

Software Engineering Department

Ort Braude College

Course 61998: Extended Project in Software Engineering

**Pre-Trained Authorship Representation Transformer**

In Partial Fulfillment of the Requirements for

Final Project in Software Engineering (Course 61998)

Karmiel – January 2023

Meitar El-Ezra 304680218

Tuval Zitelbach 316219633

**Supervisor**:

Mr. Zeev Volkovich

Mrs. Renata Avros

Contents

[1. ABSTRACT 3](#_Toc121159834)

[2. INTRODUCTION 4](#_Toc121159835)

[3. ARCHITECTURE 5](#_Toc121159836)

[3.1 PRE PROCESSING 5](#_Toc121159837)

# 1. ABSTRACT

Authorship attribution (AA) is the computational task of identifying the author of a text based on a set of possible candidates. Authors imprint intentional and unintentional traces in the form of linguistic features such as punctuation, registry and semantics which can be used to profile authorship. Previous works attempted to encode the semantic features. Those approaches led to poor results on open-set authors (authors that were not included in the training phase). The following approach, nonetheless, attempts to encode features from both context and semantics, focusing more on style rather than semantics alone, which creates what is called, in this article, **authorship embeddings**. PART: contrastively Pre-Trained Authorship Transformer uses zero shot generalization capabilities in authorship identification to compute authorship embeddings with the assistance of the state-of-the-art contextual based models. Analyzing those attribution can contribute to several fields such as forensics, social networks analysis, identifying fake news or profiles and more.

**Keywords:** Authorship attribution, Neural networks, Transformers, Contrastive pretraining.

# 2. INTRODUCTION

Authorship identification of handwritten documents started approximately in the late 19th century; however most textual data is digital. Typically, the input for AA models is a set of candidates and number of texts for each candidate. The objective is to assign text to one of the candidates. AA is divided to three forms. First, closed-set attribution, where the true author of a given text must be included in the set of suspects. Second, open-set attribution, where the true author of a given text could be excluded from the set of candidates. The third approach, the author verification where there is only one candidate author.

In real life scenarios, the training set of texts and the test set of texts may be differed in the topic and genre. Those situations are examined in the Cross-Domain AA. Most of the previous works in AA focus on the closed-set AA. Nevertheless, the **authorship embedding,** that is presented in this paper, is calculated with zero-shot generalization capabilities in authorship identification, namely, open-set AA form. In this case, the model gets test set of texts, from both new and known authors, topics, genres, etc.

# 3. ARCHITECTURE

## 3.1 PRE PROCESSING