DIRECT-INFO: MEDIA MONITORING AND MULTIMODAL ANALYSIS FOR TIME CRITICAL DECISIONS

Herwig Rehatschek

JOANNEUM RESEARCH Forschungsgesellschaft mbH, Institute of Information Systems & Information Management, Steyrergasse 17, A-8010 Graz, Austria.
E-mail: Herwig.Rehatschek@joanneum.at

ABSTRACT

DIRECT-INFO aims to create a basic system for semi-automatic extraction of consistent and meaningful semantic information from multimedia content. Its main goal is to offer an integrated system combining the output of basic media analysis modules to semantically meaningful trend analysis results which shall give executive managers and policy makers a solid basis for their strategic decisions.

1 OBJECTIVES

DIRECT-INFO deals with media monitoring and multimodal analysis for time critical decisions and aims to create a basic system for semi-automatic extraction of consistent and meaningful semantic information from multimedia content. Its main goal is to offer an integrated system combining the output of basic media analysis modules to semantically meaningful trend analysis results which shall give executive managers and policy makers a solid basis for their strategic decisions.

The DIRECT-INFO system shall provide an innovative solution in the area of decision support systems. It will significantly ease necessary information filtering and analysis steps that are important in decision making processes. DIRECT-INFO will address concrete usage scenarios that are common to virtually all advertising and media monitoring industries such as company image analysis and introduction of new product / competitor analysis, while it will be designed in such a way as to cover further scenarios in additional sectors such as government, policy/PR analysis and decision making, financial (decision making, trend analysis) and other corporate and media scenarios.

Technically the system shall integrate the entire information value chain, audio/visual, text, audio and video data, with a comprehensive, end-to-end solution for the collection, identification and, analysis of media information. It shall *utilise the multimodality and event-based characteristic* of multimedia content for a wide range of applications as for example the extraction and annotation of stories from broadcasts, the detection of highlights and summarisation of sport events or monitoring solutions.

The goal is to *first detect and extract logical entities (context)* from the data by pre-specified schemes for different kind of information, specific broadcasts, and events of interest or general descriptions of different genres. The underlying context-specification model comprises then the pre-stored information. In the second step, the extracted entities will then be semantically correlated by taking advantage of *predefined and learned contextual knowledge and semantic information* and finally fused to a for the end-user meaningful summary.

The outstanding differentiation factors of DIRECT-INFO are as follows:

1. Ability to monitor and analyze audio, image, video and audio-visual information across media channels including radio, TV and the Internet;
2. Applicable to numerous industries, also beyond the media-information market;
3. Non-intrusive solution allowing instant capture of information;
4. System ‘learns’ as data and information is acquired and compared to existing data;
5. Access to information integrated across channels and markets;
6. Comprehensive end-user access to real-time global information; and,
7. Reliable information that represents the trends and is not statistically based to human error.

The result of the project will be a flexible system – flexible in terms of: easy to re-configure according to user needs – comprising several components that implement the needed functionality for analysing of multimedia data taking advantage of combining

complementary analysis results of different content to semantic meaningful entities.

2 INNOVATION

2.1 Semantic correlation & fusion of multi-modal content analysis

The main innovation in the knowledge management area comes from the integration of semantic web technology into a content management system (CMS). In the DIRECT-INFO project significant development will be carried out in three areas. First, techniques from knowledge representation and knowledge querying will be integrated into the existing indexing and search framework of the CMS. Second, the simple keyword-based techniques for categorisation of content will be extended in a way that builds on the experience of web ontologies. Third, it will be ensured that content that is made available to other DIRECT-INFO components, and to external software, comes tagged with a suitably rich set of meta-data encoded according to W3C standards. A further novel aspect of the project, will be extending the CMS to handle streamed multi-media formats. Some basic functionality already exists, but not enough to cover the sophisticated needs of DIRECT-INFO. Significant innovation will be found in the integration of the multi-modal content analysis components into the enhanced CMS. This could be the basis of a powerful system that will be useful in many areas outside of those of image and competitor analysis that are trialled within DIRECT-INFO itself.

2.2 Identification of semantic objects within A/V content

Generating a high-level semantic description, as it is required in DIRECT-INFO, requires the recognition of objects and a description of their relationships. The central content focused within DIRECT-INFO project is advertisement and news content. It is a matter of fact that within advertisements the object of main interest are often moving because the advertiser wants to attract the attention of the customer on the object he wants to sell. Based on JRS's already existing expertise in fully automatically segmentation of moving objects, an object recognition module will be developed, which is capable of segmenting and recognizing both moving and static objects from a given set of object classes. Moving objects will be tracked throughout the video sequence, so that a description of their motion trajectory can be generated. Because of the fact, that moving objects can be automatically segmented, efficiency and robustness of recognition can be improved beyond the constraints of general object recognition systems. Furthermore, unrecognized moving objects can be presented to the user to add them as new object classes, so that the system is capable of learning new object descriptions. An object recognition module will generate a description of the objects in the video (identification, shape and motion trajectory, if applicable). This information will be stored in a standardised way (MPEG-7) which enables interoperability across components and systems.

2.3 Multi-modal event modelling for videos

DIRECT-INFO will be completely based on multi-modal analysis in contrast to most solutions for video analysis, which are still focusing on one modality. There are four different kinds of information channels regarding video multimodality, visual modality including everything that is visible in the video scene, including artificial (graphics) and natural content (video), audio modality including environmental sounds as well as music, jingles etc., speech modality including the spoken language in the video, which could already provide semantic information about the content of the video, and text modality including text overlays, which also provide semantic information already.

The usage of multi-modal analysis in video raises the question about what should be analysed in the video stream and how could it be done. Regarding human beings the process of perception is a pre-conscious level of cognition ("signal level"); it organizes the incoming sensory signals (for instance, visual light waves or auditory sound waves) into information instances such as objects and events. This perceptual organization is then taken over by higher cognitive levels in order to be enriched by knowledge, so that we can become aware of what is present in the world around us. Because object recognition is still a hard task, event detection and modeling is the more promising way towards automatic semantic annotation and description of multi-modal broadcasts.

DIRECT-INFO will first aim on the identification of "master events" representing logical units of coherent content, starting from basic events in video corresponding to the pure perceptual level as shots, noise, music, text-overlays, etc.,. Master events could occur on different stages or levels of the cognitive analysis process as they could be used for the description of compact entities as single news stories, trailers, interview situations, etc. as well as for complete broadcasts, which itself contain several master events. To achieve the goal of multi-modal integration in terms of event detection, classification, or

identification, different approaches could be followed as for example Hidden Markov Models that are frequently used as a statistical classification method for multi-modal integration. A clear advantage of this framework is its capability to integrate multi-modal features as well as to include sequential features as they occur in videos.

2.4 Video Segmentation

DIRECT-INFO's aim is to segment the broadcasted video stream into logical story units first, and then to annotate them with semantic information. The segmentation will be achieved through the identification of master events in the broadcasted video stream. One specific aim is to provide a description model that allows a consistent and efficient description of master events and its visual and auditory content for later identification purposes. For the representation of the event model, available standards as MPEG-7 will be evaluated and eventually extended in order to achieve the necessary description capabilities. The goal is to create an adequate and process-able representation for all kinds of events in the field of broadcasted video. The description model will represent single or combined events, which are related to specific TV channels, broadcasts, special events of interest as dialogue scenes in movies, story entities in news broadcasts, etc. or general descriptions of genres as news, talk shows or sport.

DIRECT-INFO will provide a novel automatic semantic enhanced indexing method that currently does not exist. Common indexing technologies only use low-level features or manual extracted semantic annotations for indexing. The segmented units in DIRECT-INFO will be annotated using predefined semantic attributes that are derived from the underlying event model. The pre-stored information will therefore include for example contextual knowledge including general attributes as purpose, genre or sub-genre. For a further semantic description of the segmented video units additional functionality could be included as for example the detection of people, which could be found in the visual modality by means of their faces or body parts as well as in the auditory modality by the presence of speech. This information could be used for the annotation of the video as well as information that is derived by the classification of the video setting, where we could differentiate between indoor scenes and outdoor scenes. Outdoor scenes could be further classified into city and landscape shots.

Beside these global attributes additional semantic information could be added through the specific detection of named events, which could classify certain events or master events in the video stream. This could include e.g. explosions in action movies, interviews in news broadcasts, highlights in sport broadcasts or simply the weather-forecast in a news broadcast.

In addition to the general semantic attributes, which could be derived by contextual knowledge and specific event detection, the system will also apply speech recognition and transcription as well as character recognition (OCR) for text overlays in order to gain additional high-level semantic information. Therefore the trans-scripted text fragments will be further analysed using state of the art techniques for text mining and summarization. The goal will be to extract meaningful and reliable keywords for the annotation of the video segments.

3 BENEFITS

DIRECT-INFO aims to offer cost reducing and efficiency enhancing solutions by automating the currently mostly manually performed media monitoring process as much as possible.

The distinct impact of DIRECT-INFO will be an overall increase in competitiveness due to the rapid access to relevant information with minimal background ‘noise’ and interference – or in other words a drastic reduction of information overload.

Primarily targeted end-users:

Media Information Companies, active in the advertising, news and music information markets, that capture, monitor, archive, and analyze media information to serve clients’ needs. The system will provide:

- Enhanced productivity – fewer operators are required per monitored channel
- Greater information accuracy
- Shorter time-to-information
- Re-purposing of information and content
- Creation of new business opportunities

Even though not primarily addressed also the following sectors will benefit from the DIRECT-INFO solutions:

1. Government Sector. The needs of this sector range from the monitoring and analysis of broadcast content for PR and compliance purposes for government and parliaments, all the way through to military intelligence garnered from public sources. The DIRECT INFO solution can offer the following added value:

- Increase the number of monitored information source
- Efficient "focus in" to events of interest"
- Immediate distribution of relevant broadcast material allowing timely response

2. Financial Sector. For equity, foreign exchange and futures trading rooms the delivery of relevant information, derived through continuous monitoring of broadcast content, can provide the competitive edge that can make the difference between loss and profitability. Financial PR organizations constantly seek to enhance their access to 'financial intelligence' in order to stay one-step ahead of the market and enhance the value of their service. Investment bankers vying for business that can provide extremely lucrative transaction fees (Mergers & Acquisitions, Syndicated Finance, etc.) must have complete access to information that can help them identify/recommend suitable targets, assess risk and determine other challenges or competitive threats. The accuracy and extent of such information (presented in a 'pitch book') is critical in closing deals against tough competition.

DIRECT-INFO will allow amalgamation of content from various multimedia sources of relative events, and will provide true value to the above sectors.

4 PROJECT DATA

Starting date: 01/01/2004, duration 24 months

Project value: 4,072,879 Euro total budget (2,099,959 Euro funding), 356.5 person months total

Consortium: 8 partners from 5 European states, 3 R&D and 5 industrial

5 PARTNERS AND ROLES

JOANNEUM RESEARCH Forschungsgesellschaft mbH (A) (http://www.joanneum.at/iis), Project management and co-ordinating partner, Metadata management system based on MPEG-7 for storage of all metadata relevant for the A/V subsystem, infrastructure for content based queries in the area of video and still images, Content analysis modules in connection with generation of summaries for video content, and object recognition, provision of a content analysis framework for system integration

Nielsen Media Research (I) (http://www.nielsenmedia.it), End-User partner, test content provider, evaluation and testing of system, Exploitation partner for the integrated system

Softeco Sismat SpA (I) (http://www.softeco.it), Media acquisition component and media repository system for managing digital content (audio, video, text, still images), media explorer browsing and organizing content, Exploitation partner for the integrated system

Idioma ltd. (IL) (http://www.idiomasolutions.com), Acquisition of multiple-channel digital broadcast feed, A/V comparison algorithms and infrastructure for content identification, search and retrieval of A/V content, Development of speech-based 'topic detection' of video and audio content, System evaluation, Business model and exploitation partner for the integrated system

HS-Art Digital Service GmbH (A) (http://www.hs-art.com), System preparation and bundling and semantic block classification.

DFKI Deutsches Forschungszentrum für Künstliche Intelligenz (D) (http://www.dfki.de), Knowledge management component, data fusion of the text-subsystem and the A/V subsystem to semantically meaningful pieces to be used by the delivery system. Text content analysis modules → filtering of company relevant information

Fraunhofer IGD (D) (http://www.igd.fhg.de), Multimodal event modelling for videos, Combined A/V segmentation, Organisation and dissemination of an international workshop, Query & result tool for the scenario 1 "company image analysis" (what was positively mentioned, what negatively)

Promitheas Business Innovation Centre ltd. (CY) (http://www.promitheas.com), Business model, exploitation and dissemination manager

6 ACKNOWLEDGEMENTS

The R&D work carried out for the DIRECT-INFO project is partially funded under the 6th Framework Programme of the European Commission within the strategic objective "Semantic-based knowledge management systems" of the IST Workprogramme 2003 (IST FP6-506898). Further information on the project can be obtained from the public project website: http://www.direct-info.net