FutureTDM
Explore . Analyse . Improve

IMPROVING UPTAKE OF TEXT AND DATA MINING IN THE EU

OpenMinTeD
Open Mining INfrastructure for TExt and Data

What is OpenMinTeD?

OpenMinTeD is an ongoing Horizon 2020 project (01/06/2015 - 31/05/2018) which aims to create an open, service oriented, infrastructure for text and data mining (TDM) of scientific and scholarly content.

How is OpenMinTeD related to FutureTDM?

FutureTDM is defining the challenges of TDM and proposing a legal and policy framework. OpenMinTeD is focusing on building an infrastructure to facilitate TDM in Europe. Both projects want to increase awareness of TDM, showcase best practices and build a community.

Why is such an infrastructure necessary?

Digital research data is being produced in massive amounts and it cannot be harnessed or utilised unless appropriate technologies exist. This combination is provided by infrastructures which serve as data treasuries and offer the necessary tools for its processing.

The infrastructure proposed by OpenMinTeD will enable researchers to collaboratively create, discover, share and re-use knowledge from a wide range of text-based scientific related sources in a seamless way. Through the analysis of data at multiple levels, unstructured data is turned to structured and further processed so as to detect relations and
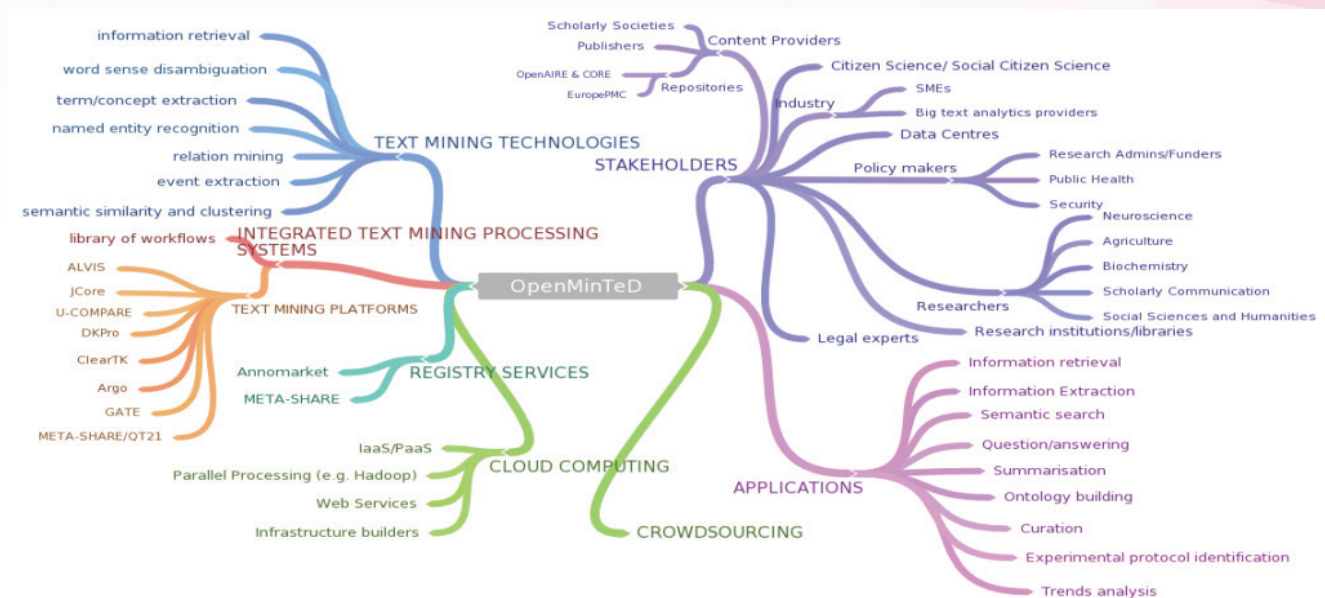

patterns, discover hidden and new knowledge and reveal meaning. To do so, OpenMinTeD makes use of pre-existing TDM tools, services, resources and content which are all integrated into one platform.

How will this integration be achieved?

Four groups of experts, internal and external, have been assigned specific problems which must be resolved in order to have technologies and content all integrated into one platform.

The Resource Metadata Working Group, has to compile and maintain an inventory of existing metadata schemas and their representation (XML, RDF) for documenting scholarly communication/ literature content, language resources, text mining/ language processing services and platforms. Their focus is on content and service discoverability and interaction, as well as on metadata quality. Their objective is to create a draft specification catering for metadata descriptions and standards.

Source: OpenMinTeD

The Language Resources Group is responsible for defining a specification for the representation of text, lexicons, terminologies, thesauri and ontologies, and their interoperability across different resources/tools for the purpose of their exploitation by TDM applications. Their focus is on ensuring the interoperability, consistency and discoverability of linguistic, terminological and ontological content at the granular representation level of individual knowledge elements.

The IPR and licensing Group, has been studying and identifying copyright and related rights (e.g. sui generis database right) restrictions and exceptions to the use and reuse of sources (both textual sources and text-mining services) in TDM activities. On this basis, they will also identify contractual tools and schemes (e.g. licenses) that can best serve the needs of TDM services.

The Annotation and Workflow Services Group has to define annotation models/ type systems, input-output representation formats (e.g., XML, XMI, RDF, JSON, NIF, W3C Annotations, efficient wire formats like UIMA Binary CAS), appropriate annotation service input-output conversions, type system alignment, annotation tag sets, confidence scores and workflow persistence formats (e.g. UIMA aggregate engine descriptor).

Which are the domains of interest for OpenMinTeD?

OpenMinTeD covers a wide range of scientific fields, starting from the broad categories of life and social sciences, agriculture and biodiversity, etc. and extending to areas of special scientific interest such as Neuroinformatics.

Whom does the project address?

OpenMinTeD addresses stakeholders from both the public and private sector: researchers, universities, research institutions, libraries and computing centres, publishers, OA repositories, scholarly or learning societies and content providers, data and computing centres, legal experts, infrastructure developers, industrial players, and SMEs.

Who is involved in the project?

The OpenMinTeD consortium consists of groups specializing in TDM and/or research in general. These are either independent organizations/ companies or centres based in academic institutes around Europe.

For a full list of the consortium members, please visit http://openminted.eu/about/partners

Discover more
VISIT OUR COLLECTION
STORIES · PROJECTS · ORGANISATIONS · TOOLS · CHALLENGES