

Hugo Gonçalo Oliveira, Livy Real, and Erick Fonseca (Eds.)

Proceedings of the

ASSIN 2 Shared Task
Evaluating Semantic Textual Similarity
and Textual Entailment in Portuguese

Copyright © 2020 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0
International (CC BY 4.0).

Preface

ASSIN 2 is the second edition of the Evaluation of Semantic Similarity and Textual Inference (*Avaliação de Similaridade Semântica e Inferência textual*) in Portuguese, that took place as a parallel event with the STIL conference in 2019. Like its previous edition, it proposed a shared task on Semantic Similarity and Text Entailment; with the former ranking pairs of sentences from 1 to 5, and the latter labeling them as either entailment or non-entailment (but not paraphrases, in contrast with the first edition).

There are some notable differences between the first and second edition of the shared task. Concerning the data, a new corpus of 10 thousand sentences was presented, but instead of text extracted from news articles, it contains much simpler sentence pairs, modeled after the SICK corpus. With sentences written on purpose for this task, some linguistic phenomena could be directly controlled. As a result, a word overlap baseline is not so powerful on ASSIN 2 as it was on ASSIN 1.

On the side of systems, we saw a reflection of the recent development of neural networks. While hand-engineered features and lexical resources are still useful, pretrained language models proved themselves as very helpful for both tasks evaluated.

This volume presents the main findings of the shared task organizers, and the descriptions of the strategies developed by the participants. With a total of nine of them, we are happy with the results of ASSIN 2. We leave a new dataset as a benchmark to evaluate the progress of this area in Portuguese, as well as the reflections upon its research directions.

February, 2020

Erick Fonseca
Livy Real
Hugo Gonçalo Oliveira

Organization

Livy Real	B2W Digital/Grupo de Linguística Computacional – USP, Brazil
Hugo Gonçalo Oliveira	CISUC / DEI, Universidade de Coimbra, Portugal
Erick Fonseca	Instituto de Telecomunicações, Lisboa, Portugal

Reviewers

Ana Alves	CISUC and ISEC, Polytechnic Institute of Coimbra, Portugal
Evandro Fonseca	Stilingue, Brazil
Irene Rodrigues	Laboratório de Informática, Sistemas e Paralelismo (LISP), Departamento de Informática, Universidade de Évora
Jéssica Rodrigues	Department of Computer Science, Federal Univer- sity of São Carlos, Brazil
Lucas Oliveira	Graduate Program on Health Technology (PPGTS), Pontifical Catholic University of Paraná (PUCPR). Curitiba, Brazil
Marco Sobrevilla Cabezudo	NILC - Interinstitutional Center for Computational Linguistics, ICMC, Universidade de São Paulo, São Carlos SP, Brazil
Nádia Félix Felipe da Silva	Institute of Informatics, Federal University of Goiás, Brazil
Rui Rodrigues	Centro de Matemática e Aplicações (CMA), FCT, UNL Departamento de Matemática, FCT, UNL
Valeria de Paiva	Samsung Research America, USA

Table of Contents

Organizing the ASSIN 2 Shared Task	1
<i>Livy Real, Erick Fonseca, Hugo Gonçalo Oliveira</i>	
ASAPPpy: a Python Framework for Portuguese STS	14
<i>José Santos, Ana Alves, Hugo Gonçalo Oliveira</i>	
Multilingual Transformer Ensembles for Portuguese Natural Language Tasks	27
<i>Ruan Chaves Rodrigues, Jéssica Rodrigues da Silva, Pedro Vitor Quinta de Castro, Nádia Félix Felipe da Silva, Anderson da Silva Soares</i>	
IPR: The Semantic Textual Similarity and Recognizing Textual Entailment systems	39
<i>Rui Rodrigues, Paula Couto, Irene Rodrigues</i>	
NILC at ASSIN 2: Exploring Multilingual Approaches	49
<i>Marco A. Sobrevilla Cabezudo, Marcio Inácio, Ana Carolina Rodrigues, Edresson Casanova, Rogério Figueredo de Sousa</i>	
Incorporating multiple feature groups to a Siamese Neural Network for Semantic Textual Similarity task in Portuguese texts	59
<i>João Vitor Andrioli de Souza, Lucas Emanuel Silva e Oliveira, Yohan Bonescki Gumiel, Deborah Ribeiro Carvalho, Claudia Maria Cabral Moro</i>	
Multilingual Transformer Ensembles for Portuguese Natural Language Tasks	68
<i>Evandro Fonseca, João Paulo Reis Alvarenga</i>	