Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 2293–2299
Marseille, 11–16 May 2020

A French Corpus and Annotation Schema for Named Entity Recognition and Relation Extraction of Financial News

Ali Jabbari, Olivier Sauvage, Hamada Zeine, Hamza Chergui
Skaizen Group
Paris, France
{ajabbari, osauvage, hzeine, hchergui}@skaizengroup.fr

Abstract

In financial services industry, compliance involves a series of practices and controls in order to meet key regulatory standards which aim to reduce financial risk and crime, e.g. money laundering and financing of terrorism. Faced with the growing risks, it is imperative for financial institutions to seek automated information extraction techniques for monitoring financial activities of their customers. This work describes an ontology of compliance-related concepts and relationships along with a corpus annotated according to it. The presented corpus consists of financial news articles in French and allows for training and evaluating domain-specific named entity recognition and relation extraction algorithms. We present some of our experimental results on named entity recognition and relation extraction using our annotated corpus. We aim to furthermore use the the proposed ontology towards construction of a knowledge base of financial relations.

Keywords: Annotation Schema, Named Entity Recognition, Relation Extraction

1. Introduction

Strict regulatory regimes mandate financial institutions (e.g. banks, insurance companies, investment firms) to rigorously monitor their customers' financial activities. Operating in fear of massive fines in case of undetected violation of the regulations, these institutes have set up a procedure called *Know Your Customer (KYC)*. Today KYC procedure is a time consuming effort of collecting information through declarative questionnaires and watchlist lookups. Furthermore, each customer's profile need periodical reviews to ensure the accuracy of the provided information. The regulations in force also evolve over time and legislators do not specify a standard of sufficient information in order to prevent financial institutes from enforcing a bare minimum control. This has led financial institutions to seek as much data as possible on their clients' financial activities. In this context, automated solutions can help financial institutions to continuously update the information that can impact the assessment of risk profiles, e.g. ownership and managerial changes and company's domains of activity. This goal can be achieved by means of a knowledge base (KB) of the financial relations of the clients which benefits from population through automated information extraction.

For example, the news headline *"Jean-Dominique Senard appointed as Renault Chairman"* announces a financial event (managerial change) and thus, implies that the existing KYC records of those entities need to be updated. We would therefore like to extract the entities *Jean-Dominique Senard* as a *Person*, *Renault* as *Company*, and *chairman* as a *Role* (Named Entity Recognition). We would also like to extract the relation between the person and his role, and that between the role and the affiliated company (Relation Extraction). Extracting such information from unstructured data raises several challenges. Among other determining factors, the performance of named entity recognition and relation extraction algorithms depend on the available resources in the destination language. In this regard, applications for non-English languages (including French) face extra challenges related to available tools and the amount of training data. Besides the destination language, specific domains (e.g. economy and finance) further limit the available training resources. For instance, information about companies and their executives is quite limited in freely available knowledge bases such as DBpedia (Auer et al., 2007) or Freebase (Bollacker et al., 2008). These sources are mostly based on Wikipedia which does not include commercial non encyclopedic information. Furthermore, the writing style of the target texts can impact information extraction. While the official press provides well-formed texts, other sources such as social media are often written in an informal, not always grammatically correct style. The quality of the texts (e.g. typographical errors, missing text excerpts) can also affect the information extraction algorithms.

In this work, we present an ontology of financial concepts and relations that are important from the KYC and financial compliance point of view. This ontology serves as a model for custom named entity recognition and relation extraction tasks and will eventually be used as the schema of a future knowledge base of financial relations (Jabbari et al., 2019). We have created a French-language corpus consisting of news articles published in financial press annotated according to our ontology. This corpus allows for training and evaluating named entity recognition and relation extraction algorithms.

2. Related Works

Information extraction tasks often require annotated corpora for training and evaluation. Such linguistic resources are often rare or limited for French and other non-English languages in general. WikiNER (Nothman et al., 2013) is one of the few existing multilingual works in this regard. WikiNER offers a French corpus of Wikipedia texts annotated according to Wikipedia's schema. Europeana Newspapers (Neudecker, 2016) is another project which aims to construct an annotated corpus of historic newspapers in French and other European languages. Nevertheless, like most of the existing work in Named Entity Recognition (NER), it only contains standard entity types, such as persons, organizations, and locations. However, for domain-specific application, e.g. financial compliance, we need corpora that include domain-specific entity types which are unfortunately publicly unavailable. Extraction of key financial relations and events for populating knowledge bases needs a more fine-grained model, i.e. a domain-specific ontology. Ontologies are models meant to organize the knowledge as hierarchies of concepts and the relations between them. More specifically, ontologies can be defined as models to guide the relation extraction task (Reyes-Ortiz, 2018). In order to develop specialized information extraction tools, we need corpora annotated according to domain-specific ontologies in the target language.

Extraction of financial information is not an entirely untapped research area and we have identified several economy-related ontologies in the literature. The Resource Event Agent (REA) ontology (McCarthy, 1982) is a classic ontology originally developed for accounting applications. The main concepts in REA are *resources* (e.g. money or services), *events* (e.g. transactions), and *actors* (e.g. companies or individuals). The *Ontology of Economic Events (OEE)* (Benetka et al., 2017) further extends this ontology by including a semantic hierarchy (e.g. get → earn → profit-gross). They used a semi-supervised method that starts from a set of verbs describing common transactions in economic contexts (e.g. acquire, sell). At the end of the process, they propose a hierarchy of the most common verbs describing monetary transactions organized into five levels. However, their approach is limited to texts containing a monetary value and thus only covers a subset of financial relations. Semantic-based pipeline for economic events (SPEED) (Hogenboom et al., 2010) is an another example which employs a language processing pipeline combined with an ontology of economic concepts. One of the shortcomings of SPEED is that it does not address the ambiguity problem induced by conflicting reports of a single event.

Although some ontologies exist for financial compliance (See (Wang et al., 2007) and (Gaidukovs et al., 2017)), they are not intended for automated information extraction tasks, e.g. Named Entity Recognition and Relation Extraction. An appropriate ontology should allow for annotation of all relevant information in text (e.g. domains of activity, management personnel, mergers and other events) without including excessively fine-grained concepts that do not appear explicitly in the text. Given the absence of appropriate annotation schema and lack of annotated corpora in French language, we consider that constructing a domain-specific ontology and an annotated corpus is a key effort towards building knowledge management tools for financial compliance. While the ontology is used as annotation schema in the first place, it will be served as the model for organizing the extracted information in form of a future knowledge base of financial relations. The corpus itself allows for training and evaluating information extraction tools, such as named entity and relation extraction algorithms.

3. Dataset Collection

The information important to financial compliance experts appears in various data sources, such as official reports, news articles, and social media. Each of these sources have different writing styles and their information bear varying levels of availability and reliability. On the one hand, information issued in official reports are well written, rather unambiguous and with the highest levels of reliability. On the other hand, in social media, e.g. tweets, information is often communicated through informal writing style and their reliability is unverified due to the characteristics of such media. In this corpus, we have focused on news articles as they follow the standard writing style that is common in official reports and other reliable sources. Moreover, the amount of publicly available news articles are far beyond the volume of accessible official reports.

In order to create an economically-oriented corpus, we selected forty daily French financial newspapers as our main sources. Through private APIs, first we collected a dataset of more than 1 million news articles published between April 2011 and March 2019. Next, we compiled a list of 130 keywords including major companies' names (e.g. top 40 French companies in stock value), financial interactions (e.g. mergers, acquisitions), currencies, etc. By means of those keywords, we selected 130 news articles at random for manual annotation, ensuring the sample covers diverse financial relations and events.

4. Annotation Schema

The majority of the existing entity recognition algorithms such as Stanford CoreNLP tools (Bauer et al., 2015) only classify basic entity types, i.e. *Person*, *Organization*, and *Location* or do not support French language at all. since our main goal in this work to create a corpus of annotated texts that can be used to train and evaluate custom named entity recognition and relation extraction algorithms, we decided to define a custom annotation schema.

Towards building an appropriate schema or ontology, we first identify the concepts and relations important to banking compliance. We gathered the relevant information by studying current practices in banks (e.g. KYC questionnaires) and by discussing with the experts of the compliance domain. We also investigated existing domain-specific ontologies used in knowledge bases, such as DBpedia (Auer et al., 2007) and Freebase (Bollacker et al., 2008). While some key entities and relations are present in existing models (e.g. *Person*, *Company*, and *Location*), many other concepts are not covered by those schema. We hence crafted our ontology with a few instructions in mind: our model is aimed to be used as an annotation schema and thus, should allow to annotate the information that is explicitly present in text. This makes our approach different from excessively fine-grained ontologies that are not suitable for information extraction from text. Also, we aimed to keep our ontology compatible with the existing models in order to be able to further integrate existing data in external sources (e.g. complementary data on cities and countries). The resulting ontology was first implemented in *OWL* format using *Protégé* (Musen, 2015) tool and its entities and relations were adopted as annotation schema. Through a

Entity Type	Definition	Examples (French)
Person	Physical Persons	*Emmanuel Macron, Carlos Ghosn*
Organization:Association	Unions, clubs, and NGOs. It was not used due to its juridical ambiguity.	*CGT, Greenpeace*
Organization:Company	Private or public corporation	*Total S.A., Airbus*
Organization:GPE	National/International geopolitical entities	*gouvernement français, UE*
Organization:Media	Press and broadcasts	*Le Figaro, BBC*
Location:WorldRegion	World regions and continents	*Asie, moyen orient*
Location:Country	World countries	*France, USA*
Location:LocalRegion	Local regions, states and provinces	*Californie, Normandie*
Location:City	Cities, towns and urban areas	*Grand Londre, New York*
Role	Professional roles and positions	*PDG, patron*
Currency	Currencies and their symbols	*Euro, CHF, $*
Asset:FinancialAsset	Non-physical assets	*actions, obligations*
Asset:TangibleAsset	Physical assets and goods	*immobilier, véhicules*
Asset:MoneyAmount	Monetary amount without its currency	*dix millions, 10,35*
Document	Offical documents, diplomas, etc.	*passeport, contract*
Financing	Financings and investments	*financement, investissement*
Merger	Consolidation of two or more companies into one company	*fusion, opération M&A*
Demerger	Converse of a Merger	*scission*
Acquisition	Transfer of ownership of a company (Acquiree) to another (Acquirer)	*acquisition, rachat*
Activity	Economical sector of activity	*aéronautique, énergie*
IPO	Initial Public Offering	*introduction en bourse*
Penalty:Fine	Financial punishments	*amende, contravention*
Penalty:Imprisonment	Prison sentence	*peine de deux ans, condamnation*
Penalty:Sanction	Embargo and sanctions	*sanctioné, interdit*
Shareholding	This category was finally not used in annotations	*actionnarit*
BusinessDeal	Commercial transactions, i.e. payments and purchases	*transaction, paiement*

Table 1: Definition of annotated entity types according to our ontology

phase of trial annotation, we further refined our ontology in order to cover as much relevant information as possible and reduce possible ambiguities.

4.1. Entities

The entity types currently annotated in our corpus are described in Table 2.. Wherever hierarchies of concepts are present (e.g. *Organization >Company*), annotators were instructed to annotate the entity with the more specific type (i.e. *Company*) unless it was unclear from the context. Along with custom entity types, we have incorporated classic types.

Person denotes physical persons (first and last names). *Organization* type covers all public, private, national or international institutions. *Organization* is further categorized into *Company* (corporations and businesses), *GeoPoliticalentity* (national governments and international political organizations), and *Media* (press, broadcasts and websites). *Location* is also integrated with several subtypes: *WorldRegions* (continents and other world regions, e.g. Middle East), *LocalRegion* (states and provinces), and *City* (urban areas). We still tagged capital cities as *City* where they were referring to the government of a country (e.g. *Washington* as US government).

Through initial annotation efforts, we further trimmed down ambiguous entity types by either removing or merging them. We also added entity types that we found in our text sources that were not included in the initial schema. We advised annotators to tag noun groups, proper names, and trigger words depending on the corresponding types. For instance, entity types like *Person* and *Organization* are often mentioned as proper names (e.g. *Emmanuel Macron, Total S.A.*), *Role* instances are often present in form of noun groups (e.g. *Directeur Marketing*). Others like *Acquisition* are mostly mentioned by trigger nouns which can be in turn, nouns or verbs, e.g.:

Mercredi, Total a annoncé l'acquisition du néerlandais PitPoint.
(This Wednesday, Total announced the acquisition of Dutch PitPoint)

Google a racheté le leader de la vidéo en ligne YouTube pour 1,65 milliard de dollars.
(Google has bought the leader of online video YouTube for $ 1.65 billion)

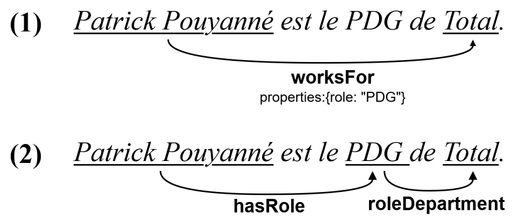

Figure 1: Two possible ways to annotate working relations in the sentence: *Patrick Pouyanné is the CEO of Total.*

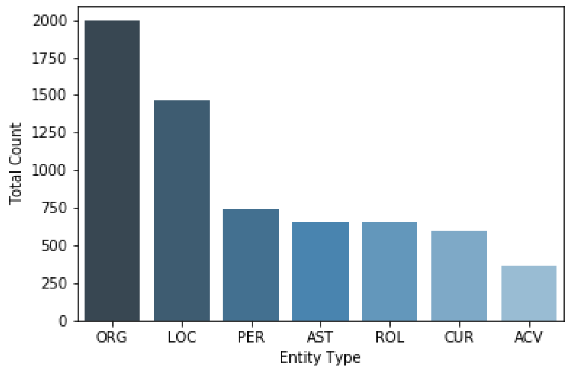

Figure 2: The distribution of annotated entities for the most frequent types: Organization, Location, Person, Role, Asset, Currency, and Activity. Together, these types represent more than 96% of the annotated entities in our corpus.

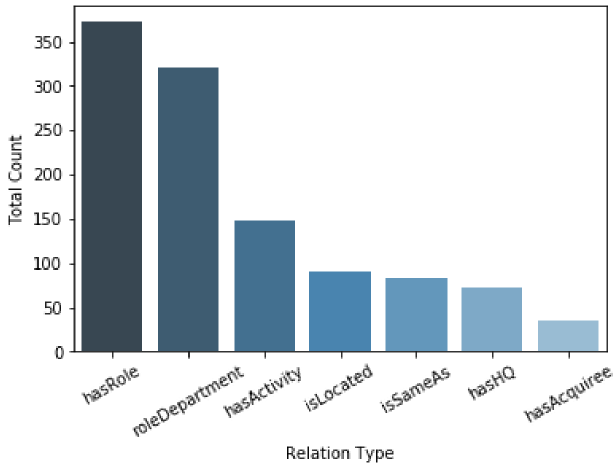

Figure 3: The distribution of annotated relations for the most frequent types. Together, these types represent more than 64% of the annotated relations in our corpus.

Compound entities (e.g. *Bank of England*) were annotated as a single entity (i.e. *Organization*) and nested entities (i.e. *England*) were intentionally left out. Finally, in case an entity was referred to with more than one mention, such as acronyms, translations, etc. every mention was annotated separately and were associated together by *isSameAs* relations.

4.2. Relations

A relation connects two or more named entities (*arguments* of the relation) and describes the type of association and interaction between those entities. Contrary to named entities which are explicit, i.e. expressed as sequences of word, relations are often implicit and can be recognized by the semantics of the context. A domain knowledge model (ontology) which consists in concepts, events, connections, and characteristics can be formalized by means of a set of both entity types and relations. While constructing our ontology, we noticed that a piece of information can be modeled in several ways. In Figure 1, the professional relationship between a company and its CEO is formalized in two fashions. The first one relates the person to his company via a direct *workFor* relation and his role as CEO is incorporated as a property of this relation. In the second one, the same information is represented by two relations instead: the *hasRole* relation denotes the position of the person and points to the mention of *CEO* in the text and the second relation (*roleDepartment*) shows in which company/department this position is held.

We decided to opt for the second approach in Figure 1 for a few reasons. First, this approach is more dependent on named entities explicitly present in text rather than complex semantic relations. Also, we know that Relation Extraction (RE) is generally a more challenging task compared to Named Entity Recognition (NER) and hence, we benefit from fine-grained NER to drive the relation extraction task. Moreover, we target text sources written in formal style and thus, this approach allows to adopt rule-based methods to relation extraction based on the types of entities and their grammatical relations in the sentence. Finally, by favoring entities to properties, we can enrich our future knowledge base with more nodes and improve querying on archives, e.g. querying all positions hold by a person through time or all persons who occupied a specific position in a company.

Table 2 lists the relation that are annotated in the current corpus. Instead of semantically complex relations to describe financial events and relations, we have adopted simpler relations which link entities as engaging parties. A financial relation is therefore composed of several basic relations. In this way, annotators were able to partially annotate a complex financial relation or event even if all ontological relations were not explicit in the text (e.g. payment parties are expressed explicitly, but the value is missing).

5. Corpus Statistics

Our corpus currently contains 130 manually annotated news article. Table 3 reports overall statistics of the corpus in terms of its volume and numbers of annotated entities and relations. The current corpus is annotated by four members of our team with engineering and financial operations backgrounds. The annotation was carried out using the open-source tool BRAT (Stenetorp et al., 2012) which allows for both entity and n-ary relation annotation based on custom schema. Each annotated document was reviewed twice to reduce error. In its current state, our corpus contains a total of 6736 named entities and 1754 annotated relations. The number of entities, however, varies greatly be-

Relation	Definition	Argument 1	Argument 2
hasRole	Occupation of a position in an organization	Person	Role
roleDepartment	Attachment of a position to an organization	Role	Organization
hasIncomingParty	Entering Parties	Merger/Demerger	Company/Organization
hasOutgoingParty	Outcome party of a merger/demerger	Merger/Demerger	Company/Organization
hasAcquirer	Initiator of an acquisition	Acquisition	Company
hasAcquiree	Subject of an acquisition	Acquisition	Company
hasActivity	Participation of an organization in a sector of activity	Organization	Activity
hasHQ	Having headquarters in a location	Organization	Location
hasNationality	Having citizenship of a country	Person	Country
isLocated	Localisation of an entity	*	Location
owns	Ownership of assets	Person/Organization	Asset
ownedBy	Converse of ownership relation	Asset	Organization/Person
hasCurrency	Corresponding currency of a money value	MoneyAmount	Currency
hasObject	Contract object of a transaction	BusinessDeal/Financing	Asset/MoneyAmount
hasCreditor	beneficiary of a transaction	BusinessDeal/Financing	Activity/Organization/Location
hasDebtor	Initiator of a transaction	BusinessDeal/Financing	Person/Organization
hasIPOCompany	Company going public in the IPO	IPO	Company
hasIPOMarket	IPO's launch market	IPO	Organization/Location
hasCondemned	The condemend party of a penalty sentence	Penalty	Person/Organization
hasIssuingAuthority	The authority issuing the penalty	Penalty	Organization

Table 2: Definition of annotated relations and the possible entity types of their arguments according to our ontology

	Total	Median per document
Documents	130	-
Sentences	3697	18
Words	64556	441
All Entities	6736	32
All Relations	1754	7

Table 3: Statistics of the manually annotated corpus

tween different types.
Figure 2 presents the distribution of annotated entities for the top seven most frequent entity types. These seven types together constitute more than 96% of all entity mentions. Frequency of each other entity types is about 1% of total counts or less. The distribution of annotated relations also varies across different types. Figure 3 shows the distribution of the top seven most frequent annotated relation types which combined, represent over 64% of all annotated relations. The presented corpus and its annotation interface is accessible on web at `http://bit.ly/CorpusFR` for viewing. Annotators with credentials can edit and add new documents to the corpus.

6. Experiments

We created the presented corpus with the goal of training named entity recognition and relation extraction models capable of retrieving the entity and relation types defined in our ontology. We hence conducted a series of experiments to evaluate baseline performance of our corpus for named entity recognition (NER) end relation extraction (RE). These experiments are meant as a proof of concept for developing information extraction tools based on our corpus. Future development of our corpus thus can improve the reported scores.

6.1. NER Experiment

We trained a NER model using the open-source NLP framework spaCy (Honnibal and Montani, 2017) which provides a training module allowing for custom NER training. Our experimental setup included spaCy version 2.1.4 on Python 3.7.3 running on a machine with Intel quad-core processor (Kaby Lake), Linux 4.4.0, and 16GB of memory.
We randomly selected 80% of the annotated data for training, keeping the remaining 20% for evaluation. We used spaCy's French model (`fr_core_news_md`) which supports *Person*, *Location*, and *Organization* entity types and we trained it with our selected custom and standard entity types, i.e. *Role*, *Currency* in addition to *Person*, *Location*, and *Organization*. Other annotated custom entity types (e.g. *Asset, Acquisition*) were not included individually due to the limited number of available annotations for these types. However, we empirically observed that including other types under a single entity type (`Miscellaneous`) can improve the NER performance for the target entity types. for the same reasons, we also combined subtypes (e.g. *City*, *Country*) into their parent class (i.e. *Location*) for training and evaluation.
The results of the NER evaluation are reported in Table 4. for each entity type, we calculated NER evaluation metrics which include precision, recall, and F1-score. We observed

Entity Type	Precision	Recall	F1-Score
Person	0.8450 (0.95)	0.7843 (0.88)	0.8135 (0.91)
Location	0.7755 (0.84)	0.7729 (0.84)	0.7742 (0.84)
Organization	0.7185 (0.81)	0.6819 (0.77)	0.6997 (0.79)
Role	0.6355 (0.78)	0.5440 (0.66)	0.5862 (0.71)
Currency	0.8067 (0.84)	0.8000 (0.83)	0.8033 (0.84)
All	0.7536 (0.90)	0.7182 (0.85)	0.7355 (0.87)

Table 4: Custom NER performance for the selected entity types using spaCy training module. Scores for partial extraction are reported in parentheses.

the best overall accuracy for *Person* (best precision score) and *Currency* (best recall score) entity types. given the limited number of world currencies, such performance can be explained by the less variability of *Currency* mentions. Next, we report the best performance for *Location*, *Person*, and *Organization* entity types. These entity types are also the most frequent types in the corpus and have certain features (e.g. initial capital letters) that may explain the superior performance. Finally, for *Role* type the average F1-score is lowest at 0.71. This score can be explained by the smaller number of *Role* mentions in the training set and also varying length of the mentions. For instance, a position title such as *"directeur adjoint en charge des finances, des systèmes d'information et de la direction industrielle"* (Deputy director in charge of finances, information systems, and industrial management) may be partially retrieved as *"directeur adjoint en charge des finances" (Deputy director in charge of finances)*. Although the retrieved text contains important information on the mentioned position, such partial extraction counts as an error. By including partial extractions, the performance levels would be considerably higher for *Role* type.

6.2. RE Experiment

Following the named entity recognition (NER) experiment, we started a series of experiments on relation extraction (RE). In this section, we report the experiment we conducted on extraction of the relations that we have defined around the entity type *Role*, i.e. *hasRole* and *roleDepartment*. Similar to entity types, the relations in our ontology are defined based on real world concepts and how they appear in text sources. We have thus defined which types of entities are accepted as arguments of each relation (Table 2). Such approach in relation extraction allows to filter out non relevant results, e.g. `Person:hasRole:Location` and helps rule-based relation extraction.

Using spaCy's dependency parser, we extract the dependency tree for each annotated phrase containing relations around *Role*. As we work with texts written in formal style, most of the mentions of these relations follow a series of dependency patterns. Once a relation is identified by a pattern, we verify the recognized type of the entities engaged in the relation in order to associate the result to the relevant relation type (e.g. a *hasRole* relation requires a *Person* and a *Role* instance as arguments). We identified and integrated 6 patterns in our relation extraction algorithm and observed strong precision rates with both exact and partial entity extraction (see Table 5). In contrast, recall rates remain lower in comparison for both tests. We hypothesize that recall rates can further be enhanced by adding additional patterns. We also look forward to examine our approach for other relation types and also, to examine statistical approaches to relation extraction.

Evaluation	Precision	Recall	F1-Score
Exact NER	0.81	0.34	0.49
Partial NER	0.92	0.39	0.55

Table 5: Rule-based relation extraction performance for *hasRole* and *roleDepartment* relations. Evaluation using exact vs. partial NER results.

7. Conclusion

In this work, we presented a domain-specific ontology of financial entities and relations that we used as an annotation schema for French news. We created a corpus of annotated texts in French that can be used to train and evaluate information extraction models. We demonstrated some of our experimental results on entity recognition and relation extraction using our corpus as training dataset. In parallel to our ongoing experiments and our efforts towards building a knowledge base of financial relations, we continue to extend our corpus by adding more annotated documents.

8. Bibliographical References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *Proceedings of the 6th International The Semantic Web and 2Nd Asian Conference on Asian Semantic Web Conference*, ISWC'07/ASWC'07, pages 722–735, Berlin, Heidelberg. Springer-Verlag.

Bauer, J., Surdeanu, M., Bethard, S., Manning, C., McClosky, D., and Finkel, J. (2015). The Stanford CoreNLP Natural Language Processing Toolkit. pages 55–60.

Benetka, J. R., Balog, K., and Norvag, K. (2017). Towards Building a Knowledge Base of Monetary Transactions from a News Collection. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. (2008). Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 1247–1250, New York, NY, USA. ACM.

Gaidukovs, A., Grjozs, P., and Kebers, G. (2017). Ontology based PEP identification. In *CEUR Workshop Proceedings*, volume 1898, pages 1–8.

Hogenboom, F., Hogenboom, A., Frasincar, F., Kaymak, U., Van Der Meer, O., Schouten, K., and Vandic, D. (2010). SPEED: A semantics-based pipeline for economic event detection. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*

Intelligence and Lecture Notes in Bioinformatics), volume 6412 LNCS, pages 452–457.

Honnibal, M. and Montani, I. (2017). spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*.

Jabbari, A., Sauvage, O., and Cabioch, N. (2019). Towards a knowledge base of financial relations: Overview and project description. In *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 313–316. IEEE.

McCarthy, W. E. (1982). the Rea Accounting Model - A Generalized Framework for Accounting Systems in a Shared Data Environment. *Accounting Review*, 57(3):554–578.

Musen, M. A. (2015). The protégé project: a look back and a look forward. *AI Matters*, 1(4):4–12.

Neudecker, C. (2016). An Open Corpus for Named Entity Recognition in Historic Newspapers. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 4348–4352.

Nothman, J., Ringland, N., Radford, W., Murphy, T., and Curran, J. R. (2013). Learning multilingual named entity recognition from wikipedia. *Artif. Intell.*, 194:151–175, January.

Reyes-Ortiz, J. A. (2018). Criminal Event Ontology Population and Enrichment using Patterns Recognition from Text. *International Journal of Pattern Recognition and Artificial Intelligence*, page S0218001419400147, dec.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations Session at EACL 2012*, Avignon, France, April. Association for Computational Linguistics.

Wang, Y., Xu, D., Wang, H., Ye, K., and Gao, S. (2007). Agent-oriented Ontology for Monitoring and Detecting Money Laundering Process. *Proceedings of the 2Nd International Conference on Scalable Information Systems*, (7001896):81:1—-81:4.