

1. Record Nr.	UNISA996464451003316
Titolo	Computational processing of the Portuguese language : 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21-23, 2022, proceedings / / edited by Vladia Pinheiro
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2022] ©2022
ISBN	3-030-98305-6
Descrizione fisica	1 online resource (447 pages)
Collana	Lecture Notes in Computer Science ; ; v.13208
Disciplina	469.0285635
Soggetti	Computational linguistics
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- Contents -- Resources and Evaluation -- UlyssesNER-Br: A Corpus of Brazilian Legislative Documents for Named Entity Recognition -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Semantic Classes -- 3.2 Annotation Process -- 4 The UlyssesNER-Br Corpus -- 4.1 PL-corpus -- 4.2 ST-corpus -- 4.3 Evaluation -- 4.4 Results and Discussion -- 5 Conclusion and Future Works -- References -- A Test Suite for the Evaluation of Portuguese-English Machine Translation -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Creation of the Test Suite -- 3.2 Limitations of the Method -- 3.3 Experimental Setup -- 4 Findings -- 4.1 Overall Performance of MT Systems -- 4.2 BLEU vs. Test Suite Scores -- 4.3 Categories -- 4.4 Phenomena -- 4.5 Qualitative Analysis -- 5 Conclusion -- References -- MINT - Mainstream and Independent News Text Corpus -- 1 Introduction -- 2 Related Work -- 3 Corpus Organization -- 3.1 MINT-articles -- 3.2 MINT-annotations -- 4 Corpus Characterization -- 4.1 Linguistic Characterization -- 4.2 Insights from Crowdsourced Annotations -- 5 Conclusion -- References -- Fakepedia Corpus: A Flexible Fake News Corpus in Portuguese -- 1 Introduction -- 2 Related Work -- 3 Our Proposal: Fakepedia Corpus -- 4 Experiments and Results -- 5 Conclusion -- References -- A Targeted Assessment of the Syntactic Abilities of Transformer Models for Galician-Portuguese -- 1 Introduction -- 2

Related Work -- 3 Materials and Methods -- 3.1 Experiments and Data -- 3.2 Models -- 3.3 Evaluation -- 4 Results and Discussion -- 5 Conclusions and Further Work -- References -- FakeRecogna: A New Brazilian Corpus for Fake News Detection -- 1 Introduction -- 2 Related Works -- 3 FakeRecogna Corpus -- 4 Methodology -- 4.1 Pre-processing -- 4.2 Text Representation -- 4.3 Classifiers -- 4.4 Evaluating Measures -- 4.5 Additional Experiments.

5 Experimental Results -- 5.1 No Removal of Words -- 5.2 Augmentation Study -- 6 Conclusions and Future Works -- References -- Implicit Opinion Aspect Clues in Portuguese Texts: Analysis and Categorization -- 1 Introduction -- 2 Related Work -- 3 Data and Methods -- 3.1 Methods -- 3.2 Datasets -- 3.3 Identification of IACs -- 3.4 Lexicons of IACs -- 3.5 Categorization of IACs -- 4 Results -- 5 Final Remarks -- References -- CRPC-DB a Discourse Bank for Portuguese -- 1 Introduction -- 2 Related Work -- 3 The CRPC-DB -- 3.1 Raw Corpus and Pre-processing -- 3.2 Annotation Scheme -- 3.3 Annotation Process -- 4 Inter-Annotator Agreement Experiment -- 5 Final Remarks -- References -- Challenges in Annotating a Treebank of Clinical Narratives in Brazilian Portuguese -- 1 Introduction -- 2 Related Work -- 3 Materials and Methods -- 3.1 Data Preparation -- 3.2 Corpus Characteristics -- 3.3 Decisions Made in the Annotation Process -- 4 Results -- 5 Discussion -- 6 Conclusion -- References -- PetroBERT: A Domain Adaptation Language Model for Oil and Gas Applications in Portuguese -- 1 Introduction -- 2 Related Works -- 3 Proposed Work -- 4 Experimental Evaluation and Discussion -- 4.1 Experiment I - NER -- 4.2 Experiment II - Sentence Classification -- 5 Conclusions and Future Works -- References -- SS-PT: A Stance and Sentiment Data Set from Portuguese Quoted Tweets -- 1 Introduction -- 2 Related Work -- 3 Corpus -- 3.1 Data Collection -- 3.2 Guidelines -- 3.3 Annotation Procedure -- 3.4 Balancing the Corpus -- 3.5 Descriptive Statistics -- 4 Baseline Experiment -- 5 Conclusion -- References -- Natural Language Processing Tasks -- ZeroBERTo: Leveraging Zero-Shot Text Classification by Topic Modeling -- 1 Introduction -- 2 Background and Related Work -- 3 Proposed Method -- 3.1 0shot-TC Task Formalization -- 3.2 ZeroBERTo -- 4 Experiments.

5 Discussion and Future Work -- References -- Banking Regulation Classification in Portuguese -- 1 Introduction -- 2 Related Works -- 3 The Application -- 3.1 The Corpus -- 3.2 The Architecture -- 4 Discussion, Results and Future Work -- 5 Conclusions -- References -- Automatic Information Extraction: A Distant Reading of the Brazilian Historical-Biographical Dictionary -- 1 Introduction -- 2 Information Extraction -- 3 Methodology -- 4 Extraction Evaluation -- 5 Distant Reading DHBB -- 6 Final Considerations -- References -- Automatic Recognition of Units of Measurement in Product Descriptions from Tax Invoices Using Neural Networks -- 1 Introduction -- 2 Related Works -- 3 Methodology -- 3.1 Materials -- 3.2 Dataset -- 3.3 Data Preparation -- 3.4 Training and Test -- 4 Results and Discussion -- 4.1 General Analysis -- 4.2 Analysis of Errors -- 5 Conclusions -- References -- Entity Extraction from Portuguese Legal Documents Using Distant Supervision -- 1 Introduction -- 2 Related Work -- 3 Entity Extraction System -- 4 Experimental Results -- 4.1 Dataset and Evaluation Metric -- 4.2 Best Model Results -- 4.3 DAM: Role-Specific Threshold -- 4.4 Assessment of DAM Components -- 5 Conclusion -- References -- Sexist Hate Speech: Identifying Potential Online Verbal Violence Instances -- 1 Introduction -- 2 Background -- 2.1 The Campos Mello Case -- 3 The Linguistic-Computational Interface -- 4 Computational Approaches to Support Hate Speech Identification

Through Linguistic Characteristics -- 4.1 Fallacies in Intolerant Speech -- 5 Final Remarks -- References -- Book Genre Classification Based on Reviews of Portuguese-Language Literature -- 1 Introduction -- 2 Related Work -- 3 Genre Classification Methodology -- 3.1 Data -- 3.2 Reviews Preprocessing -- 3.3 Book Genre Classification -- 3.4 Evaluation -- 4 Experiments and Results. 5 Conclusion and Future Work -- References -- Combining Word Embeddings for Portuguese Named Entity Recognition -- 1 Introduction -- 2 Related Work -- 3 Resources -- 3.1 Corpora -- 3.2 Word Embedding Models -- 3.3 NER Classification Model -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Experimental Results -- 5 Conclusion -- References -- BERT for Sentiment Analysis: Pre-trained and Fine-Tuned Alternatives -- 1 Introduction -- 2 Related Work -- 3 Datasets -- 4 Models -- 4.1 Pre-trained BERT -- 4.2 Fine-Tuned BERT -- 5 Results and Discussion -- 5.1 Pre-trained BERT -- 5.2 Fine-Tuned BERT -- 5.3 Cross-model Comparison -- 6 Conclusion -- References -- Fostering Judiciary Applications with New Fine-Tuned Models for Legal Named Entity Recognition in Portuguese -- 1 Introduction -- 2 Related Work -- 3 Materials and Methods -- 3.1 Fine-Tuned Legal NER -- 3.2 Prototype Application -- 4 Results and Discussion -- 4.1 Fine-Tuned Legal NER -- 4.2 Prototype Application -- 5 Conclusion -- References -- Natural Language Processing Applications -- Using Topic Modeling in Classification of Brazilian Lawsuits -- 1 Introduction -- 2 Related Works -- 3 Corpus and Data Preparation -- 3.1 Corpus and Golden Collection -- 3.2 Integration with the Brazilian Legal Knowledge Graph -- 4 Topic Modeling in Legal Documents -- 4.1 Pre-processing -- 4.2 Topic Generation -- 4.3 Converting Topics to Feature Vectors -- 5 Experiments -- 5.1 Experimental Setup -- 5.2 Models -- 5.3 Experimental Results and Analysis -- 6 Conclusions -- References -- PortNOIE: A Neural Framework for Open Information Extraction for the Portuguese Language -- 1 Introduction -- 2 Related Work -- 3 PortNOIE -- 3.1 Problem Definition -- 3.2 Architecture -- 4 Experiments -- 4.1 Datasets -- 4.2 Experimental Design -- 4.3 Results -- 4.4 Ablation and Discussion -- 5 Conclusion and Future Work -- References. Tracking Environmental Policy Changes in the Brazilian Federal Official Gazette -- 1 Introduction -- 2 Methods -- 2.1 Data Preparation -- 2.2 Experiment Description -- 3 Results -- 4 Conclusion and Future Work -- References -- A Transfer Learning Analysis of Political Leaning Classification in Cross-domain Content -- 1 Introduction -- 2 Related Work -- 3 Data Collection -- 4 Experiments -- 4.1 Congressional Speeches Classification -- 4.2 Transfer Learning Classification -- 4.3 Transfer Learning Decay over Time -- 5 Discussion -- 6 Limitations and Future Work -- References -- Integrating Question Answering and Text-to-SQL in Portuguese -- 1 Introduction -- 2 Background and Tools -- 3 Proposed Architecture -- 4 Question Answering Datasets -- 5 Experiments -- 5.1 Classifier -- 5.2 Question Answering Reasoner -- 6 Results and Analyses -- 7 Conclusion -- References -- Named Entity Extractors for New Domains by Transfer Learning with Automatically Annotated Data -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Datasets -- 3.2 BERT-Based Classifiers for Entity Detection -- 4 Results -- 5 Conclusion -- 5.1 Future Work -- References -- PTT5-Paraphraser: Diversity and Meaning Fidelity in Automatic Portuguese Paraphrasing -- 1 Introduction -- 2 Related Work -- 3 The PTT5-Paraphraser -- 4 Evaluating Paraphrasers -- 4.1 Evaluation by Computational Metrics -- 4.2 Human Evaluation -- 5 Data Augmentation Experiment -- 6 Conclusions -- References -- Speech Processing and Applications -- A Protocol for Comparing Gesture and

Prosodic Boundaries in Multimodal Corpora -- 1 Gesture and Prosody
Alignment Background -- 1.1 The Alignment and Its Types -- 1.2 The
Language into Act Theory -- 1.3 BGEST Corpus Overview -- 2 Script
Outline -- 3 Results and Discussion -- References -- Forced Phonetic
Alignment in Brazilian Portuguese Using Time-Delay Neural Networks.
1 Introduction.
