Towards a Workflow-based Design of Multimedia Annotation Systems

Cristian Hofmann, Dirk Burkhardt, Matthias Breyer, Kawa Nazemi, Christian Stab, Dieter W. Fellner
Fraunhofer Institute for Computer Graphics Research
Fraunhoferstrasse 5, 64283 Darmstadt, Germany
{cristian.hofmann, dirk.burkhardt, matthias.breyer, kawa.nazemi, christian.stab, d.fellner}@igd.fraunhofer.de

Abstract: Annotation techniques for multimedia contents have found their way into multiple areas of daily use as well as professional fields. A large number of research projects can be assigned to different specific subareas of digital annotation. Nevertheless, the annotation process, bringing out multiple workflows depending on different application scenarios, has not sufficiently been taken into consideration. A consideration of respective processes and workflows requires detailed knowledge about practices of digital multimedia annotation. In order to establish fundamental groundwork towards workflow-related research, this paper presents a comprehensive process model of multimedia annotation which results from a conducted empirical study. Furthermore, we provide a survey of the tasks that have to be accomplished by users and computing devices, tools and algorithms that are used to handle specific tasks, and types of data that are transferred between workflow steps. These aspects are assigned to the identified sub-processes of the model.

Introduction

As means of enriching digital content by additional information, annotation techniques have found their way into multiple areas of daily use. In the Web, several platforms can be identified that enable visitors to annotate multimedia content with metadata in order to organize such files in a structured form and facilitate later retrieval. To name but a few, respective popular platforms are Flickr.com, Youtube.com, or Delicious.com, which enable shared manual classification and bookmarking of contents of various media formats. In addition to these simple ways of applying annotations, more complex environments are utilized in various professional application fields. For various purposes and objectives, such as information retrieval, content analysis, or group communication, today's annotation environments support practices in computer-supported education, medicines, engineering, human science, sports science, e-Commerce, edutainment, or gaming. By analogy, beginning from the times of manual annotation of textual documents, research on annotation has long tradition and issued a large number of research projects that can be assigned to different specific subareas of annotation.

A lot of work has already been done concerning various forms of annotation of contents that are coded by differing media formats (Agosti et al., 2007). Previous research projects focus on one or only few aspects of digital annotation. However, the annotation process, as a whole operative unit which issues several different workflows, has not been sufficiently taken into consideration up to now (Agosti et al.,2007, Hagedorn et al.,2008, Hofmann et al.,2009a). The exploration of processes of digital multimedia annotation is a relevant research topic due to user-specific requirements and challenges of application design. First, from a user's point of view, problems can be identified with respect to the usage of annotations systems. Depending on the media format of the contents that are to be annotated and the objectives and purposes of the annotation, different functionalities, algorithms, etc. can be used to accomplish a large number of tasks. Commonly, the amount of available tools and services leads to multi-optional and complex user interfaces. Consequently, users struggle in understanding and learning how to use the annotation system's set of tools and have problems in recognizing their current state and the next steps within the process (Sliski et al., 2001). Furthermore, annotation can be a laborious and time-consuming work, especially when it is conducted in a "manual" way (Bertini et al., 2005). Second, from an application design point of view, relevant key features have been declared with respect to flexibility in order to support different ways of using annotations and integration of different services (Agosti et al.,2007, 14. Constantopoulos et al., 2004).

In order to take on the described problems and challenges to improve annotation workflows, it is indispensable initial groundwork to obtain detailed knowledge about the process of annotation in general. According to that, practices of annotation have to be investigated, i.e., it is to be clarified how people factual work with annotation environments. In that scope, three aspects have to be considered in order to support processes of annotation: *control flow*, *service application*, and *data flow*. Control flow specifies the order of tasks and/or subactivities. Thus, transitions between steps of the annotation process need to be supported. Different annotation-related services must be integrated in order to realize flexibility of an annotation framework to support various forms of using annotations. Depending on these services, also the transfer of (annotation) data between annotation services must be considered. Especially the two latter cases are essential for the specification of required uniform interfaces.

Based on these considerations, the main contribution of this paper is the presentation of a process model for digital multimedia annotation. The model elucidates the concrete processes taking place, splitting them up into sub-processes. Furthermore, we give a survey of tasks that have to be accomplished, types of annotation data and other information that are involved in these tasks, as well as tools and algorithms that are used to handle tasks and data. There, these several aspects are assigned to the indentified sub-processes of annotation. In section 2, related work concerning annotation variants and practices of annotation is illuminated. Section 3 gives a general overview of the types, viewpoints, and application areas of digital multimedia annotation. In section 4, an empirical study of annotation practices is presented and an explanation of the exact methodology is given. On the basis of that overview, an abstract model describing annotation processes is illustrated in section 5. The last section summarizes the results and explains the transfer of the presented model into a technical concept.

Related Work

As stated above, few research activities can be located with respect to annotation workflows. Hagedorn et al. (2008), Mikova and Janik (2006), and Seidel et al. (2005) give an overview of procedures of computer-supported video analysis and annotation conducted by expert groups. Pea and Hoffert (2007) deal with the topic video analysis from a further point of view and illustrate a basic idea of the video research workflow in the learning sciences. Users of a web-based collaborative text annotation system were interviewed in (Cadiz et al., 2000). At this, annotators and behaviors, the annotated documents, and the use of a specific notification system are investigated. Furthermore, factors influencing the usage of annotation systems are presented. Cox and Greenberg (2000) define design principles for systems that support collaborative interpretation of information by means of annotation. The results are based on identified key behaviors of people engaged in emergence situations and investigation of how shared communities collaborate over graphical interfaces. A formal model of annotation is presented by (Agosti & Ferro, 2007). Here, they illuminate the meaning of annotations, define a formal description of the temporal dimension, and illustrate how hyperspaces emerge through linking by means of annotations. As can be seen from these examples, identified work mostly refers to specific areas, structures, or usage scenarios of digital annotation. Nevertheless, annotation regarded as *process* still receives little attention.

Preliminary Studies on Annotation Systems and Practices

For model building purposes, preliminary studies were conducted. These included a feature-oriented analysis of ten annotation systems from different sub-domains of multimedia annotation, at which provided tools, functionalities, and approaches were extracted. Second, studies on annotation practices were held. Here, experts at five different German research or education institutes were interrogated by means of semi-structured interviews, and also observed while operating the respectively applied annotation software (IPN Kiel, KRMC Tübingen, TU Darmstadt Telecooperation Group, TU Darmstadt Institute of General Education Science, TU Darmstadt Institute for Sports Science). Moreover, related literature was surveyed, including existing workflow models, field reports, or proposals for procedure execution.

In the scope of the conducted analysis of different types of multimedia annotation systems, identified tools, functionalities, approaches, and types of generated data were assigned to the respective classes of tasks and activities. These classes were derived from an inductively developed category system, that is, categories were improved and modified at each analysis run by means of generalization. Additionally, in the context of the examination of annotation practices, different workflows of multimedia annotation were derived. By means of the summarization and generalization of these workflows, general phases and sub-processes, as well as sequential relations between these items were abstracted. The findings of both preliminary investigations were combined in order to construct the generic process model described in this chapter. Here, classes of tasks and activities obtained from the system analysis were assigned to tasks within phases and sub-processes of annotation identified in the conducted studies. In doing so, also indentified features and sequences were combined, so that the entirety of all imposed information could be summarized within a common model.

Generic Process Model of Multimedia Annotation

In order to consider annotation workflows, it is indispensable initial groundwork to obtain detailed knowledge about the process of annotation in general. As the main contribution of this paper, this section presents a process model for digital multimedia annotation. The model elucidates the concrete processes taking place, splitting them up into sub-processes. Furthermore, tasks that have to be accomplished by annotators and/or the application, tools and algorithms that are used to handle certain tasks, and general types of data that are transferred between tasks and services are assigned to the indentified sub-processes.

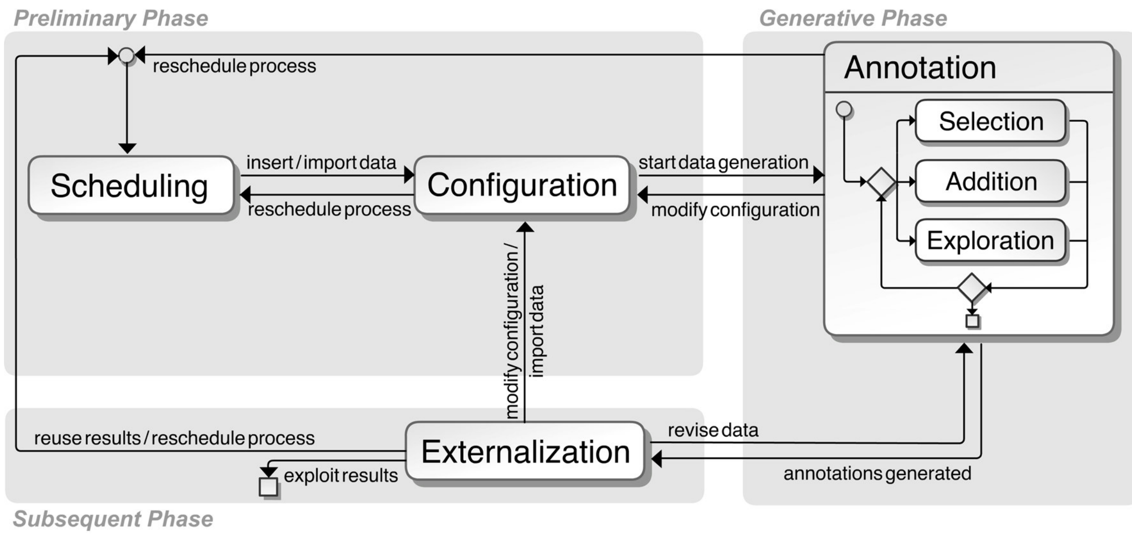

Figure 1: Generic Process Model of Multimedia Annotation.

As shown in Figure 1, the established process model not only refers to the actual activities of annotation, but also considers processes which take place before and after annotation, as well as without usage of the annotation system. Accordingly, the entire annotation process can be structured into the three superior phases *Preliminary*, *Generative*, and *Subsequent*. The initial Preliminary Phase comprises all activities that might need to be performed before users actually start annotating. Here, the sub-processes of *Scheduling* and *Configuration* are included, which refer to planning, defining strategies, gathering data, or preparation of the annotation environment. In addition to that, sub-processes can be detected that are to be assigned to the Generative Phase, in which *Annotation* is effectively done. Here, Annotation is subdivided into the partially processes *Selection* of validity areas and *Addition* of the supplementary data, which are accompanied by acts of *Exploration*. Furthermore, *Externalization* procedures refer to the further processing of already annotated data, passing into a Subsequent Phase that is not conducted by means of the annotation system and might imply switching to another project. In the following, the described sub-processes Scheduling, Configuration, Annotation, and Externalization are explained.

Scheduling

First, decisions have to be made concerning the data that has to be gathered, required, additional data, or – from a methodical view - a theoretical framework that might need to be built up. In a further step, the data has to be captured or gathered and re-edited and stored as suitable files (Brugman & Russel,2004, Hagedorn et al.,2008, Link,2006, Mikova & Janik,2006, Pea & Hoffert,2007., Seidel et al.,2005, Stahl et al.,2006).

Configuration

Before beginning with the annotation activities, content conceivably captured or gathered from specific databases or storage media has to be organized and stored, which may also require encoding and re-editing of the information to suitable (e.g. more granulated) files, depending on the given format or composition (Hagedorn et al.,2008, Mikova & Janik,2006, Pea & Hoffert,2007, Stahl et al.,2006).

In addition to that, specific project preferences can be adjusted and the graphical user interface may be customized (Brugman & Russel, 2004). Among these preferences count the involved users and user groups. For example, the general tasks of a project are assigned to predefined groups, and furthermore a group administrator is able to distribute the annotation tasks among the individual users (Sesink et al.,2007, Volkmer et al.,2005). Thus, users and groups are associated with specific roles that particularly include access rights and restrictions.

In content analysis use cases, specific classification systems, vocabularies or ontologies are applied or developed deductively (Mikova & Janik,2006, Seidel et al.,2005). These systems need to be linked or fed into the system. Since annotation processes are iterative and contain loops and re-entries to previous process states (Pea & Hoffert,2007, Stahl et al.,2006), predefined configurations of the used environment often need to be modified during a running annotation process.

Annotation

The next phases, selection, addition, and exploration, can be regarded as one operative unit. All conducted expert interviews revealed that no specific operative sequences can be identified referring to these "real" acts of annotating. Pea and Hoffert (2007) note on video analysis processes that activities of de-composition (segmentation, coding, categorization, and transcription) and re-composition (rating, interpretation, reflection, comparison, and collocation) are closely interrelated. They depict video annotation as a complex process that contains circular and recursive loops, in which the analyst alternately marks, transcribes and categorizes, analyzes and reflects, and needs to conduct searches. Marshall and Ruotolo (2002) performed a field study with respect of the annotation of digital libraries, reporting that acts of searching, reading and annotating are performed at the same time and can be done together with other activities, e.g. working with colleagues. Hence, selection, addition, and exploration, as higher-level categories, have to be considered related to each other, not able to assign specific fixed sequences. In the following, these three items are explicitly described.

Annotation: Selection

Annotators need to mark concrete contents of interest that annotations shall refer to, i.e., digital contents need first to be declared as "annotate-able". The simplest variant is marking a whole document such as a website. For example, the so-called social bookmarking platform Delicious.com enables users to classify websites via tags and bookmark them for later access by the annotator himself or an authorized group (Kolay & Dasdan, 2009). A second variant is the selection of elements contained within a document. Let us assume that a considered website consists of a text, various graphics, and a video. An annotator is able to mark each of these elements as independent units, and subsequently annotate them with different information (Finke,2005, Reif,2006). As a third variant, these single elements can be again subdivided into content-subsets, usually called segments. For example, annotations may be associated to a whole text, or to one or more special parts of the text (Constantopoulos et al., 2004). (Cadiz et al. (2000) describe such kind of annotations as "in-context-annotations". By means of the eMargo system, user communities are able to mark previously separated sections of a script and annotate shared textual contributions as part of a discourse (Sesink et al., 2007).

How exactly segments are defined depends on the media format of the original content and the specific media properties, as well as the purpose of annotation. In order to segment images or videos, annotators may use manual, semiautomatic, or automatic techniques (Finke,2005, Kipp,2008, Pea & Hoffert,2007). In the scope of video annotation purposes, existing solutions are the semiautomatic keyframe-method which is accompanied by linear interpolation, or automatic approaches like object or scene detection, scene-based event logging, or object of focus detection (Finke, 2005, Hagedorn, 2008, Pea & Hoffert, 2007). Corresponding to the time code in which an event takes place, users can define a point in time (single video frame) or a time interval (multiple following frames). Furthermore, an enrichment of this temporal information with spatial information is essential for almost every case of video "pointing" (Finke, 2005, Kipp, 2008).

The concrete segmentation methods or algorithms determine the degree of human activity. For example, in case of the aforementioned keyframe-method for video object tracking purposes, users have to define distinctive video frames with respect to a certain state of the respective object of interest. Thereto, they define wireframes enclosing the object of interest. Tracking is realized in a further step by automatic interpolation of subsequent wireframes (Finke, 2005). Depending on the duration of tracking and complexity of object motion, this approach can be by far more costly than applying automatic object tracking methods.

Segments may also be artifacts of collaborative work. In some of the investigated use cases, the segmentation task is partitioned and assigned to individual users or groups. For example, group A chunks the video according to a certain characteristic 1, group B seeks for characteristic 2, and so on. Furthermore, in content analysis use cases, users or groups may also work with different classification systems, vocabularies, ontologies, etc. (Stahl et al., 2006) report on held university courses, in which students have the task to create so-called "sensitive regions" (video segments) in collaborating groups. In doing so, each user obtain the access rights to modify segments of his own group.

Annotation: Addition

After marking the relevant documents, document subsets, or object segments, annotators continue adding annotations as additional information to these elements. Based on the classification of different forms of annotations *metadata*, *contents*, *dialog acts*, *hyperspace* (cf. Agosti & Ferro, 2007), different areas of addition tasks can be detected.

A relevant objective of annotation is the classification of given information resources or parts of it, in order to facilitate archiving and later retrieval (Bertini et al., 2005). In that context, Steimle et al. (2008) noticed that annotators use specific categorization schemes. In order to categorize contents, users may insert some kind of metadata which may range from simple features of the content to complex semantic information. For example, CoSCRIBE supports category tagging (classification using a predefined vocabulary) and free tagging (classification by selection of one or more arbitrary keywords) (Steimle et al., 2008). Video analysts at the IPN Kiel use Videograph to assign one predefined category to a current time interval via keyboard shortcuts (Rimmele,2004, Seidel et al.,2005). In particular, tagging is a useful method for collaboratively organizing large amounts of information (Baecker et al., 2007). In the described cases, tagging implies an active involvement of annotators, since they must select categories or enter keywords in a manual way. But there are also several semiautomatic or automatic approaches, e.g., (Bertini et al., 2005) presented a system that supports automatic extraction of video features in order to permit automatic summarization of sport videos. The latter example demonstrates that metadata can be simple such as classification keywords or a creation date, but also may describe more specific features like colors or patterns of images or “moving-images” (Bertini et al., 2005). A more complex approach is content description and categorization by semantic information which aims to enhance human-interpretable data with well-structured meaning in order to create computer-interpretable descriptions (Berners-Lee et al., 2001). Since annotated semantics comprise multiple information such as classes, instances, properties, or relations, more human interaction is required consisting of information assignment, presentation modification, and information re-editing (Nazemi et al., 2009). A further way of applying annotations as metadata is bookmarking (Cadiz et al.,2000, Marshall,1998). This enables users to mark and store relevant contents or portions for later inspection. In this scope, another relevant marking activity can be shared bookmarking with the objective of supporting collaborative work. In this case, users are enabled to see contents that seem to be relevant to other members of their group (Kolay & Dasdan, 2009). EMargo, for example, allows attaching of “flags” that serve as shared bookmarks with respect to sections of the script (Sesink et al., 2007). By means of CoSCRIBE, digital paper bookmarks are first digitalized and fed into the system, and afterwards presented by a collaborative visualization (Steimle et al., 2008).

In a further context, users may describe observed facts, e.g. behaviors and events of a video, objects within an image, sequences of an audio file, etc. In analysis cases, a task can be the transcription of verbal and nonverbal communication, which is often used in the context of communication or interaction analyses (Mikova & Janik, 2006). In order to explain, elucidate, interpret, or comment on the given contents, users need to give descriptions in a more free way than assigning metadata (Seidel et al., 2005). For that purpose, they can use tools allowing them to enter free textual annotations. Within the university courses attended by eMargo, students were allowed to annotate provided textual lesson scripts with questions or remarks on the course topics. These contributions are visible to fellow students and teachers (Sesink et al., 2007). At the KRMC Tübingen, video areas are annotated with textual interpretations and comments by means of the WebDIVER software. Indeed, also other types of media formats can be annotated for the same purposes. For example, the video platform Youtube.com enables not only textual commentary but also the aggregation of self-shot video comments. In the inspected video analysis use cases, the annotation phase includes interpretation, rating, and reflecting. These activities can be performed either qualitatively, e.g. in discussions, or quantitatively, by means of statistic methods provided by specialized software (Hagedorn et al.,2008, Pea & Hoffert,2007).

Like the selection task, the addition of annotations may be divided and distributed to different users and groups. For example, in the courses performed with eMargo, students are divided into groups which are assigned to specific tasks. For each task, alternately each student obtains the role “group administrator”. This user is responsible for the coordination of the collaborative elaborated tasks and unblocking of a final version for review purposes (Sesink et al., 2007). Thus, any annotator has got access to his or her group’s selections and annotations and is allowed to conduct modifications. In that case, annotated information becomes a shared contribution (Agosti & Ferro,2007, Constantopoulos et al.,2004, Zheng et al.,2006).

Communicational contributions constitute an essential kind of shared annotation with respect to collaboration. They enable the communication between co-annotators as well as the organization of common tasks. Organization and coordination of collaborative activities are essential in the context of co-writing or co-authoring with respect to granulated information exchanges between collaborators creating a shared document (Cadiz et al.,2000, Zheng et al.,2006). For that purpose, authors have to contribute of multiple categories: To-do items that determine tasks, summaries of edits which were performed by co-annotators, discussion about the content that can be subdivided into questions and general comments, and comments on exiting comments (Zheng et al., 2006). Furthermore, when users work separately, they need to discuss their annotations, conclusions, and the analysis process with other participants (synchronous and/or asynchronous) (Brugman & Russel, 2004) Current applications usually realize group communication by providing textual comments similar to web forums. Within a WebDIVER project, users at the KRMC are able to respond already submitted textual interpretations or comments. In doing so, a discourse is realized making use of a hierarchical construction of annotations. In the online-courses attended by eMargo, student groups have to distribute the set of tasks and determine a responsible group administrator for each single task by themselves. For that purpose, they use integrated commentary tools (Sesink et al., 2007). In the performed courses on hypermedia design, one of the first group tasks was to discuss the video segments that had to be specified, before actually entering that information into the used application (Stahl et al., 2006). Particularly in the context of consensual approaches applied in content analysis, discussion is a means of agreement and consistency of different annotators' results. Discussion often leads to a return to previous steps of the annotation process. In the case of video analyses performed by the IPN, several analysts first classify the same video segments independently and subsequently compare the individual results with their co-analysts. In case of disagreement, the data is reviewed and modified as a result of held discussions (Seidel et al., 2005). In addition to that, the interviewed experts at the IPN report on a training phase that is conducted before the actual video analysis on new video material (Seidel et al., 2005). This phase aims to develop basic analytic skills (Stahl et al., 2006), perform checks for objectivity and reliability, applying different annotator agreement measures (Hagedorn et al., 2008, Link, 2006, Mikova & Janik, 2006, Seidel et al., 2005), and to validate the deployed category system (Seidel et al.,2005). As a consequence, these checks lead to a return to the planning and configuration phases (Mikova & Janik, 2006, Seidel et al., 2005). In the end, the final results of the annotation project arise from iterative loops through the process, in which the data is continually modified and adjusted.

Annotations enable the creation of new relationships in the form of a link source and destination, connecting annotations with existing contents (Agosti & Ferro, 2007, Constantopoulos et al., 2004, Finke, 2005). Thus, content and annotations establish a inter-connecting hyperstructure (Agosti et al., 2007). That enables recipients to obtain multiple views and consequently new perspectives on existing information (Marshall & Brush, 2004, Stahl et al.,2006, Volkmer et al., 2005). In addition to that, further navigation and reception options, as well as enhanced search functionalities are revealed (Agosti et al., 2007, Finke, 2005). In order to create hyperstructures, annotators need to operate specific net-building tools. The concrete from of these services, as well as the associated way of interacting with the system, depends on the specific characteristics of the contents that are to be hyperlinked (Costa et al.,2002). For example, a website is referenced differently than unlike parts of it, or, a moving video object has to be linked different from an object within an image, since object motion is additionally tied to temporal information (Costa et al.,2002, Finke,2005). Beyond that, it has to be considered that also annotations can be destination of a hyperlink. Hence, annotations themselves can be "existing content" with specific properties.

Annotation: Exploration

Selection and information addition always go along with searching, browsing, and reception activities (Marshall & Ruotolo,2002, Pea & Hoffert,2007). First of all, surveying one's own data is required to properly perform digital annotation (Pea & Hoffert, 2007). As stated by Marshall and Ruotolo (2002), acts of searching, reading and annotating can also be done together activities such as working with colleagues. Consequently, especially in collaborative annotation situations, users also need to search for results of co-annotators, experts, or other sources (Hollender et al., 2008). The interview at the IPN revealed that novice annotators use already analyzed videos as training material and compare their own results with the results of their expert colleagues. CoSCRIBE enables annotators to compare own document structures with those of co-annotators (Steimle et al., 2008). Exploration of co-annotator's data also can be an issue in asynchronous collaborative projects which proceed over a long timeframe. After being absent, users may need to track the changes performed by other annotators involved in the project. In this context, they also need to browse chat or commentary histories (Baecker et al., 2007). Exploration also includes restructuring of the data representation. With regard to this, annotators are allowed to contrast relevant data with each other, or to hide less important information. Therefore, activities of searching, filtering, and sorting have to be performed. This is especially important when annotators are confronted with a large amount of annotations including those of co-annotators and other external resources (Costa et al., 2002, Hollender et al., 2008). In content analysis, pooling commonly classified information and making statistical comparisons are part of re-composition (Pea & Hoffert, 2007, Seidel et al., 2005). According to this, exploration also supports reflection. Thus, it facilitates the consideration of multiple views of the video where users are allowed to obtain perspectives on the contents beyond their subjective point of view (Marshall & Brush, 2004, Stahl et al., 2006, Volkmer et al., 2005).

Externalization

The externalization phase refers to two different aspects at the end of the annotation process: Publication of the process' results and export of data for the purpose of processing the data with external applications. As already mentioned, public annotations are treated by all participating users (Agosti & Ferro, 2007, Marshall, 1998). Normally, publication begins with editing and converting the data into several formats, and moves on to presenting this information by means of corresponding media (Pea & Hoffert, 2007). Published results can be used for demonstration purposes. Reporting on the CPV Video Study Physics, Mikova and Janik (2006) describe exclusive video sequences of filmed lessons that are shown in teacher-training as examples for "good teaching". Also databases of already annotated material can serve as digital resource for information retrieval in following annotation sessions, e.g., comparable to the preceding training phases conducted by the IPN, in which novices explore already analyzed video sequences (Seidel et al., 2005). Furthermore, a goal of annotation can be obtaining (mostly automatically) generated surveys and assemblies of similarly categorized content subsets, e.g. video summaries. Creating surveys, assemblies, and summaries is an elaborate and time-consuming work. Thus, much research is concerned with automatic approaches. (Bertini et al., 2005) present a system that supports automatic extraction of video features as basis for semantic annotation in order to permit the generation automatic summarizes of game highlights. EMargo provides functionalities to display only the contributions of one student in a single view. In doing so, teachers are facilitated by reviewing student's data in order to conduct evaluation of accomplishment (Sesink et al., 2007). Thus, if required, externalization can be means of supporting quality control for annotation. Furthermore, it is often necessary to export data for further processing by means of more specific applications. For example, experts at the IPN report on exporting data to various formats, such as tab-delimited text or transcription files in order to generate statistical calculations with SPSS (Seidel et al., 2005) Thus, further analytic activities can be executed with tools and services that are not provided by the video annotation application.

Summary and Transfer into a Technical Concept

In this paper, we pointed at a lack of research with respect to workflows of annotation. In that scope, we identified problems from a user's point of view (What is my current state within the workflow? What tools/UI components shall I use?), as well as requirements from a system design perspective (flexibility and service integration). As the main contribution of this paper, we presented a process model for multimedia annotation. In spite of the wide range of annotation, we demonstrated that a description of annotation processes from a general point of view can still be established. For that purpose, we summed up low-level items of the process such as human/system tasks or applied services to higher-level process categories. Thus, an appropriate degree of abstraction could be established. That achieved knowledge about processes of digital annotation is groundwork for a treatment of the problem areas of annotation workflow research. Figure 2 shows a summary of our results referring to assignment of phases, sub-processes, human/system tasks, existing approaches, and general forms of data which are generate within tasks.

In the scope of our research work, we developed an application which supports collaborative annotation of multimedia documents. The conceptualized framework model is described in (Hofmann & Fellner, 2010). Here, the established generic process model provided essential information which enabled the specification of interfaces and the internal control procedures within the system architecture. Thus, the system supports the integration of different annotation tools and services which can then be coordinated and executed in a specific sequential order. The latter aspect forms the basis for a visual-interactive concept which we regard as *Process-driven User Assistance*: Based on predefined process specifications, workflow control is realized. Additionally, users obtain information about the tasks to be accomplished and respective available tools. For that purpose, a specific visualization component is integrated. Furthermore, specific annotation tools and services are explicitly invoked and/or hidden, depending on the recent annotation task to be accomplished.

The mentioned multimedia annotation framework has been developed in the scope of the THESEUS program. THESEUS is a research program, initiated by the Federal Ministry of Economy and Technology (BMWi), to develop a new internet-based infrastructure in order to better use and utilize the knowledge available on the internet.

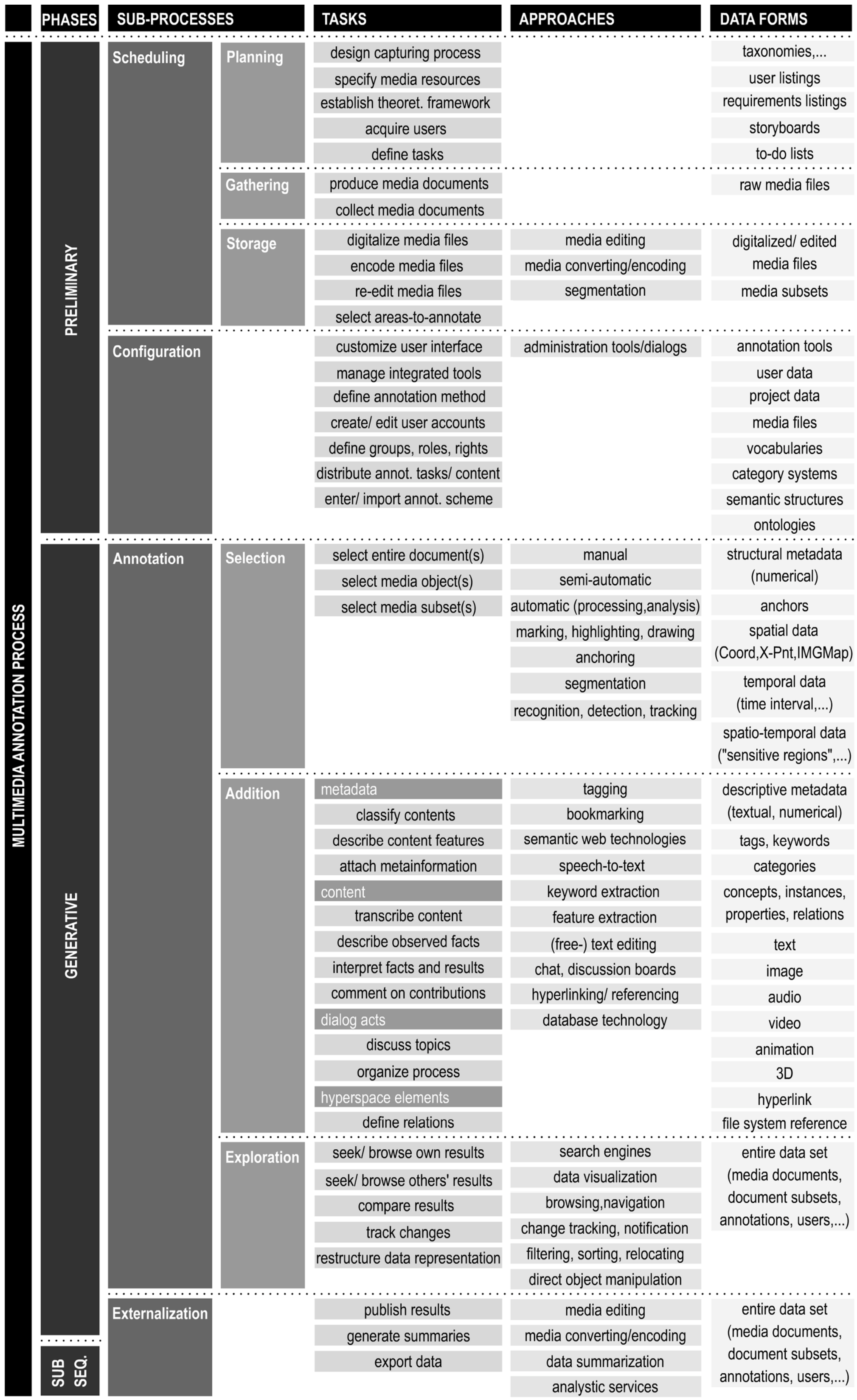

Figure 2: Tasks, Approaches, and Data Forms related to Processes of Multimedia Annotation.

References

Agosti, M., Bonfiglio-Dosio, G., & Ferro, N. (2007). A historical and contemporary study on annotations to derive key features for systems design. *International Journal on Digital Libraries* 8 (1), 1-19.

Agosti, M., & Ferro, N. (2007). A formal model of annotations of digital content. ACM Transactions on Information Systems 26 (1), 3.

Baecker, R. M., Fono, D., & Wolf, P. (2007). Toward a Video Collaboratory Video research in the learning sciences. *Video Research in the Learning Sciences*, Lawrence Erlbaum Associates, 461-478.

Bertini, M., Del Bimbo, A., Cucchiara, R., & Prati, A. (2004). Applications ii: Semantic video adaptation based on automatic annotation of sport videos. *Proceedings of MIR ´04*, ACM Press, 291-298.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 34-43.

Brugman, H., and Russel, A. (2004). Annotating multi-media / multi-modal resources with ELAN. *Proceedings of LREC 2004*, European Language Resources Association, 2065-2068.

Buneman, P., Khanna, S., Tajima, K., & Tan, W.-C. (2004). Archiving scientific data. *ACM Transactions on Database Systems* 29 (2), 2–42.

Cadiz, J., Gupta, A., & Grudin, J. (2000). Using Web Annotations for Asynchronous Collaboration around Documents. *Proceedings of CSCW '00*, ACM Press, 309-318.

Constantopoulos, P., Doerr, M., Theodoridou, M., & Tzobanakis, M. (2004). On information organization in annotation systems. *Lecture Notes in Computer Science*, Springer, 189-200.

Costa, M., Correia, N., & Guimarães, N. (2002). Annotations as multiple perspectives of video content. *Proceedings of MULTIMEDIA '02*, ACM Press, 283-286.

Cox, D., & Greenberg, S. (2000). Supporting Collaborative Interpretation in Distributed Groupware. *Proceedings of CSCW '00*, ACM Press, 289-298.

Finke, M. (2005). *Unterstützung des kooperativen Wissenserwerbs durch Hypervideo-Inhalte*. Dissertation Thesis, Technische Universität Darmstadt, Germany.

Hagedorn, J., Hailpern, J., & Karahalios, K.G. (2008). VCode and VData: illustrating a new framework for supporting the video annotation workflow. *Proceedings of AVI '08*, ACM Press, 317-321.

Hofmann, C., & Fellner, D.W. (2010). Supporting Collaborative Workflows of Digital Multimedia Annotation. *Proceedings of COOP'10*, Springer (accepted).

Hofmann, C., Hollender, N., & Fellner, D. W. (2009). Workflow-Based Architecture for Collaborative Video Annotation. *Lecture Notes in Computer Science*, Springer, 33-42.

Hofmann, C., Hollender, N., & Fellner, D. W. (2009). Task- and Process-related Design of Video Annotation Systems. *Proceedings of Mensch und Computer 2009*, Oldenbourg, 173-182.

Hollender, N., Hofmann, C., & Deneke, M. (2008). Principles to reduce extraneous load in web-based generative learning settings. *Workshop on Cognition and the Web*, 7-14.

Kipp, M. (2008). Spatiotemporal Coding in ANVIL. *Proceedings of LREC '08*, European Language Resources Association, 2042-2045.

Kolay, S. & Dasdan, A. (2009). The value of socially tagged urls for a search engine. *Proceedings of WWW '09*, ACM Press, 1203-1204.

Link, D. (2006). *Computervermittelte Kommunikation im Spitzensport.* Sportverlag Strauß, Köln.

Marshall, C. C. (1998). Toward an ecology of hypertext annotation. *Proceedings of HYPERTEXT '98*, ACM Press, 40-49.

Marshall, C. C., & Brush, A. B. (2004). Exploring the relationship between personal and public annotations. *Proceedings of JCDL '04*, ACM Press, 349-357.

Marshall, C. C., & Ruotolo, C. (2002). Reading-in-the-small: a study of reading on small form factor devices. *Proceedings of JCDL '02*. ACM Press (2002), 56-64.

Mikova, M., & Janik, T. (2006). Analysis of elements supporting health in physical education: Methodological approach of video study. *Neue Herausforderungen im Gesundheitsbereich an der Schule. Was kann der Sportunterricht dazu beitragen?*, MU. Brno, 248-260.

Nazemi, K., Ullmann, T.D., & Hornung, C. (2009). Engineering User Centered Interaction Systems for Semantic Visualizations. *Lecture Notes in Computer Science*, Springer, 126-134.

Pea, R., & Hoffert, E. (2007). Video workflow in the learning sciences: Prospects of emerging technologies for augmenting work practices. *Video Research in the Learning Sciences*, Lawrence Erlbaum Associates, 427-460.

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert. E. (2004). The DIVER™ Project: Interactive Digital Video Repurposing. *IEEE Multimedia* 11 (1), 54-61.

Reif, G. (2006) Semantic Annotation. *Semantic Web - Wege zur vernetzten Wissensgesellschaft* 5 (4), 405-418.

Rimmele, R. (2004). *Videograph*. IPN, Kiel, Germany.

Seidel, T., Prenzel, M., & Kobarg, M. (2005). How to run a video study. *Technical report of the IPN Video Study*, Waxmann, Münster, Germany.

Sesink, W., Göller, S., Rößling, G., & Hofmann, D. (2007). eMargo: eine digitale Randspalte zum Selbststudium (nicht nur) der Informatik. *Proceedings of DeLFI '07*, Logos-Verl., Berlin, Germany, 101-108.

Sliski, T. J., Billmers, M. P., Clarke, L. A., & Osterweil, L. J. (2001). An architecture for flexible, evolvable process-driven user-guidance environments. *Proceedings of ESEC/FSE-9*, ACM Press, 33-43.

Stahl, E., Finke, M., & Zahn, C. (2006). Knowledge Acquisition by Hypervideo Design: An Instructional Program for Univiversity Courses. *Journal of Educational Multimedia and Hypermedia* 15 (3), 285-302.

Steimle, J., Brdiczka, O., & Mühlhäuser, M. (2008). CoScribe: Using Paper for Collaborative Annotations in Lectures. *Proceedings of ICALT '08*, IEEE Computer Society, Washington, DC, 306-310.

Volkmer, T., Smith, J. R., & Natsev, A. (2005). A web-based system for collaborative annotation of large image and video collections: an evaluation and user study. *Proceedings of MULTIMEDIA '05*, ACM Press, 892-901.

Zheng, Q., Booth, K., & McGrenere, J. (2006). Co-authoring with structured annotations. *Proceedings of CHI '06*, ACM Press, 131-140.

Citation:

Hofmann, Cristian Erik; Burkhardt, Dirk; Breyer, Matthias; Nazemi, Kawa; Stab, Christian; Fellner, Dieter W.:
Towards a Workflow-Based Design of Multimedia Annotation Systems.
In: Association for the Advancement of Computing in Education (AACE):
Proceedings of ED-Media 2010 : World Conference on Educational Multimedia, Hypermedia & Telecommunications [online].
Chesapeake : AACE, 2010, pp. 1224-1233