

# OpenMinTeD: Facilitating Large Scale Text Mining of Open Access Corpora

**Author1, Author2, Author3**

Affiliation1, Affiliation2, Affiliation3

Address1, Address2, Address3

author1@xxx.yy, author2@zzz.edu, author3@hhh.com

{author1, author5, author9}@abc.org

## Abstract

The OpenMinTeD platform aims to bring together open-access content from a wide range of publisher and text mining tools from numerous existing natural language processing frameworks in a seamless environment to enable those with little text mining experience to access the wealth of information locked away in the vast sea of published free-text material.

**Keywords:** text mining, open access, corpora, natural language processing

## 1. Introduction

## 2. Access to Corpora

Talk about the content connectors and give an idea of how many articles publishers, journals etc. we can give access to for text mining through the platform

## 3. Components and Workflows

we support components from existing frameworks (GATE (Cunningham et al., 2013) and UIMA (Ferrucci et al., 2009)), web services, and other tools provided as docker images as long as they follow the guidelines<sup>1</sup>.

these tools are registered with the platform using a common description (OMTD-SHARE) and can then be combined into workflows

workflows are executed via Galaxy on a large cluster to provide scalability and performance

results are pushed back into the OMTD store

## 4. Seamless Integration

The platform serves two roles. Firstly as a text mining expert it's the portal through which you can register components and build workflows (see above), but also it's where non expert users can build corpora and execute pre-existing workflows to mine scholarly publications for information which they would not otherwise be able to access.

## 5. Bibliographical References

- Cunningham, H., Tablan, V., Roberts, A., and Bontcheva, K. (2013). Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS Computational Biology*, 9(2):e1002854, 02.
- Ferrucci, D., Lally, A., Verspoor, K., and Nyberg, E. (2009). Unstructured information management architecture (UIMA) version 1.0. OASIS Standard, mar.

---

<sup>1</sup><https://guidelines.openminded.eu/>