

# SEBI Regulation Biography

Sathvik Sanjeev Buggana  
IIIT Hyderabad, India  
Hyderabad, Telangana, India  
sathviksanjeev.b@research.iiit.ac.in

Deepti Saravanan  
IIIT Hyderabad, India  
Hyderabad, Telangana, India  
deepti.saravanan@research.iiit.ac.in

Shravya Kanchi  
IIIT Hyderabad, India  
Hyderabad, Telangana, India  
shravya.k@research.iiit.ac.in

Ujwal Narayan  
IIIT Hyderabad, India  
Hyderabad, Telangana, India  
ujwal.narayan@research.iiit.ac.in

Shivam Mangale  
IIIT Hyderabad, India  
Hyderabad, Telangana, India  
shivam.mangale@students.iiit.ac.in

Lini T. Thomas  
IIIT Hyderabad, India  
Hyderabad, Telangana, India  
lini.thomas@iiit.ac.in

Kamalakar Karlapalem  
IIIT Hyderabad, India  
Hyderabad, Telangana, India  
kamal@iiit.ac.in

Natraj Raman  
J.P. Morgan AI Research  
London, London, United Kingdom  
natraj.raman@jpmorgan.com

## ABSTRACT

The Securities and Exchange Board of India is the regulatory body for securities and commodity market in India. A growing number of SEBI documents ranging from government regulations to legal case files are now available in the digital form. Advances in natural language processing and machine learning provide opportunities for extracting semantic insights from these documents. We present here a system that performs semantic processing of SEBI documents using state-of-the-art language models to produce enriched regulations containing timelines of amendments and cross references to legal case files.

## CCS CONCEPTS

• **Information systems** → **Document structure.**

## KEYWORDS

regulation, SEBI, semantic tagging, knowledge graph, template creation, data mining

### ACM Reference Format:

Sathvik Sanjeev Buggana, Deepti Saravanan, Shravya Kanchi, Ujwal Narayan, Shivam Mangale, Lini T. Thomas, Kamalakar Karlapalem, and Natraj Raman. 2022. SEBI Regulation Biography. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion)*, April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3487553.3524638>

## 1 INTRODUCTION

Securities and Exchange Board of India, commonly referred to as SEBI, is the regulatory body for securities and commodity market

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*WWW '22 Companion*, April 25–29, 2022, Virtual Event, Lyon, France

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-9130-6/22/04...\$15.00  
<https://doi.org/10.1145/3487553.3524638>

in India under the ownership of Ministry of Finance, Government of India. The Indian companies are obligated to adhere to the regulations drafted by SEBI. The regulations need to be interpreted by the information technology departments and/or finance departments to diligently follow the Regulations. Moreover, the companies' lawyers interact with their SEBI counterparts about case arguments and outcomes. Manual analysis of these documents is tedious owing to their content size and suffer from the inability to discover inter-linked information. An AI system can aid and improve these interpretations and interactions by automatically processing and deriving insights from SEBI regulations, associated case files and other pertinent documents. Towards this goal, we present our work on Regulation Biography which presents a temporal understanding of a Regulation Document. We perform document pre-processing for entity extraction and entity linking. We also have the semantics extraction layer that generates the syntactic structure and performs semantic analysis of the documents. Some of the core NLP techniques like word embedding, biLSTM[4], and BERT[1] modules are used for semantics extraction. The system developed has various components

- **Semantic Extraction for Regulations:** Forms the building block of various applications
- **Regulation Timeline:** displays the amendment locations in the text, amendment type, rationale for an amendment and comments and discussion on the amendment.
- **Template Creation for Regulatory Documents:** for readability.
- **References to Regulations:** in legal case files and news articles.

The above tasks together offer a better understanding of a Regulatory document, its regulations and its metamorphosis over time as amendments. It also helps us understand regulations from the perspective of news article references and case files. A tool developed, can be viewed at our Project page<sup>1</sup> while data, code and trained models are available at Github<sup>2</sup> for public use.

<sup>1</sup>[https://jpmc.iiit.ac.in/regn\\_bio/](https://jpmc.iiit.ac.in/regn_bio/)

<sup>2</sup>[https://github.com/JPMS-DSAC/sebi\\_regulation\\_biography](https://github.com/JPMS-DSAC/sebi_regulation_biography)

## 2 REGULATION BIOGRAPHY MODULES

In this section, we shall present the modules of our system.

### 2.1 Semantic Extraction for Regulations

Unlike traditional NER tasks, we detect *overlapping named entities* from regulation texts due to the presence of overlapping tags (For example, "listed in recognised stock exchanges" is a *Transaction* while the span "recognised stock exchanges" can be a *Subject-Organisation* based on the context.). We deployed multiple named entity models to identify overlapping semantic tag occurrences in the sentence. We experimented training one NER model for each entity using customized spacy[3] NER v3.0. We used a train-dev-test split of 75-15-10 of the annotated data. Models, code and data are available at Github<sup>2</sup>. We detect semantic tags of the form *Transactions*, *Legal Term*, *Date/Time*, *Legal Document*, *Object* (An entity which is acted upon by authority, subject-Individual and subject-organization), *Subject-Individual* (specific individuals or a class of individuals who engage in the securities market in any manner and upon whom, the acts, rules and regulations apply), *Subject-Organization* (Companies or organizations, who are bound to act as per the rules, regulations, acts as well are directions of the authorities), *Authority*, etc. These tags were decided by domain legal experts.

We evaluated our work using the NER evaluation method proposed in SemEval' 13 NER task [2]. A perfect NER system should be able to detect the right word boundary of the named entity as well as classify the detected word into the right class. We define two categories of correctness namely *Strict* and *Exact*. *Strict* implies an exact boundary match with the actual tagged entities and the right entity type is assigned to the span of characters. *Exact* implies an exact boundary match regardless of the right entity type being assigned to the span of characters.

We pooled the named entities labelled by the ten NER models. Table 1 shows the precision, recall and F1 scores for the 2 categories of evaluation used for NER systems.

**Table 1: Quantitative Results for pooled ten NER model**

	Precision	Recall	F1
<b>Strict</b>	0.8715	0.6521	0.7460
<b>Exact</b>	0.8747	0.6545	0.7488

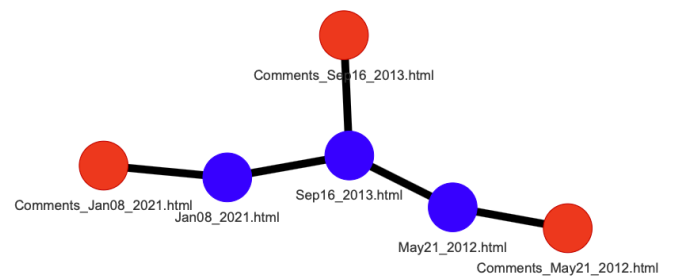
### 2.2 Regulation Over time

This section offers a view of a regulation over time and includes

- (1) visualisation that helps understand the amended versions of a regulatory document
- (2) identifies what changes have been made between two successive versions of a regulation document
- (3) provides additional information from other SEBI documents that help understand the rationale behind the amendment and other associated comments
- (4) provides tags in order to categorise the kind of amendments, etc.

Provided below are further details of the tasks.

**2.2.1 Regulation Timeline.** Our tool represents the amended versions of a document as a linear chain with nodes as regulation documents and clickable edges that provide comparison of two consecutive regulatory documents as shown in Figure 1. The blue circles represent amendments ordered over time. They can be clicked to open up the corresponding document. The corresponding red circles contain comments extracted from various supporting documents that pertain to regulations of the respective regulatory document. This match is done based on the time when the comment appeared in a document and the corresponding amended version of the regulatory document at that point in time. The edges when clicked provide a comparison between two consecutive amended versions of a Regulatory document as shown in Figure ??.



**Figure 1: Regulation timeline of Alternative Investment Funds**

The documents are compared to understand the syntactic differences between them. Note that an amended version does not maintain the regulation numbering order of previous versions due to the introduction of new regulations in between. Hence, a similarity score function is used wherein every pair of regulations (one from each document) are considered. The similarity between two regulations belonging to any regulation document in SEBI is computed by measuring the Longest Common Sub-sequence based on the Gestalt pattern matching algorithm. The pair with the highest matching score is assumed to be the same regulation version in both documents. Once the pair (one from each document) of corresponding regulations are identified; further comparison is done to identify the amendment made to it. The comparison of regulations are presented in a tabular format. Olive color implies exact matches. Blue color implies edits in a regulation. Black color text implies insertion (newer document) or deletion (older document) of a regulation. Red color text implies insertion (newer document) or deletion (older document) of a sub-regulation.

Further, the visualisation comparing two consecutive amended Regulatory documents also gives the rationale for the amendment. Rationale explaining the reason an amendment was introduced; is often made available in related SEBI documents like Annual reports and Concept papers. Additional documents are searched using multiple factors like regulatory document name, regulation name, keywords in the regulation, year, etc to identify such a piece of text. These are simple rule based searches put together to identify the right piece of texts. The rationale provided in a document about an amended Regulation is associated with the latest version of the

35 The Collective Investment Management Company shall:(a) not invest the funds of the 108[collective investment scheme] for purposes other than the objective of the 109[collective investment scheme] as disclosed in the offer document.(b) segregate the 110[collective investment scheme] assets of different s111[collective investmentscheme] s.(c) not invest corpus of a 112[collective investment scheme] in other 113[collective investmentscheme]s.(d) not transfer funds from one 114[collective investment scheme] to another 115[collective investment scheme]:Provided that inter-scheme transfer of 116[collective investment scheme] property may be permitted at the time of termination of the 117[collective investment scheme] with prior approval of the trustee and the Board.

Rationale : o. sebi (collective investment schemes) (amendment) regulations, 2014 w.e.f January 9, 2014 the securities laws (amendment) ordinance, 2013 provides for regulation of pooling of funds under any scheme or arrangement, involving a corpus amount of ` 100 crore or more, to be deemed to be a collective investment scheme, subject to sub- section (3) of section 11aa of the sebi act. accordingly, amendments were made to sebi (collective investment schemes) regulations, 1999, providing a framework for regulation of such deemed ciss. it was also provided that all collective investment management companies shall mobilize money only through cheques.

Figure 2: Sample Rationale - Collective Investment Scheme Regulations

<p>19 ( 1 ) Where any information is called for by the Board from a credit rating agency for the purposes of these regulations , including any report relating to its — ( g ) the client shall agree to obtain a rating from at least two different rating agencies for any issue of debt activities , the credit rating agency shall furnish such information to the Board – ( a ) within a period specified by the Board or ( b ) if no such period is specified , then within a reasonable time .</p> <p>(2) Every credit rating agency shall, at the close of each accounting period,</p>	<p>19 ( 1 ) Where any information is called for by the Board from a credit rating agency for the purposes of these regulations , including any report relating to its activities , the credit rating agency shall furnish such information to the Board – ( a ) within a period specified by the Board or ( b ) if no such period is specified , then within a reasonable time . , agency shall furnish such information to the Board – ( a ) within a period specified by the Board or ( b ) if no such period is specified , then within a reasonable time .</p> <p>(2) Every credit rating agency shall, at the close of each accounting period,</p> <p>Substitution Update Declaration</p>
---	--

Figure 3: Amendment tagged - Credit rating regulations

amendment published before the additional document itself. A sample rationale (in blue text) is shown in Figure ???. This was extracted from SEBI Annual Reports 2013-14<sup>3</sup>. It mentions that amendments are made to the SEBI Act, granting SEBI powers in the regulation of Collective Investment Schemes (CIS) in 2014. This was detected by our tool and matched Regulation 35 of CIS regulatory document (which stipulates the duties of a Collective Investment Management Company in regards to Collective Investment Schemes) to the aforementioned rationale information.

In order to characterise the type of amendments we built a rule based tagger for tagging an amendment with tags pertaining to the amendment’s purpose and effect. These rules have been built after analysing the types of tags provided manually to amendments by the domain legal experts. A sample tagging can be seen in blue in the Figure ???. Initial annotation of amended regulations was conducted for a sample of 50 amended regulations. Then rule based method was developed to tag the remaining amended regulations.

Finally, the system detects comments related to a Regulation which are mentions of a regulation document in other documents available at SEBI. However, additional work is required to identify which version of a regulation document is being referred to in the

comment. For this part of biography analysis, the following steps were involved.

- First, using grep and certain pattern matches, we found potential mentions of regulations.
- Second, for each mention, the name of the exact regulation document mentioned and the time of release of document, were extracted. This helps in matching each mention with the correct regulation document type and also the exact "version" of the document. For example, if a document mentions SAST regulation and is released in 2015, then, it is matched to SAST regulations released on 23 April 2014, which is the closest version of the regulation available.
- A table is generated with the document name, SEBI link to the document, and the mention of regulation, for each type of regulation and each version of regulation. The exact mentions of the regulation were highlighted and displayed using HTML.

A tool developed for the components described above can be viewed at our demopage<sup>4</sup>.

<sup>3</sup>[https://www.sebi.gov.in/reports/annual-reports/aug-2014/annual-report-2013-14\\_27807.html](https://www.sebi.gov.in/reports/annual-reports/aug-2014/annual-report-2013-14_27807.html)

<sup>4</sup>[https://jpmc.iit.ac.in/regn\\_bio/](https://jpmc.iit.ac.in/regn_bio/)

**Table 2: Structure of Regulations**

Entity	Condition	Action
Compliance officer	Upon approval of the trading plan	notify the plan to the stock exchanges on which the securities are listed
The Board	nil	designate a division to function as the independent Office of Informant Protection

**Table 3: Phrases defining condition part of the regulations**

Location	Phrases/Tokens
Before entity is mentioned	If, where, while, After, On, Upon, In (the) case
Between entity and shall/may	whose,who, while, if, having, unless, that, required
Post entity and shall/may	phrase between comma detected after shall/may and the next detected comma

### 2.3 Regulation Template Creation

Around 2500 regulations were extracted from the latest SEBI regulatory documents. These regulations follow an intrinsic pattern of the form Entity-Condition-Action, as seen in table 2. The entity is the subject of the regulation. Condition includes the phrase that the entity must satisfy in order to perform the action. The remaining part of the regulation is categorised as the Action part, that usually occur after the rule-relevant words 'shall' or 'may'. The presence of these rule-relevant words differentiates regulatory sentences from other miscellaneous sentences in the regulatory documents.

Observing the consistent syntactic pattern followed across all the regulatory documents, the regulatory sentences are transformed to simpler switch-case statements that are more comprehensible and easier to extract relevant regulations from, according to the entity of interest.

#### 2.3.1 Approach and Observations

**Entities Extraction:** Entity detection in the context of template extraction is done by leveraging the basic entity recognition model. The output of the model provides the entities to which a specific regulation pertains to.

**Conditions Extraction** Regulations may or may not have a conditional phrase. In the former case, phrases that define the conditions on the entity, that are to be satisfied before the corresponding actions are performed, were extracted. This was possible since SEBI regulatory documents follow a fixed pattern that could be used as the base rule to extract the corresponding phrases of interest. Table 3 shows the phrases defining conditions and their corresponding locations in the sentence.

**Actions Retrieval** There are two kinds of action phrases - positive and negative. Negative actions may occur in two forms:

- Negations with rule-relevant words, 'shall not' or 'may not'.
- Negations with entities, for example, 'No insider'.

If the above forms are not identified in a regulatory sentence, then the action is considered to be positive. Once the kind of the action is identified, the actions can be directly extracted by identifying the rule-relevant words and analysing the phrase following those words.

**Identification of inter-dependent regulations** Each regulatory document has a set of interconnected chapters. There exist direct or indirect relationship between two or more regulations, both within and across chapters. To extract the phrases that uniquely capture such relationships, each part of the sentence was analyzed in detail. The regulatory interconnections occur in two forms:

- Presence of the word 'such': The entity that immediately follows 'such' points to its immediate previous instance, which may be within the same regulation or in the previous regulation.
- Direct References using regulation numbers: A few regulations have the referred regulation or sub-regulation numbers mentioned, leading to direct mapping of regulations.

**Final Template Format** Regulatory Documents can be represented as switch-case statements for easier extraction and analysis for various downstream tasks. With entity as the case, the corresponding regulations are mentioned under it. The regulations with conditional phrases will have an additional if-else statement, in order to provide more emphasis on the satisfaction of required conditions.

Generic template creation has several advantages. It captures the relationship between regulations from the perspectives and hierarchies of the entities involved, while it also groups the subset of regulations corresponding to the entity of interest. In addition, conditional statements modeled as if-else format help not to miss out any important pre-requisites for the actors involved.

**Results:** For the Insider Trading document (2020 version), a manual accuracy check was done. A 76 % accuracy was obtained when a basic entity recognition model was used and 85 % accuracy was observed when the model detailed in Section - 2.1 was used. The increase in accuracy is due to multiple types of entities that are detected by the more comprehensive entity recognition model. For example, UPSI is recognised which helped in detecting UPSI related regulations.

### 2.4 Knowledge Graph Module

We have three primary forms of data pertaining to SEBI, namely, Regulation documents and other related documents, SEBI related legal case files and news articles that relate to SEBI regulations and cases filed. Further, we also have done entity extraction of regulation

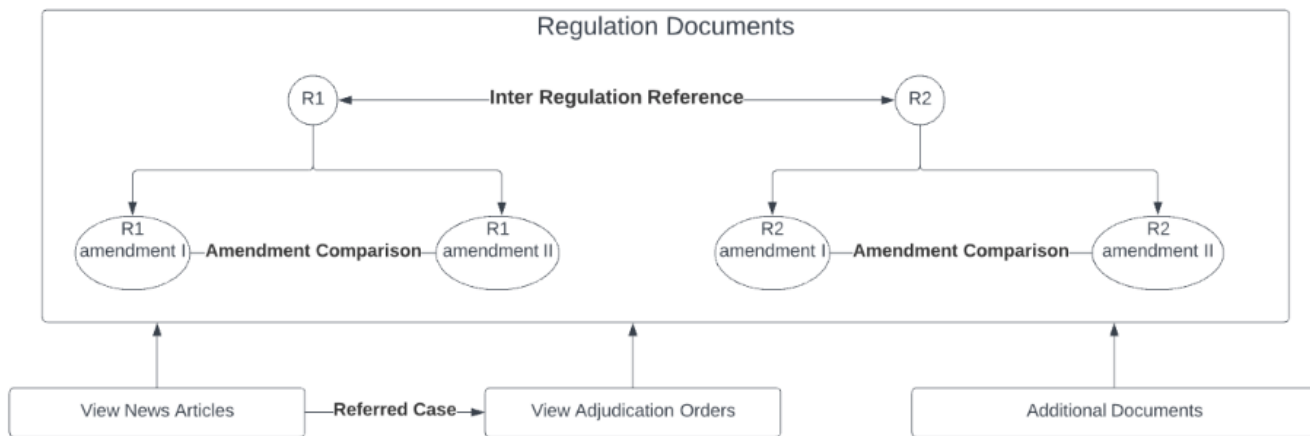


Figure 4: Knowledge Graph Constituents

39 (2) Notwithstanding such repeal:(a) Anything done or any action taken or purported to have been done or taken,including suspension or cancellation of certificate of registration, any inquiry or investigation commenced or show cause notice issued under the repealed regulations , shall be deemed to have been done or taken under the corresponding provisions of these regulations ;(b) All venture capital funds or schemes launched by such venture capital funds prior to date of notification of these regulations shall continue to be governed by provisions of Securities and Exchange Board of India (Venture Capital Funds) Regulations,1996 till the fund or Scheme is wound up:Provided that such funds shall not launch any new Scheme after notification of these regulations ;(c) Any application made to the Board under the Securities and Exchange Board of India (Venture Capital Funds) Regulations, 1996 and pending before it shall be deemed to have been made under the corresponding provisions of Securities and Exchange Board of India (Alternative Investment Funds) Regulations, 2012.

39 (3) After the repeal of Securities and Exchange Board of India (Venture Capital Funds) Regulations, 1996, any reference thereto in any other regulations made, guidelines or

Case Files that refer this Regulation Document

- [1511951662889.txt](#)
- [1540999794392.txt](#)
- [1577181091876.txt](#)

---

News Articles that refer this Regulation Document

- [89]Daily Bulletin: All four accused in Hyderabad rape, murder case killed in police encounter; Ajit Pawar gets clean chit in Vidarbha irrigation probe; day's top stories
- [414]Regulatory Updates For The Month Of June 2020 – July 2020
- [483]AIF investment committees may get some relief from Sebi's June order
- [571]Sebi writes to states to exempt staff of capital market entities from lockdown
- [3145]SEBI Decision on Recommendations of Kotak Committee
- [3702]Budget 2018: Curious Case of Amendment to Securities Laws - Finance Bill 2018

Figure 5: Inter References in Regulation document

documents which help us identify various entities, timelines, of the document. The work related to Regulation biography relates the Regulation amendment documents where multiple amendments made to a particular regulation document can be studied. Tagging to understand the amendment type was also done as part of this work. All of these can be brought under a single umbrella where the connections across documents and other mined information about these documents can be presented. The Knowledge Graph module identifies inter references to the Regulation in News Articles and SEBI related case files. It also identifies references of case files in news articles. This provides an user with the relevant News Articles when the user accesses Regulations to get information of how it was received by the media. Similarly, we process the Adjudication Orders to mine the Regulations and sub-regulations referred to in them. This provides for easier navigation and access to the Regulatory information while browsing the case file.

The data set used to build this Knowledge Graph is depicted in Table 4. The news articles are scraped and subsequently filtered from prominent News sites and Information sites such as the Hindu<sup>5</sup>, the Economic Times<sup>6</sup> and MoneyControl<sup>7</sup>. The Adjudication Orders, regulation documents and their amendments are retrieved from SEBI<sup>8</sup> and from IndianKanoon<sup>9</sup>.

The relationships between Regulation documents, additional SEBI documents, news articles and case files have been depicted in the Knowledge Graph have been presented in Figure ??.

- References across regulations: Often a regulation is found to refer to another regulation. In the Figure, the Regulation Document R1 has a reference to Regulation Document R2.

<sup>5</sup><https://www.thehindu.com>

<sup>6</sup><https://economictimes.indiatimes.com>

<sup>7</sup><https://www.moneycontrol.com>

<sup>8</sup><https://www.sebi.gov.in/>

<sup>9</sup><https://indiankanoon.org>

**Table 4: Knowledge Graph Dataset**

Document	Count
News articles relevant to SEBI	22091
News articles relevant to SEBI cases	4974
Adjudication Orders	7406

- **Comparing Amendments:** In the Regulation timeline, the amended versions of the Regulation document are shown with the ability to compare consecutive amendments. In the figure, the regulation document R1 has amendment I and II which are being compared. Additional information for the reader is provided using Tagging of Amendment types.
- **Relating comments and rationale:** The regulation document is related to certain additional documents which provide the comments and rationale pertaining to the regulation or amendment.
- **Each Adjudication Order refers to Regulatory documents that are violated or documents that are referred to for the decision making in case files.** These are captured in relating Case files with Regulation Documents.
- **Often, News Articles refer to certain Regulatory documents or regulations which are captured in the Knowledge Graph.**
- **New articles sometimes discuss particular cases, and specifically Adjudication Orders relevant to SEBI.** This is found by using a weighted overlap of entities tagged using Flair Ontonotes NER model[5].

Figure ?? provides a snapshot of the display of our tool. It shows the bottom part of the page displaying Regulations of a particular regulatory document with entities extracted in Section 2.1 highlighted in colours in the text region. At the bottom of the page are listed the links to the case files and news articles that refer to the regulations of this regulatory particular document.

### 3 CONCLUSION

In this paper we present a tool for comprehensive understanding of SEBI regulations. Overlapping named entities are extracted from regulation texts. The semantic extraction performed can find applications in various others problems such as regulation violation detection, penalty estimation, extractive question answering, regulation simplification, etc. A visualisation of Regulations amended over time along with specific changes made, rationale and comments associated with the change and tags to explain the change has been developed. Regulation Template creation has been done which makes the regulation easy to read and interpret by both human and computers. A knowledge graph module that associates various related documents has been built. Data set, code, trained models and demo have been made available for public use. A demo video can be found at the Project page<sup>1</sup>. The paper forms the foundation work to many possible extensions and applications.

### 4 ACKNOWLEDGMENTS

This work has been supported by J.P. Morgan AI Faculty Research Award. Any opinions, findings, and conclusions in this paper are

those of the authors only and do not necessarily reflect the views of the sponsors.

### REFERENCES

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL]
- [2] Paloma Martinez Isabel S. Bedmar and Maria H. Zazo. 2013. Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). Association for Computational Linguistics.
- [3] Sofie Van Landeghem Matthew Honnibal, Ines Montani and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. *Zenodo* (2020). <https://doi.org/10.5281/zenodo.1212303>
- [4] Mike Schuster and Kuldip Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45, 11 (1997), 2673–2681. <https://doi.org/10.1109/78.650093>
- [5] Stefan Schweter and Alan Akbik. 2020. FLERT: Document-Level Features for Named Entity Recognition. arXiv:2011.06993 [cs.CL]