



REASON Lighting Talk

NLP Capabilities



Accenture Applied Intelligence

Applied Intelligence (AIX) is our unique approach that combines AI/ML with data, analytics, and automation to unlock the value of client data to achieve continuous operational, strategic and tactical mission outcomes at higher quality and scale

800+

AIX AFS professionals dedicated to USG clients – including data scientists, data engineers, and industry experts

Partnerships with 16+ Universities since 2018



NLP Leader

95+ AIX AFS professionals with advanced NLP training, 300+ engaged in NLP learning circles

40+ NLP and Knowledge Graph research papers published by Accenture since 2019



Discovery Lab

70 data specialists executing prototyping, testing, evaluation of ML/AI solutions, surge support, research, and innovation across federal



Premiere Tools

Rapidly deployable PaaS for data management & analytics

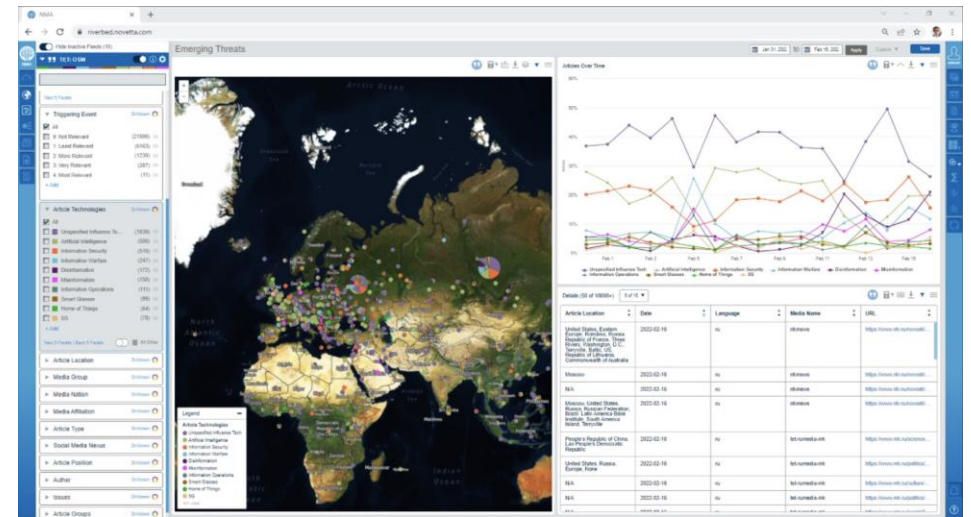
20+ alliances with highest rated industry vendors

Recognized leader in entity resolution & analytics

Mission tools, scientific focus

Accenture builds tools for analysts which integrate reliable, high performing algorithms, scale for real world use, and leverage good science.

And we do the research. We actively publish in the areas of natural language processing, graph analytics, computer vision, and geospatial analysis.



Tool Example | Novetta Mission Analytics - Augmenting Analyst Workflows with NLP

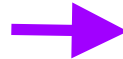
Integrating NLP applications accelerates analyst workflows, enabling them to focus on generating high-quality reports. Our process extracts text data through a unique automation process that identifies sources, selectors, topics, sentiment and geolocation under unified, scalable approach.

NATIVE LANGUAGE TEXT

Klaus Iohannis a spus,

“Aici în Craiova cunoașteți adresa acestui mare spital regional? Eu nu. Acum câteva săptămâni doamna ministru al Sănătății ne asigura că primele studii de fezabilitate vor fi comandate în această toamnă, vă dați seama ce bătaie de joc la adresa românilor? După ce au guvernat atâta timp acum ne spun că vor începe hârțile. Spitale nu avem.”

Hotnews.ro



MACHINE TRANSLATION (Optional)

Klaus Iohannis said,

“Here in Craiova you know the address of this great regional hospital? I do not. A few weeks ago, the Minister of Health assured us that the first feasibility studies will be commissioned this fall, do you realize what a gamble against the Romanians? After they have governed so long now they tell us that the papers will start. We do not have hospitals.”

Hotnews.ro

Source Identification

Track key influencers and their current and prior affiliations.

President of Romania, PNL Political Party

Selector Extraction

Target information on individuals and entities of interest.

Minister of Health, Klaus Iohannis

Topic Classification

Organize free text into customer-specific messaging buckets.

Social Welfare Policy

Sentiment Analysis

Track sentiment specific to relevant State and Non-state actors.

Negative to PSD Party

Geolocation

Identify event locations driving media conversation.

Craiova, Romania

Media Deconstruction

Uncover relationships between influencers, sentiment and the orientation of individual media outlets.

Liberal-Leaning Independent Media Outlet

Research Example | Automated Fact Checking

Automates detection of claims of fact in large-scale media data and ranks importance to a journalist to detect & counter mis/disinformation

CLEF CheckThat 2020

1st Place in English

1st, 2nd, 3rd, 4th Place in Arabic

CLEF CheckThat 2021

1st Place in Arabic



Accenture at CheckThat! 2020: If you say so: Post-hoc fact-checking of claims using transformer-based models

Evan Williams¹[0000-0002-0534-9450], Paul Rodrigues¹²[0000-0002-2151-636X],
and Valerie Novak²[0000-0001-8317-0993]

¹ Accenture, 800 N. Glebe Rd., Arlington, 22209, USA
e.m.williams@accenture.com
paul.rodrigues@accenture.com

² University of Maryland, College Park, MD, USA
³ vnovak@umd.edu

Abstract. We introduce the strategies used by the Accenture Team for the CLEF2020 CheckThat! Lab, Task 1, on English and Arabic. This shared task evaluated whether a claim in social media text should be professionally fact checked. To a journalist, a statement presented as fact, which would be of interest to a large audience, requires professional fact-checking before dissemination. We utilized BERT and RoBERTa models to identify claims in social media text a professional fact-checker should review, and rank these in priority order for the fact-checker. For the English challenge, we fine-tuned a RoBERTa model and added an extra mean pooling layer and a dropout layer to enhance generalizability to unseen text. For the Arabic task, we fine-tuned Arabic-language BERT models and demonstrate the use of back-translation to amplify the minority class and balance the dataset. The work presented here was scored 1st place in the English track, and 1st, 2nd, 3rd, and 4th place in the Arabic track.

Keywords: fact checking, fact identification, Arabic, BERT, RoBERTa

1 Introduction

Natural Language Processing (NLP) has been driving Artificial Intelligence research since the 1950s, but recently increased in distinction due to the quantity of text that can be utilized as well as new techniques to extract even more value from text. In 2018, a surge of research produced deep learning architectures in NLP which beat state of the art on a multitude of tasks, such as sentiment analysis, question answering, and semantic similarity, in a variety of languages.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessaloniki, Greece.



Accenture at CheckThat! 2021: Interesting claim identification and ranking with contextually sensitive lexical training data augmentation

Evan Williams⁰⁰⁰⁰⁻⁰⁰⁰²⁻⁰⁵³⁴⁻⁹⁴⁵⁰, Paul Rodrigues^{0000-0002-2151-636X}, and
Sieu Tran⁰⁰⁰⁰⁻⁰⁰⁰³⁻⁰⁰¹⁷⁻⁴³²⁹

Accenture, 800 N. Glebe Rd., Arlington, 22209, USA
e.m.williams@accenture.com
paul.rodrigues@accenture.com
sieu.tran@accenture.com

Abstract. This paper discusses the approach used by the Accenture Team for CLEF2021 CheckThat! Lab, Task 1, to identify whether a claim made in social media would be interesting to a wide audience and should be fact-checked. Twitter training and test data were provided in English, Arabic, Spanish, Turkish, and Bulgarian. Claims were to be classified (check-worthy/not check-worthy) and ranked in priority order for the fact-checker. Our method used deep neural network transformer models with contextually sensitive lexical augmentation applied on the supplied training datasets to create additional training samples. This augmentation approach improved the performance for all languages. Overall, our architecture and data augmentation pipeline produced the best submitted system for Arabic, and performance scales according to the quantity of provided training data for English, Spanish, Turkish, and Bulgarian. This paper investigates the deep neural network architectures for each language as well as the provided data to examine why the approach worked so effectively for Arabic, and discusses additional data augmentation measures that should be useful to this problem.

Keywords: fact-checking, claim retrieval, check-worthy, BERT, RoBERTa

1 Introduction

Labeled data for some machine learning tasks can be quite rare and valuable. Data labeling is a time consuming task, and can be expensive if subject matter or language expertise are required. Machine learning engineers know that generally, the larger the training set, the higher accuracy the classifier will have [3], so they often request more data.

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2021, 21-24 September 2021, Bucharest, Romania.



Follow up

Paul
Rodrigues, PhD

NLP Science Director

(123) 456-7890
Paul.Rodrigues@acenturefederal.com



Accenturefederal.com/NationalSecurity

Thank you

