based on the degree of atomic order: (a) single-crystal, (b) amorphous, (c) polycrystalline.

The structure of a single crystal or monocrystal (Figure 3a) is continuous and highly ordered, while an amorphous phase (non-crystal) (Figure 3b) such as glass does not display any structures, as the constituent atoms are not arranged in an ordered manner. In-between these two extremes, a polycrystal (Figure 3c) exists, which is made up of many crystallites, also referred to as grains. During the solidification of polycrystalline materials, small nuclei first form at different spots of the liquid sample and subsequently absorb atoms from the surrounding liquid to grow into larger grains. These grains vary in size from nanometers to millimeters and are randomly oriented with no preferred direction in the structure. Therefore, a large enough volume of polycrystalline material can be approximately considered isotropic. Compared to single crystals, polycrystalline materials also require less sophisticated techniques to make, significantly lowering the cost of production. As most real-world solids are polycrystalline materials, it is critical to synthesize and understand polycrystalline materials. A substantial number of studies have been done by researchers across the world to discover new materials. This work exacts knowledge from those synthesis processes and aims to guide the synthesis efforts toward the unexplored space.

B Modeling

We mainly use PURE (Zhong and Chen, 2021) as backbones for our tasks.

B.1 Sentence classification

Sentence classification is a binary text classification problem. We build one additional layer on top of the BERT and fine-tune it for another 10 epochs.

B.2 Named entity recognition

For the SciBERT model, we stack another conditional random field (CRF) (Tseng et al., 2005) layer on top of SciBERT for sequence labeling following the traditional BIO notation. For the MatBERT result, we follow the span-based approach in (Zhong and Chen, 2021) to obtain the contextualized representation for any span and feed it into another forward layer to predict the entity type.

B.3 Relation classification

We utilize the span representations of entity mentions for relation prediction with typed entity markers as proposed by the relation model in (Zhong and Chen, 2021).

B.4 Joint extraction

Following (Zhong and Chen, 2021), the predicted entities are fed into another encoder for relation prediction. And we adopt two different encoders for the joint extraction of entities and relations.

C Experimental settings

We select the best combination of hyperparameters from the development set by random search. Three random seeds are chosen for all models, and we report the results based on the median performance. The standard macro-average precision(**P**), recall(**R**), and **F1** scores are calculated.

The Adam optimizer (Kingma and Ba, 2015) is used for all models. Other parameters are selected within a range of values, for example learning rate ranges from [1e-4, 5e-5, 1e-6] and batch size of 8 or 16. The models are implemented in PyTorch[12], and a Tesla P40 with 24GB RAM is used for all experiments. The model takes around half-hour, one hour, and three hours for the training of sentence, entity, and relation tasks for 10 to 50 epochs.

[12]https://pytorch.org

Sentence Label	Number	P	R	F1
Synthesis	153	89.57	95.42	92.41
Non-synthesis	103	92.47	83.50	87.76
Overall	256	90.74	90.62	90.62

Table 10: Human agreement score on experiment-describing sentences.

C.1 Data preprocessing

Each plain text document containing the synthesis paragraphs is imported into the INCEpTION platform, which also performs the sentence segmentation and word tokenization by its built-in algorithm. After tokenization, each sentence is mapped with the corresponding entity mentions and relations, which includes the named entity type, position, token information, and the relations type, as well as left and right position information.

D Inter-annotator Agreement Study

Despite from Fleiss' kappa for measuring agreements in Table 3, we describe more details in this section.

D.1 Sentences annotation

Given a paragraph selected from a scientific publication, we first examine the synthesis-related sentences. In practice, the annotators only label synthesis-related sentences for the entity and relation information. All other sentences without labeling are considered non-synthesis sentences. To compare the model's performance with human annotation, 32 documents are labeled by two main annotators in the second round individually. Then one annotation is regarded as the ground truth and the other is treated as a prediction. A micro-average F1 score of 90.62% is calculated between the two annotators. Additional details about the precision, recall, and F1 score is shown in Table 10. In general, the main annotator selects 153 of the 256 sentences to label as synthesis-related sentences, while the second annotator chose 163 to be labeled as target sentences. The overall result demonstrates high-quality annotations and can serve as a human agreement score for further baseline.

D.2 Named entity annotation

Following the previous step, all of the entity mention boundaries are first recognized by the annotators and then one entity label is chosen from the predefined entity labels to represent the entity type.

Entity_Label	Number	P	R	F1
Brand	21	94.74	85.71	90.00
Descriptor	324	83.49	82.72	93.10
Device	79	93.67	93.67	93.67
Material − intermedium	96	87.37	86.46	86.91
Material − others	27	74.19	85.19	79.31
Material − recipe	150	86.84	88.00	87.42
Material − target	65	96.83	93.85	95.31
Operation	329	94.08	91.79	92.92
Property − pressure	41	90.00	87.80	88.89
Property − rate	15	92.86	86.67	89.66
Property − temperature	77	86.59	92.21	89.31
Property − time	72	95.71	93.06	94.37
Value	187	91.57	87.17	89.32
Overall	1483	**90.05**	**89.26**	**89.46**

Table 11: Human agreement score on NER.

Entity_Label	Number	P	R	F1
Brand − of	18	100.0	100.0	100.0
Condition − of	174	100.0	97.13	98.54
Coreference	69	81.43	82.61	82.01
Descriptor − of	256	93.94	96.88	95.38
Device − of − operation	69	98.48	94.20	96.30
Next − operation	99	98.97	96.97	97.96
Participant − material	229	94.35	94.76	94.55
Value − of	162	97.53	97.53	97.53
Overall	1076	**96.82**	**97.69**	**97.37**

Table 12: Human agreement score on RC.

Among the recognized overlap of 143 experiment-describing sentences from the previous step by both annotators, one annotator recognizes 1483 named entities while the second annotator considers 1345 entity mentions as necessary to be labeled. The agreement metric is calculated by treating one result as the true value, while the second result is used as a predictive value. The overall P, R, and F1 scores are given in Table 11 in terms of per label performance. As can be seen from the results, two of the annotators agreed on the majority of the labels, while in some circumstances (like $Material - others$), the score is relatively lower, due possibly to a different understanding of those entity mentions.

D.3 Relation annotation

Here we focus on relation annotation based on a given entity pair. When both annotators first agree on the same entity pair, the agreement F1 score is 97.37%, demonstrating the high quality of the annotation.

Figure 4 shows the confusion matrix of relations between the two lead annotators.

Journal	Train	Validation	Test
Elsevier	46	6	4
ArXiv	81	5	8
Nature family	71	13	13
ACS family	13	4	2
APS family	28	3	4
Others	4	0	0

Table 13: Document distribution among main journals: ACS: American Chemistry Society, APS: American Physical Society, and others refers to other journals not included here.

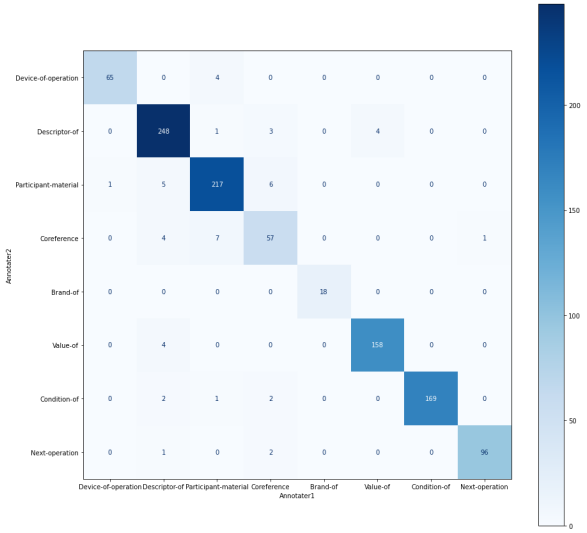

Figure 4: Confusion matrix over relations between the two lead annotators.

E Document Distribution Among Journals

Table 13 demonstrates that the source of our collected documents is distributed among different journals. Considering that the writing style and publication requirements of different journals vary a lot, we aim to include documents from a range of sources to make the dataset more diverse.

F Annotation Examples and Statistics

Common examples of entity mentions and relation triples are shown in Table 14 and Table 15, respectively. The relation triple has the form of r_i: (e_i, e_j), where r_i is one relation label, while e_i and e_j denote the entity mention within one sentence.

Entity Label	Count	Frequent mentions	Percentage
$Descriptor$	2450	Polycrystalline, quartz, polycrystalline	21.15
$Material-target$	442	Ca2CeCr2TiO9, powder, sample	3.82
$Brand$	317	Alfa Aesar, Aldrich, Sigma-Aldrich	2.74
$Device$	662	tube, crucible, glove box	5.71
$Material-intermedium$	772	pellets, mixture, samples	6.66
$Material-others$	158	water, distilled water, oxygen	1.36
$Material-recipe$	1270	Fe, As, materials, Fe2O3	10.96
$Operation$	2439	heated, sealed, mixed	21.05
$Property-pressure$	401	air, argon atmosphere, vacuum	3.46
$Property-rate$	126	heating rate, cooling rate, 1 K/min	1.09
$Property-temperature$	664	room temperature, temperature, 900 °C	5.73
$Property-time$	506	24 h, 30 min, 2 days	4.37
$Value$	1378	>99.9%, stoichiometric amounts, 10 mg	11.89
Overall	11585		100.0

Table 14: Annotated entity mention statistics in the training set.

Relation Label	Count	Frequent mentions	Percentage
$Descriptor-of$	2796	(purity, 99.6%), (Polycrystalline, materials)	25.02
$Participant-material$	2147	(Pb, melting), (SrCO3, sealed)	19.21
$Coreference$	1171	(OsO2, powder), (CuO, mixture)	10.48
$Value-of$	1737	(99.99%, Bi2O3), (50 mg, I2)	15.54
$Condition-of$	1547	(800 °C, heated), (10 hours, held)	13.84
$Next-operation$	805	(kept, heated), (sealed, evacuated)	7.20
$Device-of-operation$	637	(glove box, grinding), (calcined, ground)	5.70
$Brand-of$	336	(Aldrich, (TPrA)Br), (Alfa Aesar, ZrO2,)	3.01
Overall	11176		100.0

Table 15: Annotated relation pair statistics in the training set.