

Anomaly Detection on Scientific Publications

David-Gabriel ION

Thesis advisor:
Prof. dr. ing. Ciprian DOBRE





THE PROBLEM

Author Identification
for the CRESCDI Platform



ASSUMPTION

Authors write publications
in specific fields (topics)

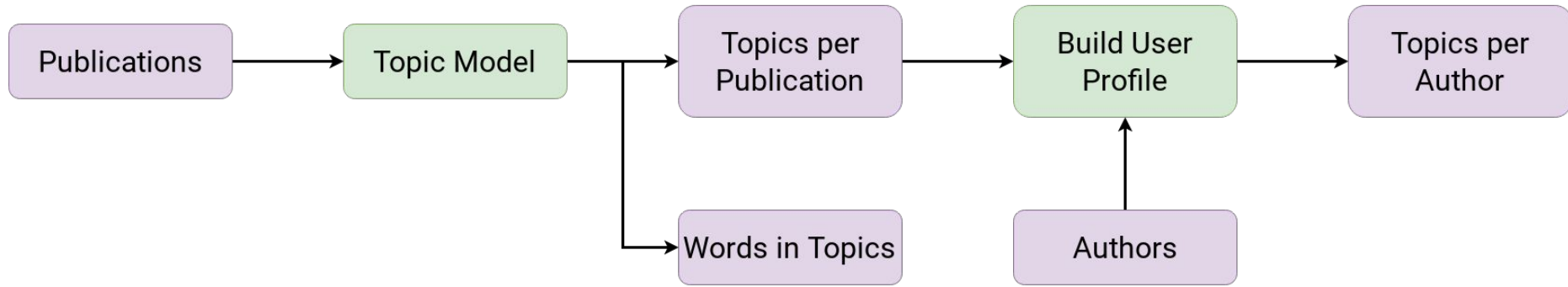


IDEA

Using Topic Models



Training



WHAT ABOUT TRANSFORMERS?

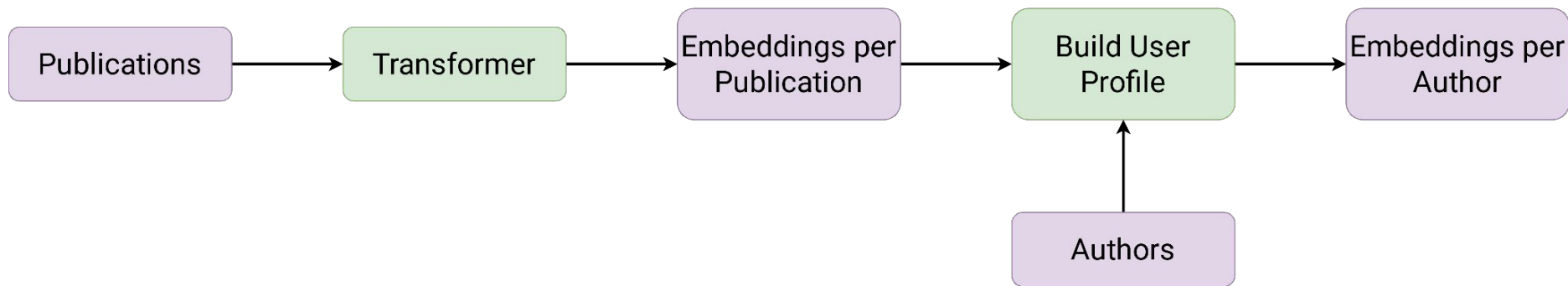
Transformers achieve state of the art performance in several NLP tasks



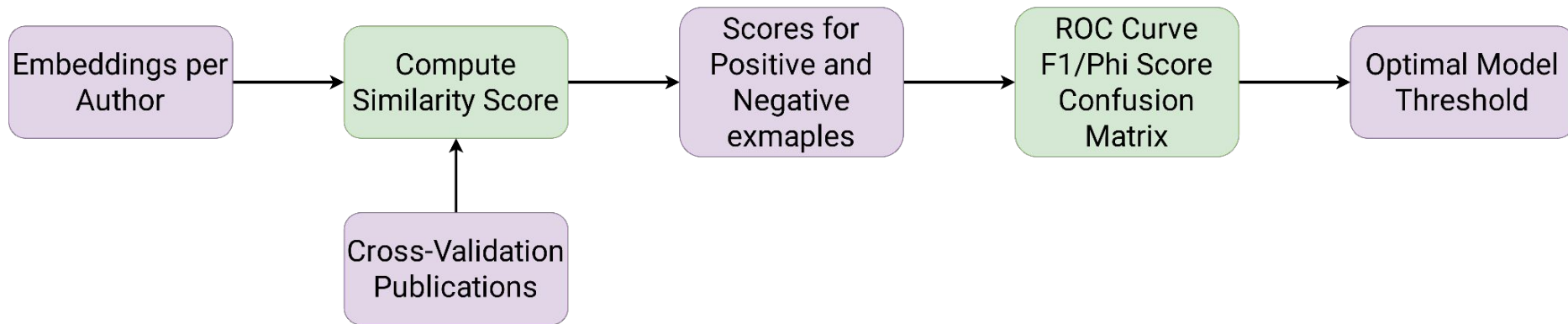
ASSUMPTION

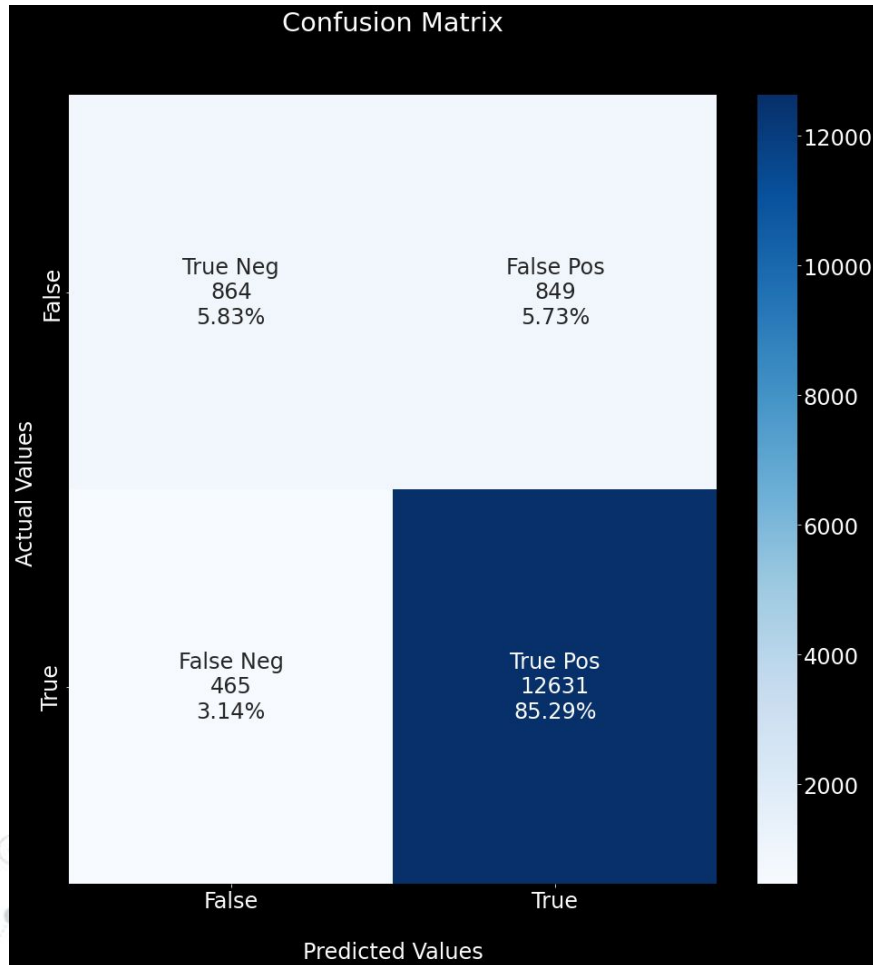
Transformers output similar embeddings
for similar documents

Training



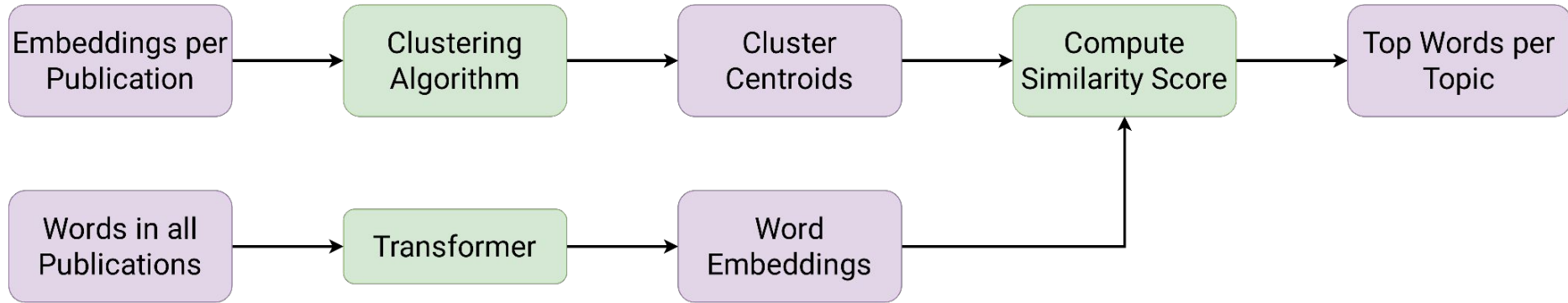
Tuning





91% Accuracy
0.95 F1 Score
15 min Train Time
40ms Prediction Time

Topic Modelling



AUTHOR KEYWORDS

Ciprian Dobre

pervasive computing
dynamic infrastructures
distributed infrastructures
distributed environments
computing service

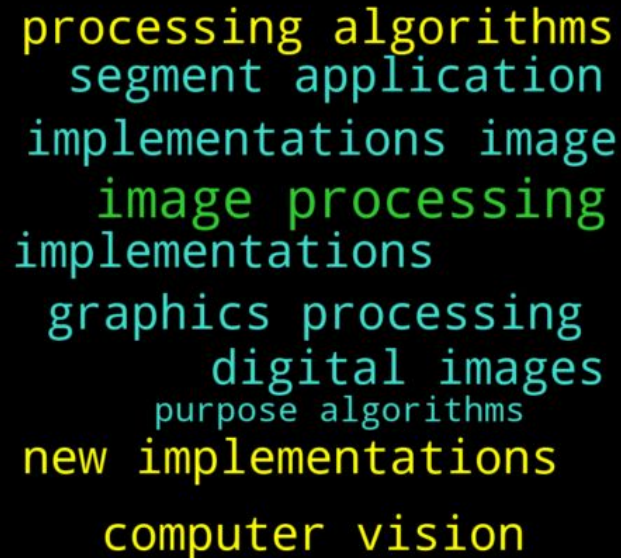
distributed environment
distributed service
infrastructure scalable
internet architecture
distributed applications

Traian Rebedea

supervised
concepts
categorization learns
learning
ai knowledge
classification
technological
wordnet

AUTHOR KEYWORDS

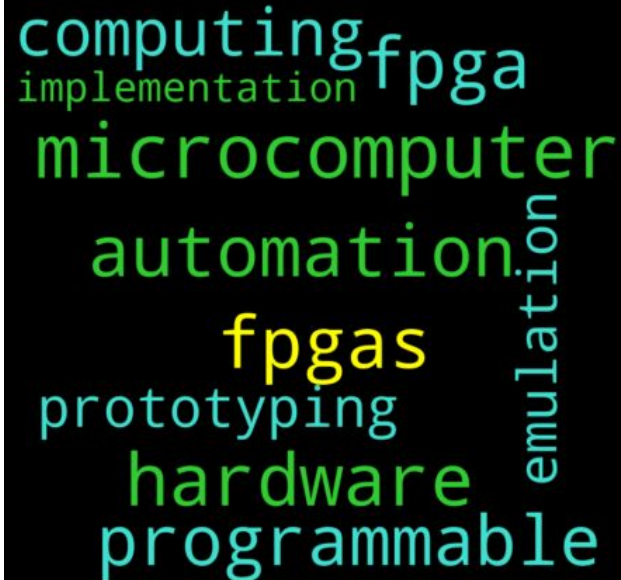
Victor Asavei



A word cloud on a black background for Victor Asavei. The words are in various colors (yellow, green, cyan) and sizes. The most prominent words are 'processing algorithms' (yellow), 'segment application' (cyan), 'implementations image' (cyan), 'image processing' (green), 'implementations' (cyan), 'graphics processing' (cyan), 'digital images' (cyan), 'purpose algorithms' (cyan), 'new implementations' (yellow), and 'computer vision' (yellow).


processing algorithms
segment application
implementations image
image processing
implementations
graphics processing
digital images
purpose algorithms
new implementations
computer vision

Cornel Popescu



A word cloud on a black background for Cornel Popescu. The words are in various colors (cyan, green, yellow) and sizes. The most prominent words are 'computing' (cyan), 'implementation' (cyan), 'fpga' (cyan), 'microcomputer' (green), 'automation' (green), 'fpgas' (yellow), 'prototyping' (cyan), 'hardware' (green), 'programmable' (cyan), and 'emulation' (cyan).

computing
implementation
fpga
microcomputer
automation
fpgas
prototyping
hardware
programmable
emulation

A decorative network diagram in the top right corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

Topic modelling
Transformers
Parameter tuning
Keyword extraction