FutureTDM

Explore . Analyse . Improve

REDUCING BARRIERS AND INCREASING UPTAKE OF TEXT AND DATA MINING FOR RESEARCH ENVIRONMENTS USING A COLLABORATIVE KNOWLEDGE AND OPEN INFORMATION APPROACH

Deliverable D4.4

Scientific Review Paper

Project

Acronym:	**FutureTDM**
Title:	Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments using a Collaborative Knowledge and Open Information Approach
Coordinator:	SYNYO GmbH
Reference:	665940
Type:	Collaborative project
Programme:	HORIZON 2020
Theme:	GARRI-3-2014 - Scientific Information in the Digital Age: Text and Data Mining (TDM)
Start:	01. September, 2015
Duration:	24 months
Website:	http://www.futuretdm.eu/
E-Mail:	office@futuretdm.eu
Consortium:	**SYNYO GmbH**, Research & Development Department, Austria, (SYNYO) **Stichting LIBER**, The Netherlands, (LIBER) **Open Knowledge**, UK, (OK/CM) **Radboud University**, Centre for Language Studies The Netherlands, (RU) **The British Library Board**, UK, (BL) **Universiteit van Amsterdam**, Inst. for Information Law, The Netherlands, (UVA) **Athena Research and Innovation Centre in Information, Communication and Knowledge Technologies,** Inst. for Language and Speech Processing**,** Greece**,** (ARC) **Ubiquity Press Limited**, UK, (UP) **Fundacja Projekt: Polska**, Poland, (FPP)

Deliverable

Number:	**D4.4**
Title:	**Scientific Review Paper**
Lead beneficiary:	**RU**
Work package:	WP4: Fields of Application, Projects, Best Practices and Resources
Dissemination level:	Public (PU)
Nature:	Report (RE)
Due date:	30.04.2017
Submission date:	02.05.2017

Authors: **Maria Eskevich**, RU
Antal van den Bosch, RU

Contributors: **Stelios Piperidis, ARC**
Kanella Pouli, ARC
Maria Gavriilidou, ARC
Dimitris Galanis, ARC

Review: **Burcu Akinci**, SYNYO

Acknowledgement: This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 665940.

Table of Contents

List of Tables

1 SUMMARY

This report provides information about the work on the scientific review paper that gives a critical review of the past, present and future of the state-of-the-art of the field of text and data mining (TDM). This publication aims to broaden perspective on the past, current, and future trends in TDM amongst audience, outlined via desktop research, consultation with the practitioners (Deliverables D.4.3. and D4.5.) and scientific communities. Report includes brief description of publication venue selection, text of the paper that can be further updated with the open accesses reviewers comments.

Additionally, the information about the update of FutureTDM knowledge base collection of relevant projects, organisations, methods and tools is provided including the current list of the collections in the Annex.

2 WORK ON THE SCIENTIFIC PAPER

2.1 Criteria for selection of potential publication venues

We have defined the two main criteria for the target publication venues:

- Diverse type of TDM applications already present in the journal
- The journal has to be Open Access, or at least to have the open access publication option

As the review of TDM work from the perspective of overall scientific development and insights on its visibility across different economic fields, that are formulated within the FutureTDM consortium, differ from the traditional type of publication content that is expected by journals, we took an initiative to contact a number of targeted journals in order to find a venue that would fit our criteria and to ensure that our content would be suitable. These were the journals to contact:

1. **Data Science Journal**
    - Link: http://datascience.codata.org
    - Aims and scope of the journal: The CODATA Data Science Journal is a peer-reviewed, open access, electronic journal, publishing papers on the management, dissemination, use and reuse of research data and databases across all research domains, including science, technology, the humanities and the arts. The scope of the journal includes descriptions of data systems, their implementations and their publication, applications, infrastructures, software, legal, reproducibility and transparency issues, the availability and usability of complex datasets, and with a particular focus on the principles, policies and practices for open data. All data is in scope, whether born digital or converted from other sources.
        - *Reviews* can cover topics such as current controversies, the current "state of the art" or the historical development of studies as well as issues of regional or temporal focus. Papers should critically engage with the relevant body of extant literature. Review articles should be no longer than 3,000 words in length.
    - Open access policy: multidisciplinary Open Access journal

2. **PLOSOne**
    - Link: http://journals.plos.org/plosone/
    - Aims and scope of the journal: PLOS ONE features reports of original research from all disciplines within science and medicine. By not excluding research on the basis of subject area, PLOS ONE facilitates the discovery of connections between research whether within or between disciplines.

- We consider publishing systematic reviews only if the methods ensure the comprehensive and unbiased sampling of existing literature.
- Submissions describing methods, software, databases, or other tools. We consider submissions describing methods, software, databases, or other tools if they follow the appropriate reporting guidelines.
- Qualitative research. We consider publishing qualitative research only if it adheres to appropriate study design and reporting guidelines.
- Studies reporting negative results.

- Open access policy: multidisciplinary Open Access journal

3. DSH (Digital Scholarship for Humanities)

- Link: https://academic.oup.com/dsh
- Aims and scope of the journal: *DSH* or *Digital Scholarship in the Humanities* is an international, peer reviewed journal which publishes original contributions on all aspects of digital scholarship in the Humanities including, but not limited to, the field of what is currently called the Digital Humanities. Long and short papers report on theoretical, methodological, experimental, and applied research and include results of research projects, descriptions and evaluations of tools, techniques, and methodologies, and reports on work in progress. *DSH* also publishes reviews of books and resources.
- Open access policy: multidisciplinary Open Access journal

4. Science Communication (SAGE)

- Web-Link: http://journals.sagepub.com/home/scx
- Aims and scope of the journal: Science Communication is an international, interdisciplinary social science journal that examines such topics as the nature of scientific expertise as represented through communication and the processes or effects characterizing the communication of science in any context. Science is broadly defined to include environmental science, health science, and technology. Science Communication welcomes submissions of empirical research from authors in all relevant disciplines (including the social sciences, the humanities, and science itself). Both qualitative and quantitative research papers with a basis in theory are acceptable. Preference is given to articles that bridge the gap between theory and practice and that will be of interest across disciplines. In addition to peer-reviewed research, Science Communication publishes commentaries that analyse issues and trends in the field – whether scholarly, professional, or policy-related – and a periodic summary of new books in the field.
- Open access policy: OA option is available

5. IEEE Access

- Web-Link: http://ieeeaccess.ieee.org/learn-more-about-ieee-access/

- Aims and scope of the journal: IEEE Access publishes articles that are of high interest to readers: original, technically correct, and clearly presented. The scope of this journal comprises all of IEEE's fields of interest, emphasizing applications-oriented and interdisciplinary articles. IEEE Access also accepts traditional technical articles, as well as reviews and surveys.
- Open access policy: Multidisciplinary Open Access Journal, fee $1,750 per article

Eventually, we have chosen the “Data Science Journal” that is an open access multidisciplinary journal run by an SME publisher “Ubiquity Press”.

2.2 Information on the submitted paper

The details of the submission are the following:

- Authors: Maria Eskevich, Stelios Piperidis, Kanella Pouli, Maria Gavriilidou, Dimitris Galanis, Antal van den Bosch.
- Title: Text and Data Mining: Past, present, and future.
- Submission date: 01.05.2017.
- Text of the paper is in the Appendix

3 WORK ON THE FUTURETDM KNOWLEDGE BASE ENRICHMENT

3.1 Collection of Projects

FutureTDM collection of projects by the Deliverable submission date contains 140 items, 37 of which are already integrated into the platform. For 7 of the projects it is not possible to assign an economic sector, such as (+)Spaces , while the other 133 can be assigned to one or even a combination of sectors due to the multidisciplinary nature of the projects. Table 1 contains the overall statistics of the number of projects in the collection.

Sector or combination of Sectors that that the project is assigned to	Number of projects
Primary-Sector, Secondary-Sector, Tertiary-Sector, Quartenary-Sector	1
Primary-Sector, Quartenary-Sector	1
Secondary-Sector	8
Secondary-Sector, Tertiary-Sector	3
Secondary-Sector, Quartenary-Sector	38
Secondary-Sector, Tertiary-Sector, Quartenary-Sector	6
Tertiary-Sector	13
Tertiary-Sector, Quartenary-Sector	24
Quartenary-Sector	38
Quinary-Sector	1
TOTAL	**133**

Table 1: Total number of projects in the FutureTDM Knowledge Base Collection structured by economic sector

3.2 Collection of Organisations

FutureTDM collection of organisations by the Deliverable submission date contains 72 items, 59 of which are already integrated into the platform. 55 of these organisations have only one EU-country location, while three of them spread across more than one country within the EU, and 6 have locations outside EU in USA.

Company locations details	Country	Number
Company is based in one country	Austria	1
	Belgium	2
	Czech Republic	2
	Denmark	2
	Finland	1
	France	4
	Germany	10
	Greece	3
	Ireland	1
	Italy	3
	Norway	1
	Slovenia	1
	Spain	4
	Sweden	2
	Switzerland	4
	The Netherlands	5
	UK	9
Company has offices internationally	Belgium, The Netherlands, Germany, USA	1
	The Netherlands,	1
	Germany/USA	1
	Ireland/Greece	1
	Spain/UK	1
	UK/USA	2
	USA, many	2
Companies outside EU	Japan	1
	USA	7
TOTAL		**72**

Table 2: Total number of projects in the FutureTDM Knowledge Base Collection structured by country

3.3 Collection of Methods

FutureTDM collection of projects splits scientific methods into 29 items, and all of them are integrated into the platform, as they are used to specify the organisations focus.

TDM Methods			
Artificial Intelligence	Machine Learning	Statistics	Kernel Methods
Association rules	Machine Translation	Summarisation	Ensemble Learning
Classification	Multimedia Processing	Term/Concept Extraction	Dimensionality Reduction
Clustering	Natural Language Processing	Text Mining	Decision Trees
Data Mining	Negation and Modality Detection	Textual Entailment	Deep Learning
Graph Mining	Predictive Analytics	Artificial Neural Networks	
Information Extraction	Regression	Bayesian Inference	
Information Retrieval	Sentiment Analysis/Opinion Mining	Instance-Based Learning	

Table 3: FutureTDM text and data mining methods collection

3.4 Collection of Tools

FutureTDM collection of projects by the Deliverable submission date contains 72 items, 59 of which are already integrated into the platform. Table contains the complete list of tools collected so far.

Name	URL
ALVEO	http://alveo.edu.au/
Alvis	https://migale.jouy.inra.fr/redmine/projects/alvisnlp
Angoss Knowledge STUDIO	http://www.angoss.com/
AnnoMarket	https://annomarket.eu/
Apache cTAKES	http://ctakes.apache.org
Apache OpenNLP	http://opennlp.apache.org
Apache UIMA	https://uima.apache.org
Argo	http://argo.nactem.ac.uk

BioCatalogue	https://www.biocatalogue.org/
BiodiversityCatalogue	https://www.biodiversitycatalogue.org
Bluima	https://github.com/BlueBrain/bluima
CLARIN-DK	http://clarin.dk/
ClearTK	https://cleartk.github.io/cleartk
Clementine	http://datamining.togaware.com/survivor/Clementine.html
ContentMine	http://www.contentmine.org/
DKPro Core	https://dkpro.github.io/dkpro-core
ELKI	http://elki.dbs.ifi.lmu.de/
FICO Data Management	http://www.fico.com/
Galaxy	https://galaxyproject.org/
GATE Embedded	https://gate.ac.uk
Heart of Gold	http://heartofgold.dfki.de
IBM SPSS Predictive Analysis	http://www-03.ibm.com/software/products/en/spss-predictive-analytics-enterprise
JCoRe	http://julielab.github.io
Jpylyzer	http://jpylyzer.openpreservation.org/
KAF	http://kyoto-project.eu/xmlgroup.iit.cnr.it/kyoto/index2091.html?option=com_content&view=article&id=504&Itemid=173
Kepler	https://kepler-project.org/
KNIME	http://www.knime.org/knime
Language Grid	http://langrid.org/en/index.html
LAPPS Grid	http://www.lappsgrid.org/
Massive Online Analysis (MOA)	http://moa.cms.waikato.ac.nz/
Matchbox	http://matchbox.openpreservation.org/
Microsoft SQL Server	https://technet.microsoft.com/en-

Analysis Service (SSAS)	us/library/ms175609%28v=sql.90%29.aspx
Neural Designer	https://www.neuraldesigner.com/
NLTK	http://www.nltk.org
Open Calais	www.opencalais.com/
Oracle Data Miner GUI	http://www.oracle.com/technetwork/database/options/odm/datamining workflow-168677.html
Orange	http://orange.biolab.si/
Pagelyzer	http://pagelyzer.openpreservation.org/
Pegasus	https://pegasus.isi.edu
Pentaho	http://www.pentaho.com/
Pipeline Pilot	http://accelrys.com/products/collaborative-science/biovia-pipeline-pilot/
QT21	http://qt21.metashare.ilsp.gr/
Rapidminer Radoop	www.rapidminer.com
Rapidminer Server	www.rapidminer.com
Rapidminer Studio	www.rapidminer.com
SAP Hana	https://hana.sap.com
SAS Data Mining	http://www.sas.com
SPSS	http://www.ibm.com/analytics/us/en/technology/spss/
Taverna	http://www.taverna.org.uk/
Think Analytics	http://thinkanalytics.com/
Think Enterprise Data Miner	www.thinkanalytics.com/
Triana	http://www.trianacode.org
TTNWW	http://yago.meertens.knaw.nl/apache/TTNWW/
Weblicht	https://weblicht.sfs.uni-tuebingen.de/
WEKA	http://www.cs.waikato.ac.nz/~ml/weka
xcorrSound	http://xcorrsound.openpreservation.org/

VisTrails	http://www.vistrails.org
Trendminer	https://www.trendminer.com/
Opendatasoft	https://www.opendatasoft.com/
Meta	http://meta.com/
TensorFlow	https://www.tensorflow.org/about/
Epinium	http://epinium.com/
Exclusivi	http://exclusivi.com/hotels/
Eparkomat	http://www.eparkomat.com/
Data Scouts	https://datascouts.eu/
TopPlace	http://www.avuxi.com/products/topplace
Influencity	https://influencity.com/en/
Plazi TreatmentBank	http://plazi.org/api-tools/api/#Plazi_API
SafeToNet	https://safetonet.com/
Egas	https://demo.bmd-software.com/egas/
Wordfreak	http://wordfreak.sourceforge.net/
Theano	https://github.com/Theano/

Table 4: Total list of tools in the FutureTDM Knowledge Base Collection

4 CONLUSIONS

This Deliverable contains the full text of submitted TDM review paper, together with the explanation on the publication venue selection.

We provide quantitative updates on the progress with filling in the four main categories of the Knowledge Base Collection, while the Deliverables 6.1, 6.2 and 6.3[1] contains the qualitative description of the technology behind the FutureTDM Platform.

[1] www.futuretdm.eu

5 ANNEX

In the following we present the scientific paper "**Text and Data Mining: Past, present, and future**" and the collections below described in the previous section. The list will be given in this order:

- Collection of Organisations
- Collection of Projects
- Collection of Tools

Text and Data Mining: Past, present, and future

Maria Eskevich[1], Stelios Piperidis[2], Kanella Pouli[2],
Maria Gavriilidou[2], Dimitris Galanis[2], Antal van den Bosch[1]

[1]*Radboud University, Nijmegen, Netherlands,*
[2]*Institute for Language and Speech Processing (ILSP), "Athena" R.C., Greece*

Abstract

In this paper we review text and data mining (TDM) as a set of data science techniques that detect information from massive amounts of data, present this information for further analysis, and support data-driven decision making. We summarise state-of-the-art TDM trends from a scientific perspective, and outline their application in industrial setting. We analyse the spread of TDM across diverse economic sectors via such parameters as publication and funding volumes in the context of research within the European Union as an example.

Keywords: Text and Data Mining, Data Science, Big Data, Review, Trends

1. Introduction

Text and Data Mining has been defined as "the discovery by computer of new, previously unknown information, by automatically extracting and relating information from different (...) resources, to reveal otherwise hidden meanings" [17], in other words, "an exploratory data analysis that leads to the discovery of heretofore unknown information, or to answers for questions for which the answer is not currently known" [17]. Nowadays, this umbrella term encompasses diverse techniques that allow for the interpretation of content of any type ranging from raw data, e.g. sensor data, text, images and multimedia, to processed content and structured data, e.g. diagrams, charts, references, maps, formulas, chemical structures, and metadata, on a large scale through the identification of patterns.

Components of TDM existed before the formal introduction of the term within the scientific community. Essentially TDM has grown as part of the general 'data science' field from a number of parent scientific disciplines, notably artificial intelligence and its subdisciplines such as machine learning, pattern recognition, and natural language processing, as well as mathematics, statistics, and information retrieval. It has benefited from multidisciplinary approaches and collaborations across these fields.

The growing quantity of digital content currently produced by society, as well as efforts on the digitization of archives, in combination with the potential ease of access to this data at any location via easy-to-use devices (with reliable Internet connection, and mobile devices) increase both the need for content mining technologies, and their proliferation in the fabric of modern society [29].

TDM is becoming increasingly ubiquitous on the global market, as each company that tracks its activities, products, and communications in some digital form has the potential to reuse this data for further analysis and improvement. For some sectors, countries and applications, it is already possible to estimate the potential profit or at least the size of the market [29, 9]. Other fields such as journalism, which is in the process of embracing the concept of Data Journalism, have to first agree on the conceptual changes required to their current work behaviour in order to merge the traditional work pattern with TDM [20].

This paper is structured as follows. We first survey the scientific trends and outline the most prominent scientific communities that develop the state-of-the-art algorithms and aim to advance beyond those (Section 2). Section 3 provides examples of the TDM scientific work in application across economic structure using the European Union as a case example. Section 4 then discusses future trends and their overall potential.

2. Scientific perspective: Research fields supporting TDM

Text and data mining has gradually evolved into a large multidisciplinary field with complex approaches for a range of applications that can be found in all types of human activities. Historically, the development of mining technologies has its roots with the invention of mathematical models [14] and statistical analysis [8]. Due to the advent of computers and systems for storing larger quantities of data, the possibility to mine large datasets and filter relevant information has greatly expanded what used to be laborious

expert analysis work. Since the beginning of the 1990s, modern data mining tools have appeared on the market and in labs, allowing more advanced analysis, to discover hidden patterns and to extract implicit knowledge in any domain, as envisaged and pioneered for the life science domain by Swanson in [41].

With the invention and adoption of the World Wide Web [4], researchers expanded their efforts to develop text and data mining methods to search and explore the web. This led to the creation of search engines, starting from simple interface-based browsers such as *Mosaic* and *Netscape*, and developed into large-scale systems such as *Google* and *Bing*. This, together with the advances in natural language processing, formed the opening gambit towards more complex approaches of text mining to different fields (such as the biomedical domain, economics, social sciences, or humanities), and types of data and content (such as scientific publications and research data, social media, multimedia, or public sector information).

The core research fields that develop, test, and publish the algorithms for TDM implementations for each specific domain of use, are

- *machine learning (ML)*, which is the automated learning of tasks and the discovery of patterns and structure in data [1];
- *information retrieval (IR) and search*, which focuses on the optimisation of search methods that, based on an index of vast amounts of content, allow for the location and extraction of the most relevant facts answering any question and user type [37]; and
- *natural language processing (NLP)* techniques, which deal with human language in its textual and spoken form [21, 27].

Research progress within TDM-related communities can be traced across two main sources: a number of high-impact yearly conferences (such as Knowledge Discovery and Data Mining (KDD), International Conference on Data Engineering (ICDE), IEEE International Conference on Data Mining (ICDM) for ML; ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) and International Conference on Information and Knowledge Management (CIKM) for IR; Annual meeting of the Association for Computational Linguistics (ACL) and Conference on Empirical Methods on Natural Language Processing (EMNLP) for NLP), and a set of prominent journals in the fields (such as IEEE Transactions on Knowledge and Data Engineering (TKDE), Information Processing Letters (IPL)

for ML; Information Retrieval Journal for IR; Transactions of the Association for Computational Linguistics (TACL) for NLP). With the speed of technology development and the race for implementation of new methods in the context of practical implementations, the publications at conferences, having the fastest turnaround, have a relatively high impact. There is also a noteworthy trend to pre-print publications via open access self-archiving services such as `https://arxiv.org/`, that originally was created for the field of physics, but gradually expanded to include mathematics, computer science, nonlinear sciences, quantitative biology, quantitative finance, and statistics. This pre-print culture has an impact on the publications acceptance at the traditional publication venues that needs to take into account these additional aspects when the reviewing process is organised [42].

2.1. Leading algorithms

A top 10 of algorithms was defined by its leading experts in 2006 [46], and they are still representative for the situation in 2017, with the notable exception of Deep Learning as a method that was dominant in the early 1990s [24, 45, 38, 23], was largely absent in the 2000s, and which returned in the 2010s [10]. We list the 2006 top 10 of methods below, and in Figure 1 we show the growth of citations for the main reference scientific articles that defined those methods using the information available via the Google Scholar service.[1] Across the board, numbers of citations have increased. Support vector machines [44] were already popular in 2006 and took the position as most frequently cited learning algorithm in 2017.

- Classification: C4.5 [36],; CART [6]; K Nearest Neighbours (kNN) [16]; Naive Bayes [49];
- Statistical Learning: SVM [44]; EM [30];
- Association Analysis: Apriori [3]; FP-tree [15];
- Link Mining: PageRank [7]; HITS [22];
- Clustering: k-Means citemacqueen1967; BIRCH [50];
- Bagging and Boosting: AdaBoost [12];

[1] `https://scholar.google.com/`

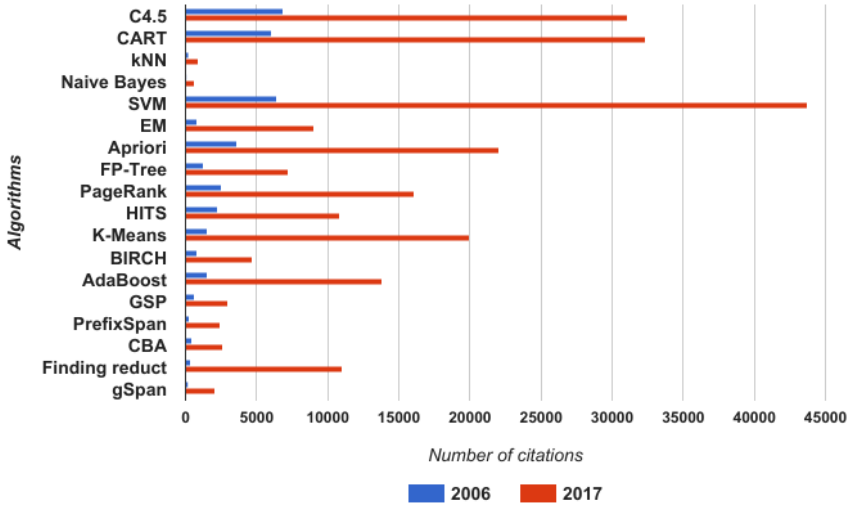

Figure 1: Comparison of the top 10 TDM papers citations: 2006 vs 2017 according to GoogleScholar.

- Sequential Patterns: GSP [39]; PrefixSpan [34];
- Integrated Mining: CBA [25];
- Rough Sets: Finding reduct [33];
- Graph Mining: gSpan [48].

2.2. NLP and IR perspective: main used algorithms and tasks

As IR has gradually developed from the techniques targeting simple keyword search [26] to more sophisticated use of language modeling [18] that requires speech and text pre-processing, and overall language modeling, the NLP and IR fields converge into sharing the algorithms used. Below we enlist the leading techniques used across those fields:

- Latent Dirichlet allocation (LDA) [5] for topic modeling and text classification;

- Hidden Markov Models (HMMs) [28], recurrent neural networks (RNNs) and Long short-term memory (LSTMs) for statistical language modeling as a component in speech recognition [19], machine translation [47], and information retrieval [18];
- Word2Vec [31] and GloVe [35] for text classification, information retrieval [13], machine translation [32];
- CRFs (conditional random fields) [40] and other sequence learners for part-of-speech tagging, Named Entity Recognition;
- CF (collaborative filtering) for recommendation [2];
- TF-IDF, BM25 for information retrieval [37, 27].

3. Tracing TDM application in different scientific areas and fields of activity

The growth of data-driven business, services and research, has led to an introduction of an extended model for the economic structure that reflects a novel type of knowledge-based economy, where knowledge and data become a separate valuable commodity (OCDE, 1996). Thus recently, the research that supports knowledge sharing and growth, as well as education of the professionals that can carry out these activities, have been termed a quaternary economic sector, to be added to the traditional primary, secondary, and tertiary ones. Moreover, the high-level decision makers in governments, large industry companies, and education, having the potential to shape the future of the entire sectors with their vision and decisions, can arguably be placed in a separate, quinary sector.

We focus our investigation on the current and potential presence of TDM across all economic sectors, and thus we envisage these relations in Figure 2. The primary, secondary, and tertiary sectors follow roughly the same type of interaction with TDM technologies. These sectors represent sources of all possible types of data that are available for mining, and at the same time, have tasks and challenges that require smart solutions. The quaternary sector is particularly central to TDM development and implementation, as it produces the TDM experts and advances the technology itself. Experts at the decision-maker level use TDM tools to test and prove their vision of

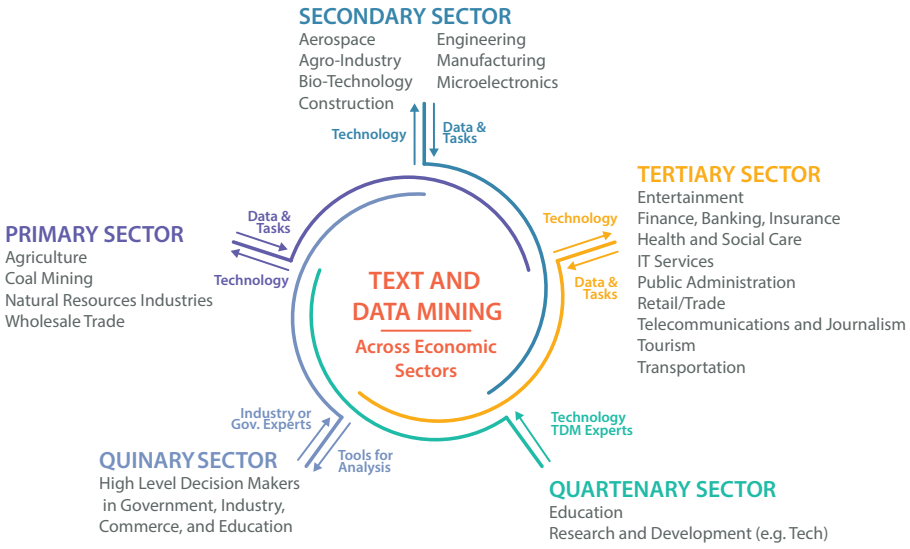

Figure 2: General economic structure and connection between TDM and all economic sectors

economic development, as well as to hypothesize where a specific decision will lead to, and to measure and assess its progress.

With the economic sectors and their relationship with TDM defined for any type of knowledge-based society, we take an example of the European Union infrastructure, and use TDM techniques to outline the scientific span of the TDM work and applications. We employ two axes, namely (i) EU funded research projects and infrastructures, and (ii) scientific articles published in journals and conferences. These two axes are considered as reflecting the strategic decisions of the EU and the policies adopted as regards the selection of specific research domains for funding, as well as the scientific trends attested in the publications.

3.1. Experimental framework

In order to estimate the distribution of scientific publications relating to Text and Data Mining per economic sector, we used 1,195,499 abstracts extracted from the latest dumps that have been made available at the CORE

(COnnecting REpositories) site[2]. From the 1,195,499 abstracts in our dataset, 11504 are related to Text and Data Mining, i.e. 0,96% of the total. This finding is commensurate to the findings of [11] and [43]. These abstracts were analysed using LDA, a probabilistic Bayesian model of text generation. We used the MALLET[3] implementation of LDA and learnt a model of 100 topics from the 1,195,499 abstracts. Each topic t (in our analysis) is represented by a small set of words; the ones that are more likely to belong to t according to the estimated LDA probabilities (word-topic distribution). Based on the word set, each topic was manually assigned an application area label from the respective classification that is depicted in Figure 2.

From all abstracts, we kept (for our analysis) only the ones in which a TDM-related term occurs (e.g. Text Mining, Text Analysis, Text Analytics, Data Mining, Natural Language Processing)[4]. We used the topic-document probabilities that are returned from LDA and indicate which topics are discussed in each abstract. The sector label (e.g. Aerospace) of the LDA topic with the highest topic-document probability is assigned to each abstract.

3.2. Primary sector

The primary sector includes all raw materials and natural resources as well as the methods used for their mining (in the prototypical sense of the word) and/or production.

The main objectives of TDM research in the primary sector include, among others:

- standardisation, understanding and forecasting of natural phenomena (e.g. earthquakes) and climate changes interacting with the lives of humans and other species;
- prediction of better or worse/problematic crops or certain climate phenomena which affect them;
- correlation of pesticide use and animal/human diseases;
- better use of water resources/materials for production optimization.

[2] `https://core.ac.uk/`

[3] `http://mallet.cs.umass.edu/`

[4] The full list of terms is in the Appendix 8

3.2.1. Funding

A total amount of approximately €98,2M, with a European Commission maximum contribution of €75,3M, has been invested in 24 projects concerning TDM in the primary sector. Most projects focus on the collection and processing of data from natural resources, oceanic data, climate data or agricultural data. The following have been selected among the top twenty projects, based on their total investment, as representative cases for the primary sector:

- MARS: Managing Aquatic ecosystems and water Resources under multiple Stress;
- ADAPTAWHEAT: Genetics and physiology of wheat development to flowering: tools to breed for improved adaptation and yield potential;
- SUSTAINMED: Sustainable agri-food systems and rural development in the Mediterranean Partner Countries;

3.2.2. Publications

The majority of the publications of TDM interest (retrieved from the CORE repository) which fall in the primary sector deal with the areas of Agriculture and Environment; other areas (e.g. Natural Resources) are comparatively less researched into while some others (e.g. Wholesale Trade) are not represented in the publications' set, see Figure 3.

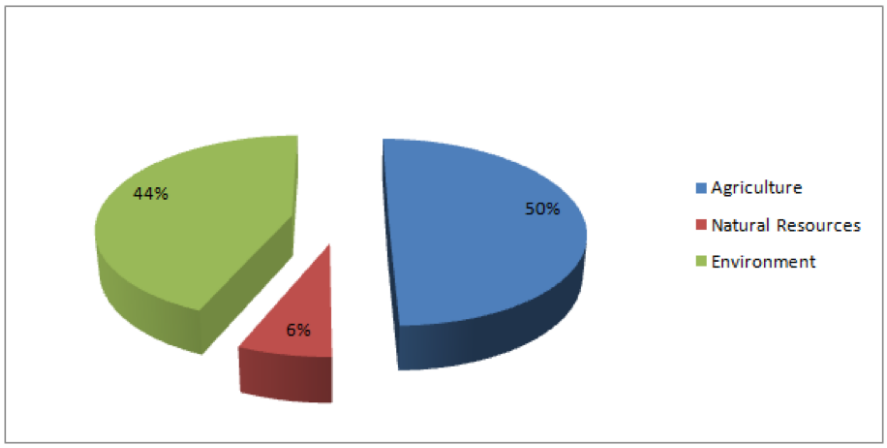

Figure 3: Publications in the primary sector.

3.3. Secondary sector

The secondary sector includes all the methods/techniques/tools/machinery used in order to transform the primary sector materials or resources into goods and products. All human activities related to manufacturing, processing and construction pertain to the secondary sector.

3.3.1. Funding

A total amount of €806,4M, with a European Commission maximum contribution of €633,8M, has been invested on 253 projects concerning TDM in the secondary sector. The majority of projects focus on the collection and processing of biological and medical data for biotechnology applications and the subsequent provision of better health care services. Many projects concern the processing of energy data and the development of green or smart energy systems. The following projects are representative cases for the secondary sector:

- GEN2PHEN: Genotype-To-Phenotype Databases: A Holistic Solution;
- LinkedDesign: Linked Knowledge in Manufacturing, Engineering and Design for Next-Generation Production;
- Fortissimo 2: Factories of the Future Resources, Technology, Infrastructure and Services for Simulation and Modelling 2.

3.3.2. Publications

Half the publications of TDM interest which fall in the secondary sector concern Engineering. The other areas of interest cover Aerospace and Energy followed by Construction and Microelectronics as depicted in Figure 4.

3.4. Tertiary sector

The tertiary sector includes all the services provided to individuals or organized groups such as businesses. All human activities related to retail and wholesale sales, transportation, entertainment, tourism, insurance, and many more pertain to the tertiary sector.

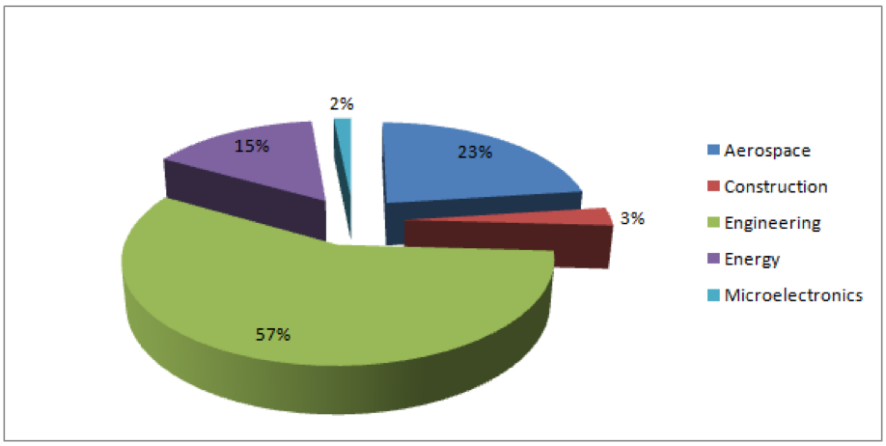

Figure 4: Publications in the secondary sector.

3.4.1. Funding

A total amount of €1,1B, with a European Commission maximum contribution of €852,4M, has been invested on 313 projects concerning TDM in the tertiary sector. Most projects focus on the collection and processing of data from medical records and biology databases. The next category in frequency is that of IT services followed by the business application area. The following projects are representative cases for the tertiary sector:

- SENSEI: Integrating the Physical with the Digital World of the Network of the Future;
- SIIP: Speaker Identification Integrated Project;
- APO-SYS: "Apoptosis systems biology applied to cancer and AIDS. An integrated approach of experimental biology, data mining, mathematical modelling, biostatistics, systems engineering and molecular medicine";
- mPlane: mPlane an Intelligent Measurement Plane for Future Network and Application Management;

3.4.2. Publications

The retrieved publications of TDM interest for the tertiary sector cover a variety of fields as depicted in Figure 5. The majority of publications (60%) pertain to the fields of Medicine, Health Care and Social Care, while 30%

are related to generic IT services. The remaining publications are related to Finance, Public Administration, Entertainment and Transportation.

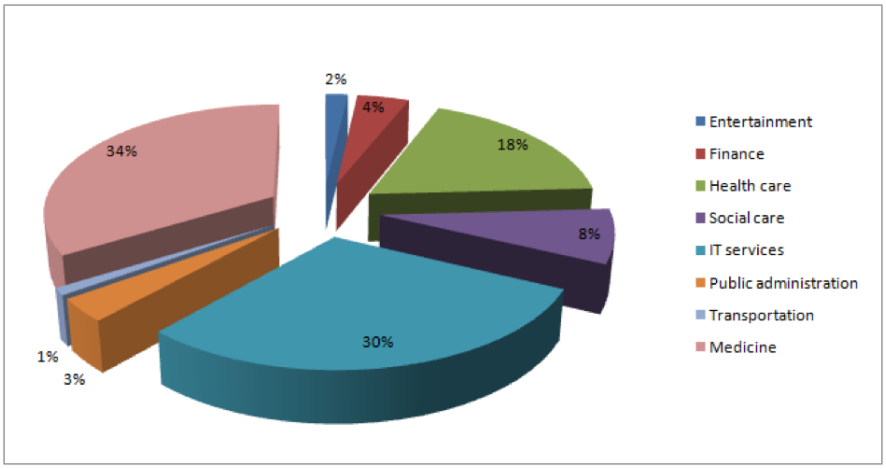

Figure 5: Publications in the tertiary sector.

3.5. Quartenary sector

The quaternary sector includes all services dealing with knowledge and information such as research, development and education among others. Research and development are terms applying to all scientific fields; therefore the following sections include projects, infrastructures, language resources and publications which could also be identified as belonging to any of the first three economic sectors but are presented here because the focus lies on the innovation, introduction, and improvement of products and processes of the various disciplines and application areas.

3.5.1. Funding

A total amount of €402M, with a European Commission maximum contribution of €319,9M, has been invested on 273 projects concerning TDM in the quaternary sector. This sector includes all these projects that are classified by the EC as scientific research under RTD horizontal topics without further distinguishing, in most cases, the particular scientific area, the development of which the particular action will affect. As a result strong infrastructural projects like EUDAT2020 and projects requiring substantial financial investment in the area of Medicine and Health Sciences, e.g. the

METACARDIS project, are grouped together. The following are the representative examples for the tertiary sector:

- METACARDIS: Metagenomics in Cardiometabolic Diseases;
- SCY: Science Created by You;
- NEXT-TELL: Next Generation Teaching, Education and Learning for Life.

3.5.2. Publications

The publications concerning TDM for the quaternary sector, as depicted in Figure 6, fall into two big categories: the first being research, which is a generic term including many scientific fields, and the second one being education. The research publications presented here are those for which no specific scientific field was denoted from the set of words of topic t. Otherwise, the classification of the topic was done in accordance to the application areas of the primary, secondary or tertiary sector.

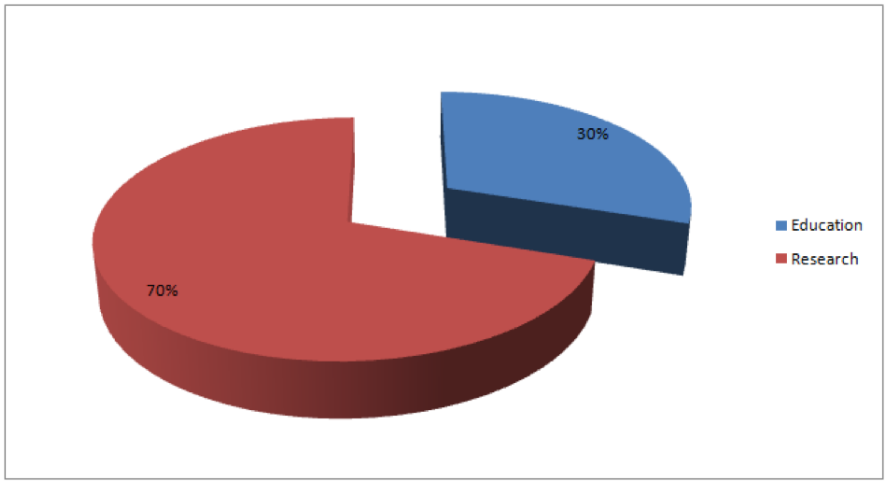

Figure 6: Publications in the quartenary sector.

3.6. Quinary sector

The quinary sector is the highest economic sector encompassing decision making for policy guidelines in Industry, Government, Science and Technology, having a profound impact on economy in general. It is the special

characteristics of this sector and the interrelation with all the other economic sectors, which make it challenging to isolate resources and material pertaining exclusively to the quinary sector itself. Decision makers need all the available data from various types of infrastructures as well as all the TDM software tools and technologies in order to analyse, deploy them, and finally make decisions affecting different application areas within all previous sectors. Taking these peculiarities into consideration, in this paper we present only the EU projects which have been funded under certain RTD Horizontal Topics, specifically those which aim at the formulation or identification of future developments in RTD and long-term strategic options.

3.6.1. Funding

A total amount of €192,1M, with a European Commission maximum contribution of €149M, has been invested on 116 projects concerning TDM in the quinary sector. Most projects focus on the evaluation or application of laws and regulations, policies, policy strategies or plans of action for research and development in science and technology. The following examples show decision making projects that are the most representative for the quinary sector:

- KHRESMOI: Knowledge Helper for Medical and Other Information users
- LOD2: LOD2 - Creating Knowledge out of Interlinked Data
- FIRST: Large scale information extraction and integration infrastructure for supporting financial decision making

4. Conclusion

In this paper we reviewed the most important algorithms that define the state-of-the-art and potential for advancement for the scientific development of TDM. Using LDA analysis on the CORE database, we highlighted the state of TDM research and funding coverage in context of the European Union test case, showcasing the spread of publications per sector and topics that are intertwined with TDM.

5. Acknowledgements

This research is funded by H2020 Project FutureTDM (call GARRI-3-2014, grant agreement 665940).

6. Competing interests

The authors have no competing interests.

7. References

[1] Glossary of terms. *Mach. Learn.*, 30(2-3):271–274, February 1998.

[2] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 17(6):734–749, 2005.

[3] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In *Proc. of 20th Intl. Conf. on VLDB*, pages 487–499, 1994.

[4] Tim Berners-Lee. The world wide web - past, present and future. *J. Digit. Inf.*, 1(1), 1997.

[5] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.

[6] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.

[7] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1):107 – 117, 1998.

[8] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference.* Springer-Verlag, 2002.

[9] Frank Buytendijk. *Hype Cycle for Big Data*. Stamford: Gartner, 2014.

[10] Li Deng and Dong Yu. Deep learning: Methods and applications. *Found. Trends Signal Process.*, 7(3–4):197–387, June 2014.

[11] S. Filippov. Mapping text and data mining in academic and research communities in europe. Technical report, Nuffield College, Oxford, UK, Discussion paper, 2014.

[12] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, August 1997.

[13] Debasis Ganguly, Dwaipayan Roy, Mandar Mitra, and Gareth J.F. Jones. Word embedding based generalized language model for information retrieval. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 795–798, New York, NY, USA, 2015. ACM.

[14] N. Gershenfeld. *The Nature of Mathematical Modeling.* Cambridge University Press, 1998.

[15] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, pages 1–12, New York, NY, USA, 2000. ACM.

[16] Trevor Hastie and Robert Tibshirani. Discriminant adaptive nearest neighbor classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(6):607–616, June 1996.

[17] Marti A. Hearst. Untangling text data mining. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 3–10, Stroudsburg, PA, USA, 1999. Association for Computational Linguistics.

[18] Djoerd Hiemstra. *Using language models for information retrieval.* Univ. Twente, 2001.

[19] Xuedong Huang, Yasuo Ariki, and Mervyn Jack. *Hidden Markov Models for Speech Recognition.* Columbia University Press, New York, NY, USA, 1990.

[20] A. B. Jones and J. M. Smith. Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3:447466, 2014.

[21] Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 2000.

[22] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, September 1999.

[23] Yann LeCun and Yoshua Bengio. The handbook of brain theory and neural networks. chapter Convolutional Networks for Images, Speech, and Time Series, pages 255–258. MIT Press, Cambridge, MA, USA, 1998.

[24] Yann Lecun, Lon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.

[25] Bing Liu, Wynne Hsu, and Yiming Ma. Integrating classification and association rule mining. In *KDD*, pages 80–86, 1998.

[26] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[27] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA, 1999.

[28] A. Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. In R. Howard, editor, *Dynamic Probabilistic Systems (Volume I: Markov Models)*, chapter Appendix B, pages 552–577. John Wiley & Sons, Inc., New York City, 1971.

[29] V. Mayer-Schónberger and K. Cukier. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. John Murray Publishers, 2013.

[30] G. J. McLachlan and D. Peel. *Finite mixture models*. Wiley Series in Probability and Statistics, New York, 2000.

[31] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.

[32] Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. Exploiting similarities among languages for machine translation. *CoRR*, abs/1309.4168, 2013.

[33] Zdzislaw Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, Norwell, MA, USA, 1992.

[34] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Meichun Hsu. Prefixspan: Mining sequential patterns by prefix-projected growth. In *Proceedings of the 17th International Conference on Data Engineering*, pages 215–224, Washington, DC, USA, 2001. IEEE Computer Society.

[35] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.

[36] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[37] C. J. Van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, Newton, MA, USA, 2nd edition, 1979.

[38] P. Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif. Intell.*, 46(1-2):159–216, November 1990.

[39] Ramakrishnan Srikant and Rakesh Agrawal. Mining sequential patterns: Generalizations and performance improvements. In *Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '96, pages 3–17, London, UK, UK, 1996. Springer-Verlag.

[40] Charles Sutton and Andrew McCallum. An introduction to conditional random fields. *Found. Trends Mach. Learn.*, 4(4):267–373, April 2012.

[41] D.R. Swanson. Fish oil, raynauds syndrome, and undiscovered public knowledge. *Perspect. Bio. Med*, 30:7–18, 1986.

[42] Andrew Tomkins, Min Zhang, and William D. Heavlin. Single versus double blind reviewing at WSDM 2017. *CoRR*, abs/1702.00502, 2017.

[43] Hsu-Hao Tsai. Global data mining: An empirical study of current trends, future forecasts and technology diffusions. *Expert Syst. Appl.*, 39(9):8172–8181, July 2012.

[44] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[45] Alexander Waibel, Toshiyuki Hanazawa, Geofrey Hinton, Kiyohiro Shikano, and Kevin J. Lang. Readings in speech recognition. chapter Phoneme Recognition Using Time-delay Neural Networks, pages 393–404. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.

[46] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg. Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14(1):1–37, December 2007.

[47] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144, 2016.

[48] Xifeng Yan and Jiawei Han. gspan: Graph-based substructure pattern mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining*, ICDM '02, pages 721–, Washington, DC, USA, 2002. IEEE Computer Society.

[49] Keming Yu and David J. Hand. Idiot's bayesnot so stupid after all? *International Statistical Review*, 69(3):385–398, 2001.

[50] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: An efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, SIGMOD '96, pages 103–114, New York, NY, USA, 1996. ACM.

8. Appendix: List of TDM related terms and their frequencies in FP7 and HORIZON2020 projects

Keywords in OBJECTIVE	**FP7**	**keywords in OBJECTIVE**	**Horizon 2020**
data analysis	240	big data	66
machine learning	156	data analysis	61
data mining	88	machine learning	51
large data	65	data analytics	21
language processing	39	data mining	14
big data	38	large data	14
information retrieval	37	linked data	6
information extraction	30	business intelligence	5
linked data	30	data science	5
text mining	17	language processing	5
data science	19	information retrieval	4
data analytics	18	predictive analytics	4
information access	17	information access	3
summarization	12	sentiment analysis	2
business intelligence	11	text and data mining	2
knowledge discovery	11	information extraction	1
multimedia processing	8	knowledge discovery	1
conversation analysis	7	summarization	1
opinion mining	7	TDM	1
sentiment analysis	5	abbreviation detection	0
semantic mining	4	competitive intelligence	0
text analytics	3	concept extraction	0
content mining	3	content classification	0
text understanding	3	content mining	0
trend detection	2	conversation analysis	0
graph mining	2	datafication	0
predictive analytics	2	graph mining	0
sense disambiguation	2	linguistic identification	0
content classification	2	modality detection	0
textual entailment	1	multimedia processing	0
datafication	0	negation detection	0
concept extraction	0	opinion mining	0
modality detection	0	semantic mining	0
negation detection	0	sense disambiguation	0
competitive intelligence	0	text analytics	0
linguistic identification	0	text mining	0
abbreviation detection	0	text understanding	0
text and data mining	0	textual entailment	0
TDM	0	trend detection	0

Collection of Organisations

Name	URL	Country	ZIP	City	Street	Type	Data Sources	Sector	Application field Sector I	Application field Sector II	Application field Sector III	Application field Sector IV	Application field Sector V
904Labs	http://904labs.com	The Netherlands	1098XH	Amsterdam	Science Park 402	Company	Business, Internet	Tertiary-Sector, Quartenary-Sector	N/A	Finance, Banking, Insurar		Research and Development (e.g. Tech)	N/A
Accenture	https://www.accenture.com	USA, many intern	N/A	N/A	N/A	Company	Business, Internet	N/A	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
ADAPT	http://adaptcentre.ie/	Ireland	Dublin 2	Dublin	Trinity College, College Green	Research	All	Tertiary-Sector, Quartenary-Sector	N/A	N/A	Entertainment, Finance,	Education, Research and Development (e.g. Tech	N/A
Agris	http://agris.fao.org/agris-search/index.do	Greece	10176	Athens	Acharnon Street 5	Company	Scientific	Primary-Sect Agriculture	N/A	N/A	N/A	N/A	N/A
Agroknow	http://www.agroknow.com/agroknow/	Greece	152 35	Athens	Grammou 17	Company	Scientific	Primary-Sect Agriculture	N/A	N/A	N/A	N/A	N/A
Algoritmica	http://algoritmica.nl/	The Netherlands	1075	Amsterdam	Oranje Nassaulaan 62-2	Company	All	Secondary-Sector	Engineering	N/A	N/A	N/A	N/A
Apteco Limited	http://www.apteco.com/	UK	CV34 4EH	Warwick	Tink-a-Tank House, Jury St 21	Company	All	Secondary-Sector, Tertiary-Sector	Manufacturing	Entertainment, Finance,	Banking, Insurance, Retail/Trade, Telecommunica	N/A	
Artelnics-Artificial Intelligence Techniques, Ltd	http://www.artelnics.com/	Spain	37900	Salamanca	Carretera de Madrid 13	Company	Business, Personal,	Secondary-Sector, Tertiary-Sector, Quarter	Bio-technology	Health and Social Care	Research and Development (e.g. Tech)	N/A	
Autonomy	http://www.autonomy.com/	UK	CB4 0WZ	Cambridge	Cambridge Business Park, Cowley Parl	Company	All	Secondary-Sector, Tertiary-Sector	Manufacturing	Finance, Banking, Insurance,	Health and Social Care, Public Administration	N/A	
AVUXI	http://www.avuxi.com/	Spain/UK	N/A	N/A	N/A	Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	
Batelle	http://www.battelle.org	USA	N/A	Columbus, Ohio	N/A	Company, Re	Business, Scientific	Secondary-Sector, Quartenary-Sector	Bio-technology, Manufacturing			Research and Development (e.g. Tech)	N/A
BG Stat	http://www.bgstats.com/index.html	Austria	1040	Vienna	Schleifmühlgasse 7	Company	Business, Public Sec	Secondary-Sector, Tertiary-Sector, Quarter	Bio-technology	Finance, Banking, Insurar	Education		N/A
Centre for Language Technology (CST)	http://cst.ku.dk/english/	Denmark	2300	Copenhagen S	Njalsgade 136, building 27	Research	Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Clinerion	http://www.clinerion.com	Switzerland	CH-4053	Basel	Margarethenstrasse 47	Company	Business, Scientific	Secondary-Sector, Quartenary-Sector	Bio-technology	N/A	N/A	Research and Development (e.g. Tech)	N/A
Copyright Clearance Center, Inc.	http://www.copyright.com	USA	MA 01923	Danvers	222 Rosewood Drive	Company	Business, Scientific	N/A	N/A	Agro-industry, Bio-technolog	Finance, Banking, Insurar	Research and Development (e.g. Tech)	N/A
CrossRef	http://www.crossref.org	UK/USA	OX1 1BY	Oxford	New Road	Research	Business, Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Data Mind	http://www.datamind.cz/	Czech Republic	170 00	Prague	U Průhonu 22 / 466	Company	All	Tertiary-Sect N/A		N/A	Health and Social Care,	Public Administration, Retail/Trade	N/A
DeepMind	https://www.deepmind.com/	UK/USA				Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Deutsche Forschungszentrum für Künstliche Intelligenz gGmbH (DFKI)	http://www.dfki.de/	Germany	D-66123	Saarbrücken	Stuhlsatzenhausweg 3	Research	Scientific	Quartenary-Sector				Education, Research and Development (e.g. Tech	N/A
Docear	http://www.docear.org	N/A	N/A	N/A	N/A	Company	All	N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
Dutch-Flemish Human Language Technology Agency	http://www.tst-centrale.org/	The Netherlands,	Belgium			Research	Scientific	N/A	N/A	N/A	N/A	N/A	N/A
eParkomat	http://www.eparkomat.com/index.html#	Czech Republic				Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Epinium	http://epinium.com/	Spain				Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Exclusivi Ltd.	http://exclusivi.com/hotels/	Ireland/Greece				Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Expert Systems S.P.A.	http://www.expertsystem.com	IT	41123	Modena	Via Virgilio 48/H	Company	All	Secondary-Sector, Tertiary-Sector	Bio-technology	Entertainment, Finance,	Banking, Insurance, Public Administration		N/A
Fondazione Bruno Kessler (FBK)	http://www.fbk.eu/	Italy	38122	Bolzano	Via Santa Croce 77	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
German National Library of Economics (ZBW)		Germany				Research	Scientific	N/A	N/A	N/A	N/A	N/A	N/A
GESIS - Leibniz-Institute for the Social Sciences	http://www.gesis.org	Germany	68159	Mannheim	Quadrat B2, 1	Research	Personal, Scientific	Tertiary-Sect N/A	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Helsinki University of Technology	http://www.aalto.fi/en/about/history/hi	Finland	2150	Helsinki	Espoo	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
Institut national de recherche en informatique et en automatique (INRIA)	http://www.inria.fr/en/	France	78153	Rocquencourt	BP 105	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
Institut Universitari de Lingüística Aplicada (IULA)	**website:** iula.upf.edu **twitter:** https://tw	Spain	8018	Barcelona	Universitat Pompeu FabraCampus del	Research	Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Institute for Language and Speech Processing (ILSP / "Athena" R.C.)	http://www.ilsp.gr/en	Greece	151 25	Athens	Marousi, Artemidos 6	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) - CNR	**website:** http://www.ilc.cnr.it/en/conten	Italy	56124	Pisa	Italian National Research CouncilPisa I	Research	Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
JISC	http://www.jisc.ac.uk/	UK	BS2 0JA	Bristol	One Castlepark, Tower Hill	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
Karlsruhe Institute of Technology	http://www.kit.edu/	Germany	76131	Karlsruhe	Kaiserstraße 12	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
Komatsu	http://www.komatsu.com.au/AboutKomz	Japan	N/A	Tokyo	N/A	Company	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Kudoos	http://www.kudoos.nl	The Netherlands				Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Linguamatics	http://www.linguamatics.com/	UK	CB4 0WG	Cambridge	Milton Road	Company	Personal, Scientific	Secondary-Sector, Tertiary-Sector, Quarter	Bio-technology	Health and Social Care	Research and Development (e.g. Tech)	N/A	
Ludwig-Maximilians-Universitaet	https://www.en.uni-muenchen.de/index.	Germany	80539	Munich	Geschwister-Scholl-Platz 1	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
Max Planck Computing and Data Facility	**website:** http://www.rzg.mpg.de/ **facebo**	Germany	85748	Garching	Gießenbachstraße 2	Research	Scientific	N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Meta	http://meta.com/	USA				Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
National Centre for Text Mining (NaCTeM)	http://www.nactem.ac.uk/	UK	M1 7DN	Manchester	Princess Street 131	Research	Internet, Personal, S	Secondary-Sector, Tertiary-Sector, Quarter	Bio-technology	Health and Social Care	Education, Research and Development (e.g. Tech	N/A	
Opendatasoft	https://www.opendatasoft.com/	USA/international				Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Paperhive	https://paperhive.org/	Germany				Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Plazi	http://plazi.org/	Switzerland				Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Pubgene	http://www.pubgene.com/	Norway	855	Oslo	Sognsveien 70A	Company	Personal, Scientific	Tertiary-Sector, Quartenary-Sector	Bio-technology	Health and Social Care	N/A	N/A	N/A
Python predictions	http://pythonpredictions.com/	Belgium	B-1150	Brussels	Avenue Roger Vandendriessche 9	Company	Business, Personal	Tertiary-Sect N/A	N/A	N/A	N/A	N/A	N/A
RapidMiner	http://rapidminer.com/	Germany	44227	Dortmund	Stockumer Str. 475	Company	Business, Internet, F	Secondary-Sector, Tertiary-Sector	Bio-technology, Manufacturi	Entertainment, Finance,	Banking, Insurance, IT Services, Retail/Trade		N/A
Royal Institute of Technology (KTH)	http://www.kth.se	Sweden	114 28	Stockholm	Brinellvägen 8	Research	Scientific	Quartenary- N/A	N/A	N/A	N/A	Education, Research and Development (e.g. Tech	N/A
SafeToNet Ltd	https://safetonet.com/	UK				Company	Business, Scientific	N/A	N/A	N/A	N/A	N/A	N/A
SAP	http://sap.com/	Germany	69190	Walldorf	Hasso-Plattner-Ring 7	Company	All	Primary-Sect Natural Resources Mining, WI	Engineering, Manufacturing	Entertainment, Finance,	Research and Development (e.g. Tech)	N/A	
Scholarly Publishing and Academic Resources Coalition (SPARC)	http://sparcopen.org	USA	N/A	Washington, DC	N/A	Research	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Scyfer	http://scyfer.nl	The Netherlands	1098XH	Amsterdam	Science Park 402	Company	Business, Internet, F	Tertiary-Sector, Quartenary-Sector	N/A	N/A	Health and Social Care,	Research and Development (e.g. Tech)	N/A
Slovenian Social Science Data Archives	http://cessda.net/	Slovenia	1000	Ljubljana	Kongresni trg 12	Research	Personal, Scientific	Tertiary-Sector, Quartenary-Sector	N/A	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A
Sparrho	https://www.sparrho.com	USA	N/A	N/A	N/A	Company	Business, Scientific	Secondary-Sector, Quartenary-Sector	Bio-technology	N/A		Education, Research and Development (e.g. Tech	N/A
Språkbanken, The Swedish language bank	http://spraakbanken.gu.se	Sweden				Research	Scientific	N/A	N/A	N/A	N/A	N/A	N/A
Squares on blue	http://www.squaresonblue.com	France	78000	Versaille	Rue Saint Honoré 13	Company	All	Tertiary-Sector	N/A	Finance, Banking, Insurance,	Health and Social Care, Public Administration	N/A	
Statoo Consulting	http://www.statoo.com	Switzerland	3018	Bern	Morgenstrasse 129	Company	Business, Internet,	Personal, Scientific	N/A	Finance, Banking, Insurance,	Health and Social Care	N/A	
Strategic Feed	http://www.strategicfeed.com/	France	92 088	Paris	Place de la Pyramide 5	Company	Business, Personal,	Tertiary-Sector, Quartenary-Sector	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A	
Swiss Centre of Expertise in the Social Sciences	http://forscenter.ch/en/	Switzerland	CH-1015	Lausanne	Bâtiment Géopolis	Research	Personal, Scientific	Tertiary-Sector, Quartenary-Sector	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A	
Synergic Partners	http://www.synergicpartners.com/en/	Spain	28006	Madrid	Calle Velázquez 92	Company	Business, Internet, F	Tertiary-Sector	N/A	Finance, Banking, Insurance,	Health and Social Care, IT Services, Public Ad	N/A	
Textkernel	http://www.textkernel.com/	The Netherlands	NL-1022 AB	Amsterdam	Nieuwendammerkade 28A-17	Company	Business, Internet, F	Secondary-Sector, Tertiary-Sector	Manufacturing	Finance, Banking, Insurance,	IT Services, Public Administration, Retail/Trad	N/A	
They say	http://www.theysay.io	UK	EC3V 3ND	London	Cornhill 24	Company	Business, Internet	Secondary-Sector	N/A	Finance, Banking, Insurance,	Retail/Trade	N/A	
Transinsight	http://transinsight.com/	Germany	D-01307	Dresden	Tatzberg 47-51	Company	Business, Internet, F	Tertiary-Sector, Quartenary-Sector	N/A	Health and Social Care	Research and Development (e.g. Tech)	N/A	
Trendminer	https://www.trendminer.com/	Belgium, The Netherlands, Germany, USA				Company	Business, Scientific		N/A	N/A	N/A	N/A	N/A
UK Data Archive	http://www.data-archive.ac.uk/	UK	CO4 3SQ	Essex	Wivenhoe Park	Research	All	Tertiary-Sector, Quartenary-Sector	N/A	Finance, Banking, Insurar	Research and Development (e.g. Tech)	N/A	
University of Bristol	http://www.bristol.ac.uk/	UK	BS8 1TH	Bristol	Tyndall Avenue	Research	Scientific	Quartenary-Sector	N/A	N/A	Education, Research and Development (e.g. Tech	N/A	
UNSILO	https://site.unsilo.com/site/	Denmark				Company	Business	N/A	N/A	N/A	N/A	N/A	N/A
Vadis	http://www.vadis.com/	Belgium	1190	Brussels	Boulevard de l'Humanité	Company	Business, Personal	Tertiary-Sector, Quartenary-Sector	N/A	Finance, Banking, Insurar	Research and Development (e.g. Tech)	N/A	
Verisk	http://www.verisk.com	USA	NJ 07310-168	Jersey City	545 Washington Boulevard	Company	All	N/A	N/A	N/A	Finance, Banking, Insurance,	Health and Social Care, Public Administration	N/A
VisTrails	http://www.vistrails.org	USA	UT 84112-905	Salt Lake City	N/A	Company	All	Quartenary- N/A	N/A	N/A	N/A	Research and Development (e.g. Tech)	N/A
Xerox Lab	http://www.erce.xerox.com/	France	38240	Meylan	6 chemin de Maupertuis	Research	Business, Internet, F	Tertiary-Sect N/A	N/A	N/A	Finance, Banking, Insurar	Research and Development (e.g. Tech)	N/A

Collection of Projects

Project Acronym	Project Title	URL	Type	Coordinator	Partners	Data Sources	Sector	Application field	Application field	Application field	Application field Sector	Application field
(+)Spaces	Policy Simulation in Virtual Spaces	N/A	Funded	Ibm Israel - Science And Technology Ltd	7	All	N/A	N/A	N/A	N/A		
ACTIVE	ACTIVE: Enabling the Knowledge Powered Enter	http://www.active-project.eu/	Funded	British Telecommunications Public Limited Company	11	Business	Tertiary-Sector	N/A	N/A	Telecommunications and Journalism		
ADMIRE	Advanced Data Mining and Integration Research	http://www.admire-project.eu/	Funded	The University Of Edinburgh	5	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
ADMIT	Agent-Oriented Distributed Data Mining using C	N/A	Funded	Technische Universitaet Clausthal	-	All	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
ADVANCE	Advanced predictive-analysis-based decision-sup	http://www.advance-logistics.eu/	Funded	Magyar Tudomanyos Akademia Szamitastechnikai Es Automatizalasi Kutatointezet	4	Business, Scientif	Secondary-Sector	N/A	Engineering	IT Services		
ALLEGRO	Active large-scale learning for visual recognition	N/A	Funded	Institut National De Recherche En Informatique Et En Automatique (Inria)	1	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
ALMOND	Algorithms and Tools for Mining Biological Sequ	N/A	Funded	Ben-Gurion University Of The Negev	-	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
Aparsen	Alliance for Permanent Access to the Records of	http://www.alliancepermanentaccess.org	Funded	Science And Technology Facilities Council	33	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
APO-SYS	Apoptosis systems biology applied to cancer and	N/A	Funded	Karolinska Institutet	23	Scientific	Secondary-Sector	N/A	Bio-technology	N/A		
ARSINFORMATICA	Artificial intelligence, branching processes and c	N/A	Funded	Politechnika Slaska	~	Scientific	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Development (e.g. Tech)	
ARTIAC	The Affective Role of the Teachers in the Inform	N/A	Funded	Tel Aviv University	~	Internet	Tertiary-Sector, C	N/A	N/A	IT Services	Education	
AUTOUNIMO	Automotive Production Engineering Unified Pers	http://autounimo.aei.polsl.pl/	Funded	Politechnika Slaska	3	Business, Scientif	Secondary-Sector	N/A	Engineering	N/A	Research and Development (e.g. Tech)	
AVERT-IT	Advanced Arterial Hypotension Adverse Event p	N/A	Funded	Pera Innovation Limited	9	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
BALTICGRID-II	Baltic Grid second phase	N/A	Funded	Kungliga Tekniska Hogskolan	12	Public Sector Info	Tertiary-Sector	N/A	N/A	IT Services		
BI4MASSES	Business Intelligence for the Masses	N/A	Funded	Ozyegin University	~	Business, Scientif	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
BigDataEurope	BigDataEurope	http://www.big-data-europe.eu/	Funded	Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V	12	All	N/A	Agriculture	Agro-industry, Bi	Health and Social Care, IT Services, Transportation		
BIOLITCONTEXTMINING	Contextual Text Mining from the Biomedical Scie	N/A	Funded	Bogazici Universitesi	-	Scientific	Secondary-Sector	N/A	Bio-technology	N/A		
BISON	Bisociation Networks for Creative Information D	N/A	Funded	Universitat Konstanz	7	Business, Scientif	Secondary-Sector	N/A	Bio-technology	IT Services	Research and Development (e.g. Tech)	
BRAVEHEALTH	Patient Centric Approach for an Integrated, Ada	N/A	Funded	Labor S.R.L.	16	Scientific	Tertiary-Sector	N/A	N/A	Health and Social Care, IT Services		
CLARIN	Common Language Resources and Technology Ir	http://www.clarin.eu	Funded	Universiteit Utrecht	36	All	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
CLARIN-PLUS	Strengthening the CLARIN Infrastructure.	http://www.clarin.eu	Funded	Clarin Eric	4	All	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
CYANOFACTORY	Design, construction and demonstration of solar	http://www.cyanofactory.eu/	Funded	Uppsala Universitet	9	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
DebugIT	Detecting and Eliminating Bacteria UsinG Inform	http://www.debugit.eu/home.html	Funded	Agfa Healthcare N.V.	13	Business, Scientif	Secondary-Sector	N/A	Bio-technology	Health and Social	Research and Development (e.g. Tech)	
Dicode	Mastering Data-Intensive Collaboration and Dec	http://dicode-project.eu/	Funded	Computer Technology Institute & Press Diophantus	7	All	N/A	N/A	N/A	N/A		
digital.me	Integrated digital.me Userware for the Intelligent, Intuitive, and Trust-Enhancing Management (		Funded	Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V	7	Business, Scientif	Tertiary-Sector	N/A	N/A	Health and Social Care, IT Services		
DIMENSION	The Role of Dimension in Metric Embedding	N/A	Funded	Ben-Gurion University Of The Negev	~	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
DISTRIBUTION TESTING	Algorithms for Testing Properties of Distribution	N/A	Funded	Tel Aviv University	~	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
DRIVER II	Digital Repository Infrastructure Vision for Euro	http://search.driver.research-infrastructures.e	Funded	National And Kapodistrian University Of Athens	28	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
e-LICO	e-Laboratory for Interdisciplinary Collaborative F	http://www.e-lico.eu/	Funded	Universite De Geneve	8	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
E-SWARM	Engineering Swarm Intelligence Systems	N/A	Funded	Universite Libre De Bruxelles	~	Scientific	Secondary-Sector	N/A	Engineering	N/A	Research and Development (e.g. Tech)	
ENANOMAPPER	eNanoMapper - A Database and Ontology Frame	http://www.enanomapper.net	Funded	Douglas Connect Gmbh	8	Business, Scientif	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
EPIHEALTH	Linking perturbed maternal environment during	http://epihealth.biotalentum.eu/	Funded	Biotalentum Tudasfejleszto Kft	6	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
EPIHEALTHNET	EpiHealthNet	N/A	Funded	Universiteit Utrecht	8	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
ERACEP	Emerging Research Areas and their Coverage by	http://www.eracep.eu	Funded	Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V	1	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
EU-ADR	Exploring and understanding adverse drug react	N/A	Funded	Erasmus Universitair Medisch Centrum Rotterdam	13	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
EUDAT2020	EUDAT: the collaborative Pan-European infrastru	https://www.eudat.eu/	Funded	Csc-Tieteen Tietotekniikan Keskus Oy	34	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
EURECA	Enabling information re-Use by linking clinical RE	http://eurecaproject.eu/	Funded	Philips Electronics Nederland B.V.	16	Scientific	Secondary-Sector	N/A	Bio-technology	Health and Social	Research and Development (e.g. Tech)	
EuroRIs-Net+	European Research Infrastructures Network of N	N/A	Funded	Ethniko Idryma Erevnon	36	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
FLARENET	Fostering Language Resources Network	www.flarenet.eu	Funded	Institute For Computational Linguistics	38	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
FORSIED	Formalizing Subjective Interestingness in Explora	http://www.interesting-patterns.net/forsied/	Funded	Universiteit Gent	1	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
Foster	Facilitate Open Science Training for European R	http://www.fosteropenscience.eu/	Funded	Universidade Do Minho	13	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
FREME	Open framework of e-services for multilingual ar	http://www.freme-project.eu/	Funded	Deutsche Forschungszentrum Für Künstliche Intelligenz Gmbh (Dfki)	8	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Development (e.g. Tech)	
Fusepool	Fusing and pooling information for product/serv	http://www.fusepool.eu/	Funded	Berner Fachhochschule	7	Business	Tertiary-Sector	N/A	N/A	Tourism		
GAMES	Green Active Management of Energy in IT Servic	N/A	Funded	Engineering - Ingegneria Informatica Spa	7	Scientific	Tertiary-Sector	N/A	N/A	IT Services		
GENODED	Genomics data mining for the genetic analysis o	N/A	Funded	Universidad Politecnica De Madrid	-	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Development (e.g. Tech)	
GEOSTREAM	Exploiting User-Generated Geospatial Content S	N/A	Funded	Athena Research And Innovation Center In Information Communication & Knowledge	15	Business, Internet	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Development (e.g. Tech)	
GLASS	Global Lakes Sentinel Services	http://www.glass-project.eu/	Funded	Water Insight Bv	7	Business, Scientif	Primary-Sector, C	Natural Resource	N/A	N/A	Research and Developm	N/A
GLODERS	The Global Dynamics of Extortion Racket System	http://www.gloders.eu/	Funded	University Of Surrey	3	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
GS-A-DM-DS	Grey Systems and Its Application to Data Mining	N/A	Funded	De Montfort University	-	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
HEALS	Health and Environment-wide Associations base	N/A	Funded	Universite Pierre Et Marie Curie - Paris 6	31	Business, Scientif	Secondary-Sector	Agriculture, Natu	Bio-technology	N/A	Research and Developm	N/A
HiPerDNO	High Performance Computing Technologies for S	N/A	Funded	Brunel University	10	Business, Scientif	Secondary-Sector	N/A	Bio-technology, E	Health and Social	Research and Developm	N/A
I3RES	ICT-based Intelligent management of Integrated	N/A	Funded	Instalaciones Inabensa Sa	7	Business, Scientif	Secondary-Sector	N/A	Engineering	N/A	N/A	N/A
IAM4MARS	Intelligent Automated Methods for Monitoring A	N/A	Funded	Uluslararasi Antalya Universitesi	~	Scientific	Secondary-Sector	N/A	Aerospace	N/A	Research and Developm	N/A
ICON	Inductive Constraint Programming	http://www.icon-fet.eu/	Funded	Katholieke Universiteit Leuven	4	Scientific	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
ICSI	Intelligent Cooperative Sensing for Improved tra	http://www.ict-icsi.eu/	Funded	Intecs Spa	9	Business, Public Sector Information		N/A	N/A	Transportation	N/A	N/A
IHEARU	Intelligent systems' Holistic Evolving Analysis of	N/A	Funded	Technische Universitaet Muenchen	~	Business, Scientif	Tertiary-Sector, C	N/A	N/A	Telecommunicati	Research and Developm	N/A
IMAGELINK	Multiparametric tumor imaging and beyond: Towards understanding in vivo signals		Funded	Eberhard Karls Universitaet Tuebingen	~	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
IMARS	iMars: Analysis of Mars multi-resolution images	N/A	Funded	University College London	5	Scientific	Secondary-Sector	N/A	Aerospace	N/A	N/A	N/A
IMPACT	IMProving ACcess to Text	http://www.csr-impact.eu	Funded	Koninklijke Bibliotheek	27	Business	Tertiary-Sector	N/A	N/A	Finance, Banking,	N/A	N/A
INSEARCH	Supporting SMEs innovation processes through	N/A	Funded	Innovation Engineering Srl	7	Business	N/A	N/A	N/A	N/A	N/A	N/A
INSIGHT	Innovative Policy Modelling and Governance To	N/A	Funded	Universidad Politecnica De Madrid	7	Scientific	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
JMUGCS	Jointly Mining User Generated Content Sources	N/A	Funded	University Of Ioannina	-	Internet	Tertiary-Sector	N/A	N/A	Entertainment, H	N/A	N/A
JOINUS	Joint Inference with the Universal Schema	N/A	Funded	University College London	~	All	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
KAP	Knowledge, Awareness and Prediction of Man, M	http://astute-project.eu/content/kap-project	Funded	Sap Ag	13	Business	Secondary-Sector	N/A	Manufacturing	N/A	N/A	N/A
KYOTO	Knowledge Yielding Ontologies for Transition-ba	http://kyoto-project.eu/	Funded	Stichting Vu-Vumc	8	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
LANGTERRA	Enhancing the Research Potential of ILSP/"Ather	http://www.langterra.eu/	Funded	Athena Research And Innovation Center In Information Communication & Knowledge T	~	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
LAST	Large Scale Privacy-Preserving Technology in the	N/A	Funded	Bar Ilan University	~	Business, Internet	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
LEARN	Leaders Activating Research Networks: Impleme	http://learn-rdm.eu/	Funded	University College London	5	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
LEAs BOX	Tools for Competence-Centred, Multi-Source Le	N/A	Funded	Technische Universitaet Graz	3	Scientific	Quartenary-Secto	N/A	N/A	N/A	Education	N/A
LinDA	Enabling Linked Data and Analytics for SMEs by	http://linda-project.eu/	Funded	National Technical University Of Athens - Ntua	7	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
Live+Gov	Reality Sensing, Mining and Augmentation\nfor	http://liveandgov.eu/	Funded	Universitaet Koblenz-Landau	5	Personal, Public S	Tertiary-Sector	N/A	N/A	Public Administra	N/A	N/A
LTE-HEALTH	Long-Term e-Health Evolution for Improving Dia	N/A	Funded	Kingston University Higher Education Corporation	~	Business, Scientif	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
MAESTRA	Learning from Massive, Incompletely annotated,	http://maestra-project.eu/	Funded	Institut Jozef Stefan	4	Internet	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
MANTRA	Multilingual Annotation of Named Entities and T	http://www.mantra-project.eu/	Funded	Universitaet Zuerich	5	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
Meta (T4ME Net)	Multilingual Europe: a Technology Alliance	http://www.meta-net.eu/events/meta-forum-2	Funded	Deutsches Forschungszentrum Fuer Kuenstliche Intelligenz Gmbh	14	Internet, Persona	Tertiary-Sector, C	N/A	N/A	Entertainment, IT	Education	N/A
MIGENOF113	MINING THE GENOME OF P. FLUORESCENS F113	N/A	Funded	Universidad Autonoma De Madrid	-	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
MIGRANT	Mining Graphs and Networks: a Theory-based ap	N/A	Funded	Katholieke Universiteit Leuven	-	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
MININEXACT	Exact Mining from In-Exact Data	N/A	Funded	Ibm Research Gmbh	-	Business, Scientif	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
MIRRI	Microbial Resource Research Infrastructure - MI	http://www.mirri.org/	Funded	Leibniz-Institut Dsmz-Deutsche Sammlung Von Mikroorganismen Und Zellkulturen Gm	15	Business, Scientif	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
MODAP	Mobility, Data Mining, and Privacy	http://www.modap.org/	Funded	Sabanci University	10	Internet, Persona	Tertiary-Sector	N/A	N/A	Transportation	N/A	N/A
MOLD4PRODE	Intelligent Mold for Productivity Enhancement	N/A	Funded	Association Pole Europeen De Plasturgie	25	Scientific	Secondary-Sector	N/A	Engineering	N/A	Research and Developm	N/A
MONOTRANS	Statistical Machine Translation Using Monolingu	N/A	Funded	Universitat Rovira I Virgili	-	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
mPlane	mPlane – an Intelligent Measurement Plane for	N/A	Funded	Politecnico Di Torino	15	Business, Scientif	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
MULTIFLOW	Multiscale dynamics of turbulent flows	N/A	Funded	Universidad Politecnica De Madrid	~	Scientific	Tertiary-Sector, C	N/A	Engineering	N/A	Research and Developm	N/A
MULTIJEDI	Multilingual Joint Word Sense Disambiguation	N/A	Funded	Universita Degli Studi Di Roma La Sapienza	-	Business, Scientif	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
MULTISENSOR	Mining and Understanding of multilinguaL conte	http://www.multisensorproject.eu/	Funded	Ethniko Kentro Erevnas Kai Technologikis Anaptyxis	8	All	Tertiary-Sector, C	N/A	N/A	Telecommunicati	Research and Developm	N/A
NANO2MARKET	Best Practices for IPR and Technology Transfer ir	http://www.nano2market.eu/	Funded	Universidad De Alicante	9	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
NETSIM	An integrated study on three novel regulatory h	N/A	Funded	The Chancellor, Masters And Scholars Of The University Of Cambridge	9	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
NHECD	Nano health-environment commented database	http://www.nhecd-fp7.eu	Funded	Tel Aviv University	3	Business, Scientif	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
NRG4CAST	Energy Forecasting	http://www.nrg4cast.org/	Funded	Institut Jozef Stefan	7	Business, Scientif	Secondary-Sector	N/A	Construction	Tourism, Transpo	Research and Developm	N/A
ONTIC	Online Network Traffic Characterization	http://ict-ontic.eu/	Funded	Universidad Politecnica De Madrid	6	Business, Scientif	Secondary-Sector	N/A	Engineering	Transportation	N/A	N/A
OpenAire2020	Open Access Infrastructure for Research in Euro	https://www.openaire.eu	Funded	Ethniko Kai Kapodistriako Panepistimio Athinon	50	All	Quinary-Sector	N/A	N/A	N/A	Research and Developm	High Level Decisio
Openminted	Open Mining Infrastructure for Text and Data	http://openminted.eu/	Funded	Athena Research And Innovation Center In Information Communication & Knowledge 1	16	Business, Public S	Primary-Sector, S	Agriculture, Natu	Agro-industry, Bi	Health and Social	Research and Developm	N/A
OraMod	VPH based predictive model for oral cancer reoc	http://www.oramod.eu/	Funded	Universita Degli Studi Di Parma	7	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
Organic.Lingua	Demonstrating the potential of a multilingual W	http://www.organic-lingua.eu/	Funded	Universidad De Alcalá	11	Business, Scientif	Tertiary-Sector, C	N/A	N/A	IT Services	Education, Research and	N/A
PASTEUR4OA	Open Access Policy Alignment STrategies for Eur	http://www.pasteur4oa.eu	Funded	Ethniko Idryma Erevnon	15	All	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	High Level Decisio
PHENOMINER	Semantic mining of phenotype associations from	N/A	Funded	European Molecular Biology Laboratory	-	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	N/A	N/A
PHIDM	Persistent Homology - Images, Data and Maps	N/A	Funded	Institute Of Science And Technology Austria	~	Scientific	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
PONTE	Efficient Patient Recruitment for Innovative Clini	http://www.ponte-project.eu/	Funded	Centre D'Excellence En Technologies De L'Information Et De La Communication	8	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
POPDAT	Problem-oriented Processing and Database Crea	http://www.popdat.org/	Funded	Technische Universitat Berlin	7	Scientific	Secondary-Sector	N/A	Aerospace	N/A	Research and Developm	N/A
PPI-MARKER	Interruption of protein-protein interaction and n	N/A	Funded	Technische Universitaet Dresden	2	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
PPIDR	Privacy-Preserving Intrusion Detection and Resp	N/A	Funded	Ecole Polytechnique Federale De Lausanne	~	Internet	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
PROLOGUE	Promoting real life Observations for Gaining Unc	N/A	Funded	Stichting Wetenschappelijk Onderzoek Verkeersveiligheid	8	Scientific	Tertiary-Sector, C	N/A	N/A	Transportation	Research and Developm	N/A
PROT-HISPRA	HIgh Speed PRoteomics Analysis (Prot-HiSPRA): !	N/A	Funded	Medizinische Universitaet Wien	6	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
PSIP	PATIENT SAFETY THROUGH INTELLIGENT PROCE	N/A	Funded	Centre Hospitalier Regional Et Universitaire De Lille	12	Scientific	Tertiary-Sector, C	N/A	N/A	Health and Social	Research and Developm	N/A
PSYCHE	Personalised monitoring SYstems for Care in me	http://www.psyche-project.org/	Funded	Universita Di Pisa	9	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
RDA Europe	Research Data Alliance - Europe 3	https://rd-alliance.org/node	Funded	Max Planck Gesellschaft Zur Foerderung Der Wissenschaften E.V.	11	Business, Scientif	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
REACTOPS	Reactive Operations Models for a Changing Wor	N/A	Funded	Universidad De Navarra	~	Business, Scientif	Secondary-Sector	N/A	Manufacturing	N/A	Research and Developm	N/A
Recode	The Policy RECommendations for Open Access t	http://recodeproject.eu/	Funded	Trilateral Research & Consulting Llp	8	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
REWIRE	Rehabilitative Wayout In Responsive home Envir	https://sites.google.com/site/projectrewire/ho	Funded	Universita Degli Studi Di Milano	11	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
Scape	Scalable Preservation Environments	http://scape-project.eu/	Funded	Ait Austrian Institute Of Technology Gmbh	21	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
SCIIMS	Strategic Crime and Immigration Information Ma	http://www.sciims.co.uk/	Funded	Bae Systems Integrated System Technologies Ltd	8	Public Sector Info	Tertiary-Sector	N/A	N/A	IT Services	N/A	N/A
SCY	Science Created by You (SCY)	http://scycom.collide.info/	Funded	Universiteit Twente	16	Scientific	Tertiary-Sector, C	N/A	N/A	IT Services	Education	N/A
SEEK	Semantic EnrichmEnt of trajectory Knowledge di	http://www.seek-project.eu	Funded	Consiglio Nazionale Delle Ricerche	2	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
SIIP	Speaker Identification Integrated Project	N/A	Funded	Verint Systems Ltd	16	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
SKIN TREAT	Novel approaches for the development of custo	http://www.skintreat.eu	Funded	Ahava Dead Sea Laboratories Ltd	14	Business, Scientif	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
SMART	Sustainable Mobility and Robust Transportation	N/A	Funded	Technion - Israel Institute Of Technology	~	Business, Scientif	Tertiary-Sector, C	N/A	N/A	Transportation	Research and Developm	N/A
SMARTMUSEUM	Cultural Heritage Knowledge Exchange Platform	http://www.smartmuseum.eu/	Funded	Apprise Ou	7	Cultural Heritage	Tertiary-Sector, C	N/A	N/A	Telecommunicati	Education	N/A
SME E-COMPASS	E-COMmerce Proficient Analytics in Security and	N/A	Funded	Exodus A.E.	12	Scientific	Secondary-Sector	N/A	Aerospace	N/A	Research and Developm	N/A
SNIPER	Statistical Network-based Inference for Proteom	N/A	Funded	Katholieke Universiteit Leuven	~	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
SPECMATE	Specification Mining and Testing	N/A	Funded	Universitaet Des Saarlandes	-	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
STREAM	Scalable Autonomic Streaming Middleware for R	http://www.streamproject.eu/	Funded	Universidad Politecnica De Madrid	5	Business	Tertiary-Sector	N/A	N/A	IT Services	N/A	N/A
SUPREME	SUstainable PREdictive Maintenance for manufa	N/A	Funded	Centre Technique Des Industries Mecaniques	9	Business, Scientif	Secondary-Sector	N/A	Manufacturing	N/A	N/A	N/A
SUSTAINABLE	Smart Distribution System OperaTion for MAxim	http://www.sustainableproject.eu/	Funded	Edp Distribuicao Energia Sa	7	Business, Scientif	Secondary-Sector	N/A	Construction	N/A	Research and Developm	N/A
SYSTEMS MICROSCOPY	Systems microscopy – a key enabling methodolo	http://www.systemsmicroscopy.eu	Funded	Karolinska Institutet	10	Business, Scientif	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
TIMET	Linking the clock to metabolism	N/A	Funded	The University Of Edinburgh	8	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
TRANSFoRm	TRANSLATIONAL RESEARCH AND PATIENT SAFET	N/A	Funded	King's College London	21	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
TRANSMED	Translating basic research findings in major hum	N/A	Funded	"Biomedical Research Foundation, Academy Of Athens"	~	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A
TREELIM	A functional explanation of low temperature tre	N/A	Funded	Universitaet Basel	~	Scientific	Tertiary-Sector, C	N/A	Bio-technology	N/A	Research and Developm	N/A
TrendMiner	Large-scale, Cross-lingual Trend Mining and Sum	http://www.trendminer-project.eu/	Funded	Deutsches Forschungszentrum Fuer Kuenstliche Intelligenz Gmbh	11	Personal	Tertiary-Sector	N/A	N/A	Entertainment	N/A	N/A
TRIBUTE	Take the energy bill back to the promised buildir	http://www.tribute-fp7.eu/	Funded	Csem Centre Suisse D'Electronique Et De Microtechnique Sa - Recherche Et Developpe	16	Business, Scientif	Secondary-Sector	N/A	Construction	N/A	Research and Developm	N/A
URBAN SENSING	Urban Sensing through User Generated Content	http://urban-sensing.eu	Funded	Accurat Srl	6	Business, Scientif	Secondary-Sector	N/A	Construction	Tourism, Transpo	Research and Developm	N/A
VERCE	Virtual Earthquake and seismology Research Cor	N/A	Funded	Centre National De La Recherche Scientifique	11	Scientific	Tertiary-Sector, C	N/A	Engineering	N/A	Research and Developm	N/A
VIALACTEA	VIALACTEA - The Milky Way as a Star Formation	N/A	Funded	Istituto Nazionale Di Astrofisica	8	Scientific	Secondary-Sector	N/A	Aerospace	N/A	Research and Developm	N/A
Visioneer	Envisioning a Socio-Economic Knowledge Collide	http://www.visioneer.ethz.ch/	Funded	Eidgenössische Technische Hochschule Zürich	~	Scientific	Quartenary-Secto	N/A	N/A	N/A	Research and Developm	N/A
ViSTA-TV	Video Stream Analytics for Viewers in the TV Ind	N/A	Funded	Universitaet Zuerich	6	Internet	Tertiary-Sector, C	N/A	N/A	IT Services	Research and Developm	N/A
X-Like	Cross-lingual Knowledge Extraction	http://www.xlike.org/	Funded	Institut Jozef Stefan	7	Business, Scientif	Tertiary-Sector, C	N/A	N/A	Telecommunicati	Research and Developm	N/A
ZHOU_PFL_POLYKETIDES	"Genome mining in Streptomyces violaceusniger	N/A	Funded	The Chancellor, Masters And Scholars Of The University Of Cambridge	-	Scientific	Secondary-Sector	N/A	Bio-technology	N/A	Research and Developm	N/A

Project Acronym	Methods	Tools	Organisations	URL Cordis	Programme	Start	End	Twitter	Facebook	LinkedIn
(+)Spaces	Data Mining, Multimedia Processing	N/A	Ibm Israel - Science And Technology Ltd	http://cordis.europa.eu/projec	FP7-ICT	01.01.2010	30.09.2012			
ACTIVE	Text Mining	N/A	British Telecommunications Public Limited Company	http://cordis.europa.eu/projec	FP7-ICT	01.03.2008	28.02.2011			
ADMIRE	Data Mining	N/A	The University Of Edinburgh	http://cordis.europa.eu/projec	FP7-ICT	01.03.2008	31.05.2011			
ADMIT	Data Mining, Statistics	N/A	Technische Universitaet Clausthal	http://cordis.europa.eu/projec	FP7-PEOPLE	01.06.2011	18.07.2014			
ADVANCE	Data Mining, Text Mining	N/A	Magyar Tudomanyos Akademia Szamitastechnikai Es Automatizalasi Kutatointezet	http://cordis.europa.eu/projec	FP7-ICT	01.10.2010	30.09.2013			
ALLEGRO	Multimedia Processing	N/A	Institut National De Recherche En Informatique Et En Automatique (Inria)	http://cordis.europa.eu/projec	ERC-2012-ADG_20120216	01.04.2013	31.03.2018			
ALMOND	Data Mining	N/A	Ben-Gurion University Of The Negev	http://cordis.europa.eu/projec	FP7-PEOPLE	15.07.2012	14.07.2014			
Aparsen	N/A	N/A	Science And Technology Facilities Council	http://cordis.europa.eu/projec	FP7-ICT-2009-6	01.01.2011	31.12.2014			
APO-SYS	Data Mining	N/A	Karolinska Institutet	http://cordis.europa.eu/projec	FP7-HEALTH	01.02.2008	31.01.2012			
ARSINFORMATICA	Artificial Intelligence, Data Mining, Mach	N/A	Politechnika Slaska	http://cordis.europa.eu/projec	FP7-PEOPLE	01.10.2012	30.09.2014			
ARTIAC	Data Mining	N/A	Tel Aviv University	http://cordis.europa.eu/projec	FP7-PEOPLE	01.10.2013	30.09.2017			
AUTOUNIMO	Artificial Intelligence, Data Mining, Predi	N/A	Politechnika Slaska	http://cordis.europa.eu/projec	FP7-PEOPLE	01.10.2013	30.09.2017			
AVERT-IT	Data Mining, Statistics	N/A	Pera Innovation Limited	http://cordis.europa.eu/projec	FP7-ICT	01.01.2008	31.07.2011			
BALTICGRID-II	Data Mining	N/A	Kungliga Tekniska Hogskolan	http://cordis.europa.eu/projec	FP7-INFRASTRUCTURES	01.05.2008	30.04.2010			
BI4MASSES	Data Mining, Text Mining	N/A	Ozyegin University	http://cordis.europa.eu/projec	FP7-PEOPLE	02.07.2010	01.07.2014			
BigDataEurope	Data Mining	N/A	Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V	http://cordis.europa.eu/projec	H2020-ICT-2014-1	01.01.2015	31.12.2017	https://twitter.com/BigData_Europe		
BIOLITCONTEXTMINING	Data Mining	N/A	Bogazici Universitesi	http://cordis.europa.eu/projec	FP7-PEOPLE	01.03.2012	29.02.2016			
BISON	Data Mining, Information Retrieval	N/A	University Of Bristol, Universitat Konstanz	http://cordis.europa.eu/projec	FP7-ICT	01.06.2008	31.05.2011			
BRAVEHEALTH	Data Mining, Text Mining	N/A	Labor S.R.L.	http://cordis.europa.eu/projec	FP7-ICT	01.03.2010	31.07.2012			
CLARIN	N/A	CLARIN-DK, TTNW	Institute For Language And Speech Processing (Ilsp / "Athena" R.C.), Universiteit Utrecht	http://cordis.europa.eu/projec	FP7-INFRASTRUCTURES-2007-	01.01.2008	30.06.2011			
CLARIN-PLUS	N/A	CLARIN-DK, TTNW	Institute For Language And Speech Processing (Ilsp / "Athena" R.C.), Clarin Eric	http://cordis.europa.eu/projec	H2020-INFRADEV-1-2015-1	01.09.2015	01.09.2017			
CYANOFACTORY	Data Mining	N/A	Uppsala Universitet	http://cordis.europa.eu/projec	FP7-ENERGY	01.12.2012	30.11.2015			
DebugIT	Data Mining	N/A	Agfa Healthcare N.V.	http://cordis.europa.eu/projec	FP7-ICT	01.01.2008	30.06.2012			
Dicode	Data Mining, Machine Learning	N/A	Computer Technology Institute & Press Diophantus	http://cordis.europa.eu/projec	FP7-ICT	01.09.2010	31.08.2013			
digital.me	Data Mining, Text Mining	N/A	Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V	http://cordis.europa.eu/projec	FP7-ICT	01.11.2010	31.10.2013			
DIMENSION	Data Mining	N/A	Ben-Gurion University Of The Negev	http://cordis.europa.eu/projec	FP7-PEOPLE	01.10.2012	30.09.2016			
DISTRIBUTION TESTING	Data Mining	N/A	Tel Aviv University	http://cordis.europa.eu/projec	FP7-PEOPLE	01.09.2008	31.08.2012			
DRIVER II	Data Mining	N/A	National And Kapodistrian University Of Athens	http://cordis.europa.eu/project/rcn/86426_en.html		01.12.2007	30.11.2009			
e-LICO	Data Mining	Rapidminer Studi	Universite De Geneve	http://cordis.europa.eu/projec	FP7-ICT	01.02.2009	31.01.2012			
E-SWARM	Data Mining	N/A	Universite Libre De Bruxelles	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.06.2010	31.05.2015			
ENANOMAPPER	Data Mining	N/A	Douglas Connect Gmbh	N/A	FP7-NMP	01.02.2014	31.01.2017			
EPIHEALTH	Data Mining	N/A	Biotalentum Tudasfejleszto Kft	http://cordis.europa.eu/projec	FP7-HEALTH	01.11.2011	31.10.2015			
EPIHEALTHNET	Data Mining, Multimedia Processing	N/A	Universiteit Utrecht	http://cordis.europa.eu/projec	FP7-PEOPLE	01.05.2013	30.04.2017			
ERACEP	N/A	N/A	Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.10.2009	28.02.2013			
EU-ADR	Data Mining	N/A	Erasmus Universitair Medisch Centrum Rotterdam	http://cordis.europa.eu/projec	FP7-ICT	01.02.2008	31.01.2012			
EUDAT2020	Data Mining	N/A	Jisc, Csc-Tieteen Tietotekniikan Keskus Oy	http://cordis.europa.eu/projec	H2020-EINFRA-2014-2	01.03.2015	01.03.2018	https://twitter.com/Eudat_eu	https://www.facebook.com/EUDAT	
EURECA	Data Mining, Multimedia Processing	N/A	Philips Electronics Nederland B.V.	http://cordis.europa.eu/projec	FP7-ICT	01.02.2012	31.07.2015	N/A		
EuroRIs-Net+	Data Mining	N/A	Ethniko Idryma Erevnon	http://cordis.europa.eu/projec	FP7-INFRASTRUCTURES	01.11.2007	31.10.2011	N/A		
FLARENET	Data Mining, Information Extraction, Tex	N/A	Institute For Computational Linguistics	N/A	ECP-2007-LANG	01.09.2008	31.08.2011	N/A		
FORSIED	Data Mining, Information Extraction	N/A	Universiteit Gent	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.05.2014	30.04.2019	N/A		
Foster	N/A	N/A	Universidade Do Minho	http://cordis.europa.eu/projec	FP7-SCIENCE-IN-SOCIETY-201	01.02.2014	01.02.2016	http://twitter.com/fosterscie	http://www.facebook.com/fosteropenscience	
FREME	Data Mining, Multimedia Processing, Tex	N/A	Deutsche Forschungszentrum Für Künstliche Intelligenz Ggmbh (Dfki)	http://cordis.europa.eu/projec	H2020-ICT-2014-1	01.02.2015	31.07.2017	https://twitter.com/H2020_FREME		
Fusepool	Data Mining, Information Retrieval	N/A	Xerox Lab, Berner Fachhochschule	http://cordis.europa.eu/projec	FP7-ICT	01.07.2012	30.06.2014	N/A		
GAMES	Data Mining	N/A	Engineering - Ingegneria Informatica Spa	http://cordis.europa.eu/projec	FP7-ICT	01.01.2010	30.06.2012	N/A		
GENODED	Data Mining	N/A	Universidad Politecnica De Madrid	http://cordis.europa.eu/projec	FP7-PEOPLE	01.11.2009	31.10.2011	N/A		
GEOSTREAM	Data Mining	N/A	Athena Research And Innovation Center In Information Communication & Knowledge Technolo	http://cordis.europa.eu/projec	FP7-SME	01.04.2013	31.03.2015	N/A		
GLASS	N/A	N/A	Water Insight Bv	http://cordis.europa.eu/projec	FP7-SPACE	01.03.2013	29.02.2016	N/A		
GLODERS	Text Mining	N/A	University Of Surrey	http://cordis.europa.eu/projec	FP7-ICT	01.10.2012	30.09.2015	N/A		
GS-A-DM-DS	Data Mining	N/A	De Montfort University	http://cordis.europa.eu/projec	FP7-PEOPLE	01.01.2015	31.12.2016	N/A		
HEALS	Data Mining	N/A	Universite Pierre Et Marie Curie - Paris 6	http://cordis.europa.eu/projec	FP7-ENVIRONMENT	01.10.2013	30.09.2018	N/A		
HiPerDNO	Data Mining	N/A	Brunel University	http://cordis.europa.eu/projec	FP7-ICT	01.02.2010	31.01.2013	N/A		
I3RES	Data Mining	N/A	Instalaciones Inabensa Sa	http://cordis.europa.eu/projec	FP7-ICT	01.11.2012	31.10.2015	N/A		
IAM4MARS	Data Mining	N/A	Uluslararasi Antalya Universitesi	http://cordis.europa.eu/projec	FP7-PEOPLE	01.08.2013	31.07.2017	N/A		
ICON	Data Mining, Machine Learning, Text Mir	N/A	Katholieke Universiteit Leuven	http://cordis.europa.eu/projec	FP7-ICT	01.01.2012	30.06.2015	N/A		
ICSI	Data Mining, Multimedia Processing	N/A	Intecs Spa	http://cordis.europa.eu/projec	FP7-ICT	01.11.2012	30.04.2015	N/A		
IHEARU	Text Mining	N/A	Technische Universitaet Muenchen	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.01.2014	31.12.2018	N/A		
IMAGELINK	Data Mining, Multimedia Processing, Tex	N/A	Eberhard Karls Universitaet Tuebingen	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.06.2013	31.05.2018	N/A		
IMARS	Data Mining	N/A	University College London	http://cordis.europa.eu/projec	FP7-SPACE	13.12.2013	01.01.2017	N/A		
IMPACT	N/A	N/A	Koninklijke Bibliotheek	http://cordis.europa.eu/projec	FP7-ICT	01.01.2008	30.06.2012	N/A		
INSEARCH	Data Mining, Information Retrieval, Text	N/A	Innovation Engineering Srl	http://cordis.europa.eu/projec	FP7-SME	01.12.2010	31.12.2012	N/A		
INSIGHT	Data Mining	N/A	Universidad Politecnica De Madrid	http://cordis.europa.eu/projec	FP7-ICT	01.10.2013	30.09.2016	N/A		
JMUGCS	Data Mining, Information Extraction	N/A	University Of Ioannina	http://cordis.europa.eu/projec	FP7-PEOPLE	01.10.2012	30.09.2016	N/A		
JOINUS	Natural Language Processing	N/A	University College London	http://cordis.europa.eu/projec	FP7-PEOPLE	01.09.2013	31.08.2017	N/A		
KAP	Data Mining	N/A	Sap Ag	http://cordis.europa.eu/projec	FP7-ICT	01.09.2010	31.12.2013	N/A		
KYOTO	Data Mining, Information Extraction, Ter	KAF	Stichting Vu-Vumc	http://cordis.europa.eu/projec	FP7-ICT	01.03.2008	28.02.2011	N/A		
LANGTERRA	Multimedia Processing, Text Mining	N/A	Institute For Language And Speech Processing (Ilsp / "Athena" R.C.)	http://cordis.europa.eu/projec	FP7-REGPOT	01.06.2012	31.05.2015	N/A		
LAST	Data Mining	N/A	Bar Ilan University	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.10.2009	30.09.2014	N/A		
LEARN	Data Mining	N/A	University College London	http://cordis.europa.eu/projec	H2020-INFRASUPP-2014-2	06.01.2015	06.01.2017	http://twitter.com/learnRDM		
LEAs BOX	Data Mining, Text Mining	N/A	Technische Universitaet Graz	http://cordis.europa.eu/projec	FP7-ICT	01.03.2014	30.11.2016	N/A		
LinDA	Data Mining	N/A	National Technical University Of Athens - Ntua	http://cordis.europa.eu/projec	FP7-ICT	01.12.2013	30.11.2015	N/A		
Live+Gov	Data Mining	N/A	Universitaet Koblenz-Landau	http://cordis.europa.eu/projec	FP7-ICT	01.02.2012	31.01.2015	N/A		https://www.linkedin.com/company/live-gov
LTE-HEALTH	Data Mining	N/A	Kingston University Higher Education Corporation	http://cordis.europa.eu/projec	FP7-PEOPLE	16.01.2012	15.01.2014	N/A		
MAESTRA	Data Mining, Machine Learning	N/A	Institut Jozef Stefan	http://cordis.europa.eu/projec	FP7-ICT	01.02.2014	31.01.2017	N/A		
MANTRA	Text Mining	N/A	Linguamatics, Universitaet Zuerich	http://cordis.europa.eu/projec	FP7-ICT	01.07.2012	30.06.2014	N/A		
Meta (T4ME Net)	Machine Learning, Machine Translation	N/A	Deutsche Forschungszentrum Für Künstliche Intelligenz Ggmbh (Dfki), Fondazione Bruno Kessle	http://cordis.europa.eu/projec	FP7-ICT-2009-4	01.01.2010	31.01.2013	N/A	http://www.facebook.com/META.Alliance	
MIGENOF113	Data Mining	N/A	Universidad Autonoma De Madrid	http://cordis.europa.eu/projec	FP7-PEOPLE	01.09.2010	31.08.2012	N/A		
MIGRANT	Data Mining, Graph Mining, Multimedia	N/A	Katholieke Universiteit Leuven	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.12.2009	31.05.2015	N/A		
MININEXACT	Data Mining	N/A	Ibm Research Gmbh	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.04.2011	31.03.2016	N/A		
MIRRI	Data Mining	N/A	Leibniz-Institut Dsmz-Deutsche Sammlung Von Mikroorganismen Und Zellkulturen Gmbh*	http://cordis.europa.eu/projec	FP7-INFRASTRUCTURES	01.11.2012	30.04.2016	N/A	N/A	N/A
MODAP	Data Mining	N/A	Sabanci University	http://cordis.europa.eu/projec	FP7-ICT	01.09.2009	28.02.2013	N/A	N/A	N/A
MOLD4PRODE	Data Mining	N/A	Association Pole Europeen De Plasturgie	http://cordis.europa.eu/projec	FP7-NMP	01.01.2011	31.12.2013	N/A	N/A	N/A
MONOTRANS	Machine Translation, Text Mining	N/A	Universitat Rovira I Virgili	http://cordis.europa.eu/projec	FP7-PEOPLE	01.03.2009	28.02.2011	N/A	N/A	N/A
mPlane	Data Mining, Multimedia Processing, Tex	N/A	Politecnico Di Torino	http://cordis.europa.eu/projec	FP7-ICT	01.11.2012	31.10.2015	N/A	N/A	N/A
MULTIFLOW	Data Mining	N/A	Universidad Politecnica De Madrid	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.02.2011	31.01.2016	N/A	N/A	N/A
MULTIJEDI	Text Mining	N/A	Universita Degli Studi Di Roma La Sapienza	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.02.2011	31.01.2016	N/A	N/A	N/A
MULTISENSOR	Machine Learning, Text Mining	N/A	Ethniko Kentro Erevnas Kai Technologikis Anaptyxis	http://cordis.europa.eu/projec	FP7-ICT	01.11.2013	31.10.2016	N/A	N/A	N/A
NANO2MARKET	Data Mining, Multimedia Processing	N/A	Universidad De Alicante	http://cordis.europa.eu/projec	FP7-NMP	01.07.2009	31.08.2010	N/A	N/A	N/A
NETSIM	Data Mining, Multimedia Processing	N/A	The Chancellor, Masters And Scholars Of The University Of Cambridge	http://cordis.europa.eu/projec	FP7-PEOPLE	01.09.2008	31.08.2012	N/A	N/A	N/A
NHECD	Data Mining, Information Extraction	N/A	Tel Aviv University	http://cordis.europa.eu/projec	FP7-NMP	01.12.2008	30.11.2012	N/A	N/A	N/A
NRG4CAST	Data Mining, Text Mining	N/A	Institut Jozef Stefan	http://cordis.europa.eu/projec	FP7-ICT	01.12.2012	30.11.2015	N/A	N/A	N/A
ONTIC	Data Mining	N/A	Universidad Politecnica De Madrid	http://cordis.europa.eu/projec	FP7-ICT	01.02.2014	31.01.2017	N/A	N/A	N/A
OpenAire2020	N/A	N/A	Institute For Language And Speech Processing (Ilsp / "Athena" R.C.), Jisc, Ethniko Kai Kapodistri	http://cordis.europa.eu/projec	H2020-EINFRA-2014-1	01.01.2015	01.07.2017	http://www.twitter.com/Ope	http://www.facebook.com/gr	N/A
Openminted	Data Mining, Text Mining	N/A	Agris, Agroknow, Institute For Language And Speech Processing (Ilsp / "Athena" R.C.), National	http://cordis.europa.eu/projec	H2020-EINFRA-2014-2	01.06.2015	01.06.2018	https://twitter.com/openmint	N/A	N/A
OraMod	Data Mining	N/A	Universita Degli Studi Di Parma	http://cordis.europa.eu/projec	FP7-ICT	01.10.2013	30.09.2016	https://twitter.com/OraMod_	https://www.facebook.com/C	N/A
Organic.Lingua	Data Mining	N/A	Agroknow, Fondazione Bruno Kessler (Fbk), Xerox Lab, Universidad De Alcalá	http://cordis.europa.eu/fp7/ict	PSP -2009.6.2	01.03.2011	28.02.2014	N/A	N/A	N/A
PASTEUR4OA	Data Mining, Text Mining	N/A	Jisc, Ethniko Idryma Erevnon	http://cordis.europa.eu/projec	FP7-SCIENCE-IN-SOCIETY-201	01.02.2014	7/31/2016	https://twitter.com/PASTEUR4	https://www.facebook.com/p	N/A
PHENOMINER	Text Mining	N/A	European Molecular Biology Laboratory	http://cordis.europa.eu/projec	FP7-PEOPLE	01.11.2012	30.12.2014	N/A	N/A	N/A
PHIDM	Data Mining, Graph Mining, Multimedia	N/A	Institute Of Science And Technology Austria	http://cordis.europa.eu/projec	FP7-PEOPLE	01.04.2014	31.03.2016	N/A	N/A	N/A
PONTE	Data Mining	N/A	Centre D'Excellence En Technologies De L'Information Et De La Communication	http://cordis.europa.eu/projec	FP7-ICT	01.03.2010	28.02.2013	N/A	N/A	N/A
POPDAT	Data Mining	N/A	Technische Universitat Berlin	http://cordis.europa.eu/projec	FP7-SPACE	01.06.2011	31.05.2013	N/A	N/A	N/A
PPI-MARKER	Data Mining	N/A	Technische Universitaet Dresden	http://cordis.europa.eu/projec	FP7-PEOPLE	01.08.2010	31.07.2014	N/A	N/A	N/A
PPIDR	Data Mining	N/A	Ecole Polytechnique Federale De Lausanne	http://cordis.europa.eu/projec	FP7-PEOPLE	01.07.2010	30.06.2012	N/A	N/A	N/A
PROLOGUE	Data Mining	N/A	Stichting Wetenschappelijk Onderzoek Verkeersveiligheid	http://cordis.europa.eu/projec	FP7-TRANSPORT	01.08.2009	31.07.2011	N/A	N/A	N/A
PROT-HISPRA	Data Mining, Multimedia Processing	N/A	Medizinische Universitaet Wien	http://cordis.europa.eu/projec	FP7-HEALTH	01.10.2011	31.03.2015	N/A	N/A	N/A
PSIP	Data Mining, Text Mining	N/A	Centre Hospitalier Regional Et Universitaire De Lille	http://cordis.europa.eu/projec	FP7-ICT	01.01.2008	31.07.2011	N/A	N/A	N/A
PSYCHE	Data Mining	N/A	Universita Di Pisa	http://cordis.europa.eu/projec	FP7-ICT	01.01.2010	31.12.2013	N/A	N/A	N/A
RDA Europe	N/A	N/A	Max Planck Gesellschaft Zur Foerderung Der Wissenschaften E.V.	http://cordis.europa.eu/projec	H2020-EINFRA-2014-2	09.01.2015	2/28/2018	https://twitter.com/resdatall	https://www.facebook.com/p	N/A
REACTOPS	Data Mining	N/A	Universidad De Navarra	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.06.2012	31.05.2017	N/A	N/A	N/A
Recode	N/A	N/A	Trilateral Research & Consulting Llp	http://cordis.europa.eu/projec	FP7-SCIENCE-IN-SOCIETY-201	01.02.2013	30.06.2015	https://twitter.com/RECODE_	N/A	N/A
REWIRE	Data Mining, Multimedia Processing	N/A	Universita Degli Studi Di Milano	http://cordis.europa.eu/projec	FP7-ICT	01.10.2011	31.12.2014	N/A	N/A	N/A
Scape	N/A	Jpylyzer, Matchbc	Ait Austrian Institute Of Technology Gmbh	http://cordis.europa.eu/projec	FP7-ICT-2009-6	01.02.2011	30.09.2014	http://twitter.com/#!/SCAPEp	N/A	N/A
SCIIMS	Data Mining	N/A	Bae Systems Integrated System Technologies Ltd	http://cordis.europa.eu/projec	FP7-SECURITY	01.11.2009	31.10.2012	N/A	N/A	N/A
SCY	Data Mining, Multimedia Processing, Tex	N/A	Universiteit Twente	http://cordis.europa.eu/projec	FP7-ICT	01.03.2008	29.02.2012	N/A	N/A	N/A
SEEK	Data Mining	N/A	Consiglio Nazionale Delle Ricerche	http://cordis.europa.eu/projec	FP7-PEOPLE	01.03.2012	31.08.2015	N/A	N/A	N/A
SIIP	Data Mining, Multimedia Processing, Tex	N/A	Verint Systems Ltd	http://cordis.europa.eu/projec	FP7-SECURITY	01.05.2014	30.04.2018	N/A	N/A	N/A
SKIN TREAT	Data Mining	N/A	Ahava Dead Sea Laboratories Ltd	http://cordis.europa.eu/projec	FP7-NMP	01.09.2008	31.08.2012	N/A	N/A	N/A
SMART	Data Mining, Multimedia Processing	N/A	Technion - Israel Institute Of Technology	http://cordis.europa.eu/projec	FP7-PEOPLE	01.05.2014	30.04.2018	N/A	N/A	N/A
SMARTMUSEUM	Data Mining, Multimedia Processing, Tex	N/A	Apprise Ou	http://cordis.europa.eu/projec	FP7-ICT	01.01.2008	28.02.2010	N/A	N/A	N/A
SME E-COMPASS	Data Mining	N/A	Exodus A.E.	http://cordis.europa.eu/projec	FP7-SME	01.01.2014	31.12.2015	N/A	N/A	N/A
SNIPER	Data Mining	N/A	Katholieke Universiteit Leuven	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.08.2014	31.07.2015	N/A	N/A	N/A
SPECMATE	Data Mining	N/A	Universitaet Des Saarlandes	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.01.2012	31.12.2016	N/A	N/A	N/A
STREAM	Data Mining	N/A	Universidad Politecnica De Madrid	http://cordis.europa.eu/projec	FP7-ICT	01.02.2008	30.04.2011	N/A	N/A	N/A
SUPREME	Data Mining	N/A	Centre Technique Des Industries Mecaniques	http://cordis.europa.eu/projec	FP7-NMP	01.09.2012	31.08.2015	N/A	N/A	N/A
SUSTAINABLE	Data Mining	N/A	Edp Distribuicao Energia Sa	http://cordis.europa.eu/projec	FP7-ENERGY	01.01.2013	31.12.2015	N/A	N/A	N/A
SYSTEMS MICROSCOPY	Data Mining	N/A	Karolinska Institutet	http://cordis.europa.eu/projec	FP7-HEALTH	01.01.2011	31.12.2015	N/A	N/A	N/A
TIMET	Data Mining, Multimedia Processing	N/A	904Labs, The University Of Edinburgh	http://cordis.europa.eu/projec	FP7-KBBE	01.03.2010	28.02.2015	N/A	N/A	N/A
TRANSFoRm	Data Mining	N/A	King'S College London	http://cordis.europa.eu/projec	FP7-ICT	01.03.2010	31.05.2015	N/A	N/A	N/A
TRANSMED	Data Mining, Multimedia Processing	N/A	"Biomedical Research Foundation, Academy Of Athens"	http://cordis.europa.eu/projec	FP7-REGPOT	01.01.2010	30.06.2013	N/A	N/A	N/A
TREELIM	Data Mining	N/A	Universitaet Basel	http://cordis.europa.eu/projec	FP7-IDEAS-ERC	01.04.2009	31.03.2014	N/A	N/A	N/A
TrendMiner	Machine Learning, Text Mining	N/A	Deutsche Forschungszentrum Für Künstliche Intelligenz Ggmbh (Dfki)	http://cordis.europa.eu/projec	FP7-ICT	01.11.2011	31.10.2014	N/A	N/A	N/A
TRIBUTE	Data Mining	N/A	Csem Centre Suisse D'Electronique Et De Microtechnique Sa - Recherche Et Developpement	http://cordis.europa.eu/projec	FP7-NMP	01.10.2013	30.09.2017	N/A	N/A	N/A
URBAN SENSING	Data Mining, Text Mining	N/A	Accurat Srl	http://cordis.europa.eu/projec	FP7-SME	01.10.2012	28.02.2015	https://twitter.com/urbansen	N/A	N/A
VERCE	Data Mining	N/A	Centre National De La Recherche Scientifique	http://cordis.europa.eu/projec	FP7-INFRASTRUCTURES	01.10.2011	30.09.2015	N/A	N/A	N/A
VIALACTEA	Data Mining	N/A	Istituto Nazionale Di Astrofisica	http://cordis.europa.eu/projec	FP7-SPACE	05.09.2013	01.10.2016	N/A	N/A	N/A
Visioneer	Data Mining	N/A	Eidgenössische Technische Hochschule Zürich	http://cordis.europa.eu/projec	FP7-ICT	01.10.2009	30.09.2010	N/A	N/A	N/A
ViSTA-TV	Data Mining, Multimedia Processing, Tex	N/A	Rapidminer, Universitaet Zuerich	http://cordis.europa.eu/projec	FP7-ICT	01.06.2012	31.05.2014	N/A	N/A	N/A
X-Like	Text Mining	N/A	Institut Jozef Stefan	http://cordis.europa.eu/projec	FP7-ICT	01.01.2012	31.12.2014	N/A	N/A	N/A
ZHOU_PFL_POLYKETIDES	Data Mining	N/A	The Chancellor, Masters And Scholars Of The University Of Cambridge	http://cordis.europa.eu/projec	FP7-PEOPLE	02.07.2012	01.07.2014	N/A	N/A	N/A

Collection of TDM Tools

Name	URL	Type	Organisation Producer	Country	Data Sources	Sector	Application fiel	Application fiel	Application fiel	Application fiel	Application fiel	Methods	Projects	Experts
ALVEO	http://alveo.edu.au/	Infrastructure, Platform	National Collaborative Research Infrastructure S	Australia	Business, Interr	Tertiary-Sector, Quartenary-Sector		Entertainment,	Research and Development (e.			Data Mining, Text Mining	N/A	N/A
Alvis	https://migale.jouy.inra.fr/redmine/projects/alv	Open Source, Component Collection, Work	N/A	N/A		N/A						N/A	N/A	N/A
Angoss Knowledge STUDIO	http://www.angoss.com/	Commercial	Angoss Software Corporation	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining, Multimedia Processing	N/A	N/A
AnnoMarket	https://annomarket.eu/	Commercial, Web Service Registry	University of Sheffield	UK	Business	Quartenary-Sector			Research and Development (e.			Data Mining, Text Mining	N/A	N/A
Apache cTAKES	http://ctakes.apache.org	Open Source, Component Collection, Work	Children's Hospital Boston, Mayo Clinic	USA	Business, Scien	Secondary-Sector, Tertiary-Sec	Bio-technology	Health and Social	Care			Information Extraction, Natural Language Processing, Text	N/A	N/A
Apache OpenNLP	http://opennlp.apache.org	Open Source, Component Collection, Work	Apache Software Foundation	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			Natural Language Processing	N/A	N/A
Apache UIMA	https://uima.apache.org	Open Source, Component Collection, Work	IBM, Apache Software Foundation	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			Information Extraction, Text Mining	N/A	N/A
Argo	http://argo.nactem.ac.uk	Open Source, Component Collection, Work	The National Centre for Text Mining, The Univer	USA	N/A	N/A						Text Mining	N/A	N/A
BioCatalogue	https://www.biocatalogue.org/	Open Source, Web Service Registry	The University of Manchester	UK	Scientific	Secondary-Sector, Quartenary-	Bio-technology		Research and Development (e.			Data Mining	N/A	N/A
BiodiversityCatalogue	https://www.biodiversitycatalogue.org	Open Source, Web Service Registry	The University of Manchester	UK	Scientific	Secondary-Sector, Quartenary-	Bio-technology		Research and Development (e.			Data Mining	N/A	N/A
Bluima	https://github.com/BlueBrain/bluima	Open Source, Component Collection, Work	EPFL - Research institution	Switzerland	Business, Scien	Secondary-Sector, Quartenary-	Bio-technology		Research and Development (e.			N/A	N/A	N/A
CLARIN-DK	http://clarin.dk/	Webapplication, Workflow	N/A	EU	All	Quartenary-Sector			Research and Development (e.			Natural Language Processing, Term/Concept Extraction	CLARIN	N/A
ClearTK	https://cleartk.github.io/cleartk	Open Source, Component Collection, Work	Center for Computational Language and Educat	USA	N/A	N/A						Machine Learning	N/A	N/A
Clementine	http://datamining.togaware.com/survivor/Clem	Commercial	IBM	USA	All	Quartenary-Sector			Research and Development (e.			Artificial Intelligence, Data Mining, Information Extraction	N/A	N/A
ContentMine	http://www.contentmine.org/	Open Source, Webapplication, Pipeline	ContentMine (non-profit)	UK/EU	Scientific	Tertiary-Sector, Quartenary-Sector			Education, Research and Devel			N/A	N/A	N/A
Data Scouts	https://datascouts.eu/	Open Source	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
DKPro Core	https://dkpro.github.io/dkpro-core	Open Source, Component Collection, Work	Ubiquitous Knowledge Processing Lab (UKP), Te	Germany	N/A	N/A						Natural Language Processing	N/A	N/A
Egas	https://demo.bmd-software.com/egas/	Open Source, Standalone	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ELKI	http://elki.dbs.ifi.lmu.de/	Open Source, Workflow	Ludwig Maximilian University of Munich	Germany	Business, Scien	Quartenary-Sector			Education, Research and Devel			Data Mining	N/A	N/A
Eparkomat	http://www.eparkomat.com/	Commercial	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Epinium	http://epinium.com/	Commercial, Workflow, Standalone	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Exclusivi	https://exclusivi.com/hotels/	Open Source, Workflow Environment	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
FICO Data Management	http://www.fico.com/	Commercial	Fair Isaac Corporation	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining, Multimedia Processing	N/A	N/A
Galaxy	https://galaxyproject.org/	Open Source, Workflow Environment	Center for Comparative Genomics and Bioinfor	USA	Business, Scien	Secondary-Sector, Quartenary-	Bio-technology		Research and Development (e.			Data Mining	N/A	N/A
GATE Embedded	https://gate.ac.uk	Open Source, Component Collection, Work	The University of Sheffield	UK	Business, Interr	Secondary-Sector, Quartenary-	Bio-technology		Education, Research and Devel			Information Extraction, Natural Language Processing, Text	N/A	N/A
Heart of Gold	http://heartofgold.dfki.de	Open Source, Component Collection, Work	DFKI GmbH, Language Technology Lab	Germany	Business	Quartenary-Sector			Education, Research and Devel			Natural Language Processing	N/A	N/A
IBM SPSS Predictive Analy	http://www-03.ibm.com/software/products/en	Commercial	IBM	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining	N/A	N/A
Influencity	https://influencity.com/en/	Webapplication, Workflow	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JCoRe	http://julielab.github.io	Open Source, Component Collection, Work	Friedrich Schiller University, Department of com	Germany	Business, Scien	Quartenary-Sector			Education, Research and Devel			Information Extraction, Information Retrieval, Natural Lan	N/A	N/A
Jpylyzer	http://jpylyzer.openpreservation.org/	Open Source, Pipeline, Standalone	N/A	EU	All	Quartenary-Sector			Research and Development (e.			Multimedia Processing	Scape	N/A
KAF	http://kyoto-project.eu/xmlgroup.iit.cnr.it/kyot	Open Source, Standalone		EU	All	Quartenary-Sector			Research and Development (e.			Data Mining, Information Extraction, Term/Concept Extra	KYOTO	N/A
Kepler	https://kepler-project.org/	Open Source, Workflow Environment	UC Davis Computer Science	USA	Business, Scien	Secondary-Sector, Quartenary-	Engineering		Research and Development (e.			N/A	N/A	N/A
KNIME	http://www.knime.org/knime	Open Source, Workflow	KNIME.com AG	Switzerland	Business, Scien	Secondary-Sector, Tertiary-Sec	Bio-technology	Finance, Banking, Insurance				Data Mining, Text Mining	N/A	N/A
Language Grid	http://langrid.org/en/index.html	Infrastructure, Platform	The Language Grid Association	Japan	Business, Scien	Quartenary-Sector			Research and Development (e.			Association rules, Multimedia Processing, Natural Langua	N/A	N/A
LAPPS Grid	http://www.lappsgrid.org/	Infrastructure, Platform	Vassar College, et al	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			Natural Language Processing	N/A	N/A
Massive Online Analysis (M	http://moa.cms.waikato.ac.nz/	Open Source	University of Waikato	New Zealand	All	Quartenary-Sector			Research and Development (e.			N/A	N/A	N/A
Matchbox	http://matchbox.openpreservation.org/	Open Source, Pipeline, Standalone	N/A	EU	All	Quartenary-Sector			Research and Development (e.			Multimedia Processing	Scape	N/A
Meta	http://meta.com/	Commercial	N/A	USA	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Microsoft SQL Server Analy	https://technet.microsoft.com/en-us/library/ms	Commercial	Microsoft Technet	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining, Multimedia Processing	N/A	N/A
Neural Designer	https://www.neuraldesigner.com/	Commercial	Artelnics	Spain	Business	Tertiary-Sector, Quartenary-Sector		Health and Soc	Research and Development (e.			Data Mining, Machine Learning, Predictive Analytics	N/A	N/A
NLTK	http://www.nltk.org	Open Source, Component Collection, Work	Team NLTK	Sweden/USA/Aust	Business, Perso	Quartenary-Sector			Education, Research and Devel			Artificial Intelligence, Information Retrieval, Machine Lear	N/A	N/A
Open Calais	www.opencalais.com/	Commercial, Open Source	Thomson Reuters	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			N/A	N/A	N/A
Opendatasoft	https://www.opendatasoft.com/	Commercial, Open Source	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Oracle Data Miner GUI	http://www.oracle.com/technetwork/database	Commercial	Oracle	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining	N/A	N/A
Orange	http://orange.biolab.si/	Open Source	N/A	N/A	N/A	N/A			Research and Development (e.			Data Mining, Machine Learning, Multimedia Processing	N/A	N/A
Pagelyzer	http://pagelyzer.openpreservation.org/	Open Source, Standalone	N/A	EU	Internet	Quartenary-Sector			Research and Development (e.			Multimedia Processing	Scape	N/A
Pegasus	https://pegasus.isi.edu	Open Source, Workflow	Information Sciences Institute, The University of	USA	Scientific	Secondary-Sector, Quartenary-	Aerospace		Research and Development (e.			Data Mining, Text Mining	N/A	N/A
Pentaho	http://www.pentaho.com/	Commercial, Open Source	Pentaho Corporation	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			N/A	N/A	N/A
Pipeline Pilot	http://accelrys.com/products/collaborative-scie	Workflow	Accelrys	USA	Scientific	Secondary-Sector, Quartenary-	Bio-technology		Research and Development (e.			Data Mining, Graph Mining, Multimedia Processing, Text M	N/A	N/A
Plazi TreatmentBank	http://plazi.org/api-tools/api/#Plazi_API	Webapplication, Workflow	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
QT21	http://qt21.metashare.ilsp.gr/	Infrastructure, Platform	N/A	EU	Scientific	Quartenary-Sector			Research and Development (e.			Machine Translation	N/A	N/A
Rapidminer Radoop	www.rapidminer.com	Commercial, Open Source	RapidMiner	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			Data Mining, Predictive Analytics, Statistics	N/A	N/A
Rapidminer Server	www.rapidminer.com	Commercial, Open Source	RapidMiner	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			Data Mining, Predictive Analytics, Statistics	N/A	N/A
Rapidminer Studio	www.rapidminer.com	Commercial, Open Source	RapidMiner	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			Data Mining, Predictive Analytics, Statistics	N/A	N/A
SafeToNet	https://safetonet.com/	Open Source, Component Collection, Work	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SAP Hana	https://hana.sap.com	Commercial	SAP America, Inc.	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining	N/A	N/A
SAS Data Mining	http://www.sas.com	Commercial	SAS Institute	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining	N/A	N/A
SPSS	http://www.ibm.com/analytics/us/en/technolog	Commercial, Workflow, Standalone	IBM	USA	All	Tertiary-Sector, Quartenary-Sector		Finance, Bankir	Education, Research and Devel			Data Mining, Information Extraction, Predictive Analytics,	N/A	N/A
Taverna	http://www.taverna.org.uk/	Open Source, Workflow Environment	School of Computer Science, University of Manc	UK	Scientific	Quartenary-Sector			Research and Development (e.			Data Mining, Text Mining	N/A	N/A
TensorFlow	https://www.tensorflow.org/about/	Commercial	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Theano	https://github.com/Theano/	Open Source, Workflow	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Think Analytics	http://thinkanalytics.com/	Commercial	ThinkAnalytics Ltd	USA	Business	Quartenary-Sector			Research and Development (e.			Data Mining	N/A	N/A
Think Enterprise Data Mine	www.thinkanalytics.com/	Open Source	Think Analytics	USA	Business, Scien	Quartenary-Sector			Research and Development (e.			N/A	N/A	N/A
TopPlace	http://www.avuxi.com/products/topplace	Open Source, Workflow	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Trendminer	https://www.trendminer.com/	Commercial, Open Source	N/A	USA	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Triana	http://www.trianacode.org	Open Source, Workflow	Cardiff University	UK	Scientific	Quartenary-Sector			Research and Development (e.			Data Mining, Multimedia Processing, Text Mining	N/A	N/A
TTNWW	http://vago.meertens.knaw.nl/apache/TTNWW	Webapplication, Workflow	N/A	EU	All	Quartenary-Sector			Research and Development (e.			Natural Language Processing, Term/Concept Extraction	CLARIN	N/A
VisTrails	http://www.vistrails.org	Commercial, Open Source	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Weblicht	https://weblicht.sfs.uni-tuebingen.de/	Webapplication, Workflow	N/A	EU	All	Quartenary-Sector			Research and Development (e.			Natural Language Processing, Term/Concept Extraction	CLARIN	N/A
WEKA	http://www.cs.waikato.ac.nz/~ml/weka	Open Source, Component Collection, Work	University of Waikato	New Zealand	Business	Quartenary-Sector			Research and Development (e.			Classification, Clustering, Data Mining, Predictive Analytic	N/A	N/A
Wordfreak	http://wordfreak.sourceforge.net/	Open Source, Workflow	N/A	EU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
xcorrSound	http://xcorrsound.openpreservation.org/	Open Source, Standalone	N/A	EU	All	Quartenary-Sector			Research and Development (e.			Multimedia Processing	Scape	N/A