Multimedia Semantic Syndication for Enhanced News Services

www.mesh-ip.eu

Annual Report 2008

MESH is an Integrated Project in the Sixth Framework Programme. Its main objectives are to extract, compare and combine multimedia content from multiple news sources, automatically create advanced personalised multimedia summaries, syndicate summaries and content based on the extracted semantic information, and provide end users with a personalised *multimedia mesh* news navigation system.

During 2008 MESH has successfully developed and integrated the second iteration of the MESH integrated prototype, which is being tested at the moment of writing this. This work has focused on the integration of more advanced features developed during the last year, while improvements have been applied to all existing modules in the first iteration.

Part of the advances incorporated into this second prototype comes from the Multimedia Analysis activities. Work was carried out on video segmentation, content representation, person detection, visual classification, application of OCR techniques on video data, automatic speech recognition, linguistic analysis of textual data, and event and highlight detection.

The multimedia analysis in MESH extracts automatically the annotations from different media and provides them to the end user. However, this analysis is generally based on algorithms which can produce errors. MESH Manual Annotation Tool (MAT) addresses this problem by allowing the user to correct the automatic annotation quick, easy, and efficiently.

MESH has organized a large-scale event on Multimedia Semantics, the Summer School or SSMS 2008. A notable novel aspect of this year programme was the objective to connect multimedia semantics researchers with the media sector professionals. MESH organized a 5-day programme centred on state-of-the-art technologies and research issues; and a 3-day programme explaining the application of semantic technologies for media sector professionals. SSMS 2008 was attended by 74 students and 30 media professionals.

Finally, we consider worth mentioning our participation in TRECVID HLFE competition, though the results were not as good as expected, it has proven an exciting co-operation within the consortium and the potential refinement of the techniques developed within MESH.

The MESH 2nd Prototype

The second iteration of the MESH prototype has focused on the integration of more advanced features, while improvements have been applied to all existing modules in the first iteration. Some of these new features include:

- the ability to order a personalised summary built according to user preferences and profile;
- an improved summarization functionality based on detected events in the video;
- improvement of video analysis results by inclusion of the OCR functionality and cross media reasoning.

A brief illustration of the platform showing some basic functionalities is given through the screenshots shown below. In the Login/Search page a vertical panel on the left allows the users to Login or Register to the Mesh platform, as seen in Figure 1. A search horizontal pane is always available, at the top, where users enter keyword-based queries and receive sorted lists of the results in the "search pane". Every news item which involves video content is viewable through streaming in an embedded rectangular window.

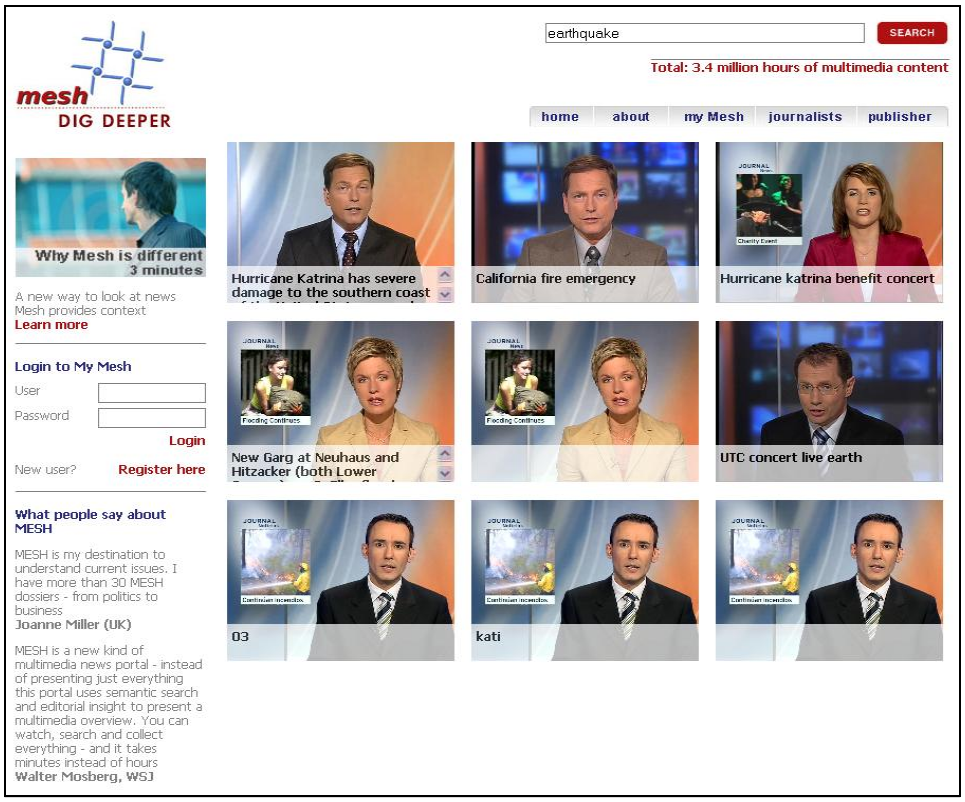

Figure 1: ***MESH Login and Search Page***

Once logged in, the user can submit a search query and view the MESH search results ranked according to her or his profile. If the user clicks on the play button the video starts on the embedded video player without switching to a new page, enabling the user to continue viewing the search results, as shown in Figure 2.

Depending on the MESH ontology a results filter is also available on the left allowing further filtering of results according to concepts found in the resulting content items. Under every video the “play”, “play summary” and “order personalised summary” buttons are always displayed. The user can order a summary that will be created according to his/her personal preferences and profile. Once the personalised summary is ready, a notification is displayed in a notifications pane under the MyMesh tab.

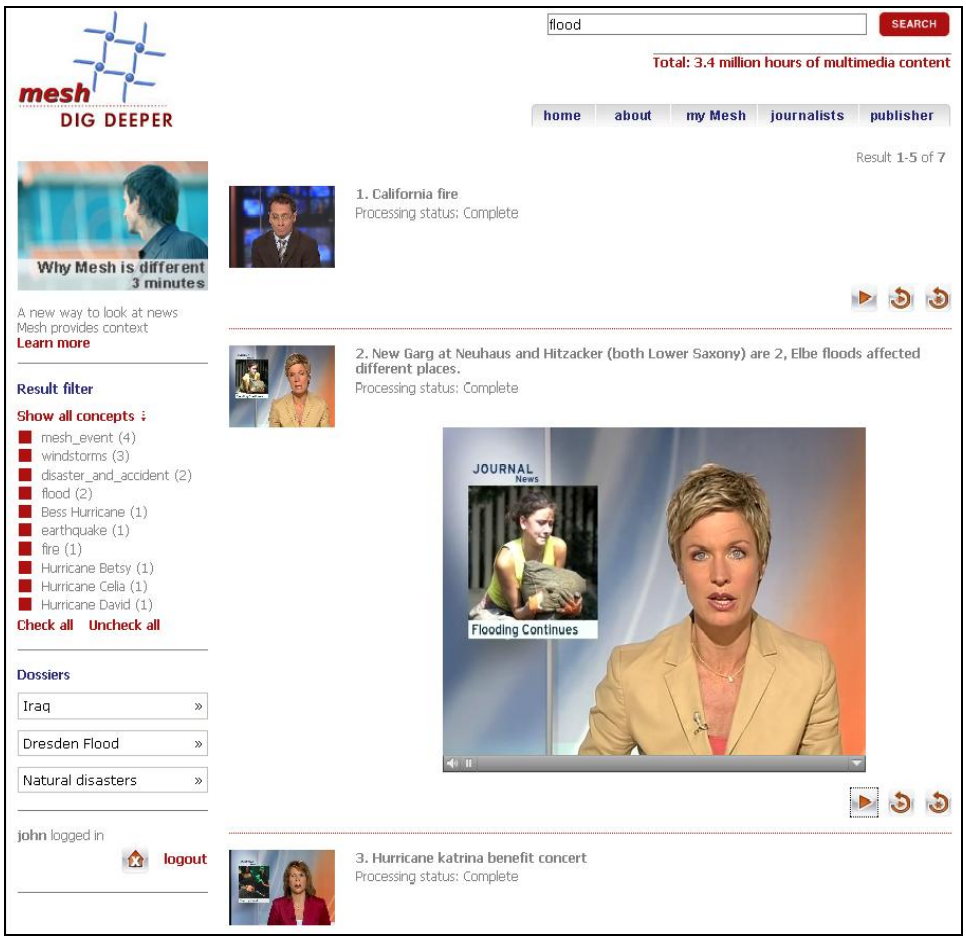

Figure 2: ***Improved Search Results page***

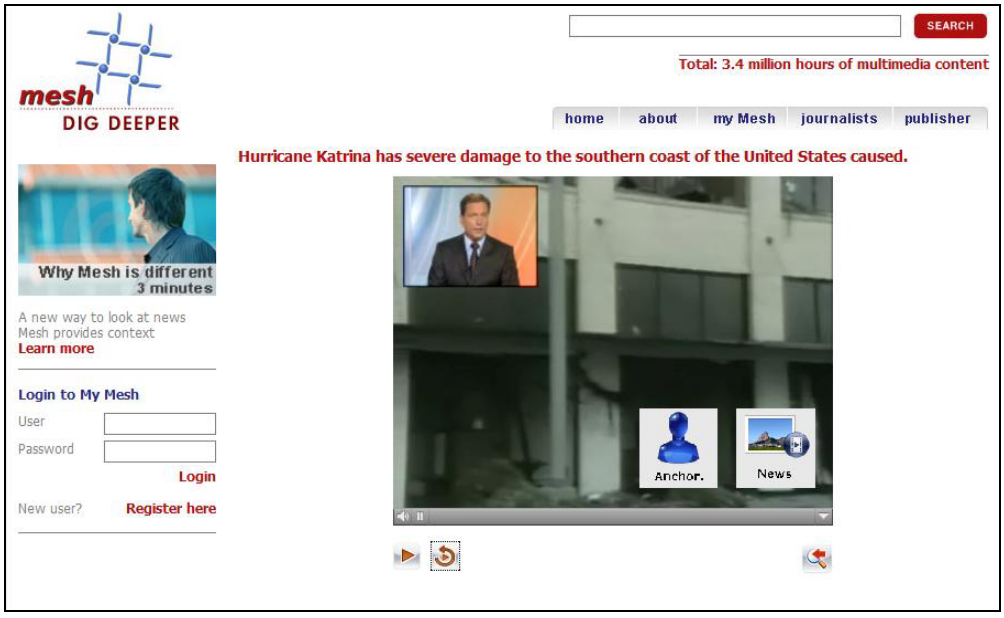

Figure 3: ***Choice between news shots and anchorman shots***

While viewing the summary of a news video the user can choose between following the link of the video that only includes the shots that involve the anchorman ('Anchor' button in Figure 3); or watching the actual shots of the event reported ('News' button in Figure 3). By default the summarised version is composed by the event shots with the video and voice of the anchorman played in an embedded video at the top left of the main video window, in a picture-in-picture style.

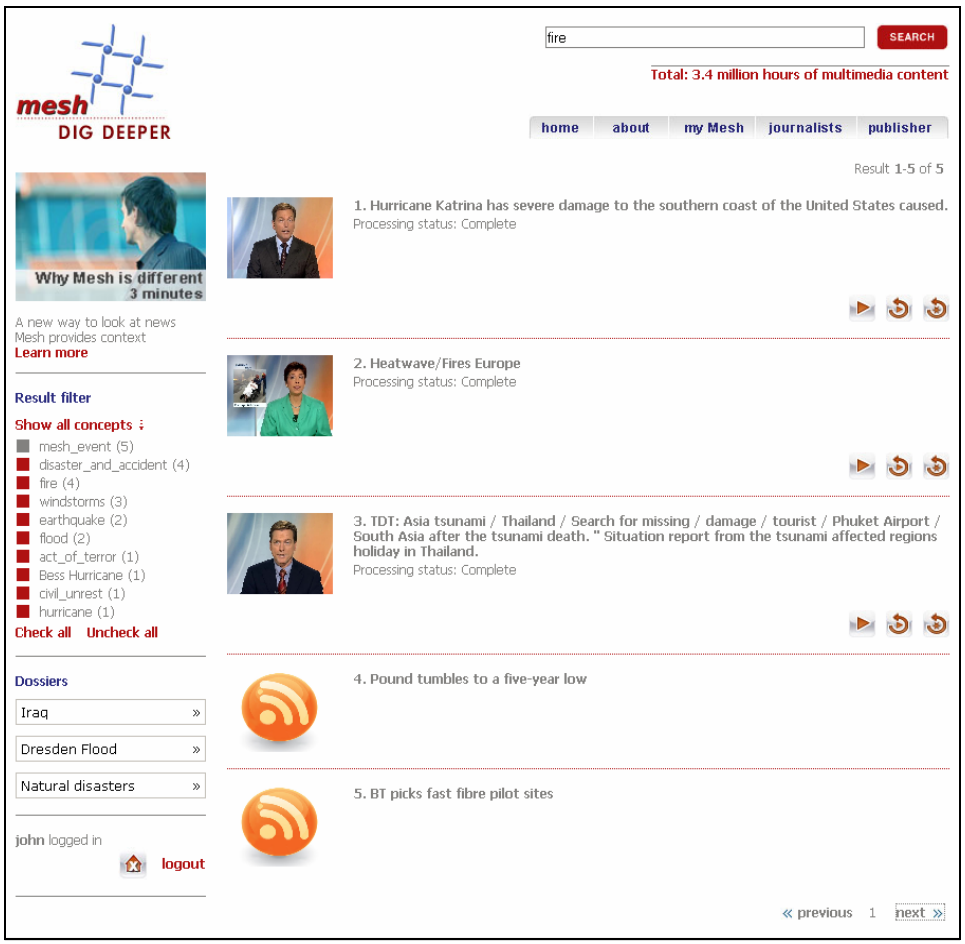

Figure 4: ***RSS Results***

Further to the video news items the MESH system is now able to update the Knowledge Base with articles from RSS feeds as they become available, and hence provide more diverse results (Figure 4).

Advances in Multimedia Analysis

The role of multimedia analysis in MESH is to automatically extract meaningful semantic-level metadata from multimedia content. This involves both the processing of the different modalities of multimedia content independently, to take advantage of the particular characteristics of each, and the combination of analysis results across modalities, to enable the reliable extraction of rich semantics from the content.

On individual modality analysis, work was carried out on video segmentation, content representation, person detection (an example of which can be seen in Figure 5), visual classification, application of OCR techniques on video data, automatic speech recognition, linguistic analysis of textual data, and event and highlight detection.

In order to produce final multimedia analysis results going beyond a simple concatenation of the individual modality results outlined above, and to serve as an error-correction mechanism that discards conflicting annotations on the basis of various criteria, work was carried out on cross-media analysis and reasoning. A cross-media analysis technique was developed to this end, based on examining the particular characteristics of the news multimedia content and using the MESH Knowledge Structures.

Frame 10 *Frame 18* *Frame 50*

(a)

(b)

Figure 5 (a) Selected frames from a video depicting two men entering the scene; (b) Corresponding saliency frames. Notice how saliency changes when those two events occur. Initially, the standing man is the most salient region. The second man becomes more salient when he enters the scene, while the same happens at the third selected frame.

Semantic Navigation of News Repositories

The MESH analysis tools result in semantic annotations to a news item. These annotations can be used to establish semantic similarity relations between news items from different sources, across language and across time as well as which items refer to the same event. A module was developed to support users in retrieving documents about specific topics from a semantically annotated repository of news items using not just text based similarity but exploiting the semantic annotations and relationships.

When the user views a document the module displays also the most relevant information about the content of the document such as topical events, persons, locations, organisations and dates mentioned.

The module integrates MESH document analysis, the MESH reasoning environment to derive semantic relationships and semantic similarity in addition to full text search.

MESH Manual Annotation Tool

The multimedia analysis in MESH (speech recognition, textual analysis, visual classification, and cross-media analysis) extracts automatically the annotations from different media and makes them available to the end user. However, this analysis is generally based on algorithms that can produce errors. The human intervention in the annotation process becomes then essential. MESH Manual Annotation Tool (MAT) addresses this problem by allowing the user to correct the automatic annotation quick, easy, and efficiently.

With MAT the annotator can quickly correct the automatic annotations by propagating the same individuals among similar shots. MAT visualizes all decompositions: textual and linguistic analysis, and cross-media decomposition, and the user can compare the intermediate annotations with the final result (cross-media) and correct or complete this later.

MAT implements also a reusability mechanism of corrected or completed annotations, thus avoiding to make the same corrections inside a video. It allows dealing with the duplication of individuals and so makes easier the annotation process.

To reach these objectives, the MAT tool communicates with other Mesh components, in particular, with the relevance feedback component (Figure 6) and with the keyword recommender (Figure 7). By means of the former we aim to propagate the same annotation among similar shots inside the video. The similarity should be based on multiple criteria: visual, semantic and face similarity. These tasks are in progress in the relevance feedback module. With the later, we provide the user with recommended (related) keywords when she or he is correcting an annotation.

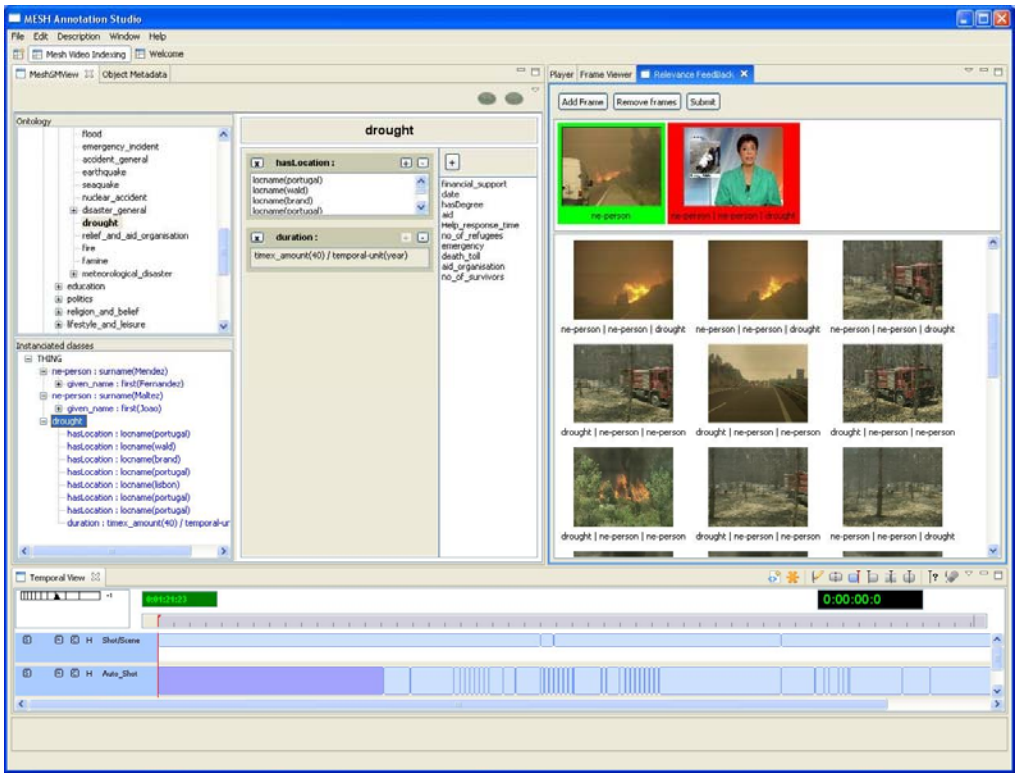

Figure 6 ***MESH Manual Annotation Tool and Relevance Feedback functionality***

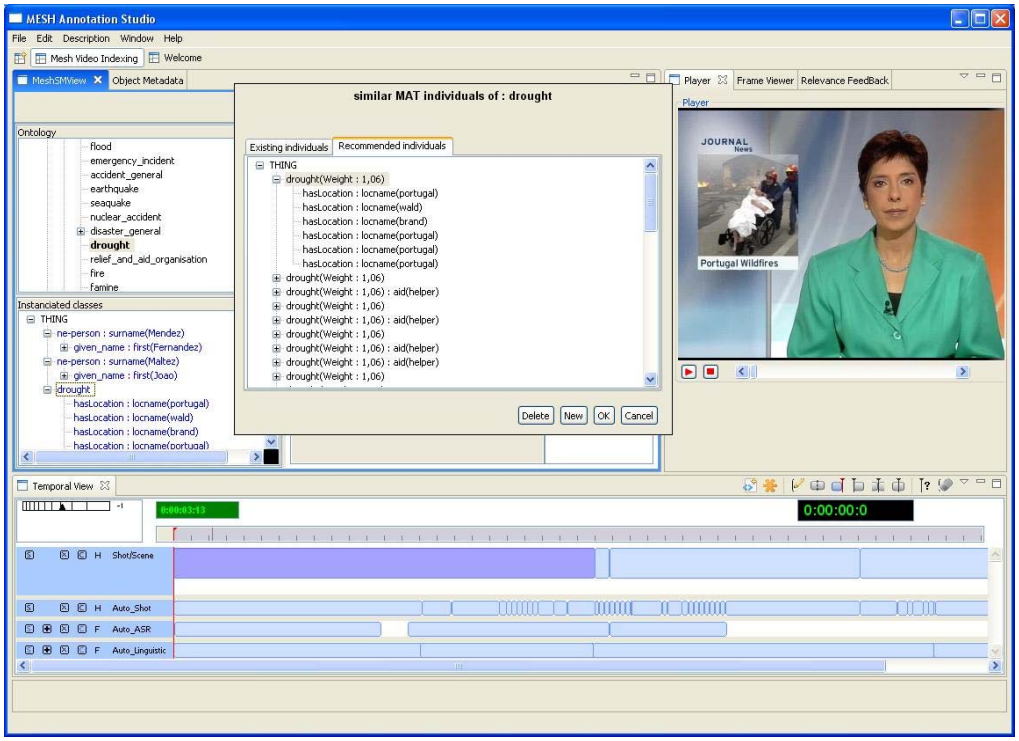

Figure 7 ***MESH Manual Annotation Tool and Keyword Recommendation functionality***

Digital Rights Management in the MESH system

The MESH Rights Management Module is an integral part of the MESH system which deals with Reasoning for syndication and access. The component was updated to allow mapping the Digital Rights Management (DRM) module to a more popular licensing schema such as Creative Commons or the licenses in use by MESH partners Deutsche Welle or DIAS Publishing House.

Together with the DRM module, the statistics component provides the information needed to generate personalized news experiences within the MESH system. The personalization modules use this module to select items that are of interest to the user. Other uses of statistics in MESH are that DRM system can check if a user has already viewed the document that she has paid for. On the server side the statistical information is used to link content by looking for documents that were viewed by the same users. The search components can use the popularity of documents in their ranking, and more popular documents can be given a higher rank than documents that have not been viewed as many times.

MESH participation to TRECVID2008 HLFE

A group of four organizations from the MESH consortium participated this year for the first time in the High Level Feature Extraction track in TRECVID, or Video Retrieval Evaluation series of the TREC conference sponsored by the National Institute of Standards and Technology (NIST) of the U.S.A.. The partners were Telefónica I+D (TID, Spain), Informatics & Telematics Institute (ITI, Greece), National Technical University of Athens (NTUA, Greece) and Universidad Autónoma de Madrid (UAM, Spain). We submitted a total of 6 runs, using different variations and configurations over a common model.

This is the first joint participation of the partners in the MESH consortium in TRECVID. With this purpose the MESH project developed a common visual analysis infrastructure to detect high level concepts in visual scenes; though the set of concepts was only partially coincident with those of TRECVID 2008 (in MESH it is tuned to news content). In the course of the development, a few new techniques, not originally present in the MESH system, were also tried.

The main architecture of the HLFE system is based on well-known paradigms in visual analysis, such as MPEG-7 descriptors, SIFT interest points and SVM classifiers. We hoped that the specifics of their combination would provide good results. Moreover, one guiding principle in the development was not to include human intervention in model selection and configuration for each individual feature. This rule stems from our aim to be able to generalize the system to any additional feature without resorting to human intelligence to select and combine adequately the available set of tools. The system, thus, gets trained blindly with a ground-truth training set, and adapts automatically to the specifics of each concept during this training phase.

With only one exception (the Combined SIFT & MPEG-7 run, which performed reasonably as compared to the median performance as computed after the submission for all participants), results obtained were below expectations, mostly due (we believe) to implementation bugs discovered afterwards. Some of those errors have already been solved and we hope to correct the rest and improve the performance of the system for future editions.

User Involvement, Promotion and Awareness

The Summer School on Multimedia Semantics 2008

Understanding and thereby manipulating multimedia content at the semantic level is the only way towards realizing the full potential of emerging digital media technologies aimed at the delivery of compelling multimedia solutions. The integration of knowledge, semantics and low-level multimedia processing for the purpose of automatic semantics extraction from multimedia content is still the subject of active research in academia and industry. The Summer School on Multimedia Semantics was established to bring the relevant academic and media professional societies closer to the latest research and commercial developments providing the opportunity to gain deeper insight into the related research challenges and the ever increasing number of new applications.

SSMS has been held on two previous occasions and SSMS2008 built upon the success of these events. A notable novel aspect of this year's programme was the objective to connect multimedia semantics researchers with the professional media sector. To this end, two groups of attendees were catered for: postgraduate students and researchers attending a 5-day programme centred around state-of-the-art technologies and research issues; and professionals attending a 3-day programme explaining the application of semantic technologies to their day-to-day activities. The fourth day of the event brought both groups together attending a series of lectures from representatives of world renowned organisations.

SSMS 2008 was attended by 74 students and 30 media professionals.

The programme for researchers included the following lectures:

- Multimedia Signal Processing, by Prof Noel O'Connor
- Machine Learning, Pattern Recognition, Cross-modal analysis and fusion, by Dr Alexander Hauptmann
- Semantic Web and Multimedia, by Prof Steffen Staab
- Multimedia Ontologies for Reasoning and Analysis, by Prof Ralf Moeller
- Multimedia Information Indexing and Retrieval, by Prof Alan Smeaton
- Flexible Interfaces for Semantically Annotated Multimedia, by Prof Lynda Hardman
- Semantics and 3D media, by Dr Michela Spagnuolo
- Social Networks and Multimedia Semantics, by Dr Peter Mika

The programme for professionals included the following presentations:

- Calais - the media and the semantic web, by Barak Pridor, CEO, ClearForest, a division of Thomson Reuters

- A News Agency in Transition, by Meinolf Ellers, Managing Director dpa-infocom GmbH, Hamburg, Germany
- Mobile Services and Mobile Devices, by Paola Hobson, Motorola UK Research Lab, Basingstoke
- Newsgaming, by Julian Kuecklich, Press Association, London
- A newspaper goes multimedia, by Robert Freeman, Head of video, Guardian, UK
- Content Management, Data Mining, CRM, by Frank Poerschmann, IBM Global Business Services, Lead Media & Entertainment GB, Lead CRM Strategy, IBM Deutschland GmbH
- Concerns on Privacy and Innovation in ICT and Media Industries, by Rainer Boehme, Faculty of Computer Science, TU Dresden
- Talking about Networked Journalism, Search, Privacy, Filtering, Personalisation. Panel including:
    - Clancy Childs, Sales Engineer, Partner Services, Google, London, UK
    - Willi Ruetten, Director European Journalism Centre, Maastricht
    - Robert Freeman, Head of video, Guardian, UK
    - Andreas Neus, IBM Germany / Karlsruhe Service Research Institute, Karlsruhe University

***More information on SSMS2008 can be found under* http://www.mesh-ip.eu/?Page=ssms08*, and, at the moment of writing this report, the audio-visual recordings of the sessions are being uploaded to videolectures website* (http://videolectures.net)**

Moreover, the MESH consortium is still making great effort in publishing the developed methods and results in academic conferences and journals. The full list of publications can be consulted at the project website: **http://www.mesh-ip.eu/?Page=results**

Future Work

The second MESH prototype has been finished in 2008. This is the MESH platform that includes all the multimedia analysis modules developed during the project, including some recent ones developed during the first half of the year 2008.

The project is now validating, as another step in the iterative testing process adopted in the project, this second prototype. The results of this validation will be incorporated in the final system, and will set up the basis for the refinements that can be planned for the MESH system in the near future. Also, in combination with efforts undertaken so far about market tracking, exploitation opportunities and planning, we will finalize the commercialization plans of both individual components or modules and the MESH system as a whole.

Further Information

The MESH consortium is formed by: **Telefónica I+D**, **Informatics and Telematics Institute**, **Athens Technology Center**, **Motorola**, **Queen Mary University of London**, **Institut National de l'Audiovisuel**, **Noterik**, **Universiteit Twente**, **Deutsche Welle**, **Deutsches Forschungszentrum für Künstliche Intelligenz**, **Universidad Autónoma de Madrid** and **DIAS Publishing**.