

OpenMinTeD Platform Training: Discover the power of TDM

Introduction

OpenMinTeD is an EC/H2020 funded project aiming to foster and facilitate the use of Text and Data Mining technologies in the scientific publications world. Recent years witness an upsurge in the quantities of digital research data which are offered under Open Access conditions and are thus eligible for mining. At the same time, though, text mining solutions are not easy to discover and use, nor are they easily combinable by end users. OpenMinTeD integrates Open Access content and existing Text Mining tools and services in a platform that builds upon them and offers end-users (technology-naïve researchers but also TDM experts) appropriate services.

OpenMinTeD has organized this training session to show how the usage of Open Access content combined with its services can offer great solutions to users. The training has taken the form of a hands-on experience where users were invited to use the platform (beta version) following the presenter's instructions in order to get better acquainted with its functionalities. The session was interactive, since participants could navigate the OpenMinTeD platform and run the services of their choice. Participants were free to ask any questions during the demonstration and receive guidance. By the end of the session, participants were expected to feel comfortable to use the OpenMinTeD platform.

Structure of the training session

- **What is OpenMinTeD**

An informative short introduction describing OpenMinTeD to attendees., based upon the definition of OpenMinTeD as “An **open** and **sustainable** Text and Data Mining (TDM) **platform** and **infrastructure** where researchers can discover, collaboratively create, share and re-use knowledge from a wide range of text based **scientific and scholarly** related sources.” The main features were explained to the attendees.

- **More information and updates about OpenMinTeD**

Presentation of the main web site of [OpenMinTeD](#) and of the relevant pages where the visitor of the website can find e.g. Support and Training, Events, News, and of course, the link to the platform.

- **Registration and Login Process [for the platform](#)**

How to register with academic (EduGAIN, OrcID) or social media accounts (Facebook, LinkedIn) or Gmail. Live registration by users and verification of their credentials as a security process of the OpenMinTeD infrastructure that allows them to successfully log in to the [OpenMinTeD platform](#).

- **How to find/upload content**

A tour of the OpenMinTeD platform menus and pages to show the users its functionalities, organized in the form of user-oriented tasks.

Task 1: discover content that is already uploaded in OpenMinTeD in a form that can be input to the TDM tools, or build a corpus (dataset) of Open Access publications retrieved via APIs from [OpenAIRE](#) and [CORE](#). Participants were shown the relevant pages in OpenMinTeD. They were trained on how to use free text queries or facets (e.g. date of publication, language, publication type) to select subsets of the content.

Task 2: upload content in OpenMinTeD directly. First, the presenter explained that users should follow the OpenMinTeD [guidelines](#) in order to be able to upload content which is compatible with the

OpenMinTeD registry requirements. A demonstration of uploading content via the OpenMinTeD editor form followed. The form includes mandatory and recommended fields which were explained to the users. Attention was drawn to the fields related to licensing, since open access conditions are a prerequisite for the corpora to be uploaded. Users were also asked to check the OpenMinTeD [Terms and Conditions with which they must agree in order to use the platform](#).

The presenter also explained that potential content providers (e.g. publishers, repositories, journals etc.) can make their content available to OpenMinTeD indirectly via [OpenAIRE](#) and [CORE](#), by following again the relevant [guidelines](#).

- **How to find/upload TDM tools/services**

The focus now is on the TDM tools and services offered via the OpenMinTeD platform.

Task 3: Attenders got to know how to browse through the catalogue of tools and services, select one of them and view its description in order to ensure that it fits their needs. The process is similar to the one for discovering content, as the user can search with a free text query or make use of specific facets (e.g. licence, language) to refine the search.

Task 4: The next step had to do with uploading TDM tools in OpenMinTeD. There are three ways to do so:

- i) directly registering a tool by using a form
- ii) by entering the [Maven](#) coordinates of the tool, if it is already uploaded in the Maven repository
- iii) by uploading an XML description of the tool compatible with the [OpenMinTeD schema](#).

Participants were given the opportunity to explore all options.

- **How to run TDM tools/services**

Task 5: Once users have discovered the content they wish to mine and the TDM tools/services they want to use, they can proceed to the actual operation. The presenter demonstrates this through three different use cases:

- i) A sample corpus from OpenAIRE consisting of 4 pdf files + An extractor of funding information
- ii) A sample corpus from OpenAIRE with publications that include citations to datasets + an extractor of dataset citations
- iii) A sample Bio-domain corpus consisting of 2 pdfs randomly selected from PMC + A Named Entity Recognizer for metabolites

In parallel, participants were invited to perform the same tests or select any of the available tools/services and run them on any of the available corpora.

- **Get the result!**

Finally, after the successful termination of the execution of the three use cases, users are shown the results. Additionally, users are advised to view the progress of their actions on the pages that show their resources and their profile.

- **What is next in OpenMinTeD**

At the end, all participants are informed that OpenMinTeD is in a beta version and in a few months they will be able to enjoy all services that are now under development such as: Workflows editor, Workflow viewer, Annotation viewer.

The training closes with a Questions and Answers section, a summary, an informative slide regarding the [Open Call Tender for content providers](#) and a reminder to comment on the OpenMinTeD [guidelines](#).