PROJECT FINAL REPORT

Grant Agreement number: 610411

Project acronym: MULTISENSOR

Project title: Mining and Understanding of multilinguaL contenT for Intelligent Sentiment Enriched coNtext and Social Oriented inteRpretation

Funding Scheme: STREP

Period covered: **from** 1st November 2013 **to** 30th October 2016

Name of the scientific representative of the project's co-ordinator[1], Title and Organisation: Dr. Yiannis Kompatsiaris, Senior Researcher, Informatics and Telematics Institute, Centre for Research & Technology Hellas

Tel: +30 2311 257 774

Fax: +30 2310 474 128

E-mail: ikom@iti.gr

Project website address: http://multisensorproject.eu/

[1] Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

4.1 Final publishable summary report

4.1.1 Executive summary

MULTISENSOR is a 3-year integrated project being co-financed by the European Commission's Seventh Research Framework Programme. It started on 1st November 2013 and ended on 31th October 2016. The main objective of MULTISENSOR was to research and develop a unified platform for enabling the multidimensional content integration from heterogeneous sensors, with a view to providing end-user services such as journalism, commercial media monitoring, and decision support for SME internationalisation.

In this document, the work performed during the MULTISENSOR project is reported. The project overall research and development approach was iterative and organised in three main cycles, each one involving research, integration, prototype building and evaluation phases. In this context, research focused on the following scientific objectives: multilingual and multimedia content extraction, user and context-centric content analysis, multidimensional content integration and retrieval, semantic reasoning and decision support, summarisation and content delivery. Much effort has been spent in developing innovative technologies, as well as in integrating them into the MULTISENSOR System. During the three years of the project's lifetime, three applications were developed based on MULTISENSOR technologies, with each application addressing one of the three use cases considered in MULTISENSOR. The first provides search and exploratory functionalities for journalists, the second aims at supporting a media monitoring company to monitor specific profiles for their clients, while the third one provide decision support for SME internationalisation. In total, 3 prototypes have been implemented and thoroughly evaluated during the three development cycles of the project.

There has also been a successful dissemination aspect of the project's results, with the organisation of several events. The consortium has published 4 articles in scientific journals, as well as 44 papers in conferences and workshops. The MULTISENSOR User Group (UG) comprises 38 members both from industry and academia. In the exploitation level, MULTISENSOR has demonstrated its achievements by participating in scientific and industrial exhibitions and events. Finally, 6 inputs to standardisation and 16 open-source project components have been achieved.

Overall, the project has built on its strong scientific results, and with enhanced communication and interaction between partners, it has resulted to a system able to support the three use cases, but most importantly capable of extending the application of the developed technologies to various other market, societal and scientific domains.

4.1.2 Project context and objectives

During the past decade, the rapid development of digital technologies and the low cost of recording media have led to a great increase in the availability of multilingual and multimedia content worldwide. In the best case, this content is repetitive or complementary across political, cultural, or linguistic borders. However, the reality shows that it is also often contradictive and in some cases

unreliable. The consumption of such large amounts of content regardless of its reliability and cross-validation can have important consequences on the society. An indicative example is the current crisis of the financial markets in Europe, which has created an extremely unstable ground for economic transactions and caused insecurity in the population. The consequence is an extreme uncertainty and nervousness of politics, and economy on the one side, which makes national and international investments (e.g. SME internationalisation) really risky, and on the other side, the inability of journalism and media monitoring to equally consider all the media resources leaves the population in each of these encapsulated areas in its own perspective–without the realistic opportunity to understand the perspective developed in the other areas in order to adjust the own.

To break this isolation, there is a need for technologies that provide unified access to multilingual and multicultural economic, news story material across borders, that ensure its context-aware, spatiotemporal, sentiment-oriented and semantic interpretation, and that correlate and summarise the content into a coherent whole. These technologies should be capable to capture, interpret and relate economic information and news from various subjective views as disseminated via TV, radio, newspapers, blogs and social media. On top of this, semantic integration of heterogeneous media including computer-mediated interaction is required to gain a usable understanding based on social intelligence, while a correlation with relevant incidents with different spatiotemporal characteristics would allow for extracting hidden meaning.

Hence, MULTISENSOR aims at providing an integrated view of heterogeneous resources sensing the world (i.e. sensors), such as international TV, newspapers, radio and social media. The approach of MULTISENSOR considers the following dimensions for mining, correlating, linking, understanding and summarising heterogeneous material: language, multimedia, semantics, context, emotion, as well as time and location. The overall objective of MULTISENSOR is to research and develop a unified platform for enabling the multidimensional content integration from heterogeneous sensors, with a view to providing end-user services such as journalism, commercial media monitoring, and decision support for SME internationalisation. More specifically, potential investors can benefit from integration and context-aware interpretation of complementary and contradicting multilingual and multimedia information and get decision support for international investments, including internationalisation taking also into account the opinion trends and the psychological and sentimental status of the people. Media companies and archives can also benefit from the spatiotemporal integration and sentiment-oriented interpretation of heterogeneous content both for media monitoring and for journalism purposes. Finally, the European public can benefit from this integration and context-aware interpretation in the sense that it learns and comes to understand the views, fears and worries of the citizens all over Europe and get support for forming an objective opinion with respect to the state of affairs.

In this context, the following scientific objectives with respect to the individual MULTISENSOR research areas can be identified:

- **SO1**. Performing mining and content distillation of unstructured heterogeneous and

distributed multimedia and multilingual data. In this context, MULTISENSOR attempts to facilitate the data mining from several international resources, including news articles, audiovisual content (TV, radio), blogs and social media and provide intelligent mechanisms for the distillation of information. This objective included low- and high-level content analysis, as well as advancements in speech recongition, text analysis, machine translation (MT) and event detection from audiovisual content modules.

- **SO2**. Performing a user- and context-centric analysis of heterogeneous multimedia and multilingual content. Here, the focus is on analysing content from the user perspective to extract sentiment and context, analysing computer-mediated interaction in the web and specifically in social media, as well as generating high-level information based on the outcome of SO1 (e.g. concepts, entities). The aim was to develop and integrate into the MULTISENSOR platform research techniques for context extraction, sentiment extraction and social media mining (influential user detection and community detection).
- **SO3**. To enable the semantic integration and context-aware interpretation over the spatiotemporal and psychological dimension of heterogeneous and spatiotemporally distributed multimedia and multilingual data as audio, video, text and social content interaction. This included multidimensional content correlation and alignment based on reasoning techniques developed in SO4, as well as on multimodal vector-based representation and topic-based modelling. The multimodal integration was performed on top of the low- and high-level content extracted in SO1 and SO2 respectively.
- **SO4**. To provide semantic reasoning and intelligent decision support services. The purpose here is to make sense of very large amounts of heterogeneous data by providing diverse analytics, contextualised decision-making support for different situations to enable objectivisation of the information and view from multiple perspectives. In this context, MULTISENSOR researched and developed advanced reasoning techniques that abide to requirements for scalability and usability. Specifically, MULTISENSOR experimented with hybrid reasoning techniques, that address the correlation between fast query-answering and optimal data modifications; multi-threaded reasoning allowing inferencing to take place concurrently, using multiple simultaneous processors; temporal and geo-spatial reasoning providing the basics for powerful data analytics for decision support.
- **SO5**. To facilitate context-aware multimodal aggregation and multilingual summarisation and adequate presentation of the information to the user. This also included context-aware interpretation of news by examining their impact on the news consumers in the light of cultural aspects, user experience and engagement, as well as impact on affect and its condensed presentation along with the content summary.

MULTISENSOR's overall technological goal was the development of an operational platform for automatic multidimensional content integration of data and content, problem-oriented retrieval and assessment of this content, and targeted provision of information pertinent for the needs of users in a

given context. This platform aims at serving as a showcase for testing and evaluating the viability of the techniques developed in MUTLISENSOR for operational use. The aforementioned goal translates into the following technological objectives:

- **TO1**. To develop crawlers and data delivery channels necessary for the collection of data.
- **TO2**. To validate the technologies developed in MULTISENSOR by implementing a number of representative scenarios of pilot use cases that target journalism, media monitoring and international investment decision support.
- **TO3**. To deliver an operational demonstrator. This technological objective concerned the implementation (in three cycles) of the three use case prototypes, using the available services integrated into the MULTISENSOR architecture.

Finally, the main business objective (**BO1**) of MULTISENSOR was to define a business model for the exploitation of the project's results by the partners and the consortium as a whole. Other dissemination and exploitation activities concerned the project publications, the organisation of events and workshops, participation in events to advertise the project's results, concrete efforts to find opportunities for establishing liaisons with other related projects, as well as the provision of inputs to standardisation and the opening of project components under open-source licensing.

4.1.3 Main S&T results

The project started by performing empirical studies of the material and setting up the technical infrastructure for all modules. Three main use cases were designed relevant to journalism, commercial media monitoring and SME internationalisation and based on these the user requirements were extracted. Based on the user requirements, the MULTISENSOR technical specifications were set and the platform architecture was designed. This architecture was realised by an operational prototype, which integrates the skeleton versions of the research modules and served as fundament for the subsequent work in the project. In addition, the crawling architecture of MULTISENSOR was defined in order to facilitate content extraction from heterogeneous resources (e.g. news articles, data from financial portals) and the retrieval of web data (e.g. blogs, social media). The next step was the development of the MULTISENSOR first prototype version integrating the initial versions of the research modules, respectively. The evaluation of the first prototype version, which was conducted according to a pre-formulated evaluation strategy, provided the consortium partners with a number of meaningful hints and insights as well as updated requirements, thus establishing a good basis for the next development cycle, allowing in addition the formulation of an evaluation plan for the second prototype. The implementation of the MULTISENSOR second prototype version allowed the integration of the advanced versions of the research modules and showed many improvements to already existing functionalities based on the findings from the evaluation of the first prototype. The feedback from the second prototype's evaluation was very constructive and led to very helpful and important insights on how users perceive specific functionalities with regard to their specific use case scenario. The analysis of the evaluation results by the consortium led to the decision to establish dedicated work groups with

selected user and technical partners in order to drive the final development of the project's modules/functionalities. Thus, the research and development, as well as the integration process during the final project phase were streamlined. The implementation of the MULTISENSOR Final System, integrating the final versions of the research modules was followed by the final evaluation round, which showed very positive results for the three MULTISENSOR use case applications. A higher number of participants were involved in this third and final evaluation round than in the previous ones. Generally, the GUI and the interface of the different modules were seen as user-friendly and easy to use. Despite having quite differentiated platforms, the users from the different professional areas were satisfied with the MULTISENSOR usability look and feel. Finally, it should be noted that for the final evaluation a summative approach was followed, which validated whether the finished product met the user requirements. On the other hand, during the evaluation of the first and second prototype versions a formative approach was followed, which mainly focused on identifying and fixing problems during the development phases.

The project's results in the context of the individual scientific Work Packages (WPs) of MULTISENSOR are presented in detail in the next Sections.

WP2 (Multilingual and Multimedia Content Extraction)

The purpose of this WP is to perform mining and content distillation of unstructured heterogeneous and distributed multimedia and multilingual data, by applying tasks like named entity extraction, concept extraction and concept linking and relation extraction, audio transcription and analysis, machine translation, as well as concept and event detection from multimedia files.

In the context of this WP, an empirical study of the data sources considered important for the MULTISENSOR use cases was conducted early in the project. The study came up with some interesting findings with respect to the data sources and the most relevant information to be extracted for the purposes of the project. Specifically, the most relevant information can be divided into three groups:

- Factual information to be extracted from the "content" part of the source, such as names, other textual and numerical concepts and relationships between them, as well as their roles, functions and attributes.
- Metainformation about the sources, such as language, media type, author, date, followers, likes, views, etc.
- Derived metainformation, i.e. our interpretation of the content, by means of identifying the main topic, the tonality and sentiment, or key messages.

With respect to the named entity extraction task, a module that identifies names of persons, locations, companies and institutions, as well as amounts and dates, was researched and developed. The module has also the ability to recognise, whenever possible, additional attributes such as first name/ surname for persons, type of location, prices and subclass of amounts, etc. The initial approach that was followed was knowledge-driven, based on language-dependent lexicons and grammars, while the final approach extended the capabilities of the module by integrating contextual coverage and by adapting it to the use cases of the MULTISENSOR project.

Regarding the concept extraction task, a hybrid system was implemented, which combines linguistic detection of candidate nominal phrases, followed by a statistical scoring and a module based on BabelFy. In the concept linking and relation extraction task, a multilingual analysis pipeline that is able to produce abstract structures in the languages considered in MULTISENSOR was implemented. This was achieved through the development of cutting-edge neural network dependency parsers, a new kind of deep-syntactic transducer and a fast rule-based frame-semantics parser that were integrated into the MULTISENSOR's Content Extraction Pipeline (CEP).

In the audio transcription and analysis task, a scalable and versatile speech recognition system for German and English was developed. The work in this task started by developing a baseline version based on the open-source framework RWTH-ASR. Then, many post-processing steps towards the development of the final system were done. In particular, work focused on the reconstruction of orthography (correct casing instead of all-lowercase), reconstruction of numbers and symbols (their most usual writing form instead of the spelled one), insertion of utterance boundaries (breaking up the long recognition string into shorter phrases for better text analysis), and on joining compound parts into whole words. The final system accepts audio and video input in multiple formats, thanks to powerful format converters and the transcription results of the system support all major standards for time-coded information (e.g. for subtitling and media mining purposes). In addition, the system has a modular architecture and can be adapted to different workload requirements.

Regarding machine translation, a statistical-based module based on the open-source framework Moses was implemented. In addition, the phrase-based approach was adopted for the implementation of the module. Eight direct translation directions were developed (German, French, Spanish and Bulgarian from and into English). Many tuning and post-processing steps were done towards the optimisation of the machine translation module's performance. In particular, work was done on the data homogenisation (better normalisation and tokenisation), reduction of unknown words, and tuning for better translation in the general journalistic domain.

The work in the multimedia concept and event detection task started by developing a baseline framework for concept detection, which incorporated a video decoding step, a feature extraction step and a classification step. Regarding the video decoding and feature extraction steps, the most recent approaches in the relevant literature were used, while for the classification step the well-established SVM classifiers were tested, as well as late fusion for fusing the prediction results of the classification models per concept. Then, a more advanced version of the aforementioned framework was implemented, where alternative local descriptors were utilised in its feature extraction step (compared to the baseline framework) and a different late fusion strategy was applied in the classification step. The final version of the framework for the detection of predefined concepts/events, with a specific focus on video files, introduced the use of deep convolutional neural networks (DCNNs), in order to extract more sophisticated visual representations, compared to the local descriptors utilised in the previous versions of the framework.

Finally, one additional component in this context was implemented, although not foreseen in the

Description of Work (DoW). More specifically, a language identification module was developed in order to meet the need for annotating all incoming texts by correct language code, before storing them into the news repository. The module supports the identification of all project languages, and also of several other languages, for better discrimination.

WP3 (User and Context-centric Content Analysis)

The purpose of this WP is to perform a user- and context-centric analysis of heterogeneous multimedia and multilingual content. The tasks considered in this WP are context extraction, sentiment extraction and social interaction analysis (influential user detection and community detection).

Another task with respect to this WP dealt with the development of indicators for measuring aspects of journalism, commercial media monitoring and internationalisation within MULTISENSOR. In this context, a report was drafted, identifying the most important indicators for the journalism, media monitoring and SME internationalisation use cases of MULTISENSOR. Specifically with respect to the SME internationalisation use case, these indicators would be used by the reasoning and the decision support modules especially for assisting the decision making process. During the process of the indicator definition, the work focused on conceptual indicators. It appeared that in some cases some indicators where highly conceptual (especially in the journalism use case) and a mapping of them onto a range of values was not possible. Towards the definition of measurable indicators, the application of specific mathematical models was also investigated. To this end, the report came to a conclusion to reuse as much information available as possible by exploiting existing public data, web portals and APIs.

Regarding the context extraction task, a study with respect to context representation and the set of contextual features that would be used to characterise media items within MULTISENSOR was conducted. This study described a number of contextual features inspired by the requirements of the project's use cases and discussed the basic approaches for extracting or deriving these features, exploiting both the textual content of each media item along with its associated metadata. In addition, the study described how these properties can be represented using an ontology and stored into the semantic repository and queried in an effective an efficient way.

The context extraction module that was researched and developed for the purposes of MULTISENSOR requires as input the textual content and the metadata that is stored in the html source of a media item. The module extracts either from the text or the metadata the following information: *author* (or creator of the content item); a set of *keywords* characterising the content item; the *genre* of the item (if found in the metadata), and other features like *date*, *location*, and *source*. In addition, the context extraction module offers valuable insights with respect to what constitutes an engaging, good quality news article by identifying benchmarks for characterising editorial-based news article quality. More specifically, it identifies the following proxies that can be learned and predicted in an automatic and scalable manner: *Fluency*: Fluent articles are built from sentence to sentence, forming a coherent body of information; consecutive sentences are

meaningfully connected; similarly, paragraphs are written in a logical sequence. *Formality*: Formal articles are written by following certain writing guidelines; they are more likely to contain formal words and obey punctuation/grammar rules. *Richness*: The vocabulary of rich articles is perceived as diverse and interesting by the readers; rich articles are not written in a plain and straightforward manner. For the modelling process, a Generalised Linear Model (GLM) was learned, using a diverse set of novel features, such as bag-of-words, shallow lexical, and syntactic (cohesion, word vectors, generative features).

The goal of the sentiment extraction task in MULTISENSOR was to provide an analysis of the sentiment that is expressed in news articles. The baseline version of the module that was implemented utilised a dictionary-based method (SentiStrength), while the final version used a machine learning domain-specific classifier that was trained using syntactical features, such as standard BoW (e.g., uni-grams, bi-grams, tri-grams) and shallow (Shallow Kincaid, Coleman-Liau) features, extracted from an annotated, in-domain news corpus. The module has the ability to extract sentiment from both the body- and sentence-level of the news articles.

The social interaction analysis task involves a set of processes related to analysis of social network data stored into the MULTISENSOR repositories, namely crawled Twitter data. In MULTISENSOR, two subtasks were considered in this task, namely influential user detection and community detection. In the social interaction analysis task, the goal is to first build a topic-dependent network of contributors based on the mentions in the set of monitored tweets, then compute retweet probabilities between users in this network. The goal of the influential user detection subtask is to provide a ranked list of users by decreasing order of influence based on the aforementioned network of mentions and retweet probabilities. The goal of the community detection subtask is to make use of crawled Twitter posts in order to detect online dynamic communities by means of an appropriate community detection algorithm, which is applied to each graph snapshot defined by the user network of mentions. The influential user detection module developed within MULTISENSOR utilises a measure called Consistency of Sphere of Influence (CSI), a metric that quantifies the consistency of information propagation cascades in a social graph for a given user. In other words, it measures the variability of the set of users influenced by the targeted user on different instances. In general, the work conducted within MULTISENSOR in this subtask was the first to show consistent improvement in terms of quality over the theoretically optimal greedy algorithm for influence maximisation. For the MULTISENSOR community detection module, the expected codelength minimisation of the Infomap method was adopted. The Infomap method tackles the problem of finding a community structure in networks as a coding problem. The goal of a coding problem is to minimise the information required for the transmission of a message. Initially, Infomap employs the Huffman code in order to give a unique name (codeword) in every node in the network and then minimises the Shannon information required to describe the trajectory of a random walk on the network.

WP4 (Multidimensional Content Integration and Retrieval)

The purpose of this WP is to enable the semantic integration and context-aware interpretation over the spatiotemporal and psychological dimension of heterogeneous and spatiotemporally distributed multimedia and multilingual data as audio, video, text and social content interaction. In this context, topic-based modelling and representation of the content is applied, by classifying it and extracting topics. In addition, content integration in the semantic dimension is performed, by applying ontology alignment techniques. Finally, a vector-based indexing representation is implemented, as well as a retrieval system based on the aforementioned representation.

The topic-based modelling task of this WP is divided into two subtasks: i) category-based classification and ii) topic-event detection and tracking. Regarding category-based classification, a framework for the classification of news items within MULTISENSOR was researched and implemented. This framework assumes that each news item is represented by a number of modalities (for instance, textual modality, visual modality etc.). By applying certain procedures, a number of features from the raw data of each modality are extracted, thus formulating the corresponding feature vectors that serve as input for the construction of the classification models. For the classification process, the framework relies on the well-known Random Forests (RF) algorithm, as well as on late fusion strategies based on the operational capabilities of RF in order to compute weights for the outputs of each modality's trained RF model and fuse them. Specifically, three different methods that exploit the operational procedures of RF were applied for the computation of modality weights:

- *OOB error estimate*: RF has the ability to provide an estimation of its generalisation error through an internal mechanism called Out-Of-Bag (OOB) error estimate. During the construction of each tree, only 2/3 of the original data's cases are used in that particular bootstrap sample. The rest 1/3 of the instances (OOB data) are predicted by the constructed tree and thus, used for testing its performance. The OOB error estimate is the averaged prediction error for each training case y, using only the predictions of the trees that do not contain y in their bootstrap sample. The underlying notion here is that if a RF model is able to predict the OOB cases for one or more classes efficiently, it is expected to perform equally well on unknown cases. Therefore, from each modality's RF model, the corresponding OOB accuracy values are computed. This is done for each class separately. Then, the accuracy values are normalised (by dividing them by their sum) and serve as weights for the RF models' outputs.
- *Proximity ratio*: For the second weighting strategy, the proximity matrix of a RF model is taken into consideration. The proximity matrix contains proximities between the training cases. This is achieved by putting all the training cases down each tree and based on the frequency that pairs of cases end up in the same terminal nodes, this proximity matrix is computed. For each RF model, the proximity matrix between all pairs of data cases is constructed. Next, the proximity ratio values between the inner-class and the intra-class proximities (for each class) are computed. Finally, for each modality and for each class, the

proximity ratio values are first averaged and then normalised (by dividing them by their sum), in order to be used as modality weights for the RF models.

- *Adjusted proximity ratio*: This approach takes into account the two aforementioned weighting strategies (OOB error estimate and proximity ratio). It is used for adjusting the proximity ratio values, in cases where one or more classes for a modality's RF model exhibit high averaged proximity ratio values but disproportionally low OOB accuracy values. As a result, the weights assigned to these classes will be biased towards the "worse" modality (in terms of accuracy performance) and this will affect the late fused RF outputs. To overcome this, for each class and for each modality, the averaged proximity ratio values are multiplied by the corresponding OOB accuracy values, in order to formulate the adjusted proximity ratio values and thus, the modality weights for the RF models.

The initial version of the category-based classification module relied on two modalities in order to suitably represent each news item for the classification process, namely a) N-gram textual features extracted from the textual description of each item and b) low-level visual features extracted from the biggest image (assumed to be the representative one) of each item. For the final version of the module, due to the fact that i) one main finding from the initial experiments conducted in the context of this subtask was that the textual modality is more reliable and suitable for the category-based classification task than the visual one and b) it is not guaranteed that all news items contain one or more images, in order to be able to extract visual features from them, it was decided to replace the visual modality with another textual-based modality, which exploits a recently introduced methodology called word2vec that is used for producing word embeddings.

Regarding topic detection, the challenges in the context of this subtask are summarised into two main directions: (a) discover the correct number of topics and (b) group the most similar news articles into clusters. These challenges were faced under the following assumptions: Firstly, it was taken into account that real data is highly noisy and the number of clusters is not known. Secondly, it was assumed that there is a lower bound for the minimum number of documents per news cluster. Thirdly, the names/labels of the topics were considered unknown. Consequently, a novel hybrid topic detection framework was developed, which combines automatic estimation of the number of clusters and assignment of news articles into topics of interest. The estimation of the number of clusters is done by the novel DBSCAN-Martingale method that was proposed in the context of MULTISENSOR, which can deal with the aforementioned assumptions. The main idea of the framework is to progressively extract all clusters (extracted by a density-based algorithm) by applying Doob's martingale and then apply a well-established method for the assignment of news articles to topics, namely Latent Dirichlet Allocation. The proposed hybrid framework does not consider known the number of topics to be discovered, but requires only a lower bound for the number of documents per topic. Each realisation of the DBSCAN-Martingale provides the number of detected topics and due to randomness this number is a random variable. As the final number of detected topics, the majority vote over 10 realisations of the DBSCAN-Martingale is used.

The mapping discovery and validation task of this WP aimed at developing a content alignment framework based on ontology alignment and knowledge integration techniques, for the discovery of candidate semantic correspondences between heterogeneous information descriptions/terminologies and for the automatic verification of the correctness and consistency of the discovered mappings. In this context, a novel visual-based algorithm for ontology alignment, named LexiVis, was developed. The algorithm makes use of state-of-the-art low-level visual feature extraction methods in order to evaluate concept similarity. In addition, it integrates a novel metric that combines visual similarity and processing based on Wordnet. Furthermore, a methodology to combine different matching algorithms for achieving increased precision and recall in mapping discovery was implemented. The methodology consists of an adaptive weighting scheme for individual matchers and a fusion method for aggregating the matchers' results. All the above were integrated into a convenient and intuitive user interface for performing the matching of ontologies in MULTISENSOR.

The content alignment and integration task extends the functionality of the mapping discovery and validation task by introducing semantic processing in order to fully integrate the extracted knowledge and populate the Knowledge Base (KB) of MULTISENSOR. This task involves the further refinement of the multimodal descriptions through the removal of noise, the checking and identification of inconsistencies among them, as well as the resolving of conflicts with already available knowledge, among other things. In this context, an approach for performing validation of the RDF content during its extraction in the MULTISENSOR Content Extraction Pipeline (CEP) was developed. This approach is based on RDFUnit, a unit test for RDF that allows for identifying errors or incoherence in the RDF content, thus going beyond basic syntactic validation. The baseline version of the MULTISENSOR Content Alignment Pipeline (CAP) that was implemented followed a query-based approach in order to retrieve content relevant to the user queries. For the final CAP version, the notion of similar and contradictory content search was introduced. This procedure involves searching through the KB and identifying similar and contradictory content, based on measures for similarity and contradiction that were defined for this purpose, comparing each article against a large number of articles, and finally transforming the results to RDF and storing them back to the KB. The following similarity measures were defined: Framenet-based similarity (structure), concept similarity (comparison of concepts) and named entity comparison (entity matching). Similarly, the following contradiction measures were defined: concept contradiction (concept antonyms) and sentiment contradiction (different polarity in similar articles). For the combination of the aforementioned measures, the weighting approach introduced in the context of the mapping discovery and validation task was utilised.

Finally, in the context of the multimodal indexing and retrieval task of this WP, a multimedia data representation model that supports multimedia indexing and retrieval was implemented. The model is called SIMMO, which stands for Socially Interconnected MultiMedia-enriched Objects. SIMMO integrates in a unified manner the representation of multimedia and social features in online environments. Its flexibility and expressive power allow it to embrace the heterogeneity of

multimedia content and its interconnections, thus making it unique in its ability to support a wide range of multimedia information processing, analysis, and access applications. In addition, a novel framework for multimedia retrieval which is based on the fusion of multiple modalities by means of an approach called Weighted Multimodal Contextual Similarity Matrix (W-MCSM), was researched and developed. Specifically, three modalities were selected for the application of this framework within MULTISENSOR, namely textual concepts, visual features and visual concepts.

WP5 (Semantic Reasoning and Decision Support)

The purpose of this WP is to provide semantic reasoning and intelligent decision support services. More specifically, the goal here is: i) to provide the infrastructure which serves as the storage layer for the (meta)data of the MULTISENSOR platform, ii) to develop reasoning techniques beyond state of the art allowing for efficient information selection from heterogeneous data pools, e.g. hybrid reasoning, multi-threaded reasoning, temporal reasoning, geo-spatial reasoning and iii) to produce a decision support mechanism based on the developed reasoning techniques, listed above, and on cognitive techniques for context aware graph navigation, such as spreading activation.

The basic semantic infrastructure that was defined and that would reflect the business cases of MULTISENSOR made use of two upper ontologies, namely PROTON and DOLCE, for knowledge modelling of the relevant information. The most important requirements for the selection of these specific ontologies were the possibility for cleaning the unnecessary statements, the connectivity to linked open datasets, such as DBPedia, Freebase, etc. and the possibility of applying a good inference mechanism over the data modelling flow. The infrastructure also relied on the OWLIM repository for the semantic representation of the data.

Early in the project, a decision was made to capture all data exchanged among the project partners in the RDF JSONLD format. The final data was stored in a semantic repository and was used by various User Interface components for end-user interaction. The semantic data flexibility allowed the accommodation of a huge variety of data in the same extensible model. A number of ontologies were used for representing this data: NIF and OLIA for linguistic info, ITSRDF for NER, DBpedia and Babelnet for entities and concepts, MARL for sentiment, OA for image and cross-article annotations, W3C CUBE for statistical indicators, etc. In addition to reusing existing ontologies, they were extended with the MULTISENSOR ontology, which introduced some innovative concepts, such as embedding FrameNet conceptualisations in NIF. A document[2], named “RDF Application Profile” was drafted, which describes the different RDF patterns used by MULTISENSOR and how the data fits together. An example-based approach was used, instead of the more formal and laborious approach being standardised by the W3C RDF Shapes working group. The following areas are covered in the “RDF Application Profile” document:

- Linguistic Linked Data in NLP Interchange Format (NIF), including Part of Speech (POS), dependency parsing, sentiment, Named Entity Recognition (NER), etc.

[2] https://github.com/VladimirAlexiev/Multisensor

- Speech recognition and translation.
- Multimedia binding and image annotation.
- Statistical indicators and similar data.
- Social network popularity and influence, etc.

With respect to the reasoning techniques task, MULTISENSOR initially developed and implemented four different types of reasoning techniques, namely hybrid reasoning, parallel multi-thread reasoning, SPARQL-MM and GeoSPARQL. Hybrid reasoning combines the strengths of forward- and backward-chaining. It is useful for scenarios where a goal-driven reasoning is applied over a large amount of data. Hybrid reasoning is particularly common in decision support systems and semantic data integration from multiple sources, involving multiple ontologies. The parallel multi-threaded reasoning reduces the loading times from 50% to 200% compared to loading with a serial inference engine. SPARQL-MM is a proposed SPARQL extension that enables reasoning and query answering over media files. GeoSPARQL defines a powerful language for expressing geo-spatial constraints, e.g. overlapping or proximity of two geographical regions. The MULTISENSOR GeoSPARQL implementation is dynamic, conforms to the standard and at the same time allows for efficient query execution plans. Finally, a hybrid search framework, combining three different approaches, namely full-text search, faceted search and novel concept search based on semantic reasoning, was developed for the purposes of MULTISENSOR. The hybrid search is a data retrieval method, which uses the three aforementioned approaches in combination and in this way, the MULTISENSOR system is able to provide its users with advanced searching capabilities.

The first version of the Decision Support System (DSS) that was implemented for the SME internationalisation use case of MULTISENSOR was query-based and provided comparison between different European countries based on loading and using statistical indicators, retrieved from World Bank and Eurostat. In MULTISENSOR, an indicator is defined as a measurable quantity (e.g., a number or a ratio) or any piece of conceptual information (i.e., non-measurable) derived from a series of observed facts, serving as a guide to the process of decision-making. In this case, measurable indicators were mainly selected, as they serve the purpose of delivering an assessment support better. Measurable indicators are computable units derived from collections of data. Such examples include a country's GDP, total population, and the percentage of people who have access to the Internet. These indicators are very useful for establishing a comparison between countries and, thus, generating a preliminary assessment of the differences the countries present. The final version of the MULTISENSOR DSS introduced a new decision generation process based on specific user input. The input is based on country of origin, destination country and the product type. Based on that information, the system gets all available data from the statistical indicators, performs calculations and generates a list of the most appropriate countries for exporting a specific product.

WP6 (Summarisation and Content Delivery)

The purpose of this WP is to facilitate context-aware multimodal aggregation and multilingual summarisation and adequate presentation of the information to the user.

Early work in the context of this WP focused on the implementation and set up of a basic summarisation architecture and the compilation of a dataset for the summarisation task. An initial architecture based on extractive automatic summarisation methods was defined for both single and multiple document summarisation of texts. The summarisation dataset that was compiled for abstractive summarisation comprised a Text Planning and Lexicalisation Dataset and a Multilingual Surface Generation Dataset. The first is a NIF-encoded LOD corpus that was produced automatically by the multimedia analysis pipeline of MULTISENSOR. This corpus was used to help develop methods for the content selection and lexicalisation tasks of the Natural Language Generation (NLG) pipeline in charge of generating abstractive summaries. The other dataset is a manually annotated multilingual corpus used to train the MATE statistical generator for the surface generation task of the NLG pipeline.

Subsequent work focused on the content selection metrics and content delivery procedures tasks of this WP. A formal description of the metrics for use by the text planning module of an abstractive summariser, as well as a description of the way these metrics would be applied for the selection and ordering of contents were provided. In addition, the pipeline for extractive summarisation was developed and extended to include some of the information extracted by the MULTISENSOR content extraction pipeline.

The final phase of the work in the context of this WP concerned the development of the summarisation functionalities in the scope of the MULTISENSOR project and their integration into the final version of the MULTISENSOR summarisation system. This included both the result of research into state-of-the-art extractive methods and a fully functional abstractive system capable of producing multilingual summaries.

MULTISENSOR crawling infrastructure

The purpose of the crawling infrastructure that was developed within MULTISENSOR was to facilitate content extraction from heterogeneous resources (e.g. news articles, data from financial portals) and the retrieval of web data (e.g. blogs, social media). It was decided to crawl three sources of data: real-time sociometric counts from social media platforms, Twitter streaming data, and web pages. The crawling architecture consists of three main components, namely a collector for supporting real-time aggregation of sociometric counts, a crawler for crawling web pages using a seed list and API wrappers for retrieving structured data from knowledge bases. The web crawler is based on the large scale Nutch open source crawler and is used both for focused crawling of web pages as well as for crawling multimedia objects. The crawlers run on a server and are based on the Apache Hadoop open framework. At some point, during the development of the MULTISENSOR second prototype version, in order to further facilitate the crawling of multimedia content and social information, it was decided to implement a dedicated API (for the crawling of multimedia content), as well as utilise the Twitter collector that was implemented during a previous project (for social media crawling). More specifically, the Twitter collector comes from the previous SocialSensor (FP7-287975) project, during which Stream Manager was implemented. Stream Manager contains a

number of APIs that collect incoming content relevant to a keyword, a user or a location from a set of social streams (Twitter, Facebook, Instagram, etc.). The Twitter collector gathers Twitter posts for a set of pre-specified high-activity hashtags for each separate use case, as well as information regarding the author, the associations found inside the posts (between users, between user and webpages, etc.). It should be noted that the Twitter posts gathered by the collector were fed as input to the WP3 influential user detection and community detection services.

The MULTISENSOR System

The integration of the MULTISENSOR System involved iterative cycles of development and evaluation, constantly monitoring the integration plan and work, tracking all the issues with relevant tools and assigning the resolution of problems to the responsible module developing partners.
The MULTISENSOR logical architecture (see Figure 1) is based on a state-of-the-art approach and encompasses two discrete modalities: offline - asynchronous processing of harvested data, and online - synchronous retrieval, delivery and exploitation of the analytical data.

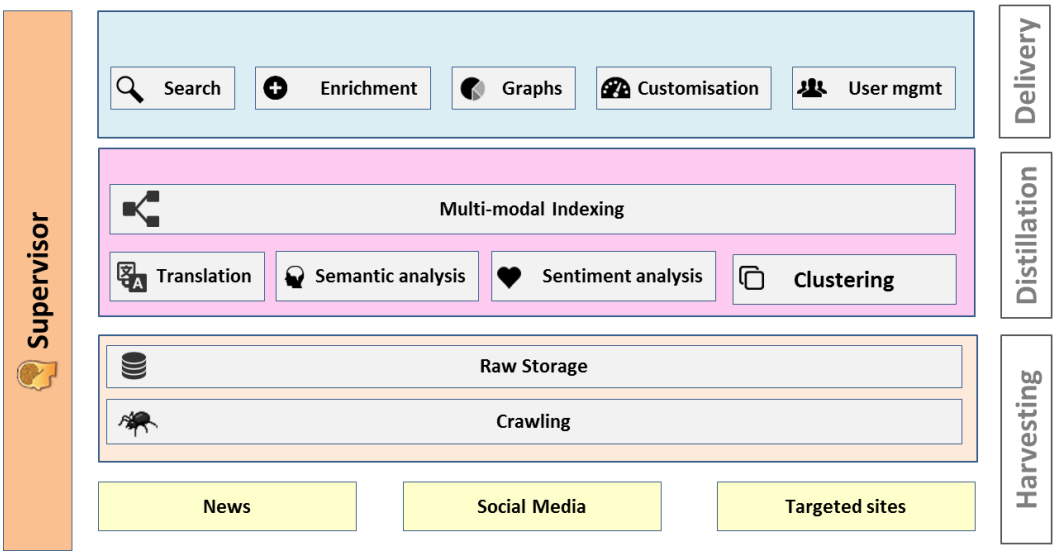


Figure 1: MULTISENSOR logical architecture

After the implementation of the operational prototype of the MULTISENSOR platform, integrating the skeleton of the content analysis services developed in the context of SO1-SO5, came the next development phases with the implementation of the 1st and 2nd prototypes of the MULTISENSOR platform, integrating the baseline and the advanced services respectively in the context of SO1-SO5. The architecture that was defined for the final MULTISENSOR System is presented from a high-level view in Figure 2.

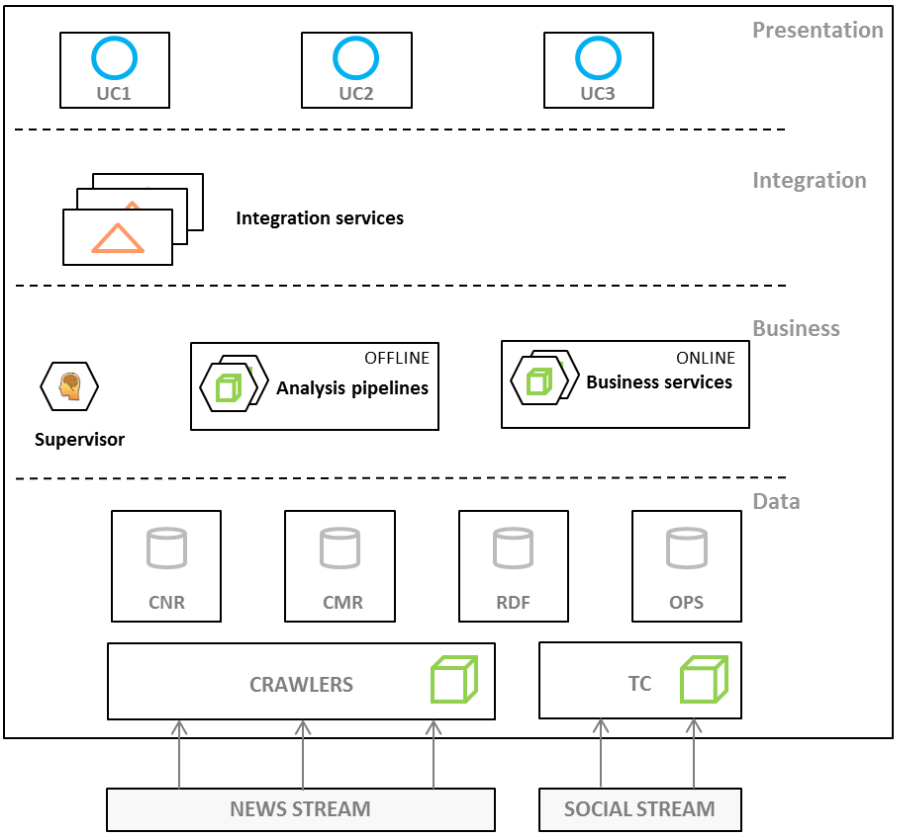


Figure 2: Final MULTISENSOR System high-level view

During the final phase, work was focused on continuous integration and collaboration with the technical partners for the implementation of the MULTISENSOR Final System, integrating the final versions of the services in the context of SO1-SO5, as well as for the optimisation of the Final System's modules performance.

The Journalism Use Case

The journalism use case application is an application that should support media professionals (e.g. journalist, media expert) to find relevant information in different formats, coming from different sources, and according the social activities that were produced around.

The access to the application is managed by a user profile service, which controls the user account (credential and preferences). The user has to login with his/her credentials to the application. When the user is logged in the system, he can access his folder that contains his/her favourite documents.

An overview of the application's different sections, as well as a description of their functionalities is provided below.

Search section: With a simple selection of keywords and filtering criteria, the user can make a textual search by querying the Semantic Search online service. The available search methods are:

- **Main Semantic Search**: This is the basic possibility to search for relevant content using some keywords.

- **Hybrid Search**: When the user starts writing the query, some entities and concepts are suggested in real time with auto-completion mechanism. Even if some entities and concepts are selected, some keywords can be added as well to complement the query (see Figure 3).
- **Multimedia Search** consists in the retrieval of textual articles, which possess at least one multimedia element (image, audio or video).

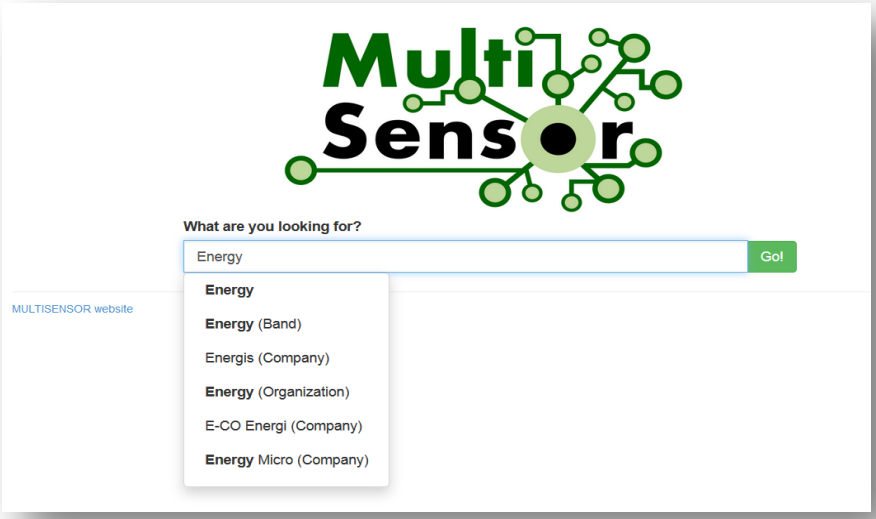


Figure 3: Main page Hybrid Search

Results section: The results page has two parts; it also has dynamic header and static footer. The header of the page displays user-related information and depends on the status of the user. A non-authorised user will have access to authorisation dialog, while an authorised one will be able to access his "Portfolio".

On the left side, advanced search features are available (a search field that supports hybrid functionality). Search can be done on full text, entities or both. Bellow, filters like searching language, date, as well as multimedia filter, can be seen (see Figure 4).

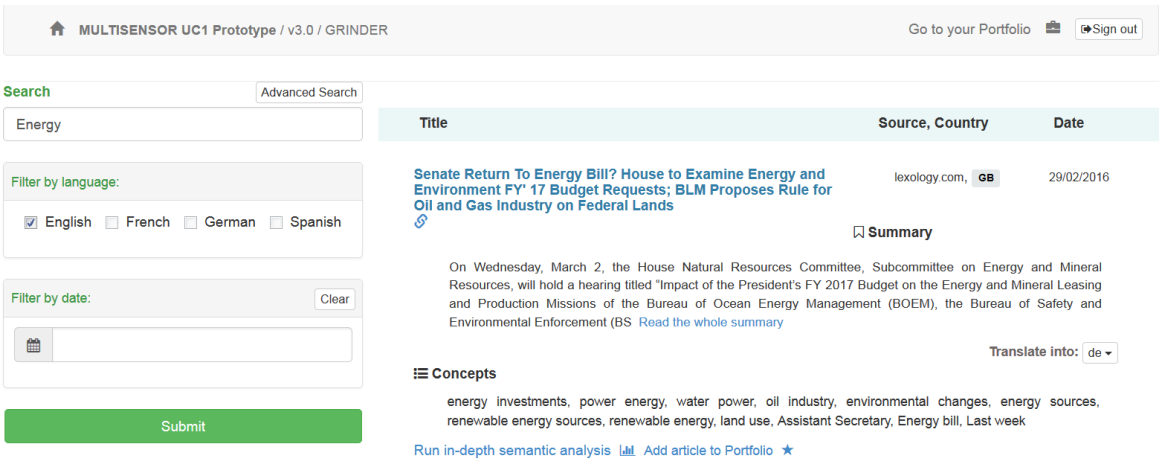


Figure 4: Result page (Header, Advanced search and result listing)

Further bottom on the left side, search related entities are displayed. By clicking on an entity it will be added to the search query. Then these entities can be used to extend the search query. Also, each entity has a link to DBpedia to obtain more information about the entity: An example is shown in Figure 5.

The right side displays the following information:

- **Context**: Contextual features per article (title, source, etc.).
- **Summarisation**: Display of the output of the summarisation service.
- **Translation**: The online machine translation service operates on this functionality in order to translate a summary to one of the five available languages (English, French, Spanish, German and Bulgarian).
- **Run in-depth semantic analysis**: Displays semantic page view.
- **Add article to Portfolio**: The link to add an article to the portfolio, for further analysis.

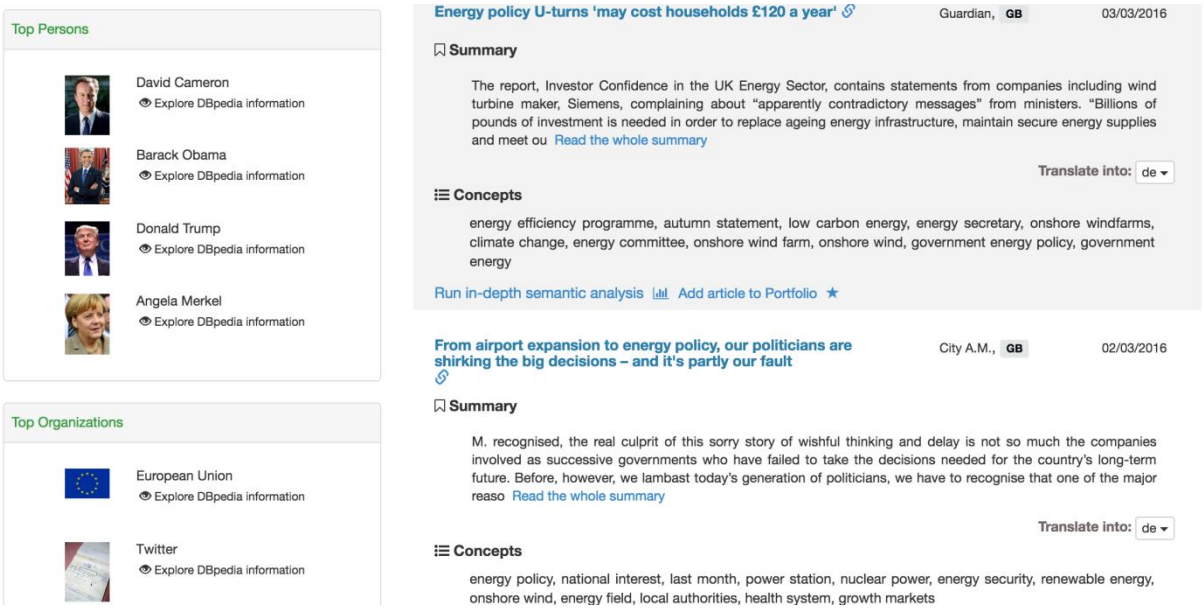


Figure 5: Result page (Entities and result listing)

Semantic analytics section: The result listing on the right side displays processed information like context (source, language, time of publishing), title, summary, translation, specific concepts and also a link to a more deep analytics page called the "semantic analytics" (see Figure 6). On this page, more information extracted from the text is displayed (the list of named entities, the sentiment polarity, a cloud of specific concepts and the related articles are listed at the bottom of the page). In addition, there is a link to add an article to the portfolio, for further analysis.

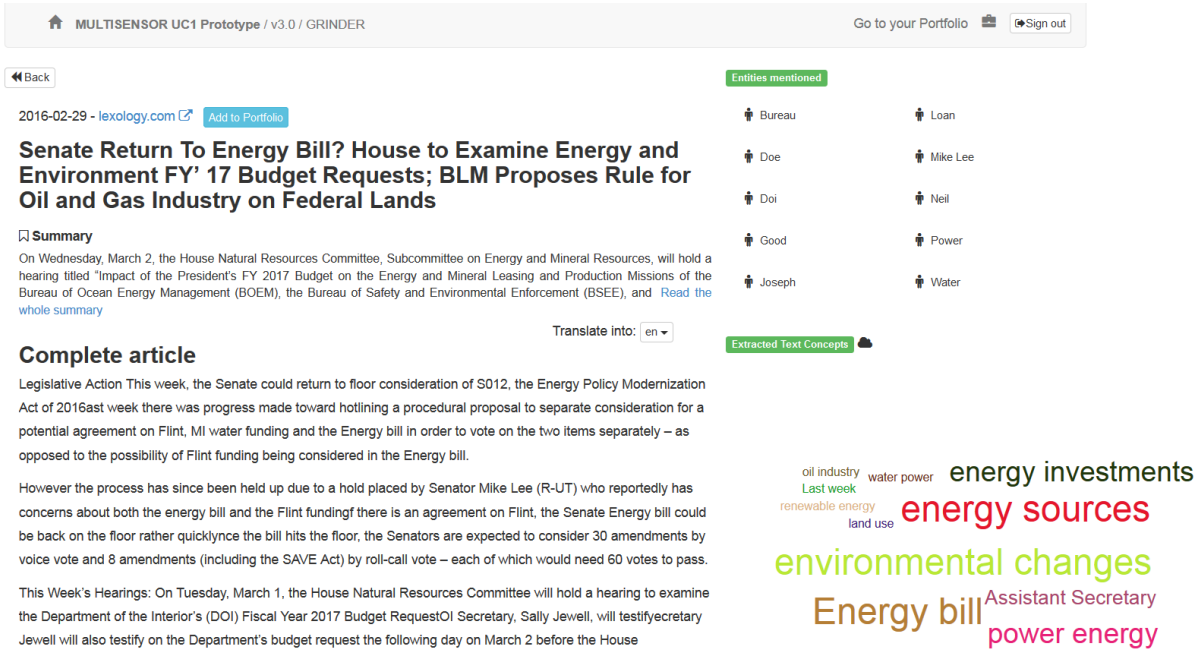


Figure 6: Semantic analytics page (textual dimension)

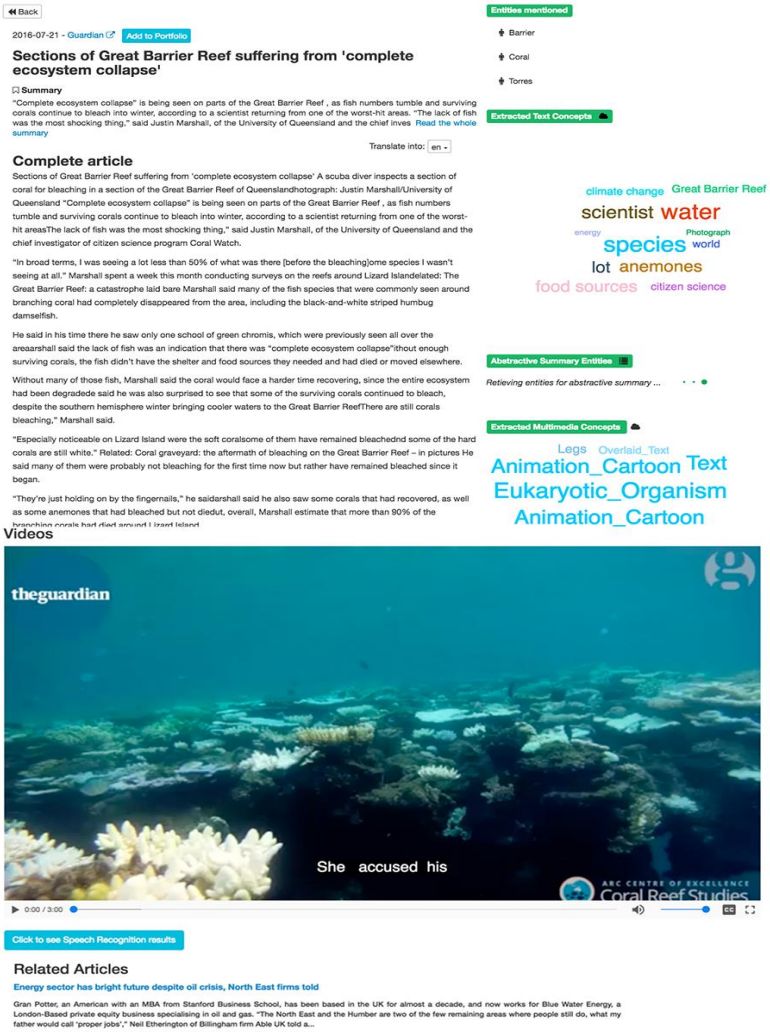


Figure 7: Semantic analytics page (multimedia dimension)

For the multimedia content, the semantic analytics page (see Figure 7) contains the video player of the images portfolio and the detected multimedia concepts, displayed as a key cloud (under the specific concept one). The ASR transcript is displayed as subtitles to the video.

Portfolio analysis section: During the search process, any article can be added to the "Portfolio" for further analysis (see Figure 8). The "Portfolio" can be accessed by clicking on the "Go to portfolio" link on the top left corner of the header.

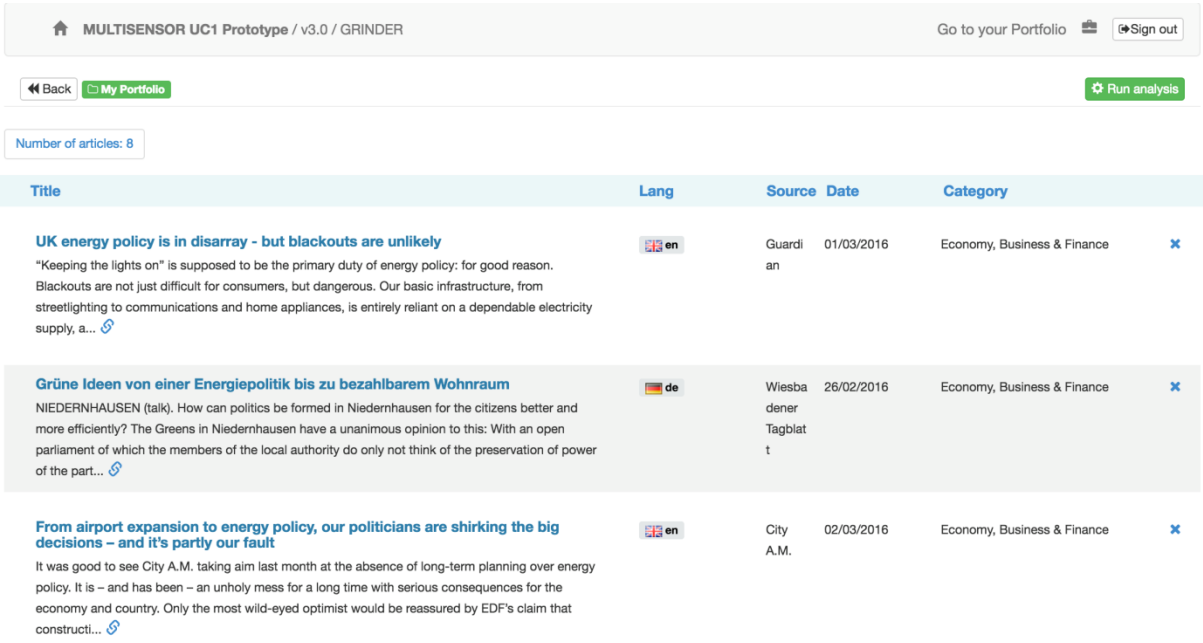


Figure 8: Portfolio home page with selected articles

By clicking the "Run analysis" button, the aggregated analytical view of the portfolio content can be generated (as shown in Figure 9), where the entities, the most frequent words, the extracted topic and similar articles that the topic from this analysis contains can be seen.

- **Tag cloud**: The system analyses the term frequency over all texts of the selected articles in the folder. It displays a graphical summary of the folder content.
- **Entity aggregates**: The analysis services retrieve all the entities that are present in the folder's documents.
- **Topic and event detection**: Extracted topics from the portfolio articles.

Firstly, the named entities of all the documents present in the portfolio are aggregated. The same approach is followed for the specific concepts. Secondly, the results of the topic and event detection service are aggregated and displayed. For this, the labels of the clusters are shown in the keyword cloud and the list of related articles is presented on the right.

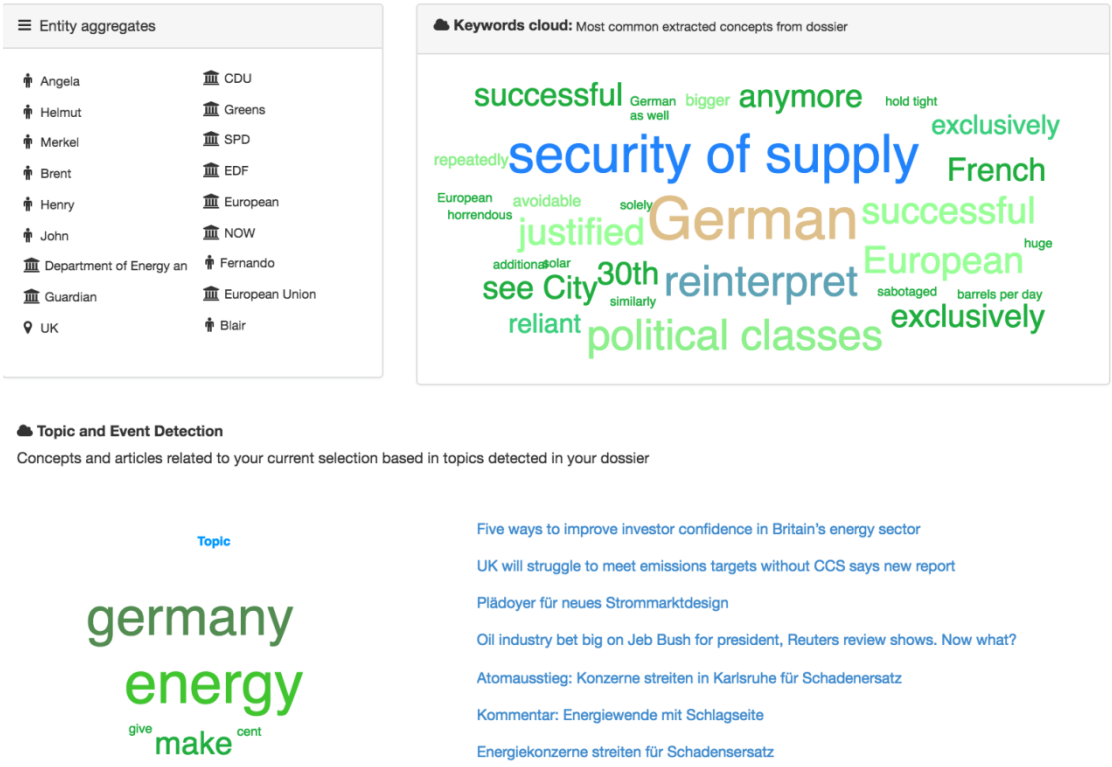


Figure 9: Portfolio aggregated portfolio view

The Media Monitoring Use Case

The media monitoring use case application replicates the workflow of a media monitoring professional to execute an analysis for a client.

This includes checking articles for relevance by various indicators and saving the relevant articles for a client's profile. The relevant articles will then be analysed, so that conclusions can be drawn from this analysis.

An overview of the application's different sections, as well as a description of their functionalities is provided below.

Search section: After logging in into the application, the user is presented with a view to search for keywords and filter languages and countries (see Figure 10). Alternatively, he can select a profile from the upper dropdown menu. Profiles have search settings stored for recurring searches in order to quickly populate the search mask.

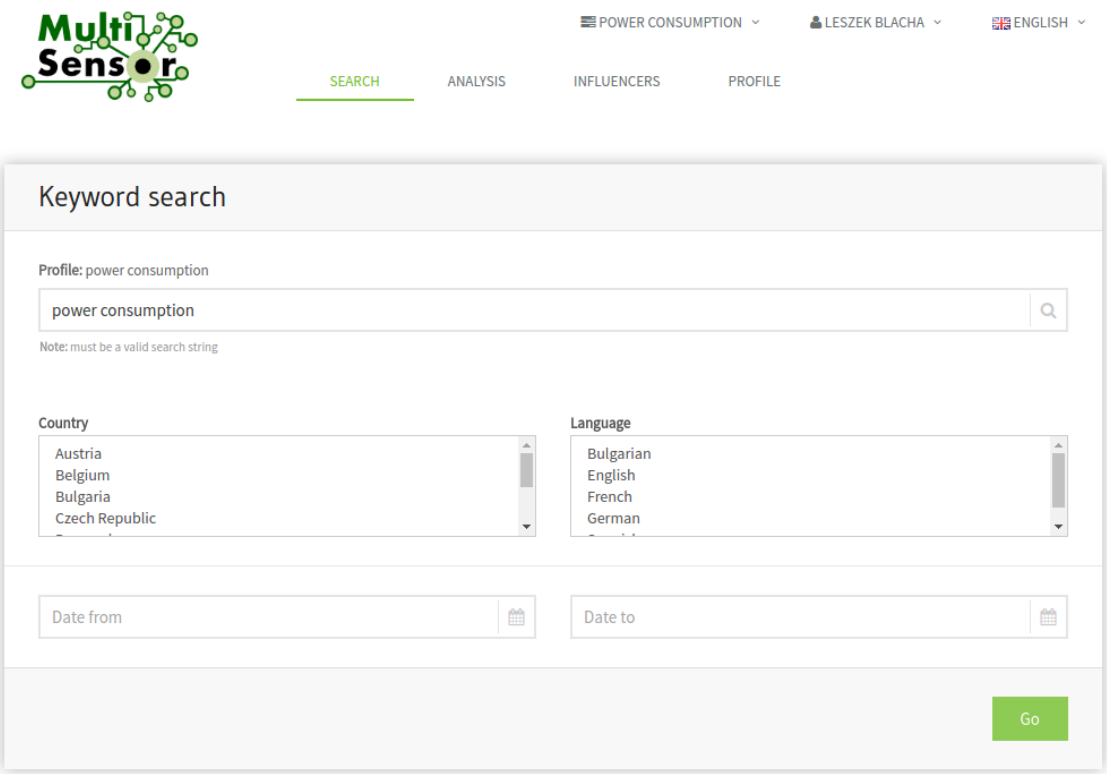


Figure 10: Search section

The search includes semantic features, so that not only articles that contain the search term are returned, but also articles that are semantically relevant. The results of a search query are displayed in a single article view (see Figure 11).

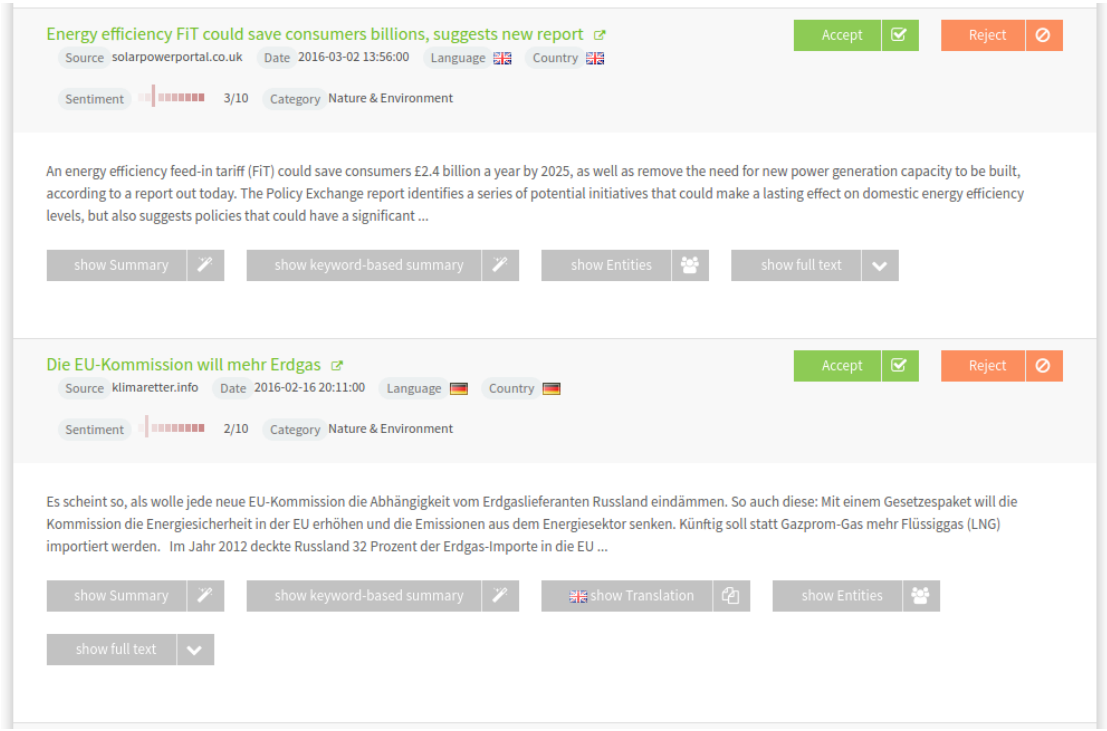


Figure 11: Results page

In order to evaluate whether an article is relevant for the client, the user can use additional functionalities, such as calling the summarisation and/or translation service. In addition, he can take a look at the entities extracted from the text and read the article's full text.
The feature "keyword-based summarisation" creates not only extractive summaries, but also summaries that are tailored to the clients' needs. When creating a keyword-based summary, detected entities and search terms can be selected. These terms are then regarded more significant and sentences including them are ranked higher when creating the summary (see Figure 12).

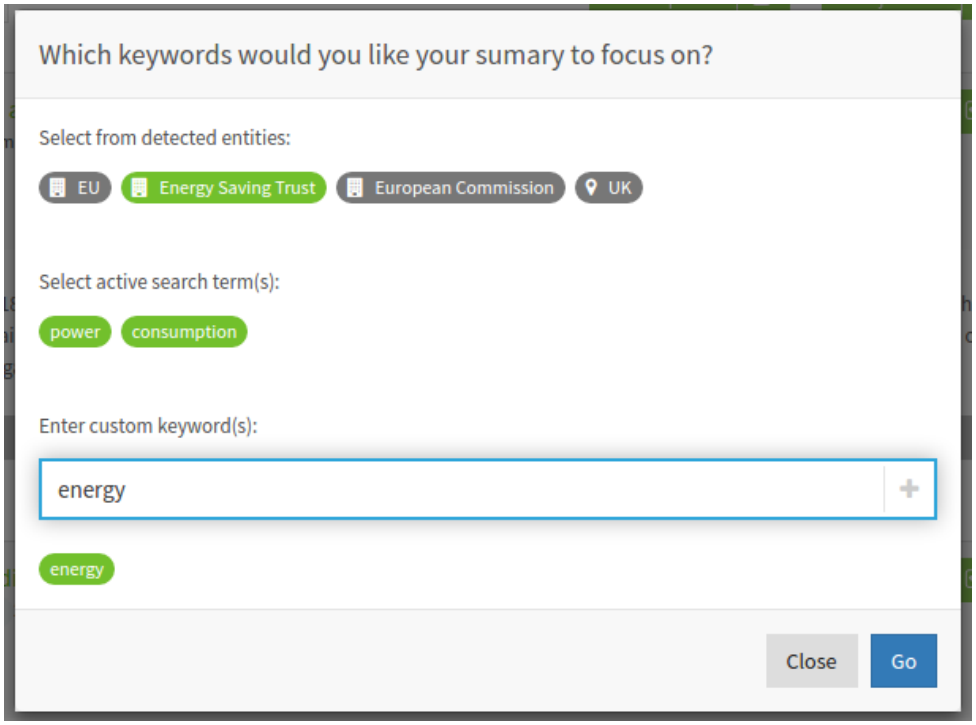


Figure 12: Keyword summary dialog

In addition to the entities and search terms, the user can also add individual terms (e.g. "energy").
In the single article view (see Figure 13), information for sentiment and category is displayed.

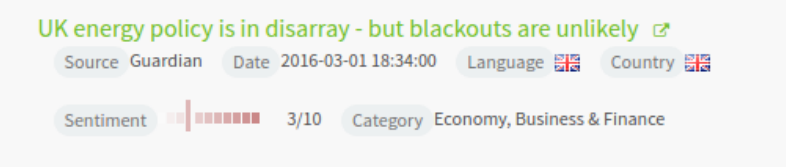


Figure 13: Single article view

Grouping articles by category (see Figure 14) is an easy and convenient way to mark groups of articles as relevant or irrelevant.

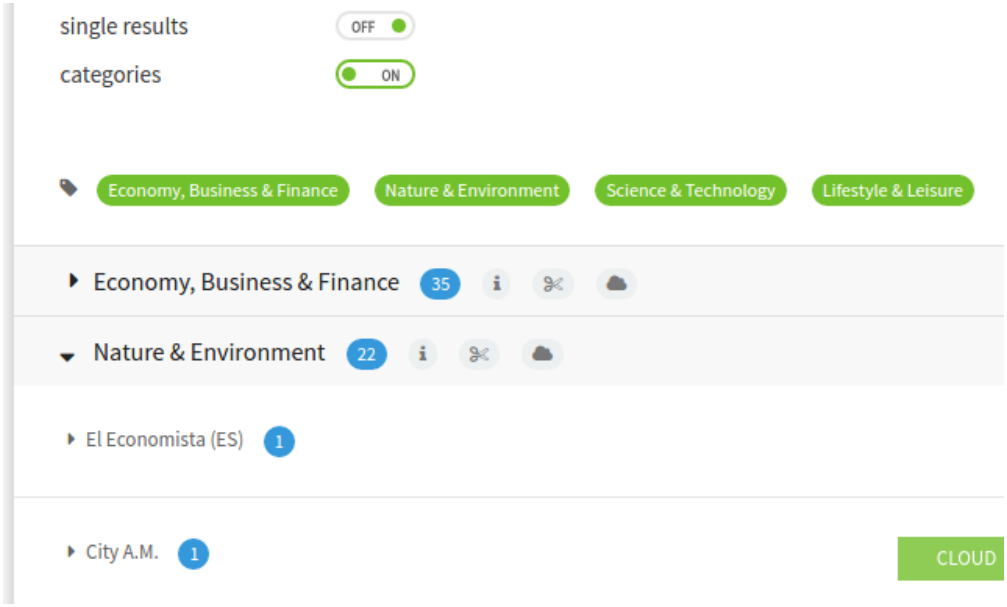


Figure 14: Article grouping

For each category, a short explanation can be displayed by clicking on the information icon (see Figure 15).

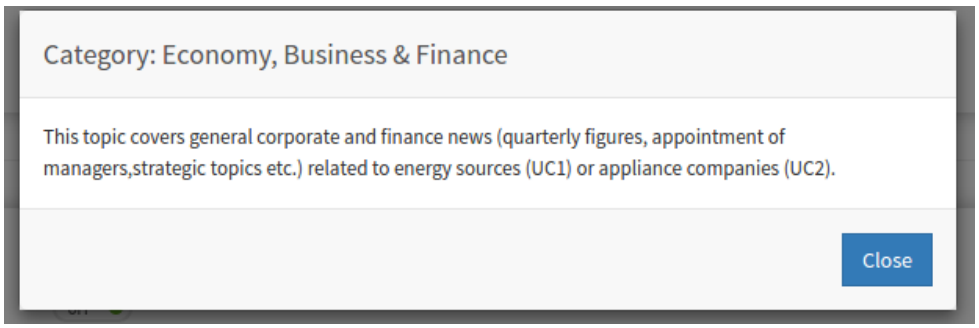


Figure 15: Category explanation

Clicking on the scissor icon will create a multi-document extractive summary (see Figure 16).

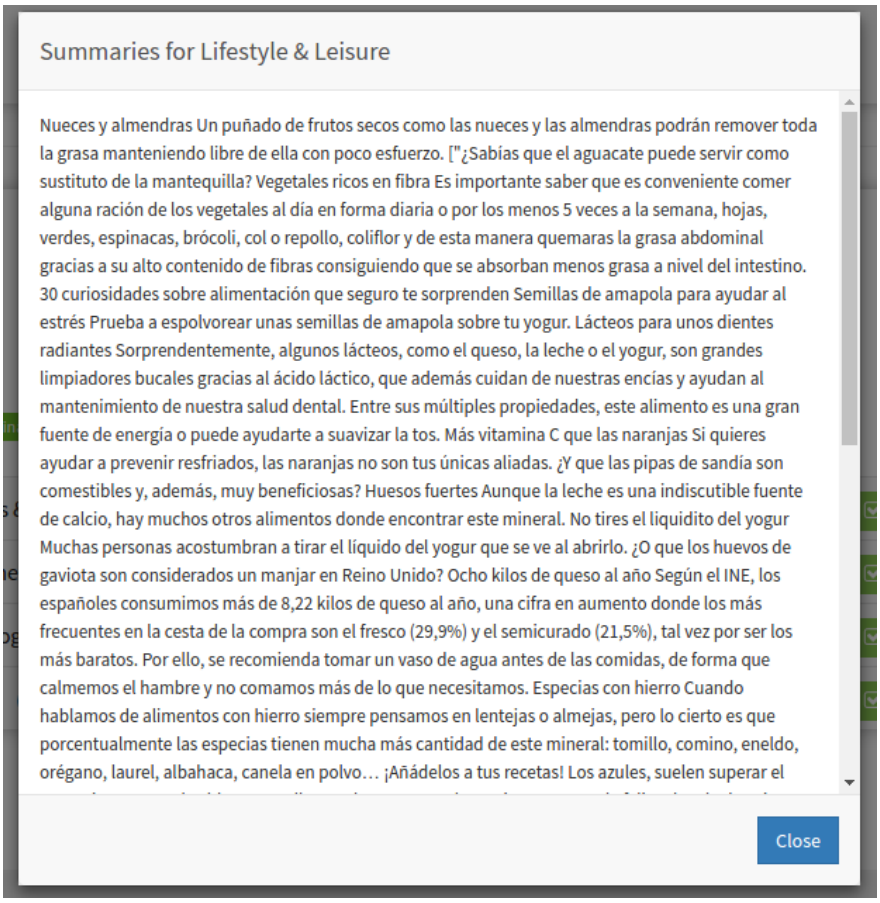


Figure 16: Multi-document extractive summary

If a user clicks on the cloud icon, the user will be presented with the main keywords from all texts within that set of articles. It is visualised in the form of a tag cloud (see Figure 17).

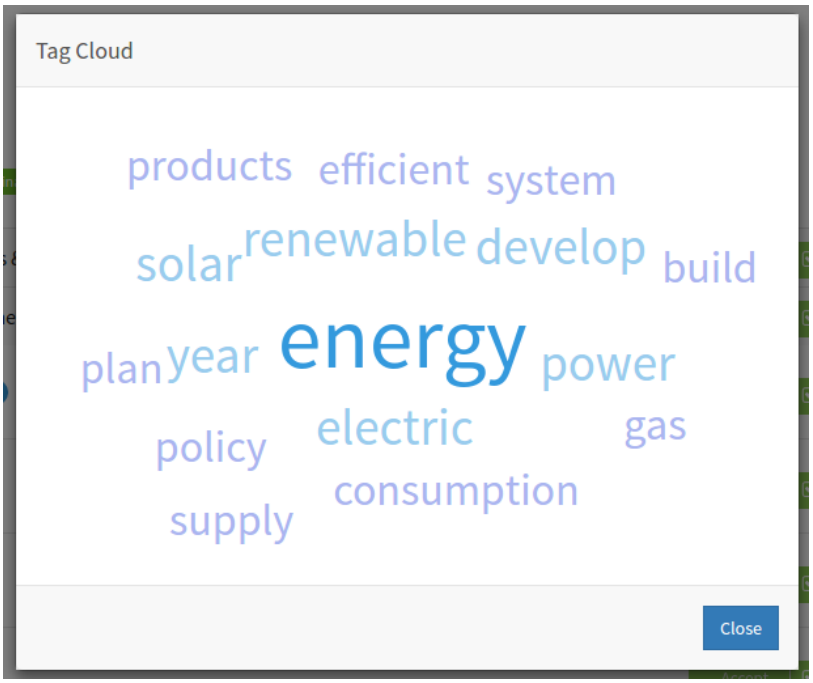


Figure 17: Aggregated most frequent words cloud

Analysis section: In the analysis section, visual results in the form of bar charts are shown for all articles that have been marked as relevant in the search section (see Figure 18). Rejected articles are not taken into account.

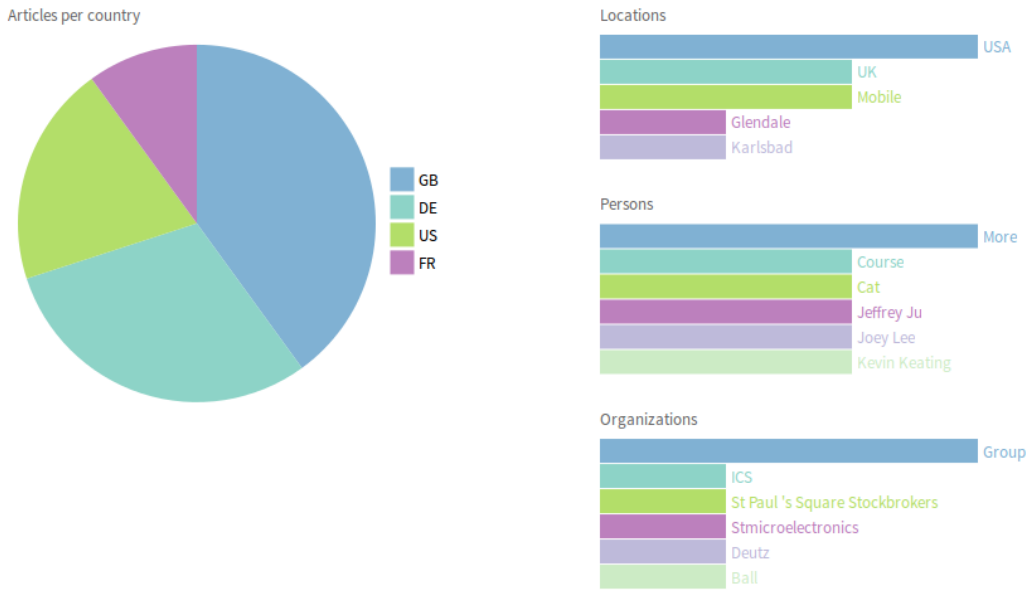


Figure 18: Analysis section

In addition, there is a chart that displays the scatter of sentiment over time in order to see how the sentiment of the set of articles has developed (see Figure 19).

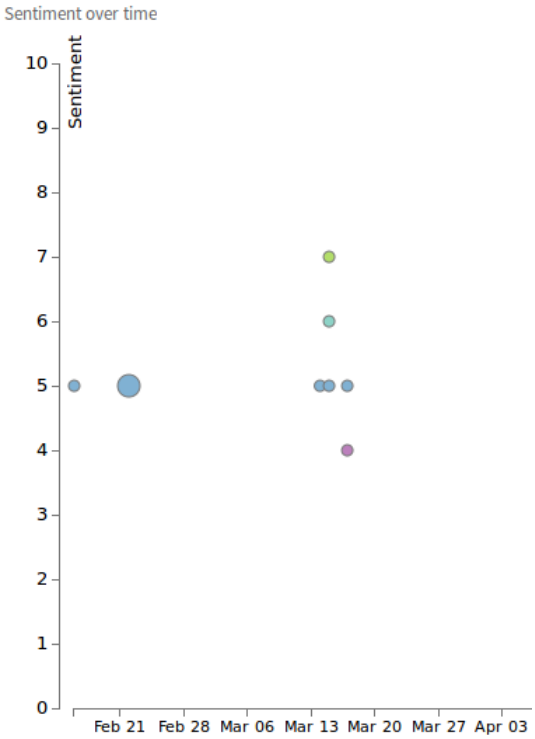


Figure 19: Scatter of the sentiment chart

By clicking on any chart section, the user is provided with the articles that fulfil the criteria, e.g. you can click on the section “GB” in the countries pie chart to see the articles from Great Britain. These are shown below the charts. Additionally, a multi-document summary is provided for these articles (see Figure 20).

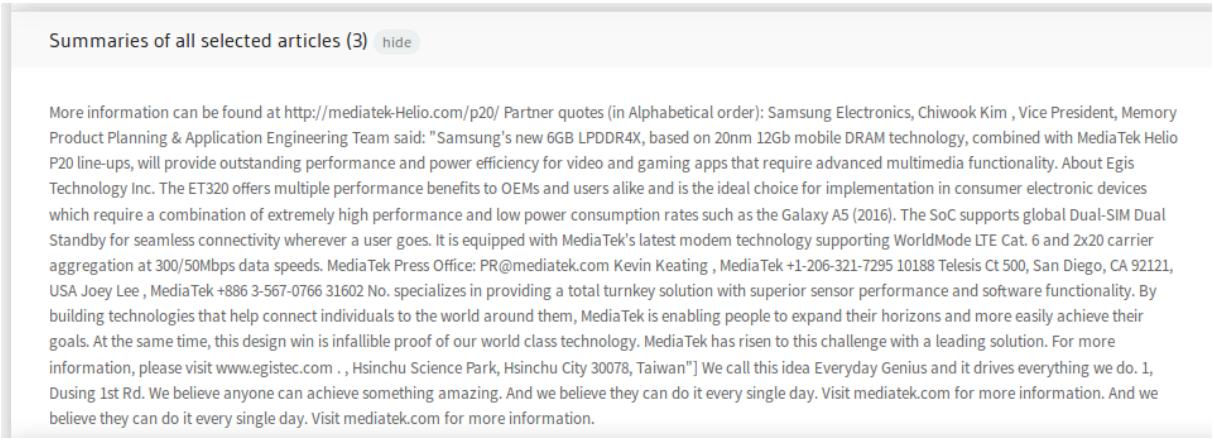


Figure 20: Multi-document summary

Influencer section: The influencer section displays the information that is extracted from the social interaction analysis services of MULTISENSOR (influential user detection and community detection). The relevant hashtags for the users, as well the 10 most influential Twitter users for a given set of hashtags are displayed in a tree map (see Figure 21). Furthermore, a grid called “Influencer meta-data” contains and provides additional data, such as the influence score, number of tweets, number of persons following and followers of each user. By clicking on the avatar/image of the Twitter user, the corresponding Twitter page will be opened to see directly the latest activities of the particular Twitter user. Finally, there is a chart that visualises the results of the community detection service (see Figure 22), where Twitter users are linked through mentioning and the more a user is mentioned by other users, the more edges target the specific user.

Image	Username	Name	Influence	Tweets	Following	Followers
	DiegoCusano_	Diego Cusano	0.099	1857	1041	3669
	gabrielesalari	Gabriele Salari	0.079	10874	1135	2120
	CNET	CNET	0.074	119670	295	1256593
	HWarlow	helen warlow	0.057	43413	13137	14786
	PaperGeekFr	PaperGeek	0.044	847	29	666

Figure 21: Twitter most influential users

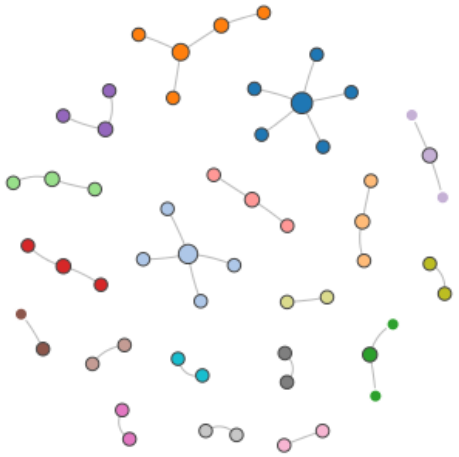


Figure 22: Detected communities chart

The SME Internationalisation Use Case

The SME internationalisation use case application supports SMEs in order to start a process of internationalisation with any kind of product. Relevant information related to the countries, the economic situation of the market, the legal information, and the exportation/importation conditions can be retrieved easily to support decision making. The list of the sectors and products considered in this application is structured in the following Table.

Sector category	**Sectors**	**Products**
C - Manufacturing	C10 - Manufacturing of Beverages	Tea Coffee Beer Soft drinks Juice
	C11 – Manufacturing of Food products	Dairy products Cheese Meat Ice cream Olive oil Bakery Vegetables Sugar Chocolate
	C13 - Manufacturing of Textiles	Animal Plant

		Mineral Synthetic

Table 1: List of the sectors and the corresponding products

When a user selects a specific sector, articles about that sector are shown. After the selection of a product, the search will contain specific information about it. The user is able to search for any keywords to retrieve the list of relevant articles. The display format of the articles allows seeing the summary with the "Read the whole summary/Read less" function that hides or shows part of it. After the summary, there is a drop list that allows choosing a language (FR, EN, DE, ES and BG) to translate the text of the article. At the bottom of every article, specific extracted concepts (if any) are displayed. The whole procedure is displayed in the following Figures (Figures 23 and 24).

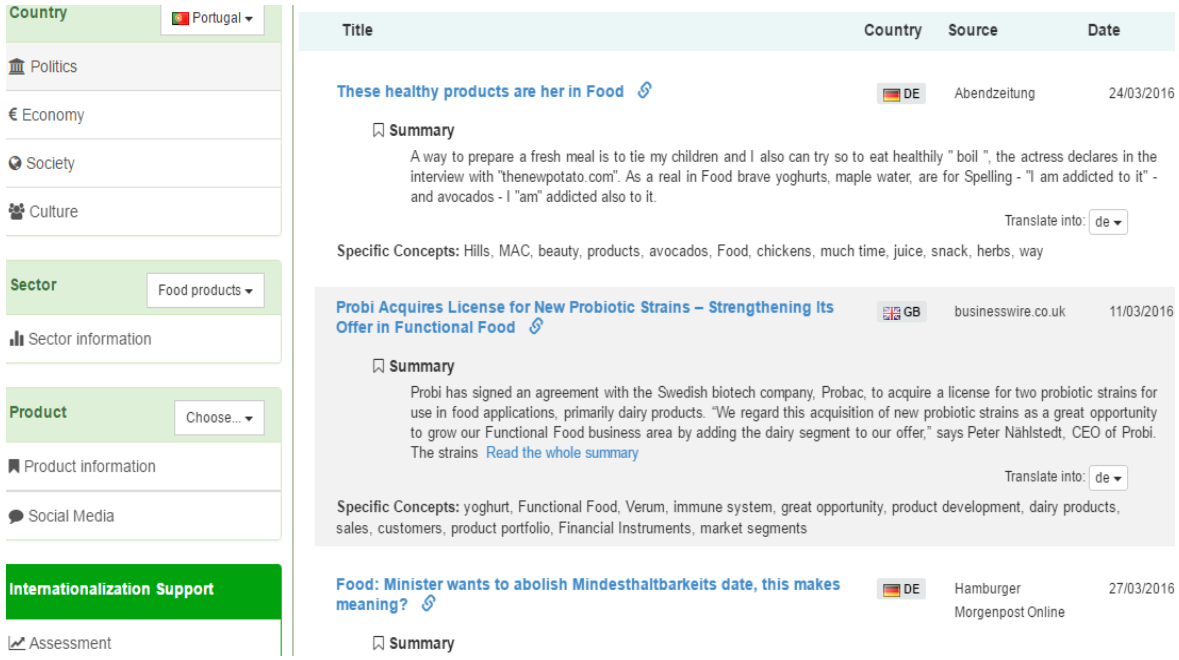


Figure 23: Selection of a specific sector

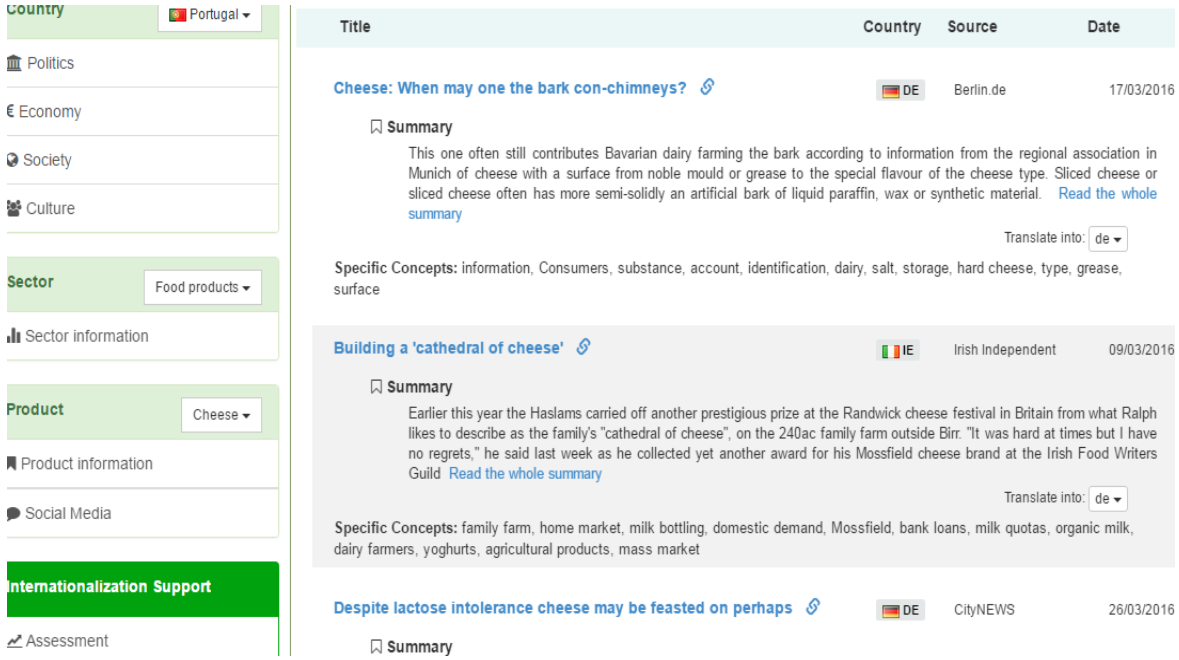


Figure 24: Selection of a specific product

The Social Media view displays the information that is extracted from the social interaction analysis services of MULTISENSOR (influential user detection and community detection). The visualisation of their results is provided in Figure 25.

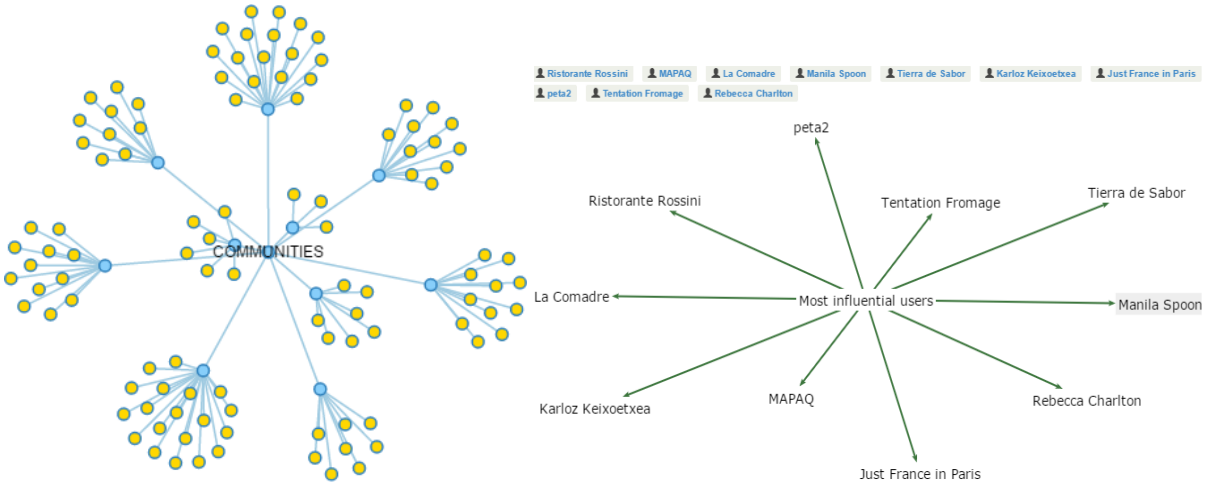


Figure 25: Social Media view: Community & Influential user detection

Another important functionality of this application is browsing specific information to a certain country for internationalisation support purposes, based on a number of indicators. The considered indicators have been selected and organised by categories to depict the relevant information related to a target country: Politics, Economy, Society and Culture. All the categories and the corresponding indicators are presented in Table 2.

Category	Sub-category	Indicators	Graphical representation
Economic indicators	GDP	GDP growth	Line chart
		Real GDP growth rate – volume (tec00115)	Line chart
		GDP per capita in PPS (tec00114)	Line chart
		GDP per capita – quarterly Data (namq_aux_gph)	Line chart
		Exports of goods and services in % of GDP (tet00003)	Line chart
		Imports of goods and services in % of GDP (tet00004)	Line chart
		Export to import ratio (tet00011)	Line chart
		Inward FDI stocks in % of GDP (tec00105)	Line chart
	Importation / exportation	Customs and tariffs	Multidimensional lines chart
		Structure of taxes by economic function (gov_a_tax_str)	Multidimensional lines chart
		Export and Import	Multidimensional lines chart
		Current account – quarterly data (ei_bpca_q)	Line chart
		Harmonised indices – monthly data (ei_cphi_m)	Line chart
		Foreign Direct Investment	Line chart
Political indicators	---	Government type	Bar chart
		Political instability index	Bar chart
		Corruption perception index	Bar chart
		General government deficit (-) and surplus (+) – quarterly data (ei_nagd_q_r2)	Bar chart
Social indicators	Population	Life table (demo_mlifetable)	Bar chart vertical
		Human Development Index	Line chart
		Population with tertiary education attainment by sex and age (edat_lfse_07)	Bar chart with age groups
	Work	Unemployment rate	Line chart
		Harmonised unemployment rates (%) – monthly data (ei_lmhr_m)	Line chart

	Health	Life expectancy	Bar chart with age groups
		Life expectancy by age and sex (demo_mlexpec)	Bar chart with age groups
		Population distribution	Line chart
Cultural indicators	Urbanisation	Distribution of population by degree of urbanisation, dwelling type and income group (source: SILC) (ilc_lvho01)	Bar chart
	Consumption habits	Economic sentiment indicator (teibs010)	Line chart
		Households having access to the internet at home (isoc_pibi_hiac)	Histogram
		Easiness of doing business	Bar chart

Table 2: List of the indicators displayed per category

Furthermore, the application offers the comparison of several indicators between two countries through the Decision Support System of MULTISENSOR. The SME professionals are interested in targeting specific countries to establish new commercial activities. For this, the "Assessment" view permits them to compare the indicators between two targeted countries. The user interface of the "Assessment" view is user-friendly, with a step-by-step guide in order to select the appropriate information (see Figure 26).

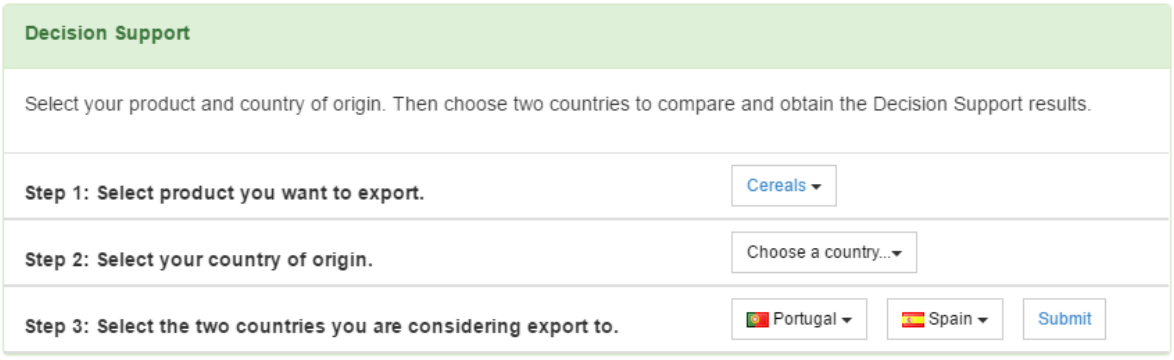


Figure 26: Step-by-step user guide in Decision Support, Assessment view

In the Comparison Table, some of the indicators mentioned previously appear (see Figures 27-29). There is also a service that returns a suggestion of the 3 best countries to introduce the selected product from the country of origin. The result of this service is displayed in the "Final results" section (see Figure 30).

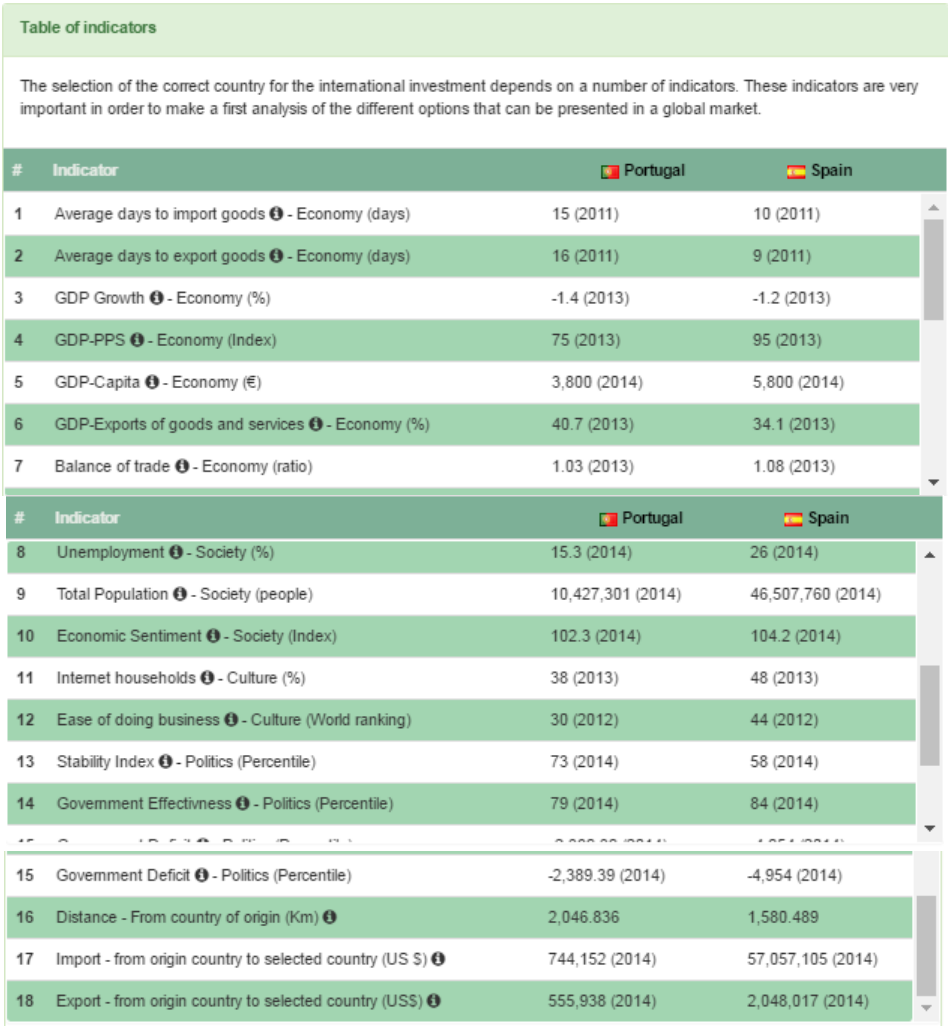


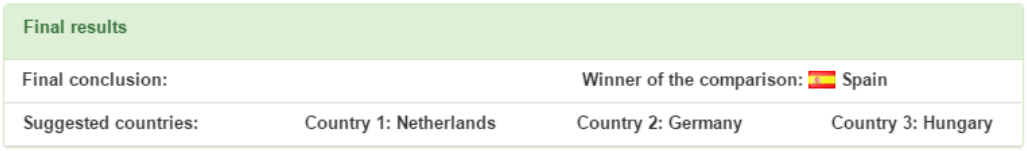
Table of indicators

The selection of the correct country for the international investment depends on a number of indicators. These indicators are very important in order to make a first analysis of the different options that can be presented in a global market.

#	Indicator	Portugal	Spain
1	Average days to import goods - Economy (days)	15 (2011)	10 (2011)
2	Average days to export goods - Economy (days)	16 (2011)	9 (2011)
3	GDP Growth - Economy (%)	-1.4 (2013)	-1.2 (2013)
4	GDP-PPS - Economy (Index)	75 (2013)	95 (2013)
5	GDP-Capita - Economy (€)	3,800 (2014)	5,800 (2014)
6	GDP-Exports of goods and services - Economy (%)	40.7 (2013)	34.1 (2013)
7	Balance of trade - Economy (ratio)	1.03 (2013)	1.08 (2013)

#	Indicator	Portugal	Spain
8	Unemployment - Society (%)	15.3 (2014)	26 (2014)
9	Total Population - Society (people)	10,427,301 (2014)	46,507,760 (2014)
10	Economic Sentiment - Society (Index)	102.3 (2014)	104.2 (2014)
11	Internet households - Culture (%)	38 (2013)	48 (2013)
12	Ease of doing business - Culture (World ranking)	30 (2012)	44 (2012)
13	Stability Index - Politics (Percentile)	73 (2014)	58 (2014)
14	Government Effectivness - Politics (Percentile)	79 (2014)	84 (2014)
15	Government Deficit - Politics (Percentile)	-2,389.39 (2014)	-4,954 (2014)
16	Distance - From country of origin (Km)	2,046.836	1,580.489
17	Import - from origin country to selected country (US $)	744,152 (2014)	57,057,105 (2014)
18	Export - from origin country to selected country (US$)	555,938 (2014)	2,048,017 (2014)

Figures 27-29: List of indicators – Assessment – Comparison Table



Final results

Final conclusion: Winner of the comparison: Spain

Suggested countries: Country 1: Netherlands Country 2: Germany Country 3: Hungary

Figure 30: Final results example

At the end of the "Assessment" view, there is a graphic summary of the indicators about the two targeted countries and the two best suggested countries. Here, main indicators representing Economy, Culture, Politics, Import and Export are displayed in a spider chart, as it can be seen in Figure 31.

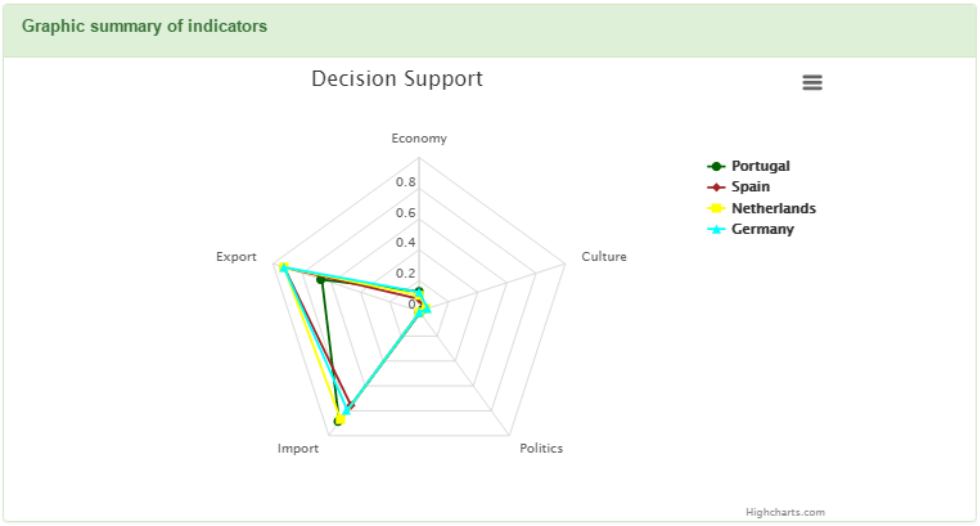


Figure 31: Example of Graphic summary of indicators

4.1.4 Achievement of the goals, portability and scalability assessment

This Section reviews the extent to which the stated MULTISENSOR goals with respect to the project's scientific objectives and WPs have been achieved. In addition, the Section assesses the portability of the results arrived at and their scalability and suitability for other tasks and domains.

SO1/WP2 (Multilingual and Multimedia Content Extraction): The indicators defined in deliverable D1.2 (Self-assessment plan v2) for assessing the tasks considered in this WP were achieved to the highest expectation for the named entity extraction, concept extraction, audio transcription and analysis and multimedia concept and event detection tasks. For the concept linking and relation extraction task, all indicators defined were achieved, very close to the highest expectations. Finally, regarding the machine translation task, depending on the language direction, the corresponding indicator defined in deliverable D1.2 was achieved to the highest expectation, while in some other cases, to the lowest expectation.

SO2/WP3 (User and Context-centric Content Analysis): Regarding the tasks of WP3, all indicators defined in deliverable D1.2 were fully achieved to the highest expectation.

SO3/WP4 (Multidimensional Content Integration and Retrieval): All indicators defined in deliverable D1.2 were achieved for all WP4 tasks to the highest expectation, with the exception of the category-based classification subtask of the topic-based modelling task, whose indicator was achieved to the lowest expectation.

SO4/WP5 (Semantic Reasoning and Decision Support): All WP5-related tasks were achieved, based on the corresponding indicators defined in deliverable D1.2.

SO5/WP6 (Summarisation and Content Delivery): Compared to the expectations set in deliverable D1.2 (Self-assessment plan v2), in most cases the significant improvements over the baseline that were set were not reached.

Regarding the assessment of the project results' portability, scalability and suitability for other tasks and domains, the following can be reported:

WP2 (Multilingual and Multimedia Content Extraction)

Named Entities Extraction Component: Because of its modular architecture, the component can be adapted to other domains and languages with only limited effort. The software components are basically language-independent, so that only the resources need to be adapted, i.e. Named Entity Recognition (NER) lexicon (annotated gazetteer) and grammar rules.

Dependency Parsing Component: The dependency-parsing module itself is language-independent. To apply it to other languages than those in MULTISENSOR, text corpora in the corresponding languages must be annotated with surface- and deep-syntactic structures. If a surface-treebank is already available, a semi-automatic mapping of the surface structures to deep-syntactic structures can be performed. The effort for the annotation may range from 3 PMs to 12 PMs.

Concept Extraction Component: In order to adapt the concept extraction module to new domains, sufficient language-specific training material is required. The annotation effort depends on the language and linguistic difference of the new domain with the MULTISENSOR domains.

Automatic Speech Recognition Component: Adaptation to other domains requires large in-domain data to train appropriate language models. Adaptation to other languages is substantially more complex, as it involves extensive recordings of native speakers together with precise transcriptions.

Multimedia Concept and Event Detection Component: The component is language independent and given a reasonable amount of time (approximately 3 PMs/10 concepts), it could be trained and extended/adapted to other concepts and events. In order to extend it to additional concepts, the module requires annotated data (i.e. images and videos that depict this concept/event and training of the predictive models).

Machine Translation Component: Adaptation to other domains requires bilingual in-domain texts to train appropriate phrase tables. In addition, adaptation to other languages requires substantially larger bilingual training texts. By using English as a pivot language, however, it is possible to support 12 additional translation directions (e.g. French-German, Bulgarian-Spanish).

WP3 (User and Context-centric Content Analysis)

Context Extraction and Representation Component: For the features extracted from the metadata, the module can be considered language-independent. For the features extracted from the text, we assume that the text would be in English (either originally or translated by the relevant component). However, the module can be extended to other languages given the availability of annotated training data.

Polarity and Sentiment Extraction Component: Contrary to lexicon-based, domain-independent solutions, which are not straightforward in how they can be effectively extended to other domains, the proposed module is a tailor-built, machine learning-based solution that targets specific domains

of interest. In principle, the Polarity and Sentiment Extraction module is language-independent and it is possible to extend to new domains, given the availability of annotated text corpora with sentiment features. However, certain techniques and features (e.g., syntactic features) may introduce language dependencies.

Social Interaction Analysis Component: The Social Interaction Analysis component is domain- and language-independent and can be expanded in several ways, like i) integration of additional signals in the computation of authority score and ii) refinement of the interest analysis. Moreover, the module may find applications in other contexts, such as for example, in the viral marketing context. In this scenario, having the spheres of influence pre-computed and stored in an index, may provide a direct solution to several variants of influence maximisation (for instance in the case where different segments of the market audience have different values for a viral marketing campaign). Then when the next campaign is run, and the users have different values, the same spheres of influence can be reused. Other examples might include viral marketing campaigns under different types of constraints, such as, e.g. when different nodes have different costs to become a seed. Outside of viral marketing, we may consider the application of the module in contagion problems, or in the vaccination problem.

WP4 (Multidimensional Content Integration and Retrieval)

Multimodal Indexing and Retrieval Component: The component is language and domain independent. If the data representation requirements that stem from the characteristics of the SIMMO model are taken into account and met, then it is plausible to extend the component to include additional fields of information with relative ease (e.g. 1-2 PMs).

Topic-based Modelling Component: The component, if trained accordingly, can receive as input features from any number of modalities. The extension/adaptation procedure of the component to other domains is considered relatively easy (about 2 PMs for any new domain set (regardless of the domain number)). It must also be taken into account that the component receives its input from the indexing structure of the multimodal indexing and retrieval component, therefore any extension/adaptation to other domains should also consider the indexing component. The approach is language agnostic and it assumes that each multimedia document is represented with vectors (e.g. bag of words).

Mapping Discovery and Validation Component: The component can be extended by integrating new similarity algorithms that take advantage of different metrics. For instance, in order to include a new existing matching algorithm, a work of 2-3PMs would be required. This application is domain and language independent.

Content Alignment and Integration Component: The component can be adapted to run on any semantic repository, while it can be tuned to specific domains of interest. In order to adapt it for a different domain, an effort of 2-3PMs is expected. The component is language independent.

WP5 (Semantic Reasoning and Decision Support)

Data Infrastructure Component: GraphDB can be adapted to work on the cloud very easily. ONTO already has such kind of project named S4, where we offer Data-as-a-Service. GraphDB has the ability to work with many different languages, so all the innovations support these languages tool.

Semantic Representation Infrastructure Management System Component: Currently, the GraphDB Workbench is available only in English. However, GraphDB Workbench can be easily extended depending on a specific use case.

Decision Support System Component: Currently, GraphDB Workbench that takes the role of management application is available only in English language. GraphDB Workbench can be easily extended depending on a specific use case, but precise estimation can't be given because it highly depends on the use case.

WP6 (Summarisation and Content Delivery)

Extractive Summarisation Component: To adapt the extractive summarisation module to new domains, training material consisting of a sufficient number of sample summaries (along with the original texts) of high quality is needed. No estimation of the effort of the compilation of such summaries can be given, since it depends on the existence of quality-annotated summaries for specific domains and languages. The effort for retraining of the summarization module is minimal.

Abstractive Summarisation Component: The content selection component of the module can be operated on any RDF/OWLIM content structures without the need of adaptation. The discourse structure module will require some adaptation to new domains; it is, however, language-independent. The adaptation of the rule-based text/sentence generator presupposes the compilation of language-specific grammatical and lexical resources at different levels of the linguistic description. The size (and thus coverage) of these resources depends on the nature and verbosity of the targeted summaries. The adaptation of the stochastic sentence generator requires the annotation of text corpora with linguistic structures at different levels of abstraction (surface-syntactic, deep-syntactic, and semantic) for each language that is to be covered by the summariser (this was done for English, French, German, Spanish in the context of the project). For an acceptable performance of the stochastic sentence generator, training treebanks with between 3,500 sentences (with a very high quality of annotation) and 10,000 sentences are needed. The estimated cost is about 1PM per 1000 sentences of high quality annotation.

WP7 (System Development and Integration)

Crawlers and data channels infrastructure Component: This component may be adapted to other "Collector" sub-components. If these respect the interface of the Crawler, integration should be perfectly feasible.

Content Extraction Pipeline (CEP): The current version of the CEP is not easily adapted and

extensible. The generic approach to extend the CEP integrating new services will require several improvements, such as the possibility to select an ontological data model that should be used to produce the output. The ontology used to structure the content should be compliant with the other services used in the pipeline. Also, the output of the new service should be automatically validated.

Final System: With respect to the adaptability of the MULTISENSOR Final System to new domains or languages, the System depends on all the aforementioned modules, for which specific details have been provided with respect to their extensibility and adaptability. With respect to scalability, it should be noted that during the project, a cloud-based infrastructure approach was adopted. The main reason for this is that a flexible solution was needed to accommodate the system's evolving requirements. At the beginning of the project, it was rather difficult to foresee the required capacity, and therefore a flexible solution was essential. The aforementioned fact that the MULTISENSOR infrastructure is cloud-based makes scalability possible without major challenges. On the other hand, it is also possible to implement the System in a local infrastructure. This approach will not limit the System's functionalities in any way, provided the infrastructure is properly dimensioned. However, the performance may be slower due to many factors like: network capacity, server potentiality, average content size especially applicable to multimedia, etc.

4.1.4 Potential impact, dissemination activities and exploitation results

Impact

At the beginning of the project we have defined several key performance indicators regarding project operation (efficiency, progress tracking and quality) and impact. To assess the project impact, we use a number of those KPIs, as well as some additional ones. Those metrics have been derived based on the dissemination and exploitation activities, and the recording of the project IPs. Those KPIs were estimated at the end of 2016, thus reflecting the impact of the project shortly after its completion. It is expected that several of them will be increased in the coming period (e.g. some more journal articles and conference papers will be accepted, papers will attract more citations, etc.). In most cases, the values of KPIs are based on manual inspection of online services (e.g. number of app downloads in app store, number of Twitter followers, Google analytics, Google scholar).

Table 3, presents the obtained scores along with the goals when those were specified in the beginning of the project. For several of those metrics, no concrete goals had been specified at the project outset. Looking at the mere numbers, MULTISENSOR has fulfilled all the goals set in the beginning of the project sometimes even exceeding the numbers, e.g. publishing more articles to scientific papers or participating in more conferences than planned. While this allows us to call it a successful dissemination campaign, the real success lies somewhere beyond the numbers.

Impact Area	KPI	Goal	Value
Community Involvement	A. Number of Joint Workshops	2	4
	B. Number of Participants in (A)	30 to 40 (avg.)	16 to 34 (avg)
	C. Number of Initiatives, Events and Conferences	9	34
	D. Number of Meetings with related projects	9	16
	E. User Group size	-	38
	F. Number of press releases	3	5
	G. Number of newsletter subscribers (PIMEC, PR)	-	>27,550
	H. Number of User Days and Open Door Days	4	5
	I. Number of Participants in User and Open Door Days	160	244
	J. Number of Demonstrations to Interested Stakeholders	>20	30
	K. Number of Project Presentations	>10	14
	L. Number of Twitter followers	-	225
	M. Number of LinkedIn members	-	79
	N. Number of unique website visitors	-	11,000
Impact on research	A. Number of accepted journal articles	-	4
	B. Average impact factor	-	
	C. Number of conference and workshop papers	-	44
	D. Number of downloads of project results	-	831
	E. Number of datasets made available	-	6
Impact on network search technologies & applications	A. Number of listed IPs	-	21
	B. Number of inputs to standardisation	-	6
	C. Number of open-source projects	-	16
	D. Number of modules integrated to user partners	3	3

Table 3: Project impact KPIs (November 2016)

Social Impact

As can be seen by the respective KPIs, the project has been very successful in stimulating community involvement. For instance, a considerable number of interested stakeholders (over 200) have attended both user days and open door days co-organised by MULTISENSOR. In terms of social media audience for the official project accounts, the recorded numbers may be considered moderately successful, amounting to approximately 225 Twitter followers and 79 LinkedIn Members. The connections made on LinkedIn or Twitter, will now allow us to reach out and further explain, where we would like to go with this idea. Or maybe set up a follow up project with a new contact coming through the successful dissemination activities of this project.

The primary communication channel for the project was the website. The website was setup as a central hub for the project to publish information about the work. It served as a reference for people at conferences to find more information about the project, the partners involved and the progress as well as a way to get in touch with the consortium to talk about MULTISENSOR and the research done in the project. It was hence important to us to keep the information updated and at hand for

everyone coming to the page. Looking at the numbers, it clearly shows that there was an interest in the project and the information published on the website. Overall, throughout all three years, the website reached a total of around **11.000 visits**, accumulating to a total of **43.658 page views** (32.959 unique) with an average duration of around **4 minutes**.

Among the most visited website pages (after the home page) are the deliverables[3], code[4] and publications[5] pages. Most visited webpages also include popular blog posts[6], the partners' section[7] and the related projects' section[8].

Scientific impact

The project had considerable impact from a scientific perspective. Regarding content analysis, MULTISENSOR pushed the state-of-the-art in concept and concept relation extraction, by developing a hybrid concept extraction system and a multilingual analysis pipeline based on cutting-edge neural network dependency parsers and a new kind of deep-syntactic transducer and a fast rule-based frame-semantics parser. In addition, there were advancements in the context of speech recognition, by implementing a scalable and versatile system, able to accept audio and video input in multiple formats by means of powerful format converters. Regarding summarisation, MULTISENSOR has worked on making extractive summaries more accurate and more coherent and cohesive to the final user, increasing in this way their readability and usability. Finally, MULTISENSOR has developed novel technologies in the context of semantic integration, indexing and retrieval of heterogeneous data, as well as in the context of topic-based modelling, with a focus on the task of topic detection by researching and developing novel techniques.

In total, 44 conference and workshop papers and 4 journal articles were accepted, while a number of submissions are still in review. Out of those, 3 were published in 2013, 16 in 2014, 15 in 2015 and 14 in 2016. Based on a manual data collection using Google Scholar in the beginning of December, the total number of citations that these publications gathered amounted to 407.

Table 1 presents the top 10 most highly cited publications in the beginning of 2016. It appears that several of the results in the areas of natural language processing, social network analysis and multimedia analysis have already made considerable impact (given the typical numbers of citations attracted by works in those areas), and we expect that even more project results will create similar and even bigger impact in the years to make.

---

[3] http://www.multisensorproject.eu/achievements/deliverables/

[4] http://www.multisensorproject.eu/achievements/code/

[5] http://www.multisensorproject.eu/achievements/publications/

[6] http://www.multisensorproject.eu/influence-indirect-ties-social-network/, http://www.multisensorproject.eu/machine-translation-between-dream-and-reality/

[7] http://www.multisensorproject.eu/project/partners/

[8] http://www.multisensorproject.eu/related-projects/

#	Publication Title	Area	Citations
1	Transition-Based Dependency Parsing with Stack Long Short-Term Memory	Natural language processing	137
2	Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs	Natural language processing	54
3	Who to follow and why: link prediction with explanations	Social network analysis	36
4	Online Topic-aware Influence Maximization Queries	Social network analysis	26
5	On the Feasibility of Predicting Popular News at Cold Start	Recommender systems	18
6	Influence-based Network-oblivious Community Detection	Social network analysis	17
7	Deep-syntactic parsing	Natural language processing	16
8	Influence Maximization with Viral Product Design	Social network analysis	11
9	ITI-CERTH participation to TRECVID 2013	Multimedia analysis	11
10	Data-driven sentence generation with non-isomorphic trees	Natural language generation	10

Table 1: High-impact research based on MULTISENSOR research (Google scholar, Dec 16)

Market impact

The project had considerable impact on the sector of multilingual media analysis technologies. In particular, there have been already 21 IPs generated in total by project partners and listed and described as part of the exploitation plan. Based on the generated IPs, business partners have already integrated MULTISENSOR modules to their workflow (pressrelations), have added and/or consulted MULTISENSOR technologies to advance their product suite (PIMEC) and are also planning to commercialise a number of promising MULTISENSOR technologies. In particular:

- PR has already **integrated** several MULTISENSOR exploitable modules into its media monitoring workflow, namely the summarisation and translation module and the influencers' detection.
- PIMEC already **used and consulted** the MULTISENSOR for an SME that was interested in exporting cosmetics both for professional (e.g. physiotherapist) and aesthetic use. It is a familiar SME that never exported before and required of PIMEC services to elaborate an initial International Plan to define the countries to export to then start the action.

In addition, MULTISENSOR has contributed to a small number of standardisation activities, relating to well-known standard bodies, mainly the W3C. By doing this, the project has helped the scientific and technological society to further develop and enhance existing tools and frameworks. More specifically, six (6) standard initiatives were addressed. MULTISENSOR, represented through its partners CERTH, ONTOTEXT, UPF and EURECAT, contributed to

- The MPEG-7 Standard (W3C)
- Linguistic Linked Data

- JSON-LD (W3C)
- RDF/OWL/SPARQL, CUBE ontology (W3C)
- The Emotion Incubator Group, Emotion Markup Language (EmotionML) (W3C)

Finally, MULTISENSOR managed to publish 16 different frameworks and code fragments, that were (are) made available through the projects website[9], for everyone to download. In total so far 90 downloads of different elements have been registered through the website (the actual numbers might be a bit larger, as the code can also be directly accessed through the publication pages, e.g. github, directly). As not all code elements were available before the end of the project, there were no download numbers available for some of them. But the code elements will be published through the website beyond the end of the project, as long as the website stays online and accessible (for another two years beyond the end of the project.) The following table gives an overview of all the code elements made available through the project and, where already applicable, the number of downloads tracked through the website of those elements.

No.	Code Element	Description	Partner	Number of downloads
1	VERGE	Hybrid interactive video retrieval system, capable of searching into video content by integrating different search modules that employ visual- and textual-based techniques.	CERTH	4
2	Socially interconnected/ interlinked and multimedia-enriched objects	Model for representing multimedia content in the context of the Web and Social Media.	CERTH	25
3	Mate Tools surface statistical dependency parser	Tool converting plain text to dependency structures annotated with surface-syntactic relations, lemmas, part of speech, and morpho-syntactic features	UPF	15
4	Character-based Stack-LSTM surface parser	Tool to convert plain text to dependency structures annotated with surface-syntactic relations.	UPF	Not available yet

[9] http://www.multisensorproject.eu/achievements/code/

No.	Code Element	Description	Partner	Number of downloads
5	Deep statistical dependency parser	Parser converting surface-syntactic dependency structure as produced, e.g., by the MATE Tools parser, to dependency structures annotated with deep-syntactic relations in the sense of the Meaning-Text Theory	UPF	12
6	DSynt Converter (ENG)	Tool to convert reference surface-syntactic annotation of English (Penn TreeBank) into its corresponding deep-syntactic annotation in the CoNLL'09 format	UPF	5
7	Frame Semantics parser (ENG)	Tool to produce structures as found in FrameNet.	UPF	Not available yet
8	Deep statistical text generator	Tool converting deep-syntactic dependency structures in the sense of the Meaning-Text Theory (in the CoNLL'09 format) to linearised structure with all the words of the sentence.	UPF	Not available yet
9	Twitter Crawler	Crawler for Contributor Analysis and Name Search	EURECAT	12
10	Topic detection	Framework for topic detection as a clustering problem and a hybrid clustering approach for assigning news articles into topics.	CERTH	9
11	Category-based classification	Framework for classification of news articles into a predefined set of generic categories	CERTH	6
12	Multimedia retrieval	Multimedia language-independent retrieval framework fusing multiple modalities, integrating high-level information.	CERTH	2

No.	Code Element	Description	Partner	Number of downloads
13	Multimedia concept and event detection	Framework for video concept and event detection	CERTH	Not available yet
14	Community detection	Module for the detection of Twitter communities, given a list of desired keywords/hashtags	CERTH	Not available yet
15	Ontology alignment	Ontology alignment algorithm for computing a visual-based similarity metric for entity matching between two ontologies	CERTH	Not available yet
16	User and Context-centric Content Analysis	Code for the implementation of models for representing contextual, sentiment and online social interaction features, as well as deploying linguistic processing at different levels of accuracy and completeness	EURECAT	Not available yet

Table 4: Open-source projects developed by MULTISENSOR.

Dissemination activities

Consortium partners have been engaged in the pursuit of market exploitation of the project's new technologies and applications. MULTISENSOR tried a varied approach in order to reach out to the target audiences with the necessary repetition. This was to ensure that MULTISENSOR could reach everyone with a targeted message. In general, dissemination activities included publications, organisation and participation in events, organisation of user and open door days and collaboration with projects. Exploitation activities included the creation of the MULTISENSOR User Group, participation in exhibitions and industrial events, integration of modules in actual business workflows, discussions about opportunities for joint exploitation schemes, and contact with companies in order to establish common exploitation strategies.

The MULTISENSOR consortium has successfully participated in and organised a number of dissemination events and participated in industrial events in order to promote the project results. A summary is depicted in Table 5 and Table 6.

Furthermore, users awareness is pursued through the issuing of pressreleases and newsletters, presenting all the latest news and events of the project, useful information and tips. Finally the MULTISENSOR video showcase has been produced and is publicly released. The project is also

actively present in Web 2.0, notably Facebook, Twitter and LinkedIn.

Activity	Number	Notes
Workshops	5	- Joint workshop with EUMSSI in Barcelona (11/03/2015) - Joint evaluation workshop with EUMSSI in Bonn (23-27/11/2015) - Workshop with EUMSSI on Multimodal media data analytics (MMDA) (http://mklab.iti.gr/mmda2016/) in conjunction with the 22nd European Conference on Artificial Intelligence (ECAI) 2016, in The Hague (30/08/2016) - Joint evaluation workshop with EUMSSI in Barcelona (22/09/2016) - Workshop organization in ICMR 2016: 1st International Workshop on Multimedia Analysis and Retrieval for Multimodal Interaction (MARMI). ((http://mklab.iti.gr/marmi2016/))
User Days and Open Door Days	3	- User & Open Door Day combined with the consortium meeting in Bonn (23-27/11/2015) - Evaluation day in collaboration with Media Informatics Lab of School of Journalism & MC in Thessaloniki (15/09/2016) - User Day combined with the consortium meeting in Barcelona (22/9/2016)
Publicity	5	- 1 dissemination video - Social media activity (Facebook, Twitter, LinkedIn) - Newsletters - Press Releases - Website
Clustering	16	- Liaising with EUMSSI (7 activities) - Meeting with SocialSensor project - Meetings with Reveal project (2 activities) - Meetings with KRISTINA project (3 activities) - Participation to "Cracking the language Barrier" Initiative - Joined European Centre of Social Media - MULTISENSOR collaborated with TENSOR project in SIMMO model and in particular community detection, abstractive summarisation and other modules
Exploitation	5	- Market Analysis for all 3 use cases (Journalistic, Media Monitoring, Internationalisation) - Standardisation

		- Open sourcing - Integration of MULTISENSOR modules in media monitoring Business Workflows (pressrelations) - Consultancy and Use of Specific Modules (PIMEC)

Table 5: Dissemination and exploitation activities co-organised by MULTISENSOR.

Activity	Number	Notes
Journals	4	- JASIST 2016 - Multimedia Tools and Applications 2016 (2 papers) - Natural Language Engineering
Conferences	34	- INNS Big Data 2016 - EISIC 2016 - 54th Annual Meeting of the Association for Computational Linguistics - 3rd International Conference on Internet Science - 12th International Conference on Machine Learning and Data Mining - ICMR 2016 - KDIR'15 - FDIA 15 - MMM ’16 (2) - EMNLP 2015 - CNLL 2015 - ACL-IJCNLP 2015 - KEOD 2015 - IKC 2015 - ICME 2015 - NAACL HLT 2015 (2) - MMM 15 (2) - International Conference on Data Mining 2014 - 6th International Conference on Social Informatics 2014 - SocInfo 2014 - ACM SIGKDD 2014 - IRFC2014

		- COLING 2014 - INLG 2014 - ECIR 2014 - Video Browser Showdown (VBS) 2014 - IEEE International Conference on Data Mining, 2013 (2) - SIAM International Conference on Data Mining – 2014 (2) - International Conference on Extending Database Technology - 2014.
Workshops	10	- 11th International Workshop on Semantic and Social Media Adaptation and Personalisation (SMAP 2016) - 1st International Workshop on Multimodal Media Data Analytics (MMDA 2016) - 14th International Workshop on Content-based Multimedia Indexing (CBMI) (2) - International Workshop on Multimedia Forensics and Security (MFSec 2015) - Environmental Multimedia Retrieval Workshop (EMR 2015) - 3rd Workshop on Vision and Language (VL) 2014 - COLING'14 Workshop on Vision and Language (VL'14) - Environmental Multimedia Retrieval Workshop (EMR 2014) - TRECVID 2013 Workshop
EU-organised events	2	- European Data Forum ‘16 - Horizon 2020 ICT-16 Big Data
Standardisation	6	- The MPEG-7 Standard (W3C) - Linguistic Linked Data - JSON-LD (W3C) - RDF/OWL/SPARQL, CUBE ontology (W3C) - The Emotion Incubator Group, Emotion Markup Language (EmotionML) (W3C)
Industrial exhibitions	7	- CeBIT ‘14 - Global Media Forum ’14, ‘16 - Technology Forum (Greece) ‘14

		- FIBEP Congress - Biz Barcelona 2016 - AMEC International Summit ’15, ‘16 - LT Innovative Summit 2013
Other events and activities	9	- Kommunikationskongress - ESSIR 2015 - ICME 2015 - Riga Summit, META-FORUM ‘15 - Virolai school, Big data project ‘15 - BBC NewsHack ‘14 - Big Bang Data exhibition at the CCCB in Barcelona ‘14 - ESCW ‘14

Table 6: Dissemination activities with MULTISENSOR participation.

Organisation and participation in dissemination and exploitation events

During the project, the project’s consortium partners have successfully participated in and organised a number of dissemination events. More specifically:

MULTISENSOR collaborated with the EUMSSI project and organised a total of two (2) User and Open Door Days with many attendants. MULTISENSOR managed to set up both, the two large user days, as well as the two open door days, bringing in a large number of potential users from the target audiences to both discuss as well as evaluate the project’s prototypes. All meetings worked out well, giving the project valuable feedback on its development as well as a good publicity. In order to evaluate the final journalistic demonstrator an evaluation workshop was also organised by CERTH and AUTH in order to receive feedback from expert users. In particular, the evaluation of the MULTISENSOR Final System’s journalistic demonstrator was carried out at the premises of the Department of Journalism & Mass Communications, Aristotle University of Thessaloniki, Greece. It consisted of a group session, which lasted about 90 minutes and in which around 20 persons (both Master and PhD students) participated. After a brief demonstration of the journalistic interface’s functionalities, the participants were asked to perform specific tasks and evaluate their general experience when using the MULTISENSOR Final System.

Specifically with respect to the journalistic demonstrator we aimed to evaluate the usability, user experience, effectiveness, efficiency and user satisfaction. Upon the completion of the session, we analysed the participant group’s responses. The overall results from the evaluation group session are included in the final evaluation deliverable of MULTISENSOR (D8.5). Overall the expert users expressed their high interest in MULTISENSOR results and they found the functionalities of the system very useful for their work.

Throughout the duration of the project, the partners organised and attended a number of scientific and industrial events relevant to the project: the ICT in Lisbon, the ICME 2015 in Torino, the

Kommunikationskongress in Berlin, the FIBEP Congress in Vienna and the European Data Forum in Eindhoven are among the major events in which MULTISENSOR has been advertised by means of presentations, demos, booths, etc. In 2016, MULTISENSOR also participated in CeBIT in Hanover advertising the project outcomes and exploring collaboration possibilities with companies in bi-lateral meetings, the ESWC Conference and the Global Media Forum in Bonn.

To strengthen the connection of the project with potential stakeholders, beyond the pre-planned joint EUMSSI Meetings and the conferences mentioned before, MULTISENSOR also tried to collaborate with other additional EU-funded research projects. With a target of three (3) meetings per year with related ICT projects, so 9 in total, the project wanted to make sure to go beyond the own development silo and get more outside views on the development, while at the same time widening its own horizon and giving valuable feedback to others. As was reported earlier, the project managed to meet a total of 14 times, with other projects including SocialSensor, EUMSSI, REVEAL, KRISTINA and TENSOR. Exchanges that proved to be very fruitful and helpful to both sides.

Ad hoc-events with specific stakeholders have also been targeted. MULTISENSOR was demonstrated and/or presented to other projects including ENGINE, MUMIA, WikiRate. Pressrelations also presented the project outcomes in media monitoring companies including Newsaccess and OBSERVER and PIMEC demonstrated the prototype to export managers and freelancers.

Many more conferences and events have been attended by several members of the consortium. The most relevant conferences in which scientific contributions have been presented include the Annual Meeting of the Association for Computational Linguistics (ACL), the International Conference on Machine Learning and Data Mining (MLDM), the ACM International Conference in Multimedia Retrieval (ICMR), the International Conference on Multimedia Modeling (MMM) and the IEEE International Conference on Multimedia and Expo (ICME).

Exploitation activities and first results

Regarding exploitation activities, apart from the project participation in industrial events, as mentioned above, MULTISENSOR has formed and populated its User Group. The group started out with experts known to the consortium in the areas covered by the project. They covered user topics like media monitoring and internationalisation as well as technical area of the project, such as sentiment analysis or language technologies.

The members of the user group were contacted upfront regarding their interest in being involved in the project's work. They were then contacted individually for evaluation sessions, the user days and open door days, as well as participating on the consortium meetings of MULTISENSOR. Starting out rather small, the project managed to grow the User group up to 38 members.

No.	Organisation/Expert	Description	Country
1	Data Scouting	ICT company providing media monitoring solutions	Greece
2	JRC	Research centre working on media monitoring	Italy
3	Ekonm	ICT company providing semantics and data mining solutions	Israel
4	DOTSOFT	ICT company providing data mining solutions	Greece
5	Mozaika	ICT company providing data science solutions	Bulgaria
6	SEERC	Research centre working on knowledge management	Greece
7	Fraunhover MOEZ	Research centre	Germany
8	Fraunhofer FAME	Research institute working on the future of media	Germany
9	ATC	Software Company, Media research and development	Greece
10	European Journalism Center	Journalism Training and Research Organisation	Netherlands
11	Beeld & Geluid	Cultural-historical Media Organisation	Netherlands
12	IRT	Research Institute for Mediatechnology	Germany
13	CASMAR	SME (security systems)	Spain
14	Zebra Design & Retail	SME (design & Retail)	Spain
15	Jordi Mallorquí	Export Manager freelance	Spain
16	Jordi Planas	Export Manager freelance	Spain
17	Míriam Sabaté	Export Manager	Spain
18	Joan Carles Espigol	Export Manager freelance	Spain
19	Jerusalem College of Technology	College of Technology	Israel
20	Aii Data Processing Ltd	Market-Expert for monitoring, measurement and analytics of mainstream and social media with focus on emerging markets of Central and Eastern Europe	Bulgaria

No.	Organisation/Expert	Description	Country
21	QMUL, Multimedia and Vision Research Group	Multimedia applications	UK
22	University St. Gallen	Media & Communications Department	CH
23	Clipit	Media Monitoring	NL
24	eMedia Monitor GmbH	Media Monitoring (RTV)	Austria
25	Media Monitoring Project Zimbabwe	Media Monitoring	Zimbabwe
26	R-Media	Media Monitoring	Italy
27	Media Image Group	Media Monitoring	Romania
28	Euregio Srl GmbH	Media Monitoring	Italy
29	Media Tenor	Media Monitoring	Czech Republic
30	Media Informatics Lab	Media Research	Greece
31	IALE	ICT company providing data mining solutions	Spain
32	Neptuno Films	SME	Spain
33	Aquarius Cosmetics	SME	Spain
34	Ricard Navàs	Export Manager freelance	Spain
35	Pere Duran	Export Manager freelance	Spain
36	Marta Sánchez-Pol	Export Manager freelance	Spain
37	Serstem (Transformados Termoplasticos SL)	SME	Spain
38	Alba Lara	Expert in Internationalisation	Spain

Table 7: Overview of the MULTISENSOR User Group Members

Further, new collaborations between MULTISENSOR and IT companies have started to emerge, so as to make the project results exploitable in a broader context. Discussions between IT companies and initiatives (namely for example Media Cloud, MIT Media Lab and A Data Pro) are going on with individual partners regarding possible collaboration and exploitation of the project results.

Actual exploitation of project outputs is already taking place within pressrelations. The company has integrated a number of MULTISENSOR's capabilities and exploitable modules into its media monitoring workflow, namely the summarisation and translation module and the influencers' detection and management module and has an ongoing collaboration with Linguatec and UPF.

PIMEC used the Decision Support System for an SME that was interested in exporting cosmetics both for professional use (e.g. physiotherapist) and aesthetic use. It is a familiar SME that never

exported before and required of PIMEC services to elaborate an initial International Plan to define the countries to export to then start the action.
Moreover, a consortium-wide exploitation strategy is formulated around the most valuable asset of the project (exploitation-wise) namely the Decision Support module of the final integrated system. The strategy is complemented with a proposed exploitation roadmap from project ending onwards as well as with an estimation of PMs and costs necessary to bring the prototype to TRL9. The next table provides an assessment of the exploitation tasks based on the indicators defined in D1.2:

Objective	Results
Developing plans for exploitation and use of knowledge, based on market watch by performing an external environment analysis of STEP factors and competitive analysis.	Achieved to the highest expectations: • experts (AB) attest the full adequacy of the market and competitive analysis.
Designing a business model including the project results and assets produced that can be exploited, products and services that can be developed, as well as the market segments to be addressed.	Achieved to the highest expectations since: • Evaluating experts (AB) attest the full adequacy of business model. Expected numbers for module exploitation: • More than 3 MULTISENSOR modules integrated by user partners. Specifically 3 modules where integrated within PR (translation, summarisation, influencer detection), while MULTISENSOR platform is already used by PIMEC. • More than 6 MULTISENSOR modules compatible with existing well known platforms, easily adaptable through APIs. All MULTISENSOR modules are/can be provided as APIs since we followed a service oriented architecture.

Table 8: Exploitation assessment

4.1.6 Contacts and Information

For further information visit the project web site http://www.multisensorproject.eu/ or send e-mail to the Project Coordinator Dr. Ioannis Kompatsiaris, ikom@iti.gr or to the deputy Project Coordinator & Scientific Manager Dr. Stefanos Vrochidis, stefanos@iti.gr. You are also welcome to join us in social media:
Facebook: https://www.facebook.com/pages/Multisensor/1481238748826033
Twitter: https://twitter.com/multisensor
LinkedIn: http://de.linkedin.com/in/multisensor

4.2 Use and dissemination of foreground

Section A (public)

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
#	**Title**	**Main author**	**Title of the periodical or the series**	**Number, date or frequency**	**Publisher**	**Place of publication**	**Year of publication**	**Relevant pages**	**Permanent identifiers[10] (if available)**	**Is/Will open access[11] provided to this publication?**
1	On the Feasibility of Predicting Popular News at Cold Start	I. Arapakis, B. B. Cambazoglu, and M. Lalmas	JASIST 2016 International Conference on Social Informatics		Springer		2016			
2	Gaze Movement-driven Random Forests for Query Clustering in Automatic Video Annotation	S. Vrochidis, I. Patras and I. Kompatsiaris	Multimedia Tools and Applications	Volume 75, Number 24	Springer		22 January 2016	pp 1–29	DOI: 10.1007/s11042-015-3221-1	http://link.springer.com/article/10.1007/s11042-015-3221-1
3	Data-Driven Deep-Syntactic Dependency Parsing	M.Ballesteros, B.Bohnet, S.Mille, and L.Wanner	Natural Language Engineering	22 (6)	Cambridge University Press 2015		August 2015	pp. 939–974	DOI: 10.1017/S1351324915000285 •	https://www.cambridge.org/core/services/aop-cambridge-core/content/view/S1351324915000285

[10] A permanent identifier should be a persistent link to the published version full text if open access or abstract if article is pay per view) or to the final manuscript accepted for publication (link to article in repository).

[11] Open Access is defined as free of charge access for anyone via Internet. Please answer "yes" if the open access to the publication is already established and also if the embargo period for open access is not yet over but you intend to establish open access afterwards.

4	Focussed Crawling of Environmental Web Resources Based on the Combination of Multimedia Evidence	Theodora Tsikrika, Anastasia Moumtzidou, Stefanos Vrochidis, Ioannis Kompatsiaris	Multimedia Tools and Applications	Volume 75, Issue 3	Springer		31 May 2015	pp 1563–1587	DOI: 10.1007/s11042-015-2624-3	http://link.springer.com/article/10.1007/s11042-015-2624-3
5	Community Detection in Complex Networks Based on DBSCAN* and a Martingale Process	I. Gialampoukidis, T. Tsikrika, S. Vrochidis, I. Kompatsiaris	11th International Workshop on Semantic and Social Media Adaptation and Personalisation (SMAP 2016)		IEEE		October 20-21, 2016		DOI: 10.1109/SMAP.2016.7753375	https://www.researchgate.net/publication/309761211_Community_Detection_in_Complex_Networks_Based_on_DBSCAN_and_a_Martingale_Process
6	Query-based Topic Detection Using Concepts and Named Entities	I. Gialampoukidis, D. Liparas, S. Vrochidis, I. Kompatsiaris	1st International Workshop on Multimodal Media Data Analytics (MMDA 2016)		ECAI 2016	The Hague, Netherlands	August 30, 2016			https://www.researchgate.net/publication/308049816_Query-based_Topic_Detection_Using_Concepts_and_Named_Entities
7	Incremental estimation of visual vocabulary size for image retrieval	I. Gialampoukidis, S. Vrochidis, I. Kompatsiaris	Proc. INNS Big Data 2016	Volume 529 of the series Advances in Intelligent Systems and Computing	Springer	Thessaloniki, Greece	Oct. 23-25, 2016	pp 29-38	**DOI** 10.1007/978-3-319-47898-2_4	https://www.researchgate.net/publication/308942099_Incremental_Estimation_of_Visual_Vocabulary_Size_for_Image_Retrieval
8	Key player identification in terrorism-related social media	I. Gialampoukidis, G. Kalpakis, T. Tsikrika, S. Vrochidis, I.	European Intelligence and Security Informatics			Uppsala, Sweden	August 17-19, 2016		DOI: 10.1109/EISIC.2016.38	https://www.researchgate.net/publication/308049766_Key_player_identification_in_terroris

	networks using centrality measures	Kompatsiaris	Conference (EISIC 2016)							m-related_social_media_networks_using_centrality_measures
9	Linguistic Benchmarks of Online News Article Quality	Ioannis Arapakis, Filipa Peleja, B. Barla Cambazoglu, Joao Magalhaes	In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics			Berlin, Germany	August 7-12, 2016	pages 1893–1902		http://www.aclweb.org/anthology/P16-1178
10	Semantic integration of web data for international investment decision support	Boyan Simeonov, Vladimir Alexiev, Dimitris Liparas, Marti Puigbo, Stefanos Vrochidis, Emmanuel Jamin and Ioannis Kompatsiaris	3rd international conference on Internet Science	Volume 9934 of the series Lecture Notes in Computer Science	Springer	Florence, Italy	25 August 2016	pp 205-217	**DOI** 10.1007/978-3-319-45982-0_18	http://vladimiralexiev.github.io/pubs/INSCI2016.pdf
11	A hybrid graph-based and non-linear late fusion approach for multimedia retrieval	I. Gialampoukidis, A. Moumtzidou, D. Liparas, S. Vrochidis, I. Kompatsiaris	14th International Workshop on Content-based Multimedia Indexing (CBMI)		IEEE	Bucharest, Romania	June 15-17, 2016		**DOI:** 10.1109/CBMI.2016.7500252	https://zenodo.org/record/160505#.WEV__KJ95PM
12	A Multimedia Interactive Search Engine based on Graph-based and Non-linear Multimodal Fusion	A. Moumtzidou, I. Gialampoukidis, T. Mironidis, D. Liparas, S. Vrochidis, I. Kompatsiaris	14th International Workshop on Content-based Multimedia Indexing (CBMI)		IEEE	Bucharest, Romania	June 15-17, 2016		**DOI:** 10.1109/CBMI.2016.7500276	https://www.researchgate.net/publication/303288183_A_Multimedia_Interactive_Search_Engine_based_on_Graph-based_and_Non-linear_Multimodal_Fusion
13	I. Gialampoukidis, S. Vrochidis, I. Kompatsiaris	A hybrid framework for news clustering	12th International Conference on	Volume 9729 of the series Lectu	Springer	New York	16-21 July 2016	pp 170-184	**DOI** 10.1007/978-3-319-41920-	https://www.researchgate.net/publication/304480189_A_Hybrid_Fra

		based on the DBSCAN-Martingale and LDA	Machine Learning and Data Mining	re Notes in Computer Science					6_13	mework_for_News_Clustering_Based_on_the_DBSCAN-Martingale_and_LDA
14	Retrieval of Multimedia objects by Fusing Multiple Modalities	I. Gialampoukidis, A. Moumtzidou, T. Tsikrika, S. Vrochidis and I. Kompatsiaris	International Conference on Multimedia Retrieval (ICMR)		ACM	New York	June 6-9, 2016	Pages 359-362	doi>10.1145/2911996.2912068	https://www.researchgate.net/publication/302593430_Retrieval_of_Multimedia_Objects_by_Fusing_Multiple_Modalities
15	Learning Text Patterns to Detect Opinion Targets	F. Peleja and J. Magalhães	Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KDIR'15)		IEEE	Lisbon, Portugal	November, 2015		**Electronic ISBN:** 978-9-8975-8164-9	http://www.academia.edu/19422216/Learning_Text_Patterns_to_Detect_Opinion_Targets
16	Explanatory opinions: to whom or what is all the fuss about?	F. Peleja and I. Arapakis	Sixth BCS-IRSG Symposium on Future Directions in Information Access (FDIA'15)		ACM	Thessaloniki, Greece	August, 2015		DOI: http://dx.doi.org/10.14236/ewic/FDIA2015.19	https://www.researchgate.net/publication/283205682_Explanatory_opinions_to_whom_or_what_is_all_the_fuss_about
17	VERGE: A Multimodal Interactive Search Engine for Video Browsing and Retrieval	A. Moumtzidou, T. Mironidis, E. Apostolidis, F. Markatopoulou, A. Ioannidou, I. Gialampoukidis, K. Avgerinakis, S. Vrochidis, V. Mezaris, I.	Proc. Video Browser Showdown (VBS'16) at the 22nd Int. Conf. on MultiMedia Modeling (MMM'16)	Volume 9517 of the series Lecture Notes in Computer Science	Springer	Miami, USA	4 January 2016.	pp 394-399	**DOI** 10.1007/978-3-319-27674-8_39	http://link.springer.com/chapter/10.1007/978-3-319-27674-8_39

		Kompatsiaris, I. Patras								
18	Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs	M.Ballesteros, C. Dyer, N. Smith	In proceedings of EMNLP (EMNLP 2015)		Association for Computational Linguistics.	Lisbon, Portugal	September 2015	pages 349–359	DOI: 10.18653/v1/D15-1041	https://www.aclweb.org/anthology/D/D15/D15-1041.pdf
19	Transition-Based Spinal Parsing	M.Ballesteros, X. Carreras	In proceedings of CoNLL (CoNLL 2015)		Association for Computational Linguistics	Beijing, China	July 30-31, 2015	pages 289–299		http://www.aclweb.org/anthology/K15-1029
20	Transition-Based Dependency Parsing with Stack Long Short-Term Memory	C.Dyer, M.Ballesteros, W.Ling, A.Matthews, N. Smith	In proceedings of ACL (ACL-IJCNLP 2015)			Beijing, China	August 2015		arXiv:1505.08075 [cs.CL]	http://www.cs.cmu.edu/~lingwang/papers/acl2015.pdf
21	Fast Visual Vocabulary Construction for Image Retrieval using Skewed-Split k-d trees	I. Gialampoukidis, S. Vrochidis and I. Kompatsiaris	Proc. 22nd Int. Conf. on MultiMedia Modeling (MMM16)	Volume 9516 of the series Lecture Notes in Computer Science	Springer	Miami, USA	Jan. 2016	pp 466-477	**DOI** 10.1007/978-3-319-27671-7_39	https://www.researchgate.net/publication/282654009_Fast_Visual_Vocabulary_Construction_for_Image_Retrieval_using_Skewed-Split_k-d_trees
22	Exploiting visual similarities for ontology alignment	C. Doulaverakis, S. Vrochidis, I. Kompatsiaris	7th International Conference on Knowledge Engineering and Ontology Development (KEOD 2015)		Scitepress	Lisbon, Portugal	12-14 November, 2015	pages 29-37	DOI: 10.5220/0005588200290037	http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005588200290037
23	Classification using various ML Methods and Combinations of Key-Phrases and Visual Features	Y. Hacohen-Kerner, A. Sabag, D. Liparas, A. Moumtzidou, S. Vrochidis and I. Kompatsiaris	1st KEYSTONE Conference (IKC2015)		Springer	Coimbra, Portugal	September 8-9, 2015			

24	Concept Detection on Multimedia Web Resources about Home Made Explosives	George Kalpakis, Theodora Tsikrika, Foteini Markatopoulou, Nikiforos Pittaras, Stefanos Vrochidis, Vasileios Mezaris, Ioannis Patras, and Ioannis Kompatsiaris	In Proceedings of the International Workshop on Multimedia Forensics and Security (MFSec 2015)		IEEE	Toulouse, France	24-27 Aug. 2015		**DOI:** 10.1109/ARES.2015.85	http://ieeexplore.ieee.org/document/7299974/
25	MULTISENSOR: Development of Multimedia Content Integration Technologies for Journalism, Media Monitoring and International Exporting Decision Support	S. Vrochidis, I. Kompatsiaris, G. Casamayor, I. Arapakis, R. Busch, V. Alexiev, E. Jamin, M. Jugov, N. Heise, T. Forrellat, D. Liparas, L. Wanner, I. Miliaraki, V. Aleksic, K. Simov, A. M. Soro, M. Eckhoff, T. Wagner, M. Puigbó	IEEE International Conference on Multimedia and Expo (ICME 2015)		IEEE	Turin, Italy	29 June-3 July 2015		**DOI:** 10.1109/ICMEW.2015.7169818	https://www.researchgate.net/publication/310160858_MULTISENSOR_DEVELOPMENT_OF_MULTIMEDIA_CONTENT_INTEGRATION_TECHNOLOGIES_FOR_JOURNALISM_MEDIA_MONITORING_AND_INTERNATIONAL_EXPORTING_DECISION_SUPPORT
26	Visualizing deep-syntactic structures	J.Soler-Company, M.Ballesteros, B. Bohnet, S. Mille, and L. Wanner	In Proceedings of the Demonstrations of the North American Chapter of Computational Linguistics (NAACL HLT 2015)		Association for Computational Linguistics	Denver US	June 2015	pages 56–60		http://www.aclweb.org/anthology/N15-3012

27	Data-driven sentence generation with non-isomorphic trees	M.Ballesteros, B. Bohnet, S. Mille, and L. Wanner	In Proceedings of the North American Chapter of Computational Linguistics (NAACL HLT 2015)		Association for Computational Linguistics	Denver US	June 2015	pages 387–397		http://www.aclweb.org/anthology/N15-1042
28	Discovery of Environmental Web Resources Based on the Combination of Multimedia Evidence	T. Tsikrika, A. Latas, A. Moumtzidou, E. Chatzilari, S. Vrochidis, and I. Kompatsiaris	In Proceedings of the Environmental Multimedia Retrieval Workshop (EMR 2015)		ACM	Shanghai, China	June 23 - 26, 2015	Pages 27-32	doi>10.1145/2764873.2764876	http://dl.acm.org/citation.cfm?id=2764876
29	VERGE: A Multimodal Interactive Video Search Engine	A. Moumtzidou, K. Avgerinakis, E. Apostolidis, F. Markatopoulou, K. Apostolidis, T. Mironidis, S. Vrochidis, V. Mezaris, Y. Kompatsiaris, I. Patras	Proc. 21st Int. Conf. on MultiMedia Modeling (MMM15)		Springer	Sydney, Australia	January 2015	249-254	DOI: 10.1007/978-3-319-14442-9_23	https://www.researchgate.net/publication/271049975_VERGE_A_Multimodal_Interactive_Video_Search_Engine
30	Modeling adoptions and the stages of the diffusion of innovations	Y. Mehmood, N. Barbieri, F. Bonchi	Proceedings of the International Conference on Data Mining	Volume 48, Issue 1,	Springer		02 November 2015	pp 1–27	DOI: 10.1007/s10115-015-0889-5	http://www.francescobonchi.com/MASD.pdf
31	A Unified Model for Socially Interconnected Multimedia-Enriched Objects	T. Tsikrika, K. Andreadou, A. Moumtzidou, E. Schinas, S. Papadopoulos, S. Vrochidis, Y. Kompatsiaris	21st MultiMedia Modelling Conference (MMM2015)	Volume 8935 of the series Lecture Notes in Computer Science	Springer	Sydney, Australia	5-7 January	pp 372-384	**DOI** 10.1007/978-3-319-14445-0_32	

32	The Influence of Indirect Ties on Social Network Dynamics	X. Zuo, J. Blackburn, N. Kourtellis, J. Skvoretz, A. Iamnitchi	Proceedings of the 6th International Conference on Social Informatics (SocInfo 2014)			Barcelona, Spain	10-13th of November 2014	Social Informatics, pp.50-65	DOI: 10.1007/978-3-319-13734-6_4	https://www.researchgate.net/publication/278707014_The_Influence_of_Indirect_Ties_on_Social_Network_Dynamics
33	Who to follow and why: link prediction with explanations	N. Barbieri, F. Bonchi, G. Manco	Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining		ACM		August 2014	Pages 1266-1275	doi>10.1145/2623330.2623733	http://www.francescobonchi.com/frp1266-barbieri.pdf
34	News articles classification using Random Forests and weighted multimodal features	D. Liparas, Y. Hacohen-Kerner, A. Moumtzidou, S. Vrochidis and I. Kompatsiaris	3rd Open Interdisciplinary MUMIA Conference and 7th Information Retrieval Facility Conference (IRFC2014)	Volume 8849 of the book series Lecture Notes in Computer Science (LNCS)	Springer	Copenhagen, Denmark	November 10-12, 2014	pp 63-75	DOI: 10.1007/978-3-319-12979-2_6	https://www.researchgate.net/publication/266141433_News_Articles_Classification_Using_Random_Forests_and_Weighted_Multimodal_Features
35	Key-phrase Extraction using Textual and Visual Features	Y. HaCohen-Kerner, S. Vrochidis, D. Liparas, A. Moumtzidou and I. Kompatsiaris	3rd Workshop on Vision and Language (VL)			Dublin, Ireland	August 23-29, 2014	pages 121–123	DOI: 10.3115/v1/W14-5421	http://www.aclweb.org/anthology/W14-5421
36	Concept-oriented labelling of patent images based on Random Forests and proximity-driven generation of synthetic data	D. Liparas, A. Moumtzidou, S. Vrochidis, I. Kompatsiaris	COLING'14 Workshop on Vision and Language (VL'14)			Dublin, Ireland	August 23, 2014	pages 25–32	DOI: 10.3115/v1/W14-5404	http://aclweb.org/anthology/W14-5404

37	"Deep-syntactic parsing"	M.Ballesteros, B. Bohnet, S. Mille, and L. Wanner	Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland			Dublin Ireland	23/08/2014 – 29/08/2014	pp 1402–1413		http://www.aclweb.org/anthology/C14-1133
38	Classifiers for Data-driven Deep Sentence Generation	M. Ballesteros, S. Mille and L. Wanner	In Proceedings of the 8th International Natural Language Generation Conference (INLG)		Association for Computational Linguistics	Philadelphia, USA	June 2014	pages 108–112		http://www.aclweb.org/anthology/W14-4416
39	Multi-evidence User Group Discovery in Professional Image Search	T. Tsikrika, C. Diou	In Proceedings of the 36th European Conference on Information Retrieval (ECIR 2014)	Volume 8416 of the book series Lecture Notes in Computer Science (LNCS)	Springer	Amsterdam, The Netherlands	13-16 April, 2014	pp 693-699	DOI: 10.1007/978-3-319-06028-6_78	http://theodoratsikrika.info/papers/ECIR2014tsikrika.pdf
40	Focussed Crawling of Environmental Web Resources: A Pilot Study on the Combination of Multimedia Evidence	Theodora Tsikrika, Anastasia Moumtzidou, Stefanos Vrochidis and Ioannis Kompatsiaris	In Proceedings of the Environmental Multimedia Retrieval Workshop (EMR 2014)		t http://ceur-ws.org	Glasgow, UK	April 1st, 2014	pp. 61-68		http://ceur-ws.org/Vol-1222/paper9.pdf
41	VERGE: An Interactive Search Engine for Browsing Video Collections	A. Moumtzidou, K. Avgerinakis, E. Apostolidis, V. Aleksic, F. Markatopoulou, C. Papagiannopoulou	Video Browser Showdown (VBS) 2014	LNCS vol. 8326	Springer	Dublin, Ireland	January 2014	pp. 411-414		https://www.iti.gr/~bmezaris/publications/mmm_vbs14_preprint.pdf

		, S. Vrochidis, V. Mezaris, R. Busch, I. Kompatsiaris								
42	Influence-based Network-oblivious Community Detection	N. Barbieri, F. Bonchi, G. Manco	Proceedings of the IEEE International Conference on Data Mining, December 2013		IEEE	Dallas, Texas	03 February 2014		**DOI:** 10.1109/ICDM.2013.164	https://www.researchgate.net/publication/261364625_Influence-Based_Network-Oblivious_Community_Detection
43	Mining Summaries of Propagations	L. Macchia, F. Bonchi, F. Gullo, L. Chiarandini	Proceedings of the IEEE International Conference on Data Mining		IEEE	Dallas, Texas	03 February 2014		**DOI:** 10.1109/ICDM.2013.163	http://francescobonchi.com/icdm13-agony.pdf
44	Influence Maximization with Viral Product Design	N. Barbieri, F. Bonchi	Proceedings of the SIAM International Conference on Data Mining			Athens, Greece	April 2014		DOI: http://dx.doi.org/10.1137/1.9781611973440.7	http://www.francescobonchi.com/how-to-design-sdm14.pdf
45	Privacy Preserving Estimation of Social Influence	T. Tassa, F. Bonchi	Proceedings of the SIAM International Conference on Data Mining			Athens, Greece	April 2014		10.5441/002/edbt.2014.50	http://www.francescobonchi.com/ppvm_EDBT14.pdf
46	Online Topic-aware Influence Maximization Queries	C. Aslay, N. Barbieri, F.Bonchi. R.Baeza-Yates	Proceeding of the International Conference on Extending Database Technology				March 2014		ISBN 978-3-89318065-3	http://www.francescobonchi.com/inflex.pdf
47	ITI-CERTH participation to TRECVID 2013	F. Markatopoulou, A. Moumtzidou, C. Tzelepis, K. Avgerinakis, N.	Proceedings of TRECVID 2013 Workshop			Gaithersburg, MD, USA.	November 2013			http://www-nlpir.nist.gov/projects/tvpubs/tv13.papers/iti-certh.pdf

		Gkalelis, S. Vrochidis, V. Mezaris, I. Kompatsiaris								
48	On the Feasibility of Predicting Popular News at Cold Start	I. Arapakis, B. B. Cambazoglu, and M. Lalmas	JASIST 2016 International Conference on Social Informatics SocInfo 2014: Social Informatics	Volume 8851 of the book series Lecture Notes in Computer Science (LNCS)	Springer		2014	pp 290-299	DOI: 10.1007/978-3-319-13734-6_21	http://link.springer.com/chapter/10.1007%2F978-3-319-13734-6_21

TEMPLATE A2: LIST OF DISSEMINATION ACTIVITIES

NO.	Type of activities[12]	Main leader	Title	Date	Place	Type of audience[13]	Size of audience	Countries addressed
1	Presentation	DW, CERTH, all partners	Project Presentation	Created at the beginning of the project (2013), was updated twice during the project's lifetime	EU	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias	N/A	EU
2	Poster	CERTH, all partners	MULTISENSOR Poster	Created at the beginning of the project (2013),	EU	Scientific Community (higher	N/A	EU

[12] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[13] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias ('multiple choices' is possible.

				was updated twice during the project's lifetime		education, Research), Industry, Civil Society, Policy makers, Medias		
3	Flyer	CERTH, all partners	MULTISENSOR flyers	Created at the beginning of the project (2013), was updated four times during the project's lifetime	EU	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias	N/A	EU
4	Website	All partners	The official MULTISENSOR website	Created at the beginning of the project (2013), was updated constantly during the project's lifetime	EU	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias	11,000 unique visitors	EU
5	Social Network Activities	DW, contributions from all partners	Twitter, Facebook, LinkedIn	Accounts Created at the beginning of the project (2013), and updated constantly during the project's lifetime	EU	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias	Twitter: 225 Facebook: 31 LinkedIn: 79	EU
6	Video	DW, CERTH, PIMEC, PR	Project Video Showcase	October 2016	EU	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias	Available in YouTube and the project website (more than 36 views so far)	EU
7	Press Release	CERTH, PIMEC and pressrelations	Project Press Releases (5)	28/11/2013 25/02/2014 08/07/2015	Germa ny, Spain,	Scientific Community (higher	N/A	Germany, Spain, Greece

				19/09/2016 20/09/2016	Greece	education, Research), Industry, Civil Society, Policy makers, Medias		
8	Newsletter	PIMEC, pressrelations input from all partners	Newsletters (19)	05/12/2013 30/1/2014 13/03/2014 10/04/2014 30/10/2014 30/10/2014 10/04/2015 15/07/2015 29/07/2015 14/10/2015 29/10/2015 26/11/2015 01/12/2015 12/01/2016 28/01/2016 17/03/2016 20/07/2016 06/09/2016 21/10/2016	Germany, Spain	Industry, Civil Society, Medias	>27550 readers	Germany, Spain
9	Datasets	All partners	WikiRef220 WikiRef150 ArticlesNews SitesData_1043 ArticlesNews SitesData_2382 MULTISENSOR_ NewsArticlesData_12073 YahooNewsQualityDataset	N/A	EU	Scientific Community (higher education, Research)	More than 30 downloads	EU
10	Open Source Code	Technical Partners	VERGE Socially interconnected/ interlinked and multimedia-enriched	N/A	EU	Scientific Community (higher education, Research),	>90 downloads (website) (the	EU

			objects Mate Tools surface statistical dependency parser Character-based Stack-LSTM surface parser Deep statistical dependency parser DSynt Converter (ENG) Frame Semantics parser (ENG) Deep statistical text generator Twitter Crawler Topic detection Category-based classification Multimedia retrieval Multimedia concept and event detection Community detection Ontology alignment User and Context-centric Content Analysis			Industry	actual numbers might be a bit larger, as the code can also be directly accessed through the publication pages, e.g. github, directly	
11	Workshop	All partners	Joint workshop in Barcelona with EUMSSI	11/03/2015	Spain	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias	29	Spain, EU
12	Workshop	All partners	Joint workshop in Bonn with EUMSSI	23-27/11/2015	Germany	Scientific Community (higher education, Research), Industry, Civil	34	Germany, EU

						Society, Policy makers, Medias		
13	Workshop	All partners	Workshop with EUMSSI on Multimodal media data analytics (MMDA) in conjunction with the 22nd European Conference on Artificial Intelligence (ECAI) 2016, in The Hague	30/08/2016	The Netherlands	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Media	16	EU
14	Workshop	All partners	Joint evaluation workshop with EUMSSI in Barcelona	22/09/2016	Spain	Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Media	32	Spain, EU
15	Presentation	CERTH	MULTISENSOR presentation at LT Innovative Summit 2013, Brussels, Belgium	26/06/2013 – 27/06/2013	Belgium	Scientific Community (higher education, Research), Industry		Belgium, EU
16	Presentation	CERTH	MULTISENSOR presentation at 6th IRF conference for Science and Industry, Limassol, Cyprus	07/10/2013 – 09/10/2013	Cyprus	Scientific Community (higher education, Research), Industry	40	Cyprus, EU
17	Exhibition	LT	MULTISENSOR presence at CeBIT 2014 and 2016	14/03/2014 14/03/2016 - 18/03/2016	Germany	Scientific Community (higher education, Research), Industry, Medias	Thousands	EU
18	Congress	pressrelations	MULTISENSOR	12/03/2014 - 15/03/2014	Dubai	Scientific Community	Thousands	Worldwide

			presence at 46th FIBEP Congress, Dubai	17/11/2015 - 20/11/2015		(higher education, Research), Industry, Civil Society, Policy makers, Media		
19	Presentation	CERTH	MULTISENSOR presentation at ESCW 2014, Anissaras/Hersonisso u, Crete	27/05/2014	Greece	Scientific Community (higher education, Research)	50	Greece, EU
20	Presentation	EVERIS	MULTISENSOR presentation at CEN/BII in Paris, France	03/06/2014 - 05/06/2014	France	Scientific Community (higher education, Research)	100	France, EU
21	Other (Summit)	pressrelations	MULTISENSOR presence at AMEC International Summit, Amsterdam, Netherlands and 2016 London	11/06/2014 - 12/06/2014 15/06/2016 - 16/06/2016	Netherl ands London	Scientific Community (higher education, Research), Industry, Policy makers, Media	100	EU
22	Other (training)	ONTOTEXT	MULTISENSOR presence at 7th GATE training course, Sheffield, UK	09/06/2014 - 13/06/2014	UK	Scientific Community (higher education, Research)	50	EU
23	Presentation	DW	MULTISENSOR presentation at Global Media Forum 2014, Bonn, Germany	30/06/2014 - 02/07/2014	Germa ny	Scientific Community (higher education, Research), Industry, Policy makers, Media	hundreds	EU
24	Presentation	CERTH	MULTISENSOR presentation at the 4th International Workshop on Cyber-Physical Cloud Computing (CPCC2014), Osaka,	27/08/2014 - 29/08/ 2014	Japan	Scientific Community (higher education, Research), Industry	50	Worldwide

			Japan					
25	Other (Congress)	pressrelations	MULTISENSOR presence at Kommunikationskongress 2014, Berlin, Germany	25/09/2014 - 26/09/2014 17/09/2015 - 18/09/2015	Germany	Industry, Media	Hunderds	Germany
26	Presentation	UPF	MULTISENSOR presentation at the Big Bang Data exhibition at the CCCB in Barcelona, Spain	21/10/2014	Spain	Scientific Community (higher education, Research), Industry, Policy makers, Media	100	Spain, EU
27	Other (Event)	ONTO	MULTISENSOR presence at BBC NewsHack, London, UK	15/12/2014 - 16/12/2014	UK	Industry, Media	50	UK, EU
28	Presentation	EVERIS	MULTISENSOR presentation at Horizon 2020 ICT-16 Big Data networking day, Brussels	16/01/2015	Belgium	Scientific Community (higher education, Research), Industry, Policy makers	50	Belgium, EU
29	Presentation	CERTH	MULTISENSOR presentation at Riga Summit, META-FORUM, Riga	27/04/2015 - 29/04/2015	Estonia	Scientific Community (higher education, Research)	100	EU
30	Presentation	EVERIS	MULTISENSOR presentation at Virolai school, Big data project, Barcelona, Spain	05/05/2015	Spain	Scientific Community (higher education, Research), Industry, Civil Society	50	Spain
31	Presentation	CERTH	MULTISENSOR presentation at	08/05/2015	Greece	Scientific Community (higher	100	Greece

			Technology Forum 2015, Thessaloniki, Greece			education, Research), Industry, Civil Society,Policy makers		
32	Other (Summit)	pressrelations	MULTISENSOR presence at AMEC International Summit, Stockholm, Sweden	03/06/2015 - 04/06/2015	Sweden	Scientific Community (higher education, Research), Industry, Media	50	Sweden, EU
33	Demonstration	CERTH	MULTISENSOR Demonstration at ICME 2015, Torino, Italy	29/06/2015 - 03/07/2015	Italy	Scientific Community	80	EU
34	Other (European Summer School)	UPF, EURECAT	MULTISENSOR presence at 10th European Summer School on Information Retrieval (ESSIR2015), Thessaloniki	31/08/2015 - 04/09/2015	Greece	Scientific Community	70	EU
35	Exhibition	CERTH, DW	MULTISENSOR presence at ICT event, Lisbon, Portugal	22/10/2015	Portugal	Scientific Community (higher education, Research), Industry,Policy makers	Hundreds	EU
36	Presentation	ONTOTEXT	MULTISENSOR presentation at the 7th International Conference on Knowledge Engineering and KEOD 2015, Lisbon, Portugal	12/11/2015 - 14/11/2015	Portugal	Scientific Community	50	EU
37	Presentation	UPF	MULTISENSOR	29/05/2016 - 02/06/2016	Greece	Scientific Community	60	EU

			demonstration at ESWC Conference, Heraklion, Greece					
38	Presentation	PIMEC	MULTISENSOR demonstration at BIZ Barcelona, Barcelona, Spain	01/06/2016 - 02/06/2016	Spain	Industry, Civil Society, Media	40	Spain
39	Other	DW	Deutsche Welle Global Media Forum, Bonn, Germany	13/06/2016 - 15/06/2016	Germa ny	Industry, Media	100	Germany
40	Exhibition	CERTH (all partners)	MULTISENSOR presentation at European Data Forum (EDF) 2016, Eindhoven, Netherlands	29/06/2016 - 30/06/2016	The Netherl ands	Scientific Community, Industry, Media	Hundreds	EU
41	Demonstration	ONTOTEXT	MULTISENSOR demonstration at Linked Open Data Workshop at SEMANTiCS 2016, in Leipzig, Germany	12/09/2016 - 15/09/2016	Germa ny	Scientfic Community, Indsutry	30	Germany, EU
42	User Day & Open Door Day	All partners	User Day combined with the consortium meeting in Bonn	23/11/2015 - 27/11/2015	Germa ny	Scientific Community (higher education, Research), Industry, Civil Society,Policy makers, Media	60-65 (incl. 19 MULTIS ENSOR)	Germany
43	User Day & Open Door Day	All partners	User Day combined with the consortium meeting in Barcelona	22/09/2016	Spain	Scientific Community (higher education, Research), Industry, Civil Society,Policy makers, Media	50-55 (incl. 17 MULTIS ENSOR)	Spain

44	User Day	CERTH	Evaluation day in collaboration with Media Informatics Lab of School of Journalism & MC in Thessaloniki	15/09/2016	Greece	Scientific Community, Media	20 (+4 CERTH)	Greece
45	Standardisation	CERTH, ONTOTEXT, UPF and EURECAT	Standardisation Bodies (6): W3C, MPEG-7 Linguistic Linked Data Linguistic Linked Data W3C, JSON-LD W3C, RDF/OWL/SPARQL, CUBE ontology W3C Emotion Incubator Group, Emotion Markup Language (EmotionML)	Throughout the project duration	EU	Scientific Community (higher education, Research)	N/A	
46	Open Door Day	All partners	Open Door Day combined with the consortium meeting in Barcelona	22/09/2016	Spain	Scientific Community (higher education, Research), Industry, Civil Society,Policy makers, Media	50-55 (incl. 17 MULTISENSOR)	Spain
47	Presentation	CERTH	MULTISENSOR presentation to the ENGINE project, Kick off meeting of ENGINE project, Wroclaw, Poland	01/10/2013	Poland	Scientific Community (higher education, Research), Industry	20	EU
48	Presentation	CERTH	MULTISENSOR presentation to the MUMIA consortium, Limassol, Cyprus	10/10/2013	Cyprus	Scientific Community (higher education, Research),	35	EU

						Industry		
49	Presentation	CERTH	MULTISENSOR presentation to the WikiRate project, Plenary meeting, Thessaloniki, Greece	11/06/2014	Greece	Scientific Community (higher education, Research), Industry	20	EU
50	Workshop	PR	Workshop with media monitoring company Newsaccess, demonstration of MULTISENSOR, Düsseldorf, Germany	26/09/2014	Germany	Scientific Community (higher education, Research), Industry	10	Germany
51	Workshop	PR	Workshop with media monitoring company push| OBSERVER, demonstration of MULTISENSOR, Düsseldorf, Germany	12/12/2014	Germany	Scientific Community (higher education, Research), Industry	5	Germany
52	Presentation		MULTISENSOR presentation to the iV&L Net members, Heidelberg, Germany	11/03/2015	Germany	Scientific Community (higher education, Research), Industry	30	Germany, EU
53	Demonstration	CERTH	Demonstration of MULTISENSOR prototype to ICME 2015	30/06/2015	Italy	Scientific Community (higher education, Research), Industry	20	EU
54	Workshop	PR	Workshop with media monitoring company Auxipress, demonstration of MULTISENSOR, Düsseldorf, Germany	06/07/2015	Germany	Scientific Community (higher education, Research), Industry	20	Germany
55	Demonstration	CERTH	Demonstrating MULTISENSOR	05/01/2016	USA	Scientific Community	40	USA

			video retrieval technologies in VBS 2016, MMM 2016, Miami, USA			(higher education, Research), Industry		
56	Demonstration	PR	Workshop with media monitoring company up2news, demonstration of MULTISENSOR, Düsseldorf, Germany	14/01/2016	Germany	Scientific Community (higher education, Research), Industry	10	Germany
57	Demonstration	PIMEC	Prototype demonstration to Pere Cots, Impuls Estratègic Consulting CEO	19/01/2016	Spain	Scientific Community (higher education, Research), Industry	20	Spain
58	Workshop	PR	Workshop with media monitoring company Media Image, demonstration of MULTISENSOR, Düsseldorf, Germany	17/02/2016	Germany	Scientific Community (higher education, Research), Industry	10	Germany
59	Demonstration	DW	Demonstration of MULTISENSOR Prototype with user Feedback, at DW GMF, Bonn, Germany	13/06/2016 - 15/06/2016	Germany	Scientific Community (higher education, Research), Industry	10	Germany
60	Presentation	PR	MULTISENSOR presentation and demonstration to national research project AGENT, Düsseldorf, Germany	12/07/2016	Germany	Scientific Community (higher education, Research), Industry	15	Germany
61	Presentation	PR	MULTISENSOR presentation and demonstration to national research project ATOM, Düsseldorf, Germany	12/07/2016	Germany	Scientific Community (higher education, Research), Industry	15	Germany
62	Presentation	CERTH	MULTISENSOR	13/07/2016	Greece	Scientific	10	USA

			presentation to Media Cloud			Community (higher education, Research), Industry		
63	Demonstration	PIMEC and All Partners	Demonstrations to researchers, industry and users	Throughout the Project Duration	EU	Scientific Community (higher education, Research), Industry	15	EU
64	Demonstration	CERTH	Demonstration of MULTISENSOR prototype to AUTH media labs.	15/09/2016	Greece	Scientific Community (higher education, Research), Industry, Media	24	Greece

Part B2

Type of Exploitable Foreground[14]	Description of exploitable foreground	Confidential Click on YES /NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[15]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Commercial Exploitation of R&D results	Multimedia Concept and Event Detection	No		Software, Algorithm	Information Service Activities	Use as tech base for startups/spinoffs	Open Source (Apache License, Version 2.0)	CERTH
Commercial Exploitation of R&D results	Community Detection	No		Software, Algorithm	Information Service Activities	Use as tech base for startups/spinoffs	Open Source (Apache License, Version 2.0)	CERTH
Commercial Exploitation of R&D results	Multimodal Indexing and Retrieval	No		Software, Algorithm, Model	Information Service Activities	Use as tech base for startups/spinoffs	Open Source(Apache License, Version 2.0)	CERTH
Commercial Exploitation of R&D results	Topic-based Modelling	No		Software, Algorithm	Information Service Activities	Use as tech base for startups/spinoffs	Open Source (Apache License, Version 2.0)	CERTH
Commercial Exploitation of R&D results	Mapping Discovery and Validation	No		Software, Algorithm	Information Service Activities	Use as tech base for startups/spinoffs	Open Source (Apache License, Version 2.0)	CERTH
Commercial Exploitation of R&D results	Content Alignment and Integration	No		Software, Algorithm	Information Service Activities	Use as tech base for startups/spinoffs	Open Source(Apache License, Version 2.0)	CERTH
General Advancement of Knowledge	Concept extraction	Yes		Software, Algorithm	Information Service Activities	Use in other projects	Proprietary Use	UPF
General Advancement of Knowledge	Concept relation extraction	Yes		Software, Algorithm	Information Service Activities	Use in other projects	Proprietary use	UPF
Commercial Exploitation of R&D results	Extractive summarisation	Yes		Software, Algorithm	Information Service Activities	Use as commercial product, collaboration with SMEs, and in	Comm. License	UPF

[19] A drop down list allows choosing the type of foreground: General advancement of knowledge, Commercial exploitation of R&D results, Exploitation of R&D results via standards, exploitation of results through EU policies, exploitation of results through (social) innovation.

[15] A drop down list allows choosing the type sector (NACE nomenclature) : http://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Type of Exploitable Foreground[14]	Description of exploitable foreground	Confidential Click on YES /NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[15]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
						other projects		
General Advancement of Knowledge	Syntactic parsing	No		Software, Algorithm	Information Service Activities	Use in collaborations and in other projects	Open Source(Apache License, Version 2.0)	UPF
General Advancement of Knowledge	Deep Parsing	No		Software, Algorithm	Information Service Activities	Use in collaborations and in other projects	Open Source (Apache License, Version 2.0)	UPF
General Advancement of Knowledge	Syntactic statistical generation	No		Software, Algorithm	Information Service Activities	Use in collaborations and in other projects	Open Source (Apache License, Version 2.0)	UPF
General Advancement of Knowledge	Statistical punctuation generation	No		Software, Algorithm	Information Service Activities	Use in collaborations and in other projects	Open Source (Apache License, Version 2.0)	UPF
General Advancement of Knowledge	Lexical resources	Yes		Content	Information Service Activities	Use in other projects	Proprietary use	UPF
Commercial Exploitation of R&D results	Context Extraction and Representation	No		Software, Algorithm	Information Service Activities	Use as tech base for startups/spinoffs, collaborations with SMEs, and other projects	Open Source (Apache License, Version 2.0)	Eurecat
Commercial Exploitation of R&D results	Polarity and Sentiment Extraction	No		Software, Algorithm	Information Service Activities	Use as tech base for startups/spinoffs, collaborations with SMEs, and other projects	Open Source (Apache License, Version 2.0)	Eurecat
Commercial Exploitation of R&D results	Social Media Mining Module	No		Software	Information Service Activities	Use as tech base for startups/spinoffs, collaborations with SMEs, and other projects	MIT free software licence	Eurecat
Commercial Exploitation of R&D results	Named Entity Recognition	Yes		Software	Information Service Activities	Internal tool, Use as tech base for startups/spinoffs, collaborations with SMEs, and other projects	Comm. License	Linguatec

Type of Exploitable Foreground[14]	Description of exploitable foreground	Confidential Click on YES /NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[15]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Commercial Exploitation of R&D results	Automatic Speech Recognition	Yes		Software	Information Service Activities	Internal tool, Use as tech base for startups/spinoffs, collaborations with SMEs, and other projects	Comm. License	Linguatec
Commercial Exploitation of R&D results	Machine Translation	Yes		Software	Information Service Activities	Internal tool, Use as tech base for startups/spinoffs, collaborations with SMEs, and other projects	Comm. License	Linguatec
Commercial Exploitation of R&D results	Language Identifier	Yes		Software	Information Service Activities	Internal tool, Use as tech base for startups/spinoffs, collaborations with SMEs, and other projects	Comm. License	Linguatec
Exploitation of R&D results via standards	W3C, MPEG-7 Contribute to the existing multimedia modelling standards by standardizing SIMMO. CERTH developed the Socially Interconnected and Multimedia-Enriched Object (SIMMO). This model definition has been accepted for publication in the 21st Conference on Multimedia Modelling (MMM2015).	No		Methodology	Information Service Activities	Public Disclosure		CERTH
Exploitation of R&D results via standards	Linguistic Linked Data Use of ontological models for NLP (NIF,	No		Methodology	Information Service Activities	Public Disclosure		UPF/ONTOTEXT

Type of Exploitable Foreground[14]	Description of exploitable foreground	Confidential Click on YES /NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[15]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
	OLIA, OLIA constituent ontologies, NERD, MARL, FrameNet) is used for storing extracted linguistic information in the Knowledge Base and exploiting them for natural language generation. We defined various data scenarios for storing various kinds of data: Automatic Speech Recognition transcripts, captions, concepts recognised in video frames, etc. This work is leveraged by ONTO in recent commercial work in video automation in the automotive domain, also using ontologies patterns from FP7 LinkedTV.							
Exploitation of R&D results via standards	Linguistic Linked Data An approach for embedding FrameNet relation data to NIF was developed and published as a paper: "FN goes NIF: Integrating FrameNet in the NLP Interchange Format." Alexiev, V.; and Casamayor, G. In	No		Methodology	Information Service Activities	Public Disclosure		UPF/ONTOTEXT

Type of Exploitable Foreground[14]	Description of exploitable foreground	Confidential Click on YES /NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[15]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
	Linked Data in Linguistics (LDL-2016): Managing, Building and Using Linked Language Resources, Portorož, Slovenia, May 2016							
Exploitation of R&D results via standards	W3C, JSON-LD Use of JSON-LD (a W3C recommendation) for serializing RDF data. This is the easiest way to share NLP results between the services along the processing pipeline, since it's easier to parse by web applications.	No		Methodology	Information Service Activities	Public Disclosure		UPF/ONTOTEXT
Exploitation of R&D results via standards	W3C, RDF/OWL/SPARQL, CUBE ontology Adopt Semantic Web standards (RDF, RDFS, OWL) for data representation and SPARQL for querying	No		Methodology	Information Service Activities	Public Disclosure		ONTOTEXT
Exploitation of R&D results via standards	W3C Emotion Incubator Group, Emotion Markup Language (EmotionML) Adopt EmotionML (partly) in sentiment analysis specifically for representing and storing information. Selection of the	No		Methodology	Information Service Activities	Public Disclosure		EURECAT

Type of Exploitable Foreground[14]	Description of exploitable foreground	Confidential Click on YES /NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[15]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
	<dimension> element, among the various representation schemas provided by EmotionML, as the most suitable representation method describing an emotion or a related state according to an emotion dimension vocabulary.							

4.3 Report on societal implications

Replies to the following questions will assist the Commission to obtain statistics and indicators on societal and socio-economic issues addressed by projects. The questions are arranged in a number of key themes. As well as producing certain statistics, the replies will also help identify those projects that have shown a real engagement with wider societal issues, and thereby identify interesting approaches to these issues and best practices. The replies for individual projects will not be made public.

1

A General Information *(completed automatically when **Grant Agreement number** is entered.*	
Grant Agreement Number:	610411
Title of Project:	MULTISENSOR
Name and Title of Coordinator:	Dr. Yiannis Kompatsiaris, Senior Researcher

B Ethics	
1. Did your project undergo an Ethics Review (and/or Screening)? • If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports? Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 *'Work Progress and Achievements'*	***No***
2. Please indicate whether your project involved any of the following issues (tick box) :	***YES***
RESEARCH ON HUMANS	
• Did the project involve children?	No
• Did the project involve patients?	No
• Did the project involve persons not able to give consent?	No
• Did the project involve adult healthy volunteers?	Yes
• Did the project involve Human genetic material?	No
• Did the project involve Human biological samples?	No

• Did the project involve Human data collection?	No
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	No
• Did the project involve Human Foetal Tissue / Cells?	No
• Did the project involve Human Embryonic Stem Cells (hESCs)?	No
• Did the project on human Embryonic Stem Cells involve cells in culture?	No
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	No
PRIVACY	
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	No
• Did the project involve tracking the location or observation of people?	No
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	No
• Were those animals transgenic small laboratory animals?	No
• Were those animals transgenic farm animals?	No
• Were those animals cloned farm animals?	No
• Were those animals non-human primates?	No
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	No
• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	No
DUAL USE	
• Research having direct military use	No
• Research having the potential for terrorist abuse	No

C Workforce Statistics

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).

Type of Position	Number of Women	Number of Men
Scientific Coordinator		2
Work package leaders		9
Experienced researchers (i.e. PhD holders)	4	13
PhD Students		2
Other	6	17

4. How many additional researchers (in companies and universities) were recruited specifically for this project?	**10**
Of which, indicate the number of men:	5

D Gender Aspects

5. Did you carry out specific Gender Equality Actions under the project?

⊙ Yes
○ No

6. Which of the following actions did you carry out and how effective were they?

		Not at all effective				Very effective
☒	Design and implement an equal opportunity policy	○	○	⊙	○	○
❑	Set targets to achieve a gender balance in the workforce	○	○	○	○	○
❑	Organise conferences and workshops on gender	○	○	○	○	○
☒	Actions to improve work-life balance	○	○	○	○	⊙
○	Other:					

7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?

○ Yes- please specify

⊙ No

E Synergies with Science Education

8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?

⊙ Yes- please specify

Participation in European researcher's night event in Thessaloniki
Organisation of ESSIR2015

○ No

9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?

⊙ Yes- please specify

Explanatory DVDs, website

○ No

F Interdisciplinarity

10. Which disciplines (see list below) are involved in your project?

○ Main discipline[16]: 1.1

[16] Insert number from list below (Frascati Manual).

○	Associated discipline[16]:	○	Associated discipline[16]:

G Engaging with Civil society and policy makers

11a Did your project engage with societal actors beyond the research community? *(if 'No', go to Question 14)*	⊙ ○	Yes No

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?

- ○ No
- ⊙ Yes- in determining what research should be performed
- ○ Yes - in implementing the research
- ⊙ Yes, in communicating /disseminating / using the results of the project

11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?	⊙ ○	Yes No

12. Did you engage with government / public bodies or policy makers (including international organisations)

- ○ No
- ⊙ Yes- in framing the research agenda
- ⊙ Yes - in implementing the research agenda
- ⊙ Yes, in communicating /disseminating / using the results of the project

13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?

- ⊙ Yes – as a **primary** objective (please indicate areas below- multiple answers possible)
- ○ Yes – as a **secondary** objective (please indicate areas below - multiple answer possible)
- ○ No

13b If Yes, in which fields?

Agriculture		Energy		Human rights			
Audiovisual and Media **Yes**		Enlargement		Information Society **Yes**			
Budget		Enterprise **Yes**		Institutional affairs			
Competition		**Environment**		Internal Market			
Consumers **Yes**		External Relations		**Justice, freedom and security** **Yes**			
Culture		External Trade		Public Health			
Customs		Fisheries and Maritime Affairs		Regional Policy			
Development Economic and Monetary Affairs **Yes**		Food Safety		Research and Innovation **Yes**			
		Foreign and Security Policy		Space			
Education, Training, Youth		Fraud		Taxation			
Employment and Social Affairs		**Humanitarian aid**		Transport			

13c If Yes, at which level?

- ⦿ Local / regional levels
- ○ National level
- ○ European level
- ⦿ International level

H Use and dissemination

14. How many Articles were published/accepted for publication in peer-reviewed journals?		**4**
To how many of these is open access[17] provided?		
How many of these are published in open access journals?		
How many of these are published in open repositories?		**3**
To how many of these is open access not provided?		**1**
Please check all applicable reasons for not providing open access:		
☒ publisher's licensing agreement would not permit publishing in a repository ❑ no suitable repository available ❑ no suitable open access journal available ❑ no funds available to publish in an open access journal ❑ lack of time and resources ❑ lack of information on open access ❑ other[18]: ……		
15. How many new patent applications ('priority filings') have been made? *("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).*		**0**
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	**0**
	Registered design	**0**
	Other	
17. How many spin-off companies were created / are planned as a direct result of the project?		**0**

[17] Open Access is defined as free of charge access for anyone via Internet.

[18] For instance: classification for security project.

Indicate the approximate number of additional jobs in these companies:	

18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:

☒	Increase in employment, or	☒	In small & medium-sized enterprises
❑	Safeguard employment, or	❑	In large companies
❑	Decrease in employment,	❑	None of the above / not relevant to the project
❑	Difficult to estimate / not possible to quantify		

19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (*FTE = one person working fulltime for a year*) jobs:	*Indicate figure:* *22*
Difficult to estimate / not possible to quantify	❑

I Media and Communication to the general public

20. As part of the project, were any of the beneficiaries professionals in communication or media relations?

⊙ Yes ○ No

21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?

⊙ Yes ○ No

22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?

☒	Press Release	☒	Coverage in specialist press
❑	Media briefing	☒	Coverage in general (non-specialist) press
☒	TV coverage / report	☒	Coverage in national press
❑	Radio coverage / report	☒	Coverage in international press
☒	Brochures /posters / flyers	☒	Website for the general public / internet
☒	DVD /Film /Multimedia	☒	Event targeting general public (festival, conference, exhibition, science café)

23 In which languages are the information products for the general public produced?

☒	Language of the coordinator	☒	English
☒	Other language(s)		

***Question F-10*:** Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES

1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]

1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)

1.3 Chemical sciences (chemistry, other allied subjects)

1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)

1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY

2.1 Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)

2.2 Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]

2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; the science and technology of food production; specialised technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES

3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)

3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)

3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES

4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)

4.2 Veterinary medicine

5. SOCIAL SCIENCES

5.1 Psychology

5.2 Economics

5.3 Educational sciences (education and training and other allied subjects)

5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary , methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]