Term Expansion and FinBERT fine-tuning for Hypernym and Synonym Ranking of Financial Terms

Ankush Chopra[*][†], **Sohom Ghosh**[†]

Fidelity Investments, AI CoE, Bengaluru, India

{ankush01729, sohom1ghosh}@gmail.com

Abstract

Hypernym and synonym matching are one of the mainstream Natural Language Processing (NLP) tasks. In this paper, we present systems that attempt to solve this problem. We designed these systems to participate in the FinSim-3, a shared task of FinNLP workshop at IJCAI-2021. The shared task is focused on solving this problem for the financial domain. We experimented with various transformer based pre-trained embeddings by fine-tuning these for either classification or phrase similarity tasks. We also augmented the provided dataset with abbreviations derived from prospectus provided by the organizers and definitions of the financial terms from DBpedia [Auer *et al.*, 2007], Investopedia, and the Financial Industry Business Ontology (FIBO). Our best performing system uses both FinBERT [Araci, 2019] and data augmentation from the afore-mentioned sources. We observed that term expansion using data augmentation in conjunction with semantic similarity is beneficial for this task and could be beneficial for the other tasks that deal with short phrases. Our best performing model (Accuracy: 0.917, Rank: 1.156) was developed by fine-tuning Sentence-BERT [Reimers *et al.*, 2019] (with FinBERT at the backend) over an extended labelled set created using the hierarchy of labels present in FIBO.

1 Introduction

Ontologies are rich sources of information that provide deep information about the underlying concepts and entities. This information is described for a specific domain, contains the clearly defined relationship, and organizes in a defined structure mostly as a hierarchy. These properties make ontologies a great source for getting a deeper understanding of the relationship and properties of resources from the domain in consideration.

Public knowledge graphs and ontologies like DBpedia and Yago have been shown to work on various applications like the ones described in [Kobilarov *et al.*, 2009] and [Hahm *et al.*, 2014]. This has motivated and paved ways for the creation of domain focused ontologies like FIBO[1].

Effective techniques that enable identifying lexical similarity between the terms or concepts increase the effectiveness of the ontologies. These methods not only help in building new ontologies faster or augment the existing ones, but also it helps in the effective querying and concept search.

FinSim [Maarouf *et al.*, 2020; Mansar *et al.*, 2021] competitions are being held to promote the development of effective similarity measures. In the third edition of the competition FinSim-3[2] (being held in conjunction with 30th International Joint Conference on Artificial Intelligence (IJCAI-21)), the participants are challenged to develop methods and systems to assign hypernym and synonyms to financial terms by mapping them to one of the 17 high-level financial concepts present in FIBO.

In this paper, we present the systems developed by our team Lipi for hypernym and synonym assignment. We experimented with basic featurization methods like TF-IDF and advanced methods like pre-trained embedding models. Our top 3 systems use pre-trained FinBERT [Araci, 2019] embedding model that was fine-tuned on the data specific to financial domain . We also augmented the training data by utilizing the knowledge from DBpedia, Investopedia, FIBO and text corpus of prospectus shared with us. We describe the works related to our solution in the next section. Section 3 contains the formal problem statement, followed by data description in section 4. We describe our top three systems in section 5. Section 6 contains the details of the experimentation that we performed and the results from some of them. We draw our conclusions in section 7 while giving a glimpse of things that we would like to try in the future.

2 Related Works

Hypernym-hyponym extraction and learning text similarity using semantic representations have been very challenging areas of research for the NLP community. SemEval-2018 Task 9 [Camacho-Collados *et al.*, 2018] was such an instance.

---

[*]Contact Author
[†]Equal Contribution

[1]https://spec.edmcouncil.org/fibo/
[2]https://sites.google.com/nlg.csie.ntu.edu.tw/finnlp2021/shared-task-finsim (accessed on 8th July 2021)

Team CRIM [Bernier-Colborne and Barrière, 2018] performed the best in this shared task. They combined a supervised word embedding based approach with an unsupervised pattern discovery based approach. The FinSim shared tasks [Maarouf et al., 2020; Mansar et al., 2021] deal with adopting these challenges specific to the Financial Domain. Team IIT-K [Keswani et al., 2020] won FinSim-1 using a combination of context-free static embedding Word2Vec [Mikolov et al., 2013] and contextualized dynamic embedding BERT [Devlin et al., 2019]. Anand et al. [Anand et al., 2020] from the team FINSIM20 explored the use of cosine similarity between terms and labels encoded using Universal Sentence Encoder [Cer et al., 2018]. They also tried to extract hypernyms automatically using graph based approaches. Team PolyU-CBS [Chersoni and Huang, 2021] won FinSim-2 shared task using Logistic Regression trained over word embedding and probabilities derived from BERT [Devlin et al., 2019] model. They also experimented with GPT-2 [Radford et al., 2019]. Team L3i-LBPAM [Nguyen et al., 2021] comprising Nguyen et al. performed better than the baseline by using Sentence BERT [Reimers et al., 2019] to calculate cosine similarity between terms and hypernyms. [Saini, 2020; Pei and Zhang, 2021] and [Jurgens and Pilehvar, 2016] discussed various techniques to enrich the data which was available for training. In this edition of FinSim, the number of training samples and labels (financial concepts) were more than the previous two editions.

3 Problem Statement

Given a set F consisting of n tuples of financial terms and their hypernyms/top-level concepts/labels i.e. $F = \{(t_1, h_1), (t_2, h_2), ...(t_n, h_n)\}$ where h_i represents the hypernym corresponding to the ith term t_i and $h_i \epsilon$ set of labels mentioned in Table 1. For every unseen financial term, our task is to generate a ranked list $\hat{y}_i$ consisting of these 17 hypernyms in order of decreasing semantic similarity.

Evaluation Metrics The expected output is a raked list of predicted labels for every scored instance. The proposed systems are evaluated based on Accuracy and Mean Rank metrices as per the shared task rules. Evaluation script was provided by organizers, where accuracy and mean rank were defined as:

$Accuracy = \frac{1}{n} \sum_{i=1}^{n} I(y_i = \hat{y}_i[1])$

$MeanRank = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i.index(y_i))$

where $\hat{y}_i$ is the ranked list (with index starting from 1) of predicted labels corresponding to the expected label y_i. I is an identity matrix.

4 Data

4.1 Data Description

The training dataset shared for this task has a total of 1050 single and multi-word terms tagged to 17 different classes/labels. More than 91% of the terms have 6 words or less and the longest term has 22 words. There were 10 duplicate entries, and 3 terms were assigned 2 different labels. Along with this, a corpus of prospectuses in English was provided that had 211 documents. Some of the terms mentioned

Label	Count
Equity Index	280
Regulatory Agency	205
Credit Index	125
Central Securities Depository	107
Debt pricing and yields	58
Bonds	55
Swap	36
Stock Corporation	25
Option	24
Funds	22
Future	19
Credit Events	18
MMIs	17
Stocks	17
Parametric schedules	15
Forward	9
Securities restrictions	8
Total	**1040**

Table 1: Label distribution in the training set

in the training data were present in the corpus. Table 1 shows the distribution of these labels in the training set.

4.2 Data Augmentation

Since the majority of the terms had only a few tokens, we decided to expand the terms wherever possible using various sources. This approach had also been adopted by [Saini, 2020] and [Pei and Zhang, 2021] while participating in FinSim-1 and FinSim-2 respectively.

Acronym expansion: As mentioned by Keswani et al. [Keswani et al., 2020], the presence of acronyms created a major issue in maintaining consistency. We used the abbreviation extractor available in spaCy[3][Honnibal et al., 2020] package on the corpus of the prospectus to extract all the acronyms and their expansions. Upon manual inspection of a sample output, we identified that not all the extracted items were valid acronyms and their expansions. We cleaned the extracted list by dropping the records where:

- expansion had equal or less length than the acronym.
- expansion had parenthesis
- extracted acronym was a valid English word such as "fund" or "Germany".
- the expansion had less than or equal to 5 characters.

We managed to extract 635 acronyms from the prospectus corpus after applying the above exclusions. We used this data to expand the matching terms in the given train set and test sets.

Definitions from DBpedia: We used the DBpedia search API[4] to extract the description of the terms present in the

---

[3]https://spacy.io/

[4]https://lookup.dbpedia.org/api/search

train and test sets. We present such an example in Figure 1. In addition to the description, the label was also retained from the result payload to identify the right description for the input terms. We tried token overlap-based similarity of input terms with both matching labels and descriptions. We decided to use the label to term match for description matching after going through a randomly drawn sample. We cleaned both input terms and labels from DBpedia results by converting them to lower case, replacing punctuations by space, removing repetitive spaces, and singularizing the text. We calculated the token overlap ratios for cleaned term and DBpedia labels using the formulas mentioned below: $Ratio1 = length(s_1 \cap s_2)/length(s_1)$, $Ratio2 = length(s_2/length(s_1)$ where s_1 and s_2 represents sets of tokenized cleaned terms and tokenized and cleaned DBpedia labels respectively. We empirically decided to use all the instances with $Ratio1 = 1$ and $Ratio2 <= 1.25$ for matching a DBpedia label (and hence description) to the input term.

Definitions from Investopedia and FIBO: Inspired by [Saini, 2020], we obtained definitions of the terms present in Investopedia's data dictionary[5] by crawling it. We downloaded a glossary of financial terms from the website of FIBO. We cleaned all the terms from the train and test set and also the terms present in Investopedia's data dictionary using the steps described in the above DBpedia section. We then assigned the Investopedia or FIBO definition to the terms from the train and test sets where cleaned terms from train and test data matched to cleaned Investopedia terms perfectly.

The test set which was provided to us had 326 terms. We augmented the original train and test set with the records where we could either find definition or expansion using the above sources. The train set size increased to 1801 records and the test set size increased to 607 after the data augmentation. We present an example of data augmentation for the term "callable bond" in Table 2. Table 3 states the number of instances we used from each of the sources to augment the data we had.

5   System Description

We tried to solve this problem as the term classification and term similarity problems. Two of our 3 submissions are modelled as the term classification problem, whereas the third system is designed to be a phrase/sentence similarity problem between terms (or expanded terms from the augmented dataset) and the definitions of 17 class labels that were extracted from FIBO / Internet. All the systems rely on semantic similarity and use FinBERT model to generate the term or token embedding representations. We divided the given data into training and validation sets having 841 and 209 terms respectively.

5.1   System - 1 (S1)

This is the simplest of our proposed systems, where we did not use the augmented dataset and stuck to the original set that was shared by organizers. We loaded FinBERT pre-trained

model and fine-tuned it by trying to classify the representation of [CLS] token into one of the 17 labels mentioned previously. Since the original data did not have longer terms, we kept the maximum length to 32, and train and validation batch sizes of 64. We used Adam optimizer with a learning rate of 0.00002. We ran the model for 40 epochs and picked the model saved after 18th epoch based on the performance on the validation set. Finally, we ranked the predictions based on the predicted probability of each class.

5.2   System - 2 (S2)

This system is similar to System-1 with the only difference that data being the augmented set and not the original dataset. Since the augmented dataset has the descriptions of the terms, the input is considerably longer. Hence, we increased the maximum length to 256 while keeping all the other hyperparameters the same. After, training the model for 40 epochs we selected the model saved after the 17th epoch as the best model based on validation set performance.

5.3   System -3 (S3)

We explored the FIBO ontology to understand the hierarchy [Stepišnik Perdih *et al.*, 2021] of the 17 labels as depicted in Figure 2. We used the augmented data described in section 4.2 to create a labelled dataset having similarity scores. For every term definition (T) to label definition (L) mapping which existed in the extended training set, we assigned a similarity score of 1.0 to the (T,L) pair and picked up 10 training instances randomly ensuring none of their label definition was same as L. For each of the label definitions (LL) present in this sample, we extracted its root node and first child node. We did the same for the original label definition (L). Then, we compared these nodes. If the root node and first child node of L were different from that of LL then we assigned a similarity score of 0 to the (T, LL) pair. If the root nodes were the same, we assigned a similarity score of 'k' when the first child nodes differed and a similarity score of '2k' when they were the same (where $0 < k < 1$). We empirically figured out that k=0.4 works the best. As expected, the number of instances with a similarity score equal to 0 increased substantially. We under-sampled such instances and the new training set had 30% instances with similarity score 1.0, 12% instances with similarity score 'k', 28% instances with similarity score '2k' and 30% instances with similarity score 0. After that, we fine-tuned a FinBERT [Araci, 2019] model using Sentence BERT [Reimers *et al.*, 2019] framework with this newly generated labelled data for 25 epochs with a batch size of 20. Our objective was to minimize the multiple negatives ranking loss and online contrastive loss. We used a margin of 0.5 and cosine distance as a distance metric while training this model. Finally, we converted all of the 17 labels' definitions and term definitions from the validation set to vectors using this fine-tuned model. For every such term definition, we performed a semantic search over the label vectors and ranked them in decreasing order of similarity.

System 2 and 3 take advantage of term expansion during both model training and scoring phases, which causes certain observations to appear more than once (reference: Table 3). We

```
-<ArrayOfResults>
  -<Result>
    <Label>Credit default swap</Label>
    <URI>http://dbpedia.org/resource/Credit_default_swap</URI>
    -<Description>
      A credit default swap (CDS) is a financial swap agreement that the seller of the CDS will compensate the buyer in the event of a debt
      default (by the debtor) or other credit event. That is, the seller of the CDS insures the buyer against some reference asset defaulting. The
      buyer of the CDS makes a series of payments (the CDS "fee" or "spread") to the seller and, in exchange, may expect to receive a payoff if
      the asset defaults.
    </Description>
```

Figure 1: Sample output from DBpedia search API

Expanded Term/Term Definition	Label	Source
Callable bond	Bonds	original and acronym expansion
bond that includes a stipulation allowing the issuer the right to repurchase and retire the bond at the call price after the call protection period	Bonds	FIBO
A callable bond (also called redeemable bond) is a type of bond (debt security) that allows the issuer of the bond to retain the privilege of redeeming the bond at some point before the bond reaches its date of maturity.	Bonds	DBpedia

Table 2: Result of Data Augmentation of the term "Callable bond"

Data Source	Count
Original modelling data	1040
DBpedia	257
FIBO	236
Investopedia	85
Acronym expansion	218

Table 3: Details of various data sources

derive the final prediction by averaging the output probabilities for all the 17 classes for all the occurrences of the term.

6 Experimentation and Results

We had 1040 observations after removing the duplicates. We did an 80:20 split to create a training and validation set from this. We augmented the given modelling set by incorporating definitions from DBpedia, FIBO and Investopedia. We used the list of acronyms extracted from the prospectus corpus to create a copy with acronym expansion. This helped us to increase the original data to 1836 records (mentioned in Table 1). It should be noted that we could not find the expansions for all the terms given in the modelling set. Train and validation set sizes for the original modelling set and expanded data were (832 & 208) and (1469 & 366) respectively.

We established a baseline by running the scripts provided by the organizers. Then, we considered original modelling data and fine-tuned base BERT-cased model [Devlin *et al.*, 2019] to predict the class label by taking the representation of [CLS] token while passing it through few layers of a feed-forward network. This performed better than baseline. We then tried the same BERT-base model on the expanded dataset, which gave us further performance improvement. Since the only change between these runs was the data,

the improvement can be attributed to the expanded data.

We experimented with a few of the other pre-trained models that are available on the Huggingface model repository [Wolf *et al.*, 2020]. We observed clear improvement when we used the FinBERT model which was trained on data specific to the financial domain. The model performance successively increased when we used a combination of data expansion with FinBERT. Furthermore, we tried to fine-tune FinBERT using Sentence Transformers [Reimers *et al.*, 2019] to capture semantic textual similarity. For this, we used several combinations of term and term definitions with label and label definitions.

All the hyperparameters for the final 3 models are already given in the system description. After rigorous experimentation, these hyperparameters were selected empirically based on validation set performance. The results are presented in Table 4. Since the number of submissions was restricted to 3 for each team, we do not have the performance numbers of the BERT models in the test set. Analysing the results we see that SentenceBERT trained with FinBERT at the backed as mentioned in section-5.3 performed the best.

7 Conclusion and Future Works

In this work, we attempted to solve the hypernym and synonym discovery hosted at FinSim-3. This challenge aimed to enable the better use of ontologies like FIBO using hypernyms and synonyms, and we used these ontologies themselves to develop our systems which perform significantly better than the provided baseline systems. This proves the present use of these ontologies. The presented solution is recursive in a sense as it uses knowledge from ontologies to further increase the effectiveness and use of the same. Apart from data augmentation, our solution relies upon semantic similarity learnt from pre-trained embedding models
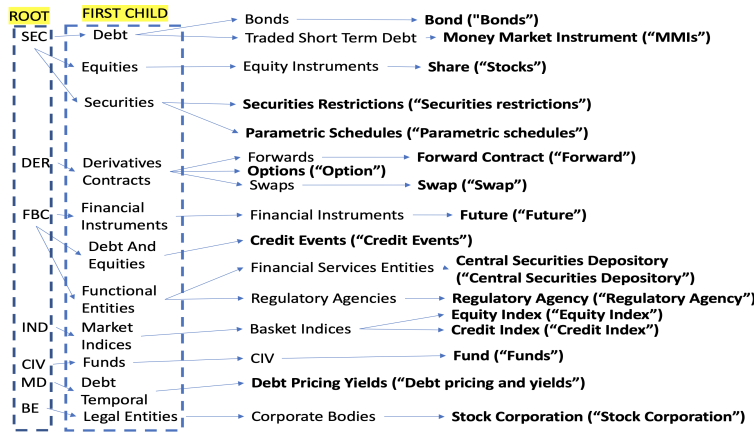
Figure 2: Label Hierarchy from FIBO. Bold (leaf nodes) denotes the labels.

Model	Data	Validation set Rank	Validation set Acc.	Test set Rank	Test set Acc.
Base-1	Org.	2.158	0.498	1.941	0.564
Base-2	Org.	1.201	0.876	1.75	0.669
BERT	Org.	1.177	0.899	-	-
BERT	Ext.	1.153	0.928	-	-
FinBERT(S1)	Org.	1.117	0.928	1.257	0.886
FinBERT(S2)	Ext.	1.110	0.942	1.220	0.895
SBERT(S3)	Ext.	**1.086**	**0.947**	**1.156**	**0.917**

Table 4: Results on validation and test set. Org. represents original and Ext. represents extended. Base refers to baseline.

that were learnt on the relevant domain. We observed the clear benefits of domain specific pretraining during the experimentation.

In future, we would like to explore Knowledge Graphs (as described in [Portisch *et al.*, 2021]) to further improve the improve performance of the models. We also want to explore other variants of FinBERT [Araci, 2019] and fine-tune them using the Masked Language Modeling technique (as mentioned by the winner of FinSim-2 [Chersoni and Huang, 2021]) and Next Sentence Prediction objective. Moreover, this research can be extended by extracting sentences present in the prospectus (similar to [Goel *et al.*, 2021]) to create positive and negative samples.

References

[Anand *et al.*, 2020] Vivek Anand, Yash Agrawal, Aarti Pol, and Vasudeva Varma. FINSIM20 at the FinSim task: Making sense of text in financial domain. In *Proceedings of the Second Workshop on Financial Technology and Natural Language Processing*, pages 104–107, Kyoto, Japan, 5 January 2020.

[Araci, 2019] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models, 2019.

[Auer *et al.*, 2007] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data, 2007.

[Bernier-Colborne and Barrière, 2018] Gabriel Bernier-Colborne and Caroline Barrière. CRIM at SemEval-2018 task 9: A hybrid approach to hypernym discovery. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 725–731, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[Camacho-Collados *et al.*, 2018] Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa-Anke, Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Saggion. SemEval-2018 task 9: Hypernym discovery. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 712–724, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[Cer *et al.*, 2018] Daniel Cer, Yinfei Yang, Sheng yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Universal sentence encoder, 2018.

[Chersoni and Huang, 2021] Emmanuele Chersoni and Chu-Ren Huang. *PolyU-CBS at the FinSim-2 Task: Combining Distributional, String-Based and Transformers-Based Features for Hypernymy Detection in the Financial Domain*, page 316–319. Association for Computing Machinery, New York, NY, USA, 2021.

[Devlin *et al.*, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina 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)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[Goel *et al.*, 2021] Tushar Goel, Vipul Chauhan, Ishan Verma, Tirthankar Dasgupta, and Lipika Dey. *TCS WITM 2021 @FinSim-2: Transformer Based Models for Automatic Classification of Financial Terms*, page 311–315.

Association for Computing Machinery, New York, NY, USA, 2021.

[Hahm *et al.*, 2014] Younggyun Hahm, Jungyeul Park, Kyungtae Lim, Youngsik Kim, Dosam Hwang, and Key-Sun Choi. Named entity corpus construction using wikipedia and dbpedia ontology. In *LREC*, pages 2565–2569, 2014.

[Honnibal *et al.*, 2020] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength Natural Language Processing in Python, 2020.

[Jurgens and Pilehvar, 2016] David Jurgens and Mohammad Taher Pilehvar. SemEval-2016 task 14: Semantic taxonomy enrichment. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1092–1102, San Diego, California, June 2016. Association for Computational Linguistics.

[Keswani *et al.*, 2020] Vishal Keswani, Sakshi Singh, and Ashutosh Modi. IITK at the FinSim task: Hypernym detection in financial domain via context-free and contextualized word embeddings. In *Proceedings of the Second Workshop on Financial Technology and Natural Language Processing*, pages 87–92, Kyoto, Japan, 5 January 2020.

[Kobilarov *et al.*, 2009] Georgi Kobilarov, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, and Robert Lee. Media meets semantic web – how the bbc uses dbpedia and linked data to make connections. In Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl, editors, *The Semantic Web: Research and Applications*, pages 723–737, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[Maarouf *et al.*, 2020] Ismail El Maarouf, Youness Mansar, Virginie Mouilleron, and Dialekti Valsamou-Stanislawski. The FinSim 2020 shared task: Learning semantic representations for the financial domain. In *Proceedings of the Second Workshop on Financial Technology and Natural Language Processing*, pages 81–86, Kyoto, Japan, 5 January 2020.

[Mansar *et al.*, 2021] Youness Mansar, Juyeon Kang, and Ismail El Maarouf. *The FinSim-2 2021 Shared Task: Learning Semantic Similarities for the Financial Domain*, page 288–292. Association for Computing Machinery, New York, NY, USA, 2021.

[Mikolov *et al.*, 2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[Nguyen *et al.*, 2021] Nhu Khoa Nguyen, Emanuela Boros, Gael Lejeune, Antoine Doucet, and Thierry Delahaut. *L3i LBPAM at the FinSim-2 Task: Learning Financial Semantic Similarities with Siamese Transformers*, page 302–306. Association for Computing Machinery, New York, NY, USA, 2021.

[Pei and Zhang, 2021] Yulong Pei and Qian Zhang. Goat at the finsim-2 task: Learning word representations of finan-

cial data with customized corpus. In *Companion Proceedings of the Web Conference 2021*, WWW '21, page 307–310, New York, NY, USA, 2021. Association for Computing Machinery.

[Portisch *et al.*, 2021] Jan Portisch, Michael Hladik, and Heiko Paulheim. *FinMatcher at FinSim-2: Hypernym Detection in the Financial Services Domain Using Knowledge Graphs*, page 293–297. Association for Computing Machinery, New York, NY, USA, 2021.

[Radford *et al.*, 2019] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners, 2019.

[Reimers *et al.*, 2019] Nils Reimers, Iryna Gurevych, Nils Reimers, Iryna Gurevych, Nandan Thakur, Nils Reimers, Johannes Daxenberger, and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2019.

[Saini, 2020] Anuj Saini. Anuj at the FinSim task: Anuj@FINSIM¡VLearning semantic representation of financial domain with investopedia. In *Proceedings of the Second Workshop on Financial Technology and Natural Language Processing*, pages 93–97, Kyoto, Japan, 5 January 2020.

[Stepišnik Perdih *et al.*, 2021] Timen Stepišnik Perdih, Senja Pollak, and Blaž Škrlj. *JSI at the FinSim-2 Task: Ontology-Augmented Financial Concept Classification*, page 298–301. Association for Computing Machinery, New York, NY, USA, 2021.

[Wolf *et al.*, 2020] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Huggingface's transformers: State-of-the-art natural language processing, 2020.