

RESSALVA

Atendendo solicitação do autor,
o texto completo desta dissertação será
disponibilizado somente a partir de
29/09/2021.



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Rio Claro - SP

Pedro Ivo Monteiro Privatto

**Uma Abordagem para Reconhecimento de
Entidades Nomeadas usando Conhecimento
Externo**

Rio Claro - SP

2020

Pedro Ivo Monteiro Privatto

**Uma Abordagem para Reconhecimento de Entidades
Nomeadas usando Conhecimento Externo**

Orientador: Prof. Dr. Ivan Rizzo Guilherme

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Rio Claro - SP.

Financiadora: FUNDUNESP

Rio Claro - SP

2020

P961a	<p>Privatto, Pedro Ivo Monteiro</p> <p>Uma abordagem para reconhecimento de entidades nomeadas usando conhecimento externo / Pedro Ivo Monteiro Privatto. -- Rio Claro, 2020</p> <p>87 f. : il., tabs.</p> <p>Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Geociências e Ciências Exatas, Rio Claro</p> <p>Orientador: Ivan Rizzo Guilherme</p> <p>1. Ciência da Computação. 2. Inteligência Artificial. 3. Processamento de Linguagem Natural. 4. Extração de Informação. 5. Reconhecimento de Entidades Nomeadas. I.</p> <p>Título.</p>
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Geociências e Ciências Exatas, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

ATESTADO DE APROVAÇÃO - DEFESA

Atestamos que **PEDRO IVO MONTEIRO PRIVATTO**, RA nº: CCO180106, RG nº 41.458.889-7, expedido pela SSP/SP, defendeu, no dia 29/09/2020, a dissertação intitulada ***Uma abordagem para reconhecimento de entidades nomeadas usando conhecimento externo***, junto ao Programa de Pós Graduação em Ciência da Computação, Curso de Mestrado Acadêmico, tendo sido 'APROVADO'.

Atestamos ainda que a obtenção do título dependerá de homologação pelo Órgão Colegiado competente.

São José do Rio Preto, 29 de setembro de 2020



Silvia Emiko Kazama
Supervisor Técnico de Seção
Seção Técnica de Pós-Graduação

Pedro Ivo Monteiro Privatto

Uma Abordagem para Reconhecimento de Entidades Nomeadas usando Conhecimento Externo

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Rio Claro - SP.

Financiadora: FUNDUNESP

Banca Examinadora

- Prof. Dr. Ivan Rizzo Guilherme (Orientador)
Departamento de Estatística, Matemática Aplicada e Computação
Universidade Estadual Paulista "Júlio de Mesquita Filho"
- Prof^ª. Dr^ª. Veronica Oliveira de Carvalho
Departamento de Estatística, Matemática Aplicada e Computação
Universidade Estadual Paulista "Júlio de Mesquita Filho"
- Prof. Dr. Evandro Eduardo Seron Ruiz
FFCLRP Departamento de Computação e Matemática
Universidade de São Paulo

Rio Claro - SP

29 de setembro de 2020

Agradecimentos

Agradeço primeiramente à minha mãe Inês Regina, meu pai Reinaldo, e meu irmão Vitor Hugo por sempre me apoiarem e me proporcionarem um ambiente propício para que eu possa trilhar meu caminho da melhor forma possível. Só consegui chegar aonde cheguei pois tive meus medos acalmados, minhas dores diminuídas, e minhas preocupações aliviadas por aqueles que fazem tudo o que podem por mim. Não conseguiria ser nada do que sou se não tivesse sido ensinado, guiado e principalmente amado por vocês. Palavras não são suficientes para expressar meu sentimento de gratidão àqueles que são a razão do meu existir.

Agradeço também ao Ivan, meu professor, orientador, e acima de tudo amigo. Agradeço pela disponibilidade em aceitar-me como orientando e pela chance de participar em outros projetos, agradeço pela paciência quando eu me enrolo na tentativa de explicar minhas ideias, agradeço pela dedicação e disposição de ler e-mails, fazer revisões e ter conversas em dias e horários completamente não convencionais. Agradeço principalmente por ser a figura de “paizão” e se importar com todos os alunos dentro e principalmente fora do laboratório.

Agradeço a todos os meus amigos que sempre me propiciam momentos incríveis, que suportam meus desabafos, ouvem minhas preocupações, se importaram comigo, divagam nos mais diversos assuntos, e tornam minha caminhada mais alegre. Agradeço tanto aos mais recentes quanto aos de mais longa data. Creio que só somos capazes de crescer como seres humanos quando partilhamos experiências e ideias, e por isso sou imensamente grato por vocês me permitirem ter a chance de tentar melhorar como pessoa.

Agradeço também ao grande Kazuo Miura (*in memoriam*), um exemplo incrível de ser humano que, apesar de eu não ter tido o privilégio de conviver por muito tempo, sempre levarei comigo como exemplo de humildade, profissionalismo, dedicação e, acima de tudo, perseverança. Sonho para que um dia eu consiga chegar a ser pelo menos uma fração da pessoa incrível que o grande Kazuo era.

Agradeço ao excelente trabalho de todas as pessoas que participaram da minha formação e espero que um dia o mundo todo reconheça e glorifique a importância dos professores.

Agradeço à FUNDUNESP pela concessão da bolsa de pesquisa, sob o processo nº Fundunesp 2014/00545-0, Fundação para o Desenvolvimento da UNESP (FUNDUNESP)

Resumo

Nas diferentes áreas do conhecimento os dados textuais são importantes fontes de informação. Neste contexto, os métodos para Extração de Informação têm sido desenvolvidos para a identificação e estruturação de informações presentes em documentos textuais. Como subárea da Extração de Informação há o Reconhecimento de Entidades Nomeadas, que consiste em definir métodos para identificar Entidades Nomeadas, tais como Pessoa, Local, Data, entre outras, em textos. Recentemente esses métodos fazem uso de técnicas provenientes do Processamento de Linguagem Natural e de Aprendizado de Máquina. O objetivo do presente trabalho é propor uma metodologia para Reconhecimento de Entidades Nomeadas considerando os seguintes aspectos: identificação e implementação de métodos para representação de características gramaticais; identificação e implementação das novas abordagens que utilizam técnicas recentes de Aprendizado de Máquina, como BiLSTM, BiGRU e CRF; e avaliar, de maneira experimental, a integração de fontes externas de conhecimento, na forma de *Gazetteers* e Grafo de Conhecimento, vindos da Freebase e YAGO. O protocolo de avaliação experimental foi composto pela definição de quatro configurações de redes neurais, duas destas fazendo uso de conhecimento externo, e sua aplicação em cinco *datasets* com diferentes características. Nos experimentos realizados, houve ganho de F1-Score em 18 dos 40 casos onde foi utilizado conhecimento externo, chegando a um ganho de até 1,3%. Além do fato de não ter apresentado ganho em grande parte dos casos, a maioria dos ganhos foi inferior a 0,5% no F1-score. Os resultados experimentais dos métodos utilizados nos *datasets* escolhidos evidenciam que as estratégias empregadas para a integração do conhecimento externo agregaram baixos ganhos aos modelos, como mostrado pelas métricas Precisão, Abrangência e F1-Score. Portanto, se a fonte de conhecimento não provê informações relevantes ao domínio da tarefa, e se a maneira de agregação do conhecimento não capta o conteúdo relevante presente no mesmo, esta adição de conhecimento externo não necessariamente é benéfica à metodologia.

Palavras-chave: Reconhecimento de Entidades Nomeadas, Processamento de Linguagem Natural, Extração de Informação, Inteligência Artificial.

Abstract

In different areas of knowledge textual data are important sources of information. In this context, Information Extraction methods have been developed to identify and structure information present in textual documents. As a subarea of Information Extraction there is Named Entity Recognition, which consists of using methods to identify Named Entities, such as Person, Place, Date, among others, in texts, using techniques from Natural Language Processing and Machine Learning. Recently, these methods use techniques from Natural Language Processing and Machine Learning. The purpose of this work is to propose a methodology for Named Entity Recognition considering the following aspects: identification and implementation of grammatical feature representation methods; identification and implementation of new approaches that use recent Machine Learning techniques, such as BiLSTM, BiGRU and CRF; and to evaluate, in an experimental way, the integration of external knowledge sources, in the form of Gazetteers and Knowledge Graph, coming from Freebase and YAGO. The experimental evaluation protocol was composed by four configurations of neural networks, two of them making use of external knowledge, and their application in five datasets with different characteristics. In the conducted experiments, there was a gain of F1-Score in 18 of the 40 cases where external knowledge was used, reaching a gain of up to 1.3%. In addition to the fact that there was no gain in most cases, the majority of the gains were lesser than 0.5% in F1-score. The experimental results of the methods applied to the chosen datasets show that the strategies used for the integration of external knowledge added low gains to the models, as shown by the metrics Precision, Recall and F1-Score. Therefore, if the source of knowledge does not provide relevant information to the task domain, and if the way of aggregating the external knowledge does not capture the relevant content present in it, this addition of external knowledge is not necessarily beneficial to the methodology.

Keywords: Named Entity Recognition, Natural Language Processing, Information Extraction, Artificial Intelligence.

Lista de Figuras

Figura 1 – Arquitetura geral de sistemas para Extração de Informação.	18
Figura 2 – Exemplo de frase com Entidades Nomeadas.	20
Figura 3 – Funcionamento de uma Rede Neural de Múltiplas Camadas.	26
Figura 4 – Exemplo de CNN aplicada à frases.	30
Figura 5 – Exemplo de CNN aplicada à palavra.	31
Figura 6 – Exemplo de funcionamento de uma RNN expandida através do tempo.	32
Figura 7 – Interação entre elementos internos de uma unidade LSTM.	33
Figura 8 – Interação entre elementos internos de uma unidade GRU.	35
Figura 9 – Exemplo de cálculo da melhor sequência de rótulos.	37
Figura 10 – Exemplo da notação IOB.	48
Figura 11 – Exemplo da seleção dos tipos usados para construção dos <i>Gazetteers</i>	56
Figura 12 – Configuração da abordagem utilizada.	61
Figura 13 – Operação de concatenação dos vetores de características.	62
Figura 14 – Mapeamento de caracteres para vetor de índices associados.	63
Figura 15 – Mapeamento de caracteres para <i>Character embeddings</i>	64

Lista de Tabelas

Tabela 1 – Síntese dos trabalhos relacionados.	46
Tabela 2 – Exemplo do padrão de dados CoNLL2003.	48
Tabela 3 – Quantificação dos conjuntos do <i>dataset</i> CoNLL2003.	49
Tabela 4 – Distribuição das entidades do <i>dataset</i> CoNLL2003.	49
Tabela 5 – Quantificação dos conjuntos do <i>dataset</i> OntoNotes 5.	50
Tabela 6 – Distribuição das entidades do <i>dataset</i> OntoNotes 5.	50
Tabela 7 – Distribuição das entidades do <i>dataset</i> GUM.	51
Tabela 8 – Quantificação dos conjuntos do <i>dataset</i> GUM.	51
Tabela 9 – Distribuição das entidades do <i>dataset</i> MIT Movies.	52
Tabela 10 – Quantificação dos conjuntos do <i>dataset</i> MIT Movies.	52
Tabela 11 – Distribuição das entidades do <i>dataset</i> MIT Restaurants.	52
Tabela 12 – Quantificação dos conjuntos do <i>dataset</i> MIT Restaurants.	53
Tabela 13 – Quantificação das sentenças e entidades contidas nos <i>datasets</i>	53
Tabela 14 – Quantificação das entidades presentes nos <i>Gazetteers</i>	57
Tabela 15 – Exemplo de uma frase com adição de informações dos <i>Gazetteers</i>	57
Tabela 16 – Parâmetros utilizados nas redes neurais durante a primeira rodada de experimentos.	70
Tabela 17 – Resultados para experimentos usando <i>Gazetteers</i> em redes com 200 unidades de memória.	71
Tabela 18 – Resultados para experimentos usando <i>Knowledge embeddings</i> em redes com 200 unidades de memória.	72
Tabela 19 – Parâmetros utilizados nas redes neurais durante a segunda rodada de experimentos.	74
Tabela 20 – Resultados para experimentos usando <i>Gazetteers</i> em redes com 400 unidades de memória.	75
Tabela 21 – Resultados para experimentos usando <i>Knowledge embeddings</i> em redes com 400 unidades de memória.	76
Tabela 22 – Síntese dos melhores resultados para cada <i>dataset</i>	78
Tabela 23 – Comparação dos melhores resultados de cada <i>dataset</i> com outros trabalhos.	79

Lista de Abreviaturas

ANN	<i>Artificial Neural Network</i>
BiLSTM	<i>Bidirectional Long Short-Term Memory</i>
BiGRU	<i>Bidirectional Gated Recurrent Unit</i>
CNN	<i>Convolutional Neural Network</i>
CRF	<i>Conditional Random Fields</i>
EI	Extração de Informação
FN	Falso Negativo
FP	Falso Positivo
GRU	<i>Gated Recurrent Unit</i>
IOB	<i>Inside-Outside-Beginning</i>
LSTM	<i>Long Short-Term Memory</i>
NER	<i>Named Entity Recognition</i>
PLN	Processamento de Linguagem Natural
RNN	<i>Recurrent Neural Network</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

Sumário

1	Introdução	13
1.1	Objetivos	16
1.2	Organização do Texto	17
2	Extração de Informação	18
2.1	Visão Geral	18
2.2	Reconhecimento de Entidades Nomeadas	20
2.2.1	Abordagens Simbólicas	21
2.2.2	Abordagens Numéricas	23
2.2.2.1	Redes Neurais Artificiais	26
2.2.2.2	Redes Neurais Convolucionais	29
2.2.2.3	Redes Neurais Recorrentes	30
2.2.2.4	Campos Aleatórios Condicionais	35
2.2.3	Abordagens Híbridas	38
3	Trabalhos Relacionados	40
3.1	Tópicos Identificados	40
3.2	Revisão Bibliográfica	40
3.3	Síntese da Revisão Bibliográfica	45
4	Materiais e Métodos	47
4.1	<i>Datasets</i>	47
4.1.1	CoNLL2003	48
4.1.2	OntoNotes5	49
4.1.3	GUM	50
4.1.4	MIT Movies	51
4.1.5	MIT Restaurants	51
4.1.6	Síntese dos <i>datasets</i>	53
4.2	Repositórios de conhecimento	53
4.2.1	Freebase	54

4.2.2	YAGO	55
4.3	Métricas de Avaliação	57
5	Abordagem Proposta	60
5.1	Avaliação Experimental	60
5.2	Representação Vetorial	61
5.2.1	<i>Word embeddings</i>	62
5.2.2	<i>Character embeddings</i>	62
5.2.3	<i>Casing embeddings</i>	64
5.2.4	Conhecimento Externo	65
5.3	Classificação	66
5.4	Resultados	68
5.4.1	Primeira rodada	69
5.4.2	Segunda rodada	73
5.5	Sínteses dos Resultados	77
6	Conclusões	80
6.1	Trabalhos Futuros	81
	REFERÊNCIAS	83

1 Introdução

A crescente evolução do uso de tecnologias de informação nas mais diversas áreas é responsável pela grande quantidade de dados gerados em diferentes formatos, sejam esses dados imagens, áudios, vídeos ou textos. Grande parte desses dados gerados não apresentam nenhum tipo de estruturação, servindo apenas como arquivo para empresas e governos. Segundo [1], um dado não estruturado é aquele que não apresenta uma estrutura clara, semanticamente evidente e de fácil processamento por máquinas. A não estruturação de informações é uma barreira para a tarefa de recuperação de dados, uma vez que toda consulta deve ser feita de maneira manual em todos os arquivos que possam conter a informação relevante em meio a tantas informações irrelevantes.

O montante de dados textuais não estruturados cresceu também devido à evolução do perfil do usuário da internet que, com a evolução das tecnologias de informação e comunicação, passou de consumidor de informações para o gerador destas, principalmente no âmbito das redes sociais, onde a interação entre usuários é seu ponto chave.

No cenário de crescimento do volume de dados textuais gerados, percebeu-se a importância da extração, estruturação e utilização desses dados estruturados. Especialmente quando se deseja filtrar por aspectos específicos em meio a enormes quantias de dados, como doenças ligadas a um certo gene em artigos científicos ou relatórios médicos, processos industriais de bombeamento com volume maior que um limiar em uma base de relatórios técnicos, atentados terroristas ocorridos num determinado país em uma coletânea de notícias de jornal, entre outros. Desta maneira, o objetivo de se estruturar dados textuais é para que estes sejam mais facilmente consultáveis e reutilizáveis.

Em meio ao contexto de fontes textuais, a estruturação dos dados faz uso de Processamento de Linguagem Natural (PLN), área a qual utiliza conceitos de Inteligência Artificial e de Linguística para processar dados e automatizar tarefas que envolvam linguagem. Dentre as tarefas que são foco de estudo do Processamento de Linguagem Natural pode-se citar Tradução de Máquina, Reconhecimento de Fala, Análise Automática de Discurso, Sumarização Automática de Textos, Extração de Informação, entre outras.

A tarefa de Extração de Informação (EI), que é o tema deste trabalho, é a subárea

de PLN que é responsável por identificar informações desestruturadas contidas no texto e estruturá-las, facilitando sua posterior recuperação [2]. No contexto da EI há a tarefa chamada de Reconhecimento de Entidades Nomeadas - inglês *Named Entity Recognition* (NER), que consiste em encontrar conceitos, formados por uma ou mais palavras, presentes nos textos e categorizá-los de acordo com seu grupo semântico.

As abordagens para Reconhecimento de Entidades Nomeadas tradicionalmente fazem uso de muitas técnicas vindas da Linguística, tais como: etiquetas sintáticas, lema das palavras, prefixos e sufixos, entre outras, para extrair as informações presentes nos textos. Todo o processo necessário para o uso das técnicas tradicionais é bastante trabalhoso, pois envolve diversas etapas de preparação dos dados de entrada.

Com a finalidade de reduzir o trabalho das abordagens tradicionais de NER, as Abordagens Numéricas têm sido desenvolvidas de forma a simplificar as etapas de preparação dos dados. Recentemente, o uso de técnicas que modelam matematicamente os aspectos linguísticos sintáticos e semânticos vêm ganhando espaço. O uso de Abordagens Numéricas vem se popularizando por estas atingirem resultados similares, algumas vezes até melhores, quando comparadas às técnicas clássicas sem necessitar de um processo extensivo de *feature engineering*, processo que tem como finalidade a seleção das características que serão utilizadas na tarefa de NER. Porém, apesar de não ser necessário, existem abordagens que realizam *feature engineering* para as técnicas de Aprendizado de Máquina com o objetivo de incluir características que não podem ser captadas pelos *Word embeddings*, por exemplo.

Dentre as modelagens que têm sido utilizadas recentemente pode-se citar os *Word embeddings*, que são mapeamentos das palavras presentes nos textos para um espaço vetorial. Ao se fazer uso dos *Word embeddings*, a tarefa de NER pode ser tratada como uma tarefa de classificação, sendo atualmente bastante empregado o uso de classificadores que aprendem a categorizar as entidades presentes nos textos.

Depois de obtidas as representações das características desejadas, ocorre a etapa na qual as entidades são de fato reconhecidas. Tradicionalmente, esta etapa de reconhecimento era composta por métodos que faziam uso de regras e padrões de palavras ou caracteres para caracterizar cada tipo de entidade. Era comum também a realização da tarefa de Reconhecimento de Entidades Nomeada fazendo uso de amplas listas com entidades previamente reunidas, chamadas de *Gazetteer lists*, tendo como exemplo listas com nomes próprios, pontos geográficos, entre outros.

As novas abordagens frequentemente fazem uso de métodos numéricos, principalmente aqueles capazes de incorporar informações episódicas ao processo de Reconhecimento de Entidades Nomeadas, se beneficiando do aspecto sequencial das palavras dentro de um texto.

Em meio a esse contexto, existem as chamadas técnicas híbridas, que fazem uso tanto de aspectos linguísticos tradicionais, quanto uso de modelos matemáticos para realizar o Reconhecimento de Entidades Nomeadas. A motivação por trás do uso das técnicas híbridas está em tentar unir os aspectos positivos dos diferentes tipos de abordagens.

Recentemente alguns autores tem realizado a integração de fontes externas para agregar mais informação semântica e contextual. Uma estratégia frequentemente utilizada é o uso de fontes externas denominadas *Gazetteers* para agregar novas características que servem como entrada para métodos numéricos. A motivação do uso de fontes externas é enriquecer as representações numéricas das amostras com informações que não são encontradas nos *datasets* que estão sendo trabalhados.

Diante deste cenário, o presente trabalho apresenta um levantamento bibliográfico acerca dos métodos de Reconhecimento de Entidades Nomeadas. O resultado deste levantamento permitiu identificar trabalhos que foram utilizados para definir a abordagem proposta para Reconhecimento de Entidades Nomeadas e encontrar bases de dados utilizadas nessa tarefa. Além disso, o levantamento bibliográfico permitiu identificar quais recursos linguísticos são comumente utilizados para a tarefa de Reconhecimento de Entidades Nomeadas.

Com essas informações foi então proposta uma abordagem híbrida para a tarefa de Reconhecimento de Entidades Nomeadas, fazendo uso de *Word embeddings* unidos à fontes externas de conhecimento para alimentar uma arquitetura de rede neural capaz de captar dependências entre episódios e rótulos de uma sequência.

A abordagem proposta tem o objetivo de incorporar informações semânticas, contidas em imensos repositórios ricos em conhecimento, às técnicas que fazem uso somente informações das palavras isoladamente. O motivo dessa combinação é avaliar a hipótese de que informações semânticas podem beneficiar os processos numéricos atuais. Esta hipótese está fundamentada no fato de que as informações contidas em fontes externas não são encontradas nos dados que estão sendo trabalhados. Deste modo, a incorporação de conhecimento externo pode vir a ajudar os modelos de Aprendizado de Máquina a realizar uma melhor distinção entre as entidades, por meio da assimilação das dependências entre essas informações externas.

Como fontes de conhecimento externo foram utilizados dois dos principais repositórios disponíveis: o Freebase e o YAGO. As informações semânticas das palavras foram unidas às respectivas representações vetoriais para servirem de entrada aos modelos de classificação.

Em relação aos modelos de aprendizado de máquina, neste trabalho foram utilizadas redes neurais recorrentes para a extração das características de cada amostra das sequências de palavras, mais especificamente as redes *Long-Short Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU). Ao analisar um episódio de uma sequência, essas redes têm a capacidade de agregar características dos episódios anteriores à representação numérica do episódio atual. Neste trabalho, essas redes estão organizadas em duas camadas, uma em cada direção de leitura, agregando assim informações dos episódios anteriores, bem como dos episódios posteriores, para desta maneira gerar representações vetoriais que caracterizem a amostra atual levando em consideração seu contexto. Como método para a classificação das palavras, neste trabalho é utilizado o *Conditional Random Fields* (CRF), o qual tem a característica de fazer uso da classe atribuída ao episódio anterior de uma sequência, agregando assim informação sequencial dos rótulos já atribuídos. Deste modo, as técnicas de Aprendizado de Máquina utilizadas neste trabalho foram adotadas justamente por fazer uso das informações sequenciais encontradas nos textos.

O desafio deste trabalho envolve a avaliação da inclusão de conhecimento externo para melhoria da tarefa de Reconhecimento de Entidades Nomeadas. Para isso, é realizada a comparação dos resultados de dois cenários: utilizando a abordagem proposta com a adição de conhecimento externo; utilizando a arquitetura da abordagem proposta sem a adição de conhecimento externo. Além disso, os resultados da abordagem proposta também foram comparados a outros trabalhos da literatura.

1.1 Objetivos

A realização deste trabalho visou cumprir com os seguintes objetivos:

- **Objetivo Principal:** Propor e implementar uma abordagem para Reconhecimento de Entidades Nomeadas que faça uso de fontes de conhecimento externo integradas a métodos recentes de Aprendizado de Máquina.
- Realizar um estudo sobre as principais abordagens que estão sendo recentemente utilizadas

para a tarefa de Reconhecimento de Entidades Nomeadas, bem como as características e recursos utilizados para essa tarefa;

- Identificar *datasets* comumente utilizados para a tarefa de Reconhecimento de Entidades Nomeadas;
- Avaliar experimentalmente a abordagem proposta, seguindo protocolos e adotando métricas comumente utilizadas pela comunidade de Reconhecimento de Entidades Nomeadas.

1.2 Organização do Texto

A organização deste trabalho se dá da seguinte maneira: o Capítulo 2 aborda a definição de Extração de Informação, além de mostrar os fundamentos teóricos dos métodos utilizados neste trabalho; no Capítulo 3 é feita uma revisão bibliográfica de trabalhos na área de Reconhecimento de Entidades Nomeadas; no Capítulo 4 são apresentadas as bases de dados utilizadas neste trabalho, os repositórios de conhecimento utilizados, e também as métricas para avaliação de resultados; no Capítulo 5 é descrito o método para Reconhecimento de Entidades Nomeadas utilizado neste trabalho, bem como resultados da aplicação desse método, além da discussão sobre estes resultados; e no Capítulo 6 são mostradas as conclusões tiradas a partir deste trabalho, além de mostrar intenções de trabalhos futuros.

6 Conclusões

Este trabalho teve o objetivo de realizar um estudo e propor um método para Extração de Informação, em especial um método comumente usado para a tarefa de Reconhecimento de Entidades Nomeadas e avaliar a introdução de conhecimento externo, tanto na forma de *Gazetteer* quanto na forma de *Knowledge embedding*. Como resultado do estudo foi identificado que boa parte das abordagens que tratam a tarefa de NER como um problema de classificação de sequências.

Uma arquitetura que utiliza os principais métodos de aprendizado de máquina foi identificada e considerada como a base para a arquitetura proposta. Para a avaliação da metodologia proposta, foram identificados os protocolos de avaliação utilizados, os *dataset* e as fontes de conhecimento externo utilizado na literatura.

Para a execução dos experimentos foram selecionados cinco *datasets* com diferentes características e quantidades de amostras. Para as fontes de conhecimento externo foram selecionadas a Freebase e YAGO.

No quesito métodos de Aprendizado de Máquina, trabalhos recentes da comunidade de NER têm utilizado redes neurais e classificadores que captam as informações sequenciais presentes nos textos. Neste contexto, os métodos utilizados foram redes neurais BiLSTM e BiGRU juntamente ao CRF.

A metodologia adotada consiste na execução de cada configuração um total de dez vezes para então se considerar a média dessas dez execuções. Isso é feito pois a inicialização dos modelos envolve fatores estocásticos. Para a avaliação dos resultados, as medidas de Precisão, Abrangência e F1-Score são adotadas pela a comunidade da tarefa de NER, e portanto são utilizadas neste trabalho. Deste modo, são executadas 10 iterações para cada tipo de rede em cada um dos *datasets* selecionados.

A realização dos experimentos utilizando o protocolo experimental descrito resultou em pouca variação da F1-Score para a maioria dos cenários onde as fontes de conhecimento externo foram inseridas. Os resultados destes experimentos poderiam ser melhores caso houvesse sido realizada a otimização dos hiperparâmetros, tais como quantidade de épocas, taxa de aprendizado, entre outros utilizados, pois com os parâmetros utilizados os modelos podem estar sofrendo de

overfitting.

Outro ponto a se considerar são os *Character embeddings* e *External embeddings*, que podem não estar sendo capazes de captar as informações extras que deveriam exprimir. O uso de *Word embeddings* que não capturam eficientemente o contexto também é um grande fator que pode ter levado aos resultados distantes do estado da arte. Por fim, o uso de repositórios de conhecimento mais volumosos e a melhor integração destes repositórios à abordagem proposta pode vir a melhorar o resultados.

Apesar da maioria dos resultados experimentais não apresentarem grandes variações, os valores resultantes dos experimentos no *dataset* OntoNotes5 foram uma exceção, pois na maioria das vezes houve ganhos expressivos e valores altos de desvio padrão. Esses ganhos podem ser explicados pela falta de otimização dos hiperparâmetros, e por ser um *dataset* muito numeroso, foram utilizadas somente 20 épocas a cada iteração, o que pode ter resultado em *underfitting* do resultado.

A utilização do conhecimento externo da maneira como foi proposta neste trabalho não trouxe grandes variações aos resultados para a maioria dos *datasets*, mostrando que a agregação das características provenientes dessas fontes não impactaram os resultados. Porém, novas estratégias para agregar o conhecimento externo podem ser definidas para a melhoria dos resultados obtidos.

Durante a realização dos experimentos deste trabalho, a maior dificuldade foi o tempo treinamento dos modelos. Essa dificuldade se mostrou a principal barreira para se testar novas técnicas e realizar a variação de parâmetros, principalmente na fase final deste trabalho.

6.1 Trabalhos Futuros

A partir dos resultados experimentais obtidos e das dificuldades encontradas durante a realização do trabalho, foram identificados os seguintes pontos de melhoria, que abrem muitos caminhos para trabalhos futuros, sendo esses pontos: i) Melhorar a quantidade e a qualidade de conhecimento externo utilizado; ii) Melhorar a incorporação do conhecimento externo; iii) Uso de *Word embeddings* mais atuais; iv) Otimização de hiperparâmetros.

Em trabalhos futuros, conforme mencionado acima, além de melhorar a qualidade dos repositórios de conhecimento utilizados, tem-se a intenção de fazer uso de mais fontes de

conhecimento externo aplicado à tarefa de NER, por meio da combinação das informações vindas de diferentes repositórios de conhecimento. Como estratégias para essa combinação de vetores obtidos a partir de fontes de conhecimento distintas tem-se: a concatenação desses vetores; a adição desses vetores; a multiplicação desses vetores; redes neurais para junção desses vetores.

Deseja-se ainda explorar outras estratégias para a incorporação de *Knowledge embeddings* aos vetores de características de cada amostra. Uma estratégia para isso é alterar o momento que ocorre a concatenação dos 4 vetores de características. Outra maneira de realizar essa incorporação é passar a fazer uso de um mecanismo de atenção. Além disso, propor uma tática para fazer uso dos vetores que correspondem a mais de uma palavra no *Knowledge embedding*, o que não foi feito neste trabalho.

No quesito de alcançar resultados comparáveis ao estado da arte, o uso de técnicas mais recentes de *Word embeddings* mostra-se como caminho um caminho promissor. Como exemplo de outra representação vetorial para as palavras pode-se citar as representações contextuais obtidas por modelos que fazem uso de *transformers*, que têm ganhado espaço por apresentarem melhoras de resultados em diversas tarefas de PLN, principalmente ao se realizar a adaptação das representações de contexto geral ao domínio trabalhado, por meio da continuação do treinamento desses modelos. Além do uso dessas técnicas mais recentes, há também a intenção de se avaliar combinações de *Word embeddings* gerados por diferentes métodos.

Em relação aos hiperparâmetros das técnicas de Aprendizado de Máquina utilizados para os experimentos há bastante espaço para melhorias. Os valores utilizados neste trabalho foram adotados com base em trabalhos que usam modelos similares, porém ao se fazer alterações a esses modelos, os parâmetros que eram utilizados deixam de ser ótimos. Com isso mente, há a intenção de realizar uma etapa de otimização de hiperparâmetros, de preferência de maneira automática, para as futuras abordagens.

Referências

- 1 MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. [S.l.]: Cambridge University Press, 2008.
- 2 JURAFSKY, D.; MARTIN, J. H. *Speech and language processing an introduction to natural language processing, computational linguistics, and speech recognition*. Não Publicado. 2018.
- 3 MAYNARD KALINA BONTICHEVA, I. A. D. Natural language processing for the semantic web. In: _____. 1. ed. San Rafael, USA: Morgan & Claypool, 2016. p. 184.
- 4 GUTIERREZ, F.; DOU, D.; FICKAS, S.; WIMALASURIYA, D.; ZONG, H. A hybrid ontology-based information extraction system. *Journal of Information Science*, v. 42, n. 6, p. 798–820, 2016. Disponível em: <<https://doi.org/10.1177/0165551515610989>>.
- 5 MARTINEZ-RODRIGUEZ, J. L.; HOGAN, A.; LOPEZ-AREVALO, I. Information extraction meets the semantic web: A survey. *Semantic Web*, p. 1–81, 10 2018.
- 6 GOYAL, A.; GUPTA, V.; KUMAR, M. Recent named entity recognition and classification techniques: A systematic review. *Computer Science Review*, v. 29, p. 21–43, 08 2018.
- 7 GORINSKI, P. J.; WU, H.; GROVER, C.; TOBIN, R.; TALBOT, C.; WHALLEY, H.; SUDLOW, C.; WHITELEY, W.; ALEX, B. Named entity recognition for electronic health records: A comparison of rule-based and machine learning approaches. *CoRR*, abs/1903.03985, 2019. Disponível em: <<http://arxiv.org/abs/1903.03985>>.
- 8 YADAV, V.; BETHARD, S. A survey on recent advances in named entity recognition from deep learning models. In: *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. p. 2145–2158. Disponível em: <<https://www.aclweb.org/anthology/C18-1182>>.
- 9 RAIS, M.; LACHKAR, A.; LACHKAR, A.; OUATIK, S. E. A. A comparative study of biomedical named entity recognition methods based machine learning approach. In: *2014 Third IEEE International Colloquium in Information Science and Technology (CIST)*. [S.l.: s.n.], 2014. p. 329–334. ISSN 2327-1884.
- 10 SEYLER, D.; DEMBELOVA, T.; CORRO, L. D.; HOFFART, J.; WEIKUM, G. A study of the importance of external knowledge in the named entity recognition task. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, Australia: Association for Computational Linguistics, 2018. p. 241–246. Disponível em: <<https://www.aclweb.org/anthology/P18-2039>>.
- 11 FREIRE, N.; BORBINHA, J.; CALADO, P. An approach for named entity recognition in poorly structured data. In: SIMPERL, E.; CIMIANO, P.; POLLERES, A.; CORCHO, O.; PRESUTTI, V. (Ed.). *The Semantic Web: Research and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 718–732. ISBN 978-3-642-30284-8.
- 12 NADEAU, D.; SEKINE, S. A survey of named entity recognition and classification. *Linguisticae Investigationes*, v. 30, n. 1, p. 3–26, January 2007. Publisher: John Benjamins Publishing Company. Disponível em: <<http://www.ingentaconnect.com/content/jbp/li/2007/00000030/00000001/art00002>>.

- 13 LI, J.; SUN, A.; HAN, J.; LI, C. A survey on deep learning for named entity recognition. *CoRR*, abs/1812.09449, 2018. Disponível em: <<http://arxiv.org/abs/1812.09449>>.
- 14 GRISHMAN, R.; SUNDHEIM, B. Message understanding conference-6: A brief history. In: *Proceedings of the 16th Conference on Computational Linguistics - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics, 1996. (COLING '96), p. 466–471. Disponível em: <<https://doi.org/10.3115/992628.992709>>.
- 15 DODDINGTON, G.; MITCHELL, A.; PRZYBOCKI, M.; RAMSHAW, L.; STRASSEL, S.; WEISCHEDEL, R. The automatic content extraction (ACE) program – tasks, data, and evaluation. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon, Portugal: European Language Resources Association (ELRA), 2004. Disponível em: <<http://www.lrec-conf.org/proceedings/lrec2004/pdf/5.pdf>>.
- 16 SANG, E. F. T. K.; MEULDER, F. D. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In: *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003. (CONLL '03), p. 142–147. Disponível em: <<https://doi.org/10.3115/1119176.1119195>>.
- 17 RAHEM, K.; OMAR, N. Drug-related crime information extraction and analysis. *Proceedings of the 6th International Conference on Information Technology and Multimedia*, p. 250–254, 2014.
- 18 SAHA, S. K.; SARKAR, S.; MITRA, P. Feature selection techniques for maximum entropy based biomedical named entity recognition. *Journal of Biomedical Informatics*, v. 42, n. 5, p. 905 – 911, 2009. ISSN 1532-0464. Biomedical Natural Language Processing. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046409000033>>.
- 19 WANG, Y.; YU, Z.; CHEN, L.; CHEN, Y.; LIU, Y.; HU, X.; JIANG, Y. Supervised methods for symptom name recognition in free-text clinical records of traditional chinese medicine: An empirical study. *Journal of Biomedical Informatics*, v. 47, p. 91 – 104, 2014. ISSN 1532-0464. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046413001494>>.
- 20 KERETNA, S.; LIM, C. P.; CREIGHTON, D.; SHABAN, K. B. Enhancing medical named entity recognition with an extended segment representation technique. *Computer Methods and Programs in Biomedicine*, v. 119, n. 2, p. 88 – 100, 2015. ISSN 0169-2607. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169260715000425>>.
- 21 CHEN, Y.; LASKO, T. A.; MEI, Q.; DENNY, J. C.; XU, H. A study of active learning methods for named entity recognition in clinical text. *Journal of Biomedical Informatics*, v. 58, p. 11 – 18, 2015. ISSN 1532-0464. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046415002038>>.
- 22 KORKONTZELOS, I.; PILIOURAS, D.; DOWSEY, A. W.; ANANIADOU, S. Boosting drug named entity recognition using an aggregate classifier. *Artificial Intelligence in Medicine*, v. 65, n. 2, p. 145 – 153, 2015. ISSN 0933-3657. Intelligent healthcare informatics in big data era. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0933365715000780>>.
- 23 BHASURAN, B.; MURUGESAN, G.; ABDULKADHAR, S.; NATARAJAN, J. Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases. *Journal of Biomedical Informatics*, v. 64, p. 1 – 9, 2016. ISSN 1532-0464. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046416301216>>.

- 24 MAJUMDER, M.; BARMAN, U.; PRASAD, R.; SAURABH, K.; SAHA, S. K. A novel technique for name identification from homeopathy diagnosis discussion forum. *Procedia Technology*, v. 6, p. 379–386, 12 2012.
- 25 GUO, J.; XU, G.; CHENG, X.; LI, H. Named entity recognition in query. In: *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2009. (SIGIR '09), p. 267–274. ISBN 978-1-60558-483-6. Disponível em: <<http://doi.acm.org/10.1145/1571941.1571989>>.
- 26 YILDIZ, B.; MIKSCH, S. Motivating ontology-driven information extraction. In: _____. [S.l.: s.n.], 2011. p. 1–19. ISBN 978-981-4307-25-3.
- 27 KIM, J.-H.; WOODLAND, P. A rule-based named entity recognition system for speech input. In: . [S.l.: s.n.], 2000. p. 528–531.
- 28 FELLBAUM, C. (Ed.). *WordNet: an electronic lexical database*. [S.l.]: MIT Press, 1998.
- 29 MIKOLOV, T.; SUTSKEVER, I.; CHEN, K.; CORRADO, G.; DEAN, J. Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546, 2013. Disponível em: <<http://arxiv.org/abs/1310.4546>>.
- 30 PENNINGTON, J.; SOCHER, R.; MANNING, C. D. Glove: Global vectors for word representation. In: *Empirical Methods in Natural Language Processing (EMNLP)*. [s.n.], 2014. p. 1532–1543. Disponível em: <<http://www.aclweb.org/anthology/D14-1162>>.
- 31 HAYKIN, S. S. *Neural networks and learning machines*. Third. Upper Saddle River, NJ: Pearson Education, 2009.
- 32 KIM, Y. Convolutional neural networks for sentence classification. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014. p. 1746–1751. Disponível em: <<https://www.aclweb.org/anthology/D14-1181>>.
- 33 HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, v. 9, p. 1735–80, 12 1997.
- 34 CHO, K.; MERRIENBOER, B. van; BAHDANAU, D.; BENGIO, Y. On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259, 2014. Disponível em: <<http://arxiv.org/abs/1409.1259>>.
- 35 CHUNG, J.; GÜLÇEHRE, Ç.; CHO, K.; BENGIO, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555, 2014. Disponível em: <<http://arxiv.org/abs/1412.3555>>.
- 36 JOZEFOWICZ, R.; ZAREMBA, W.; SUTSKEVER, I. An empirical exploration of recurrent network architectures. In: *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37. JMLR.org*, 2015. (ICML'15), p. 2342–2350. Disponível em: <<http://dl.acm.org/citation.cfm?id=3045118.3045367>>.
- 37 SCHUSTER, M.; PALIWAL, K. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*, v. 45, p. 2673 – 2681, 12 1997.

- 38 LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. (ICML '01), p. 282–289. ISBN 1-55860-778-1. Disponível em: <<http://dl.acm.org/citation.cfm?id=645530.655813>>.
- 39 LIU, K.; EL-GOHARY, N. Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports. *Automation in Construction*, v. 81, 05 2017.
- 40 NOTHMAN, J.; RINGLAND, N.; RADFORD, W.; MURPHY, T.; CURRAN, J. R. Learning multilingual named entity recognition from wikipedia. *Artificial Intelligence*, v. 194, p. 151 – 175, 2013. ISSN 0004-3702. Artificial Intelligence, Wikipedia and Semi-Structured Resources. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0004370212000276>>.
- 41 ENDARNOTO, S. K.; PRADIPTA, S.; NUGROHO, A. S.; PURNAMA, J. Traffic condition information extraction amp; visualization from social media twitter for android mobile application. In: *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*. [S.l.: s.n.], 2011. p. 1–4. ISSN 2155-6830.
- 42 NURDIN, A.; MAULIDEVI, N. U. 5w1h information extraction with cnn-bidirectional lstm. *Journal of Physics: Conference Series*, v. 978, n. 1, p. 012078, 2018. Disponível em: <<http://stacks.iop.org/1742-6596/978/i=1/a=012078>>.
- 43 LAMPLE, G.; BALLESTEROS, M.; SUBRAMANIAN, S.; KAWAKAMI, K.; DYER, C. Neural architectures for named entity recognition. *CoRR*, abs/1603.01360, 2016. Disponível em: <<http://arxiv.org/abs/1603.01360>>.
- 44 LANGE, D.; BÖHM, C.; NAUMANN, F. Extracting structured information from wikipedia articles to populate infoboxes. In: *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2010. (CIKM '10), p. 1661–1664. ISBN 978-1-4503-0099-5. Disponível em: <<http://doi.acm.org/10.1145/1871437.1871698>>.
- 45 LEK, H. H.; POO, D. C. C. An experimental study to investigate the use of additional classifiers to improve information extraction accuracy. In: *2011 10th International Conference on Machine Learning and Applications and Workshops*. [S.l.: s.n.], 2011. v. 1, p. 412–415.
- 46 RAHEM, K. R.; OMAR, N. Drug-related crime information extraction and analysis. In: *Proceedings of the 6th International Conference on Information Technology and Multimedia*. [S.l.: s.n.], 2014. p. 250–254.
- 47 FURTADO, P. H. T. *Interpretação automática de relatórios de operação de equipamentos*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro - PUC-RIO, 2017.
- 48 CHIU, J. P.; NICHOLS, E. Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, v. 4, p. 357–370, 2016. Disponível em: <<https://www.aclweb.org/anthology/Q16-1026>>.
- 49 AMARAL, D. O. F. d. *Reconhecimento de entidades nomeadas na ?rea da geologia : bacias sedimentares brasileiras*. Tese (Doutorado) — Programa de Pós-Graduação em Ciência da Computação, 2017. Escola Politécnica. Disponível em: <<http://tede2.pucrs.br/tede2/handle/tede/8035>>.

- 50 HABIBI, M.; WEBER, L.; NEVES, M.; WIEGANDT, D. L.; LESER, U. Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics*, v. 33, n. 14, p. i37–i48, 07 2017. ISSN 1367-4803. Disponível em: <<https://doi.org/10.1093/bioinformatics/btx228>>.
- 51 LING, X.; WELD, D. Fine-grained entity recognition. 2012. Disponível em: <<https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/5152/5124>>.
- 52 LIU, T.; YAO, J.-G.; LIN, C.-Y. Towards improving neural named entity recognition with gazetteers. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019. p. 5301–5307. Disponível em: <<https://www.aclweb.org/anthology/P19-1524>>.
- 53 JIE, Z.; LU, W. *Dependency-Guided LSTM-CRF for Named Entity Recognition*. 2019.
- 54 LOUVAN, S.; MAGNINI, B. Leveraging non-conversational tasks for low resource slot filling: Does it help? In: *SIGdial*. [S.l.: s.n.], 2019.
- 55 Liu, J.; Pasupat, P.; Cyphers, S.; Glass, J. Asgard: A portable architecture for multilingual dialogue systems. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 2013. p. 8386–8390.
- 56 BAEVSKI, A.; EDUNOV, S.; LIU, Y.; ZETTLEMOYER, L.; AULI, M. Cloze-driven pretraining of self-attention networks. *CoRR*, abs/1903.07785, 2019. Disponível em: <<http://arxiv.org/abs/1903.07785>>.
- 57 LI, X.; SUN, X.; MENG, Y.; LIANG, J.; WU, F.; LI, J. Dice loss for data-imbalanced nlp tasks. In: *ACL*. [S.l.: s.n.], 2020.
- 58 ZELDES, A. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, v. 51, n. 3, p. 581–612, 2017.
- 59 BORDES, A.; USUNIER, N.; GARCIA-DURÁN, A.; WESTON, J.; YAKHNENKO, O. Translating embeddings for modeling multi-relational data. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*. Red Hook, NY, USA: Curran Associates Inc., 2013. (NIPS' 13), p. 2787–2795.
- 60 HAN, X.; CAO, S.; XIN, L.; LIN, Y.; LIU, Z.; SUN, M.; LI, J. Openke: An open toolkit for knowledge embedding. In: *Proceedings of EMNLP*. [S.l.: s.n.], 2018.
- 61 CHINCHOR, N.; SUNDHEIM, B. MUC-5 evaluation metrics. In: *Fifth Message Understanding Conference (MUC-5): Proceedings of a Conference Held in Baltimore, Maryland, August 25-27, 1993*. [s.n.], 1993. Disponível em: <<https://www.aclweb.org/anthology/M93-1007>>.
- 62 TURIAN, J.; RATINOV, L.; BENGIO, Y. Word representations: A simple and general method for semi-supervised learning. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. (ACL '10), p. 384–394. Disponível em: <<http://dl.acm.org/citation.cfm?id=1858681.1858721>>.
- 63 DOZAT, T. Incorporating nesterov momentum into adam. In: . [S.l.: s.n.], 2015.

-
- 64 RODRÍGUEZ, J. D.; CALDWELL, A.; LIU, A. Transfer learning for entity recognition of novel classes. In: *COLING*. [S.l.: s.n.], 2018.
- 65 HOU, Y.; CHE, W.; LAI, Y.; ZHOU, Z.; LIU, Y.; LIU, H.; LIU, T. Few-shot slot tagging with collapsed dependency transfer and label-enhanced task-adaptive projection network. In: *ACL*. [S.l.: s.n.], 2020.
- 66 MULLER, P.; BRAUD, C.; MOREY, M. Tony: Contextual embeddings for accurate multilingual discourse segmentation of full documents. In: . [S.l.: s.n.], 2019.
- 67 PRANGE, J.; SCHNEIDER, N.; ABEND, O. Semantically constrained multilayer annotation: The case of coreference. *ArXiv*, abs/1906.00663, 2019.