1. Publishable summary

The need for annotated multimedia content is more demanding than ever before. The fully automated annotation has made recently large progress but still faces limitations due to unsolved technical issues. Thus, the machine-only annotation appears immature although presenting promising achievements. On the other hand, manual (human) annotation is very cost-prohibitive.

CASAM project is focusing on facilitating the synergy of human and machine intelligence to significantly speed up the task of human-produced semantic annotation of multimedia content. The project research focus lies in the domains of Reasoning for Multimedia Interpretation (RMI), Knowledge-Driven Multimedia Analysis (KDMA) and Human-Computer Interaction (HCI). The project has developed an annotation tool that can augment machine knowledge with human input with the target of minimizing user effort.

The annotation tool can function within the modelled domain of news production of News Agencies and Broadcasters. However, the methods that have been developed are not bound to the chosen domain, but are also applicable for the annotation of multimedia documents in a variety of contexts, ensuring generality of the system’s usage.

The main target of the CASAM project was to design, implement and validate an Annotation Toolkit that can augment machine knowledge with human input, towards the target of minimizing user effort and bridging the gap between machine-derived and human annotation. The main scientific objectives were the following,

1. To provide the technology for Multimedia Analysis
   - Knowledge-driven analysis
   - Reassessment of context and concepts based on user feedback and new knowledge

2. To develop novel methods of Reasoning for Multimedia Interpretation
   - Built around information exchange between multimedia analysis and human

3. To design unique intelligent Human-Computer Interaction methods
   - Maximization of the expected information gain from the user’s input
   - Cooperative environment to guide the knowledge aggregation process

Advances in each of the three research areas (multimedia analysis, multimedia interpretation, human-computer interaction) have been achieved throughout the span of the project in order to follow this novel approach to multimedia content annotation. The main achieved outcomes of the project include:
a. **A unifying representation for related knowledge.** The representation links domain-specific ontologies, in which concepts are represented by domain-specific terms, with multimedia interpretation and context ontologies that represent relations and constraints between concepts. The representation also includes rules that facilitate the maximization of information gain from the user’s input.

b. **A methodology for knowledge-driven multimedia analysis.** The methodology provides ways to enhance multimedia analysis when knowledge is available about the context or the probability of presence of certain entities in the content. The proposed methodology also specifies how the knowledge represented in the multimedia semantic model is used to achieve information extraction from various media (text, image, video and audio). Ontology enrichment methods have been examined for the discovery of new concepts and rules, based on the interpreted content.

c. **A methodology of reasoning for multimedia interpretation.** Development of intelligent methods that seek to acquire missing pieces of knowledge and disambiguate uncertain knowledge about multimedia documents. The algorithms are able to reason about what they know, what they don’t know, what required knowledge has the highest priority and how they acquire it. They are able to merge machine-derived and human-entered knowledge about the content and refine it through directed user feedback. The developed architecture communicates with both the knowledge-driven multimedia analysis module and the human-computer interface.

d. **Intelligent human-computer interface methodology for maximizing information gain in multimedia annotation.** Development of human-computer interface protocols and methods that query the user to acquire specific information. The HCI methods are able to transform information requirements into user effort-optimal interaction scenarios. The HCI method also maintains a model of the user and uses it to customize the queries issued in order to maximize the information value of user feedback.

e. **A toolkit for computer-aided semantic annotation of multimedia content.** The toolkit encapsulates all the methodologies developed and also supports ontology management, ontology enrichment and multimedia content management. At design time (configuration level) the expert constructs the initial domain-specific ontology. Further enrichment of the ontology with new concepts is possible and can be carried out also by the expert user. At runtime (operation level), end users are able to semantically annotate multimedia content.
1.1 Work performed and results achieved during the third and final year of the project.

Knowledge-Driven Multimedia Analysis. In the framework of knowledge-driven multimedia analysis, in the final year of the project, the KDMA module and its components were enhanced with a number of significant new features and improvements. Particular emphasis was placed on the adaptability of the automated multimedia analysis to user input and a changing view of the system about the world. Towards this goal, the textual analysis and the audio event detection components were enhanced with new methods that improve their results, based on a given context.

Additional improvements on the methods employed by KDMA components include:

- Enhancement of audio analysis with a novel algorithm for speaker clustering. The algorithm is based on optimal finding a suitable feature subspace to distinguish speakers and has won the ICMLA 2010 speaker clustering challenge\(^1\).
- New lexical sources, namely DBPedia and Wiktionary, for the calculation of semantic relatedness between the text and the ontology concepts. Extension of the relatedness methods, in order to improve probabilistic annotations.
- A new video-oriented OCR engine achieving improvements over common OCR engines that were used in the first two years. Also, a new video shot classification algorithm, based on a Support Vector Machine classifier.
- A complete framework for multimodal temporal event detection, adding probabilistic interpretations to Allen’s interval relations (Allen’s Hourglass).

Knowledge Representation and Reasoning for Multimedia Interpretation. Regarding CASAM knowledge representation and reasoning for multimedia interpretation, in the final year a stable version of the RMI module (Reasoning-based Media Interpretation) was developed. Interpretations and feedback queries to other modules are generated based on a domain model, namely an ontology and a set of weighted abduction rules. In order to be able to define the weights used in the rules, a learning component was developed.

Furthermore, during the final year of the project, the Machine Learning (ML) algorithm for the training of Markov Logic Networks (MLNs)\(^2\) was optimized. A modified Gibbs sampling method\(^3\) for the calculation of the gradient used in ML was developed that is suitable for highly complex ontologies. Two different methods that optimize the sampling algorithm for MLN cliques with many members were also developed. These methods further reduce the

---

\(^1\) Please refer to Annex - Section 6.1 for further details.
\(^3\) [http://www.jstor.org/pss/2685208](http://www.jstor.org/pss/2685208)
computational complexity of the ML method. Finally, the ML module and related web services were developed and the module was integrated into the CASAM system.

**Interactive Interface Design, Personalization and Optimization.** Within this task the remaining deliverables (D5.3 & D5.4) have been completed and submitted. The final prototype of the HCI component for integration and evaluation was completed in September 2010 and successfully integrated into the final CASAM prototype which has been used for end-user evaluations. This prototype underwent usability testing with a group of test subjects and the results of the feedback from these studies was used to specify a final set of refinements before final integration took place.

This final integrated prototype is able to receive annotations (assertions) from the RMI & KDMA components of CASAM as well as requests from these components for information to be requested from the user (queries). The volume of these assertions and queries is very great and so the HCI component needs to reason over this information and determine what to present to the user and when and how to do this. This process of managing the dialogue is driven by the need to balance the value of the information gained against the cost of obtaining the information from the user. A series of experiments were undertaken (and reported in D5.3) in order to determine the cost of different forms and timings of interruptions to the user.

The final dialogue between the system and the user is determined by a model of the user, a model of the dialogue and the value of the information to the system and its annotations.

The HCI component also uses models of the domain to predict annotations that the user might wish to add to the document. These are presented to the user as either suggested tags from which the user can choose or as explicit HCI queries, as appropriate.

A version of the HCI front end component has been implemented on a mobile device and its use embedded within a simple workflow. The example is for video capture, upload to CASAM and YouTube, annotation using the CASAM system and then the upload to YouTube of the results of the annotation process. This is intended to demonstrate the portability of the CASAM application across domains and devices and also to provide an exemplar of how the CASAM system can be embedded into a business’s workflow and integrated with other tools.

**End-User Evaluations.** The third year of the project was marked by the two rounds of user evaluations of the initial and final integrated CASAM prototypes, respectively. The first round, with a total of 28 appropriate TV and multimedia participants at different career levels, took place in May 2010, while the second round followed from mid-February 2011 and comprised 34 test users. The second time around, participants with prior exposure to the system – and thus the ability to assess overall progress and improvements – were mixed with fresh ones who could make unprejudiced observations. Testers were recruited at
international level through project partners DW, LUSA, ATC, and EJC, and included staff members as well as external contacts.

The first prototype with its still limited functionality primarily drew feedback concerning the usability of the interface and the requirements of work flow in professional environments as represented by CASAM. By the time of the second prototype many usability issues had been resolved and the full range of actual functionality could be presented to the users. This led to a stronger focus on the quality of system-generated annotations, on understanding how CASAM works and interacts with the users, and the system’s potential fit with the practical demands of media organisations.

User feedback was very rich in detail and focused. It quickly turned out that in both rounds, the vast majority of subjects commented on a rather limited number of identical issues, which accordingly were deemed the most relevant, although a great number of individual observations also provided valuable feedback for the development of the system.

In all, user evaluations came back with a general approval of the system and its objectives. Participants clearly saw the demand for and benefits of CASAM in their respective work environments. Interestingly, this held true throughout the entire range of specialisations that were represented in the tests – from cameramen to producers and video editors to commissioning editors, desk editors, archivists, documentalists, and management-level media professionals, as well as across the areas of conventional television, news agencies, and online video. This means that the relatively universal approach of CASAM successfully caters to the entire bandwidth of audiovisual activities and is capable of playing a part in all stages of a video’s life cycle.

The first round of user evaluations returned substantial feedback on the user interface and user interactions, which were by and large implemented in the final prototype. Indeed, during the second round, the number of high-priority issues with usability and work flow declined substantially. Overall, users were much more easily able to grasp the concept and workings of CASAM. This was not least due to the fact that the system could also be presented in a more elaborated context, with user management and content management modules as well as the search engine present.

The major remaining observation made by user evaluators was that the system was not yet always able to properly differentiate between annotations describing the entire video on the one hand, and annotations pertaining to individual shots on the other hand. Moreover, the nature of the actual semantic system outputs stirred up a keen user interest in how CASAM arrives at certain conclusions and from which level of analysis particular annotations were drawn (i.e., image, speech, OCR, sound, existing metadata, human input, autonomous reasoning). Accordingly, participants expressed the wish for more transparent processing, such as visual indicators of the system’s status and of the source of individual annotations.

Despite identifying further potential for improvement and fine-tuning, test users of the final prototype rated the quality of annotations quite good and conducive to successful search and retrieval of video. They also clearly recognised CASAM’s potential to save time and to
reduce tedious manual labour. In summary, and with the caveats that typically come with a prototype (as compared to a fully developed product), the system was deemed fit for purpose.

**System Design and Integration.** During the third year of the project, the final CASAM integrated system has been developed, deployed and also undergone through the final end-user evaluation process. The functionality provided by the final prototype included scheduled updates and enchantments from the various core modules of CASAM, namely HCI, RMI and KDMA, as well as additional functionalities (i.e. machine learning support, user and content management facilities, semantic searching capabilities, etc.) that were indicated and derived mostly from the outcomes of the initial end-user evaluation sessions and several other internal meetings with the users of the consortium so as to improve to the best possible extend the final end result of CASAM.

More specifically, the work conducted within the framework of the system design and integration tasks focused around the following main pillars:

- **Encasements on the Integration Platform,** in order to support updates on the core modules of CASAM with respect to the multimedia content annotation process. The main modifications and additions that have been introduced include: (a) User authentication was added at the beginning of each annotation session, (b) A session identifier accompanies all messages exchanged so that they can also be traced back to a specific session and a specific user, (c) IPTC metadata accompanies each document, (d) A new service was implemented and is provided by HCI in order to receive information (in the form of assertions sets) directly from KDMA.

- **Machine Learning (ML) support,** provided by the Integration Platform. With reference to this task, the design for the Machine Learning component of CASAM has been updated and finalised, including the services and data-types of ML. WSDL and XSD files related to Machine Learning have been produced and deployed, allowing the Integration Platform to provide access services to the Machine Learning module for the result files of the annotations sessions, along with the TBoxes which were used in those sessions. The application and web services developed in the context of Machine Learning module handles data communication between the main CASAM Annotation System and the actual Machine Learning application.

- **User and Content Management administration facilities** have been developed and deployed. These features aim to bring the final integrated CASAM Annotation System closer to the level of a commercial application. This administration management application, handles user management, access control management and content management for CASAM. At the same time, IPTC metadata management for the multimedia documents were introduced, which are made available to the Annotation User during a session, through a newly introduced web service.
- **Semantic Searching capabilities.** A semantic search module has been developed by utilizing the multimedia annotation results produced by the final integrated CASAM system. The CASAM Search Module enables users to perform high level search within the metadata document corpus and retrieves a fully ranked list of multimedia segments matching the search criteria. This novel searching approach is also taking advantage of a synergy with the text analysis module (KDTA) provided by KDMA. Toward this end, fusing semantic abilities into the CASAM Search Engine is achieved through the Semantic Search Module (SSM), which is a Java library providing a high-level interface to KDMA and RMI results.

Furthermore, following the recommendations and suggestions of the review panel as these have been portrayed in the 2nd Annual Review Report, a detailed case study has been conducted and provided on the possibilities of integrating CASAM into other 3rd Party Systems and also the specification for operating CASAM in a demanding production environment, under High Availability requirements. The aforementioned case study may be perceived as a “white paper” which describes the integration capabilities and specifications of CASAM into different levels (i.e. content level, access control level, etc.). This is intended to be utilised as a guide for business exploitation of CASAM solution to existing market players using other 3rd party systems.

Finally, in the context of the overall system evaluation of the integrated CASAM system, a detailed report has been provided including: (a) the testing methodology followed throughout the integration testing processes for both the initial and final CASAM prototype, (b) the test cases and test procedures adopted during these system test procedures, and (c) the final outcomes and results of the integration test procedures for both prototypes.
**Dissemination and Exploitation.** During the reporting period the following dissemination and exploitation activities have been performed:

With respect to dissemination, **20 scientific publications** have been successfully submitted to prestigious conferences, a special one-day session for the CASAM project has taken place in 1-3 December 2010 as part of the SAMT 2010 conference in Germany. Furthermore, several partners have presented aspects of the project in scientific and business events.

The exploitation activities included market research and the deployment of a business expert from ATC that devised two specific products from the CASAM project that were analysed and described in the updated Business and Exploitation Plan deliverable D2.5c.


1.2 Reference to the project public website and project logo

<table>
<thead>
<tr>
<th>Project website URL (<a href="http://www.casam-project.eu">www.casam-project.eu</a>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASAM</td>
</tr>
</tbody>
</table>