BETTER | PUBLICLY AVAILABLE PROGRAM ARTIFACTS

Intelligence Advanced Research Projects Activity
IARPA
Creating Advantage through Research and Technology

TITLE	DESCRIPTION	CREATOR	REFERENCE	LOCATION
ALPS (Active Learning through Processing Surprisal)	An active learning algorithm that is based on the language modeling objective, and that strives to reduce annotation costs by choosing the most critical examples to label.	JHU	*https://arxiv.org/abs/2010.09535*	*https://github.com/forest-snow/alps*
An Empirical Study of Pre-trained Transformers for Arabic Information Extraction	A pre-trained bilingual BERT encoder that is designed specifically for Arabic NLP and English-to-Arabic zero-shot transfer learning. GigaBERT's effectiveness on zero-shot transfer was studied across four IE tasks: named entity recognition, part-of-speech tagging, argument role labeling, and relation extraction. The best model significantly outperforms mBERT and XLM-RoBERTa. GigaBERT has been used as the fundamental multilingual encoder for Brown's BETTER phrase one evaluation.	Brown (Georgia Tech)	*https://arxiv.org/abs/2004.14519*	*https://github.com/lanwuwei/GigaBERT*
Arabic denormalization toolkit	A tool for training a simple sequence-to-sequence model where tokenized data is fed into the model and proper Arabic is generated.	JHU		*https://github.com/KentonMurray/ArabicDetokenizert*
Concept Wikification for COVID-19	Understanding scientific articles related to COVID-19 requires broad knowledge about concepts such as symptoms, diseases, and medicine. Given the very large and ever-growing scientific articles related to COVID19, it is a daunting task even for experts to recognize the large set of concepts mentioned in these articles. We address the problem of concept wikification for COVID19, which is to automatically recognize mentions of concepts related to COVID-19 in text and resolve them into Wikipedia titles. We develop an approach to curate a COVID-19 concept wikification dataset by mining Wikipedia text and the associated intra-Wikipedia links. We also develop an end-to-end system for concept wikification for COVID-19.	BBN	*https://aclanthology.org/2020.nlp-covid19-2.29.pdf*	*https://github.com/panlybero/Covid19_wikification*
COVID-19 Event Extraction Dataset and Demo	A manually annotated corpus of 10,000 tweets containing public reports of five COVID-19 events, including positive and negative tests, deaths, denied access to testing, claimed cures and preventions. Slot-filling questions were designed for each event type, and a total of 31 fine-grained slots were annotated on such topics as the location of events, recent travel, and close contacts. Brown's corpus can support fine-tuning BERT-based classifiers to automatically extract publicly reported events and help track the spread of a new disease. And by aggregating events extracted from millions of tweets, surprisingly high precision is achieved when answering complex queries, such as ``Which organizations have employees that tested positive in Philadelphia?''	Brown (Georgia Tech)		**Demo:** *http://kb1.cse.ohio-state.edu:8000/covid19/positive/* **Data:** *https://github.com/viczong/extract_COVID19_events_from_Twitter*
COVID-19 QA Dataset	We release a dataset of over 2,200 COVID-19 related Frequently asked Question-Answer pairs scraped from over 40 trusted websites. We include an additional 24,000 questions pulled from online sources that have been aligned by experts with existing answered questions from our dataset.m	JHU	*https://openreview.net/pdf?id=GR03UfD2OZk*	**Data:** *https://covid-19-infobot.org/data/* **Website:** *https://covid-19-infobot.org*
CrossFit: A Few-Shot Learning Challenge for Cross-Task Generalization in NLP	A task setup that aims to build few-shot learners that generalize across diverse NLP tasks. Explores whether models trained with non-classification tasks becomes good few-shot learner for classfication tasks; and whether models trained with non-MRC QA tasks become good few-shot learners for MRC QA tasks.	INK-USC	*https://arxiv.org/abs/2104.08835*	*https://github.com/INK-USC/CrossFit*
Cross-lingual transfer learning and multilingual NLP	Proposal for a new contrastive alignment objective for zero-shot cross-lingual transfer learning Data augmentation (data projection and self-training) for BETTER: ACE, NER and POS tagging, and dependency parsing	JHU	*https://arxiv.org/abs/2109.06798* *https://arxiv.org/abs/2010.02537* *https://arxiv.org/abs/2005.09093*	*https://github.com/shijie-wu/crosslingual-nlp*
Distantly-Supervised Evidence Retrieval Enables Question Answering without Evidence Annotation: DistDR	DistDR is a novel approach for open-domain question answering that iteratively improves over a weak retriever by alternately finding evidence from the up-to-date model and encouraging the model to learn the most likely evidence.	JHU	*https://arxiv.org/abs/2110.04889*	*https://github.com/henryzhao5852/DistDR*

BETTER | PUBLICLY AVAILABLE PROGRAM ARTIFACTS

Intelligence Advanced Research Projects Activity
IARPA
Creating Advantage through Research and Technology

TITLE	DESCRIPTION	CREATOR	REFERENCE	LOCATION
Everything Is All It Takes: A Multipronged Strategy for Zero-Shot Cross-Lingual Information Extraction (resulting tool) Large English-Arabic Encoder	An English–Arabic bilingual encoder with a 24-layer transformer architecture trained on additional English and Arabic data.	JHU	*https://arxiv.org/abs/2109.06798*	**Download:** *https://huggingface.co/jhu-clsp/roberta-large-eng-ara-128k* **Github:** *https://github.com/kentonmurray/arabicdetokenize* **Github:** *https://github.com/shijie-wu/crosslingual-nlp*
FourIE - Annotator for entities, relations, events and argument roles	FourIE is a neural information extraction system that can annotate texts for entity mentions (names, pronouns, nominals), relations, event triggers, and argument roles using the information schema defined in the ACE 2005 dataset. FourIE leverages deep learning and graph convolutional networks to jointly perform four tasks in information extraction, i.e., entity mention detection, relation extraction, event detection and argument role prediction in an end-to-end fashion. Our system achieves state-of-the-art performance for joint information extraction on ACE 2005.	BBN (U. of Oregon)	*https://arxiv.org/pdf/2103.09330.pdf*	*http://nlp.uoregon.edu/fourie*
Gradual Fine-Tuning for a new domain	Fine-tuning is known to improve NLP models by adapting an initial model trained on more plentiful but less domain-salient examples to data in a target domain. Gradually fine-tuning in a multi-stage process can yield substantial further gains.	JHU	*https://aclanthology.org/2021.adaptnlp-1.22.pdf*	*https://github.com/fe1ixxu/Gradual-Finetune*
Interactive Refinement of Cross-Lingual Word Embeddings: CLIME	A human-in-the-loop system that improves multilingual classifiers. Specifically, CLIME asks a bilingual user to refine pre-trained cross-lingual word embeddings.	JHU	*https://arxiv.org/pdf/1911.03070.pdf*	*https://github.com/forest-snow/clime-ui*
LEAN-LIFE: A Label-Efficient Annotation Framework Towards Learning from Annotator Explanation	A web-based, Label-Efficient AnnotatioN framework for sequence labeling and classification tasks, with an easy-to-use UI that not only allows an annotator to provide the needed labels for a task, but also enables LearnIng From Explanations for each labeling decision.	USC Ink Lab	*https://arxiv.org/abs/2004.07499* *http://inklab.usc.edu/leanlife/*	**Project website:** *http://inklab.usc.edu/leanlife/* **Github:** *https://github.com/INK-USC/LEAN-LIFE* **System demo paper:** *https://arxiv.org/abs/2004.07499*
MadDog: Web-based System for Acronym Identification and Disambiguation	Acronyms and abbreviations are short forms of longer phrases that are ubiquitously employed in various types of writing. Despite efforts from both the research community and software developers, none of the existing works provides a unified and publicly available solution capable of processing acronyms in various domains. Thus, we introduce MadDog, the first web-based acronym identification and disambiguation system which can process acronyms from various domains including scientific, biomedical, and general domains.	U. of Oregon (BBN)	*https://arxiv.org/abs/2101.09893* Paper: *https://aclanthology.org/2021.eacl-demos.20.pdf*	**Github repo (models and deployment code):** *https://github.com/amirveyseh/MadDog* **Demo website:** *http://iq.cs.uoregon.edu:5000/*
MISO for Universal Decompositional Semantic Parsing	MISO stands for Multimodal Inputs, Semantic Outputs. It is a deep learning framework with re-usable components for parsing a variety of semantic parsing formalisms.	JHU		*https://github.com/esteng/miso_uds*
Neural Semi-Markov CRF for Monolingual Word Alignment	A novel neural semi-Markov CRF alignment model that unifies word and phrase alignments through variable-length spans	Brown	*https://arxiv.org/abs/2106.02569*	*https://github.com/chaojiang06/neural-Jacana*
Neural Semi-Markov CRF for Monolingual Word Alignment	Monolingual word alignment is important for studying fine-grained editing operations (i.e., deletion, addition, and substitution) in text-to-text generation tasks, such as paraphrase generation, text simplification, neutralizing biased language, etc. In this paper, we present a novel neural semi-Markov CRF alignment model, which unifies word and phrase alignments through variable-length spans. We also create a new benchmark with human annotations that cover four different text genres to evaluate monolingual word alignment models in more realistic settings. Experimental results show that our proposed model outperforms all previous approaches for monolingual word alignment as well as a competitive QA-based baseline, which was previously only applied to bilingual data. Our model demonstrates good generalizability to three out-of-domain datasets and shows great utility in two downstream applications: automatic text simplification and sentence pair classification tasks.	Brown	*https://arxiv.org/abs/2106.02569*	*https://github.com/chaojiang06/neural-Jacana*

TITLE	DESCRIPTION	CREATOR	REFERENCE	LOCATION
Neural-CRF Word Aligner	A novel neural semi-Markov CRF alignment model that unifies both word and phrase alignments though variable length spans, calculates span-based semantic similarities, and takes alignment label transitions into consideration. It achieves state-of-the-art performance with over 92 F1 in the in-domain evaluation and demonstrates very good generalizability on three out-of-domain datasets.	Brown (Georgia Tech)		*https://github.com/chaojiang06/neural-Jacana*
NLP Research Library	An NLP research library built on PyTorch and Transformers for quick prototyping of individual IE systems. The library contains a complete training pipeline and data processing code for different types of IE tasks, including NER, POS, SRL, and sentence classification. Brown's deployed systems for BETTER evaluation were built based on this NLP research library.	Brown (Georgia Tech)		*https://github.com/edchengg/MyNLP*
Python Interface for CoreNLP Arabic Sentence Splitter	A Python interface for splitting Arabic sentences by calling the Java-based CoreNLP package. It has the advantage of fast speed and satisfactory accuracy, and can be used in the pre-processing of Arabic documents.	Brown (Georgia Tech)		*https://github.com/chaojiang06/CoreNLP_sentence_splitter*
Retrofitting Cross-Lingual Word Embeddings to Dictionaries	Retrofitting cross-lingual word embeddings to the training dictionary, which pulls training translation pairs closer in the embedding space and overfits the train- ing dictionary.	JHU	*https://aclanthology.org/2020.acl-main.201.pdf*	*https://github.com/zhangmozhi/retrofit_clwe*
Roles Across Multiple Sentences (RAMS)	RAMS is a dataset for Multi-Sentence Argument Linking. It contains 9,124 annotated events from news based on an ontology of 139 event types and 65 roles. In a 5-sentence window around each event trigger, we annotate the closest span for each role.	JHU	*https://aclanthology.org/2020.acl-main.718.pdf*	*https://nlp.jhu.edu/rams/*
SpanFinder for event structure	Tree-based model for event structure extraction. Finds and labels spans.	JHU		*https://github.com/hiaoxui/span-finder*
Textual Entailment for Relation Extraction	Our research has shown that entailment models can achieve competitive results on the Relation Extraction task with very few examples when framed properly. As a result, we developed a Framework for Zero- and Few-Shot text classification based on Textual Entailment (Ask2Transformers). The framework is built on top of HuggingFace Transformers library. It is currently under development, so we plan to add new features and support for more information extraction related tasks, as well as a demo to interactively write and test new templates (rules).	University of the Basque Country (BBN)	*https://arxiv.org/abs/2109.03659*	*https://github.com/osainz59/Ask2Transformers*
The Universal Decompositional Semantics Dataset and Decomp Toolkit	Decomp is a toolkit for working with the Universal Decompositional Semantics (UDS) dataset, which is a collection of directed acyclic semantic graphs with real-valued node and edge attributes pointing into Universal Dependencies syntactic dependency trees.	JHU	*https://aclanthology.org/2020.lrec-1.699.pdf*	*http://decomp.io/projects/decomp-toolkit/*
Trankit: A Light-Weight Transformer-Based Toolkit for Multilingual Natural Language Processing.	A lightweight Transformer-based Toolkit for multilingual Natural Language Processing (NLP). It provides a trainable pipeline for fundamental NLP tasks over 100 languages, and 90 pretrained pipelines for 56 languages. Built on a state-of-the-art pretrained language model, Trankit can process raw texts and provides state-of-the-art performance for sentence segmentation, tokenization, multi-word token expansion, lemmatization, part-of-speech tagging, morphological feature tagging, and dependency parsing over 56 languages. Trankit also achieves state-of-the-art performance for named entity recognition (NER) on 11 languages. Despite the use of a large pretrained transformer, Trankit is very efficient in memory usage and speed. This is achieved by our novel plug-and-play mechanism with Adapters where a multilingual pretrained transformer is shared across pipelines for different languages.	BBN (U. of Oregon)	*https://arxiv.org/pdf/2101.03289.pdf*	**Github:** *https://github.com/nlp-uoregon/trankit* **Demo website:** *http://nlp.uoregon.edu/trankit*
TriggerNER: Learning with Entity Triggers as Explanations for Named Entity Recognition	A model (Trigger Matching Network) that jointly learns trigger representations and soft matching module with self-attention such that can generalize to unseen sentences easily for tagging. The framework is significantly more cost-effective than traditional neural NER frameworks.	USC Ink Lab	*https://arxiv.org/abs/2004.07493*	*https://github.com/INK-USC/TriggerNER*
Universal Decompositional Semantics dataset (UDS1.0)	A collection of directed acyclic semantic graphs with real-valued node and edge attributes pointing into Universal Dependencies syntactic dependency trees—along with a toolkit for reading and querying it.	JHU	*https://arxiv.org/abs/1909.13851*	*https://github.com/decompositional-semantics-initiative/decomp*

TITLE	DESCRIPTION	CREATOR	REFERENCE	LOCATION
QueryExplorer: An Interactive Query Generation Assistant for Search and Exploration	QueryExplorer is an interactive search tool designed to help users generate, refine, and modify queries using LLMs. It supports query-by-example scenarios while addressing challenges like concept drift and retrieval sensitivity. The system integrates HuggingFace models, PyTerrier retrieval pipelines, and extensive logging of user feedback, making it a valuable research platform for Human-in-the-Loop (HITL) experiments.	Emory IR Lab	*https://aclanthology.org/2024.naacl-demo.11* (Published in NAACL 2024)	**GitHub Code:** *https://github.com/emory-irlab/query-explorer* **Video:** *https://www.youtube.com/watch?v=sXBU8-uWR3o*
An Interactive Query Generation Assistant using LLM-based Prompt Modification and User Feedback	This study presents an interactive query generation assistant that helps users refine search queries across mono-lingual and multi-lingual document collections. It leverages LLMs to enhance query formulation through iterative user feedback, making search more effective for complex information needs. The tool supports real-time query refinement, document feedback integration, and HITL experiments for evaluating retrieval and ranking models.	Emory IR Lab	*https://arxiv.org/pdf/2311.11226*	**Video:** *https://www.youtube.com/watch?v=d1bN6vcQ4Lc*
DUQGen: Effective Unsupervised Domain Adaptation of Neural Rankers by Diversifying Synthetic Query Generation	DUQGen is a novel approach to unsupervised domain adaptation for neural rankers, tackling the challenge of generating effective and diverse synthetic training data. It clusters similar documents and uses probabilistic sampling to enhance domain representation, significantly improving ranking performance	Emory IRLab	*https://aclanthology.org/2024.naacl-long.413.pdf* (Published in NAACL 2024)	**Code:** *https://github.com/emory-irlab/DUQGen*