A Novel Approach for Semantics-Enabled Search of Multimedia Documents on the Web

Lydia Weiland and Ansgar Scherp

University of Mannheim, Germany
{lydia,ansgar}@informatik.uni-mannheim.de

Abstract. We present an analysis of a large corpus of multimedia documents obtained from the web. From this corpus of documents, we have extracted the media assets and the relation information between the assets. In order to conduct our analysis, the assets and relations are represented using a formal ontology. The ontology not only allows for representing the structure of multimedia documents but also to connect with arbitrary background knowledge on the web. The ontology as well as the analysis serve as basis for implementing a novel search engine for multimedia documents on the web.

1 Introduction

Multimedia search on the web is limited to keyword search today. Search engine giants like Google simply just index the textual information encoded in the multimedia documents and the incoming and outgoing hyperlinks. Thus, Google squeezes structured multimedia documents to fit into its Page Rank model for hypertext.[1] In contrast, search for structured multimedia documents such as Silverlight presentations, Flash documents, and Adobe's Edge documents in the new W3C format HTML 5 is still very limited. Structured multimedia documents are composed of media assets like images, videos, audio, and text [6,17]. The multimedia document obtains its structure by organizing the media assets coherently in time, space, and interaction [6,17]. However, this information is not used for indexing and retrieving the content today. In addition, the arrangements of media assets as well as the media assets themselves exhibit certain semantics that is typically not explicitly encoded in the multimedia documents. This makes it hard to search for and within structured multimedia documents, which is of high benefit for a variety of reasons. Multimedia documents that can be better searched by making its media assets accessible for retrieval are better visible in the web. In addition, it enables for a better reuse of media assets in order to save costs and time, e. g., in large enterprises that professionally produce multimedia documents for e-learning, advertisement, or for creating professional websites.

In order to improve search in structured multimedia documents, we conduct an analysis of a large multimedia corpus. For the purpose of analysis, we represent

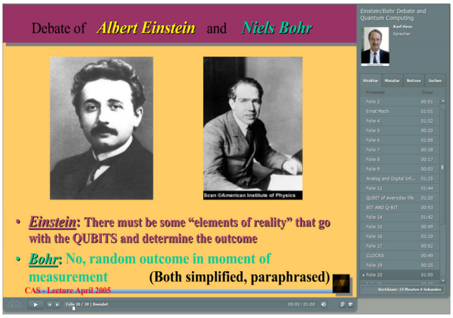
[1] http://support.google.com/webmasters/bin/answer.py?hl=en&answer=72746, access: 3/10/2013

C. Gurrin et al. (Eds.): MMM 2014, Part I, LNCS 8325, pp. 50–61, 2014.

the documents using a generic multimedia document ontology (M2DO). Besides representing media assets and their temporal, spatial, and interaction relations, the M2DO allows for a seamless integration of semantic annotations in form of background knowledge provided from the Linked Open Data (LOD) cloud.[2] On the LOD cloud, data is interlinked and provides machine readable semantics. The M2DO has been designed in terms of a backwards analysis of the existing models [18].

The remainder of the paper is organized as follows: In the following section, we present an illustrative scenario motivating the need for a multimedia document search engine. The related work is discussed in Section 3. In Section 4, we describe the requirements to our ontology, which is presented in Section 5. The results of our analysis are shown and discussed in Section 6, before we conclude the paper in the last section.

2 Scenario



Debate of *Albert Einstein* and *Niels Bohr*

Scan ©American Institute of Physics

- ***Einstein***: There must be some "elements of reality" that go with the QUBITS and determine the outcome
- ***Bohr***: No, random outcome in moment of measurement (Both simplified, paraphrased)

CAS - Lecture April 2005

Einstein/Bohr Debate and Quantum Computing

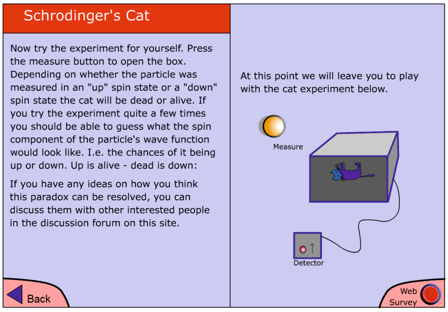
Karl Hess
Sprecher

Folientitel	Dauer
Folie 2	00:51
Ernst Mach	01:01
Folie 4	01:32
Folie 5	00:20
Folie 6	01:08
Folie 7	00:38
Folie 8	00:17
Folie 9	00:03
Analog and Digital Inf...	01:25
Folie 11	01:44
QUBIT of everyday life	01:20
BIT AND Q-BIT	00:53
Folie 14	01:42
Folie 15	00:49
Folie 16	02:10
Folie 17	00:52
CLOCKS	00:49
Folie 19	00:25
Folie 20	01:00

Restdauer: 14 Minuten 6 Sekunden

Folie 20 / 29 | Beendet 00:02 / 01:00

Fig. 1. Multimedia document with audio of a debate between Einstein and Bohr.[3]



Schrodinger's Cat

Now try the experiment for yourself. Press the measure button to open the box. Depending on whether the particle was measured in an "up" spin state or a "down" spin state the cat will be dead or alive. If you try the experiment quite a few times you should be able to guess what the spin component of the particle's wave function would look like. I.e. the chances of it being up or down. Up is alive - dead is down:

If you have any ideas on how you think this paradox can be resolved, you can discuss them with other interested people in the discussion forum on this site.

At this point we will leave you to play with the cat experiment below.

Fig. 2. Animation of Schroedinger's Cat.[4]

The scenario illustrates features of a search engine for multimedia documents. We consider the physics teacher Mr. Particle who is preparing his next lecture. He wants to get his pupils interested in quantum mechanics. Thus, he uses as introductory story the famous dispute of Albert Einstein and Niels Bohr at the Solvay physics conference back in 1930, where Einstein doubted the completeness of Bohr's quantum mechanics model. To stress his doubt, Einstein invented a thought experiment called 'photon level'. Mr. Particle wants to visualize the story and topic with an interactive multimedia presentation. Thus, he

[2] `http://www.w3.org/DesignIssues/LinkedData.html`

[3] Taken from: `http://nanohub.org/local/breeze/nt501/2005.04.01-Hess/viewer.swf`, last accessed: April 2013

[4] Taken from: `http://www.gilestv.com/tutorials/find1.swf`, last accessed: April 2013

collects appropriate content for reuse in his presentation. He searches for images of the two physicists. Mr. Particle formulates a query to find all multimedia documents, where images appear simultaneously with the terms 'Einstein' and 'Bohr' (Fig. 1). Subsequently, he specializes his query to find only multimedia documents that contain animations, videos, or serious games dealing with 'photon level'. Mr. Particle wants to explain an additional thought experiment, called 'Schroediger's cat' later in the lecture. To find multimedia documents for this experiment, he searches for documents being annotated with the category 'thought experiment'. The resulting documents should either contain embedded videos or interactive elements like the example in Fig. 2.

Overall, we can observe that Mr. Particle conducts queries along the different dimensions of multimedia documents, i.e., time, space, and interaction. In order to find multimedia presentations about persons like Einstein and Bohr and events such as the Solvey conference in 1930, he makes use of higher-level semantics associated with the multimedia documents in form of background knowledge such as the annotations 'thought experiment'. Such a query on the background knowledge allows to find documents that are annotated with the category 'thought experiment' or are annotated, e. g., with synonymous categories, subcategories, etc.

3 Related Work

Various *semantic multimedia retrieval systems* have been developed in the past like the MEMORAe project [15], where ontological knowledge is used for indexing and searching educational videos. Breaking the barrier of a single media modality, there are approaches for semantic cross-media search and retrieval like the semantic search engine Squiggle [8] for images and audio. However, the images and audios do not originate from a multimedia document and thus no spatial, temporal, or interaction relations are considered in Squiggle. The FLash Access and Management Environment (FLAME) [23] is an approach where Flash files where converted to XML in order to extract and index media assets, events, and interaction features. Although the authors recognize the importance of temporal relations between media objects, they are not considered in their work. The video search engine Yovisto [22] and work by Diemert et al. [10] allow for a semantic enrichment and search of audio-visual media content. As such they do not propose a model for representing rich, structured multimedia documents. In the area of *multimedia document models*, we find HTML 5 and Flash and abstract document models such as MM4U [18], ZyX [4], and AHM [13]. These models are targeted towards the presentation of multimedia content to the users. They are not designed for serving as internal representation model for a multimedia retrieval engine. In addition, they do not allow for a seamless integration of LOD background knowledge such as today's very popular Linked Open Data. An extensive overview is beyond the scope of this work and has been conducted earlier (see, e. g., [19,5]). In the area of *multimedia query algebras*, we find the MP7QF query language for audio-visual media content encoded in the MPEG-7 format [11]. The Unified Multimedia Query Algebra (UMQA) aims at integrating different features for multimedia querying such as content-based features, spatio-temporal relations, and traditional metadata [7]. The existing algebras either focus on spatial or temporal relations. The semantics of multimedia documents is typically not considered. Only, the query algebra EMMA [24] allows to state queries against media assets stored in Enhanced Multimedia Meta Objects (EMMOs) and the typed edges, which allow for modeling background knowledge, between EMMOs. However, there is no support for temporal, spatial, and interaction relations between media assets.

Overall, the work on multimedia retrieval is still at its beginning. So far, there is no sophisticated solution available that allows for representing and querying structured multimedia content along time, space, and interaction relations as well as background knowledge provided by the LOD cloud.

4 Requirements to a Semantic Multimedia Search Engine

From this scenario and the related work, we extract the requirements for a multimedia retrieval engine. First, the multimedia documents and its characteristics need to be represented. Multimedia documents are composed of different types of media assets. We have to be able to identify each occurrence of a media asset and assign it to its document [6]. These requirements enable a user to formulate a query like "Show me all audios within a multimedia document containing the keyword 'Solvay' (cf. Scenario)". Second, a multimedia retrieval engine needs to represent the three central relations for time, space and interaction [6,19]. These relations enable queries containing keywords as mentioned in the scenario (cf. Scenario) like 'simultaneously, after, below, on click, etc.'. Third, support for semantic annotations is required to support use of background knowledge.

5 Representing Multimedia Content with the M2DO

As prerequisite for a multimedia search engine, we first need to investigate the nature of multimedia documents on the web. To this end, we consider multimedia documents as graphs and represent them using a multimedia representation ontology. Please note that the existing multimedia document models (see discussion in Section 3) are not sufficient to this purpose as they are aimed as exchange format for the purpose of presenting the content to the users. In addition, the existing models do not support representing background knowledge. Thus, we have developed the Multimedia Document Ontology (M2DO) for the specific purpose of representing the media assets and structure of multimedia documents and its association(s) with background knowledge. The M2DO is designed such that it can be easily used in a multimedia retrieval engine. It is defined in OWL[5] and

[5] `http://www.w3.org/2001/sw/wiki/OWL`, last accessed: July 2013

axiomatized in Description Logics (DL) [2]. This has various advantages: First, it allows for checking consistency of our model due to the DL axioms. Second, existing ontologies can be easier reused. For the development of the M2DO, we made use of a scenario-based methodology, similar to the NeOn methodology[6]. We use the foundational ontology DOLCE+DnS Ultralight (DUL) [14] as basis for M2DO. DUL has proven being well suited for designing ontologies in various domains [20]. DUL makes use of ontology design patterns, allowing for a modularization of the knowledge it formalizes [21]. The M2DO is depicted in Fig. 3 and consists of three patterns. Classes taken from DUL are shown in white. Newly defined classes in M2DO have pale blue background. While the Media Asset Pattern and the Media Occurrence Relation Pattern are newly defined in the M2DO, the Media Annotation Pattern is extended from the Annotation Pattern of the M3O [16].

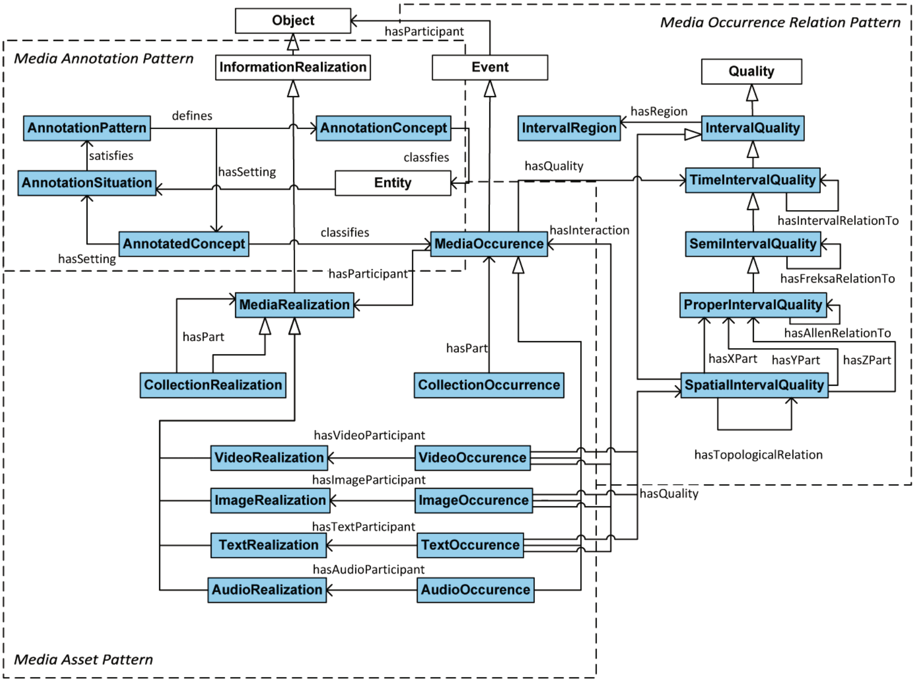


Fig. 3. M2DO Ontology

The *Media Asset Pattern* allows for representing the whole multimedia document and its assets. A Media Realization is an instance of a media type and represents a media asset. Media Occurrences are concrete occurrences of a specific

[6] http://www.neon-project.org, last accessed: July 2013

Media Realization. Thus, while one occurrence refers to exactly one realization, a realization can be associated to many occurrences.

The *Media Occurrence Relation Pattern* allows for modeling temporal, spatial, and interaction relations between media assets. The temporal relation is grouped into relations for semi-intervals and proper intervals by Freksa [12] and Allen [1], respectively. Within semi-intervals, either the starting point in time or the ending point in time is known. For proper intervals, both information is given. Spatial relations can be either topological [9], like disjoint, inside or covers, or rectangular [3]. The rectangular relations are based on the Allen calculus, where the 13 relations are pairwise connected to each other. The Cartesian product results in 169 possible combinations for spatial objects. The M2DO allows to represent interactions started by clicking on an image, video, or text occurrence. Interactions are defined using the property `m2do:hasInteraction`, where different types of interactions like a mouse click, hovering over an element, or hitting the enter key, are distinguished by specialization.

The *Media Annotation Pattern* allows for representing multimedia annotations such as the category "thought experiment" in the scenario. Annotations can be in principle any resources on the web, e. g., the Wikipedia article on thought experiments[7]. The Annotation Pattern of the M3O [16] serves as basis of this pattern. It has been connected to the M2DO via the concepts MediaRealization. With respect to our example in the scenario about thought experiments, the figure would have been connected to the AnnotatedConcept, because that is the asset which is annotated, and the AnnotationConcept is connected to the DBPedia-Link[8], because this is the annotation the asset gets.

The M2DO ontology has been created using the Protégé Ontology editor. We have checked its consistency with the FacT++ reasoner. The Ontology Pitfall Scanner (OOPS)[9] was used to detect errors in the ontology, which cannot be detected by a consistency check in order to remove missing inverse relationships or equivalent properties. The ontology together with its axiomatization in Description Logics is available as OWL file[10].

6 Analysis of a Large Corpus of Multimedia Documents

In 2012, we have crawled around 18.000 Flash files. From these, we have analyzed around 14.000 which we were able to extract with swfmill[11]. The Flash files versions range from 5 to 10 and sizes from 8 KB to 45 MB with an average of 484.75KB and a standard deviation of 1.7MB. For converting the binary flash files using swfmill an average duration of 4 seconds per file was needed.

[7] `http://en.wikipedia.org/wiki/Thought_experiment`, last accessed: April 2013

[8] `http://de.dbpedia.org/page/Kategorie:Gedankenexperiment`, last accessed: April 2013

[9] `www.oeg-upm.net/oops`, last accessed: July 2013

[10] `http://dws.informatik.uni-mannheim.de/fileadmin/lehrstuehle/ki/research/M2DO`

[11] `http://swfmill.org/`

Subsequently, we have analyzed the XML files to extract relevant information for a multimedia retrieval engine and represent the Flash file using our M2DO introduced above. The average duration for modeling a file with triples is 27 seconds, with a standard deviation of 483.

6.1 Basic Analysis of Flash File Characteristics

In order to provide insights about the nature of our data set, we have extracted the different media assets and their occurrences grouped by their media type. In Table 1, the averages and standard deviations for every media type is shown. The high standard deviation shows that there is not only a high difference in the sizes of the flash files, but also in the type of the content. Also the minimum and maximum amount of media occurrences (Table 2) per media asset grouped by type is shown. More specific characteristics of image assets within the flash files are shown in Tables 3. The assets are investigated for their widths and heights (in pixels) and their sizes (in KB).

Table 1. Mean, standard deviation, minimum, and maximum number of media assets per file

Amount of	μ	σ	min	max
Text Assets	10.78	30.87	0	276
Image Assets	6.57	15.69	0	173
Video Assets	0.08	0.27	0	1
Audio Assets	0.79	3.43	0	32

Table 2. Occurrences per media asset with mean, standard deviation, minimum, and maximum values

Amount of	μ	σ	max
Text Occurrences	0.78	1.00	5
Image Occurrences	1.25	1.61	8
Video Occurrences	0.12	0.44	3
Audio Occurrences	0.24	0.56	3

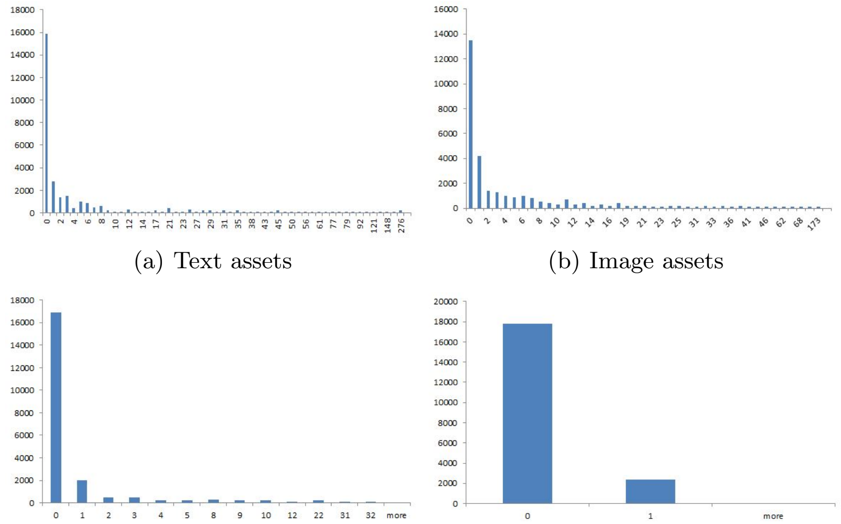
We have checked the consistency of 10% of the Flash files, which were randomly chosen from our dataset. Arguing with the assumption that 10% of 18.000 flash files covers a representative amount for all of the different types. There are around 2000 axioms per file with 1% logical axioms. The latter are those of equivalences or properties like transitivity. 30% of the axioms are inferred, e. g., the

Table 3. Characteristics of image asset with mean, standard deviation, minimum and maximum values

Characteristics of Image Occurrences	μ	σ	min	max
Width	297.5	238.73	9	1602
Height	203.1	149.07	4	1202
Size	225.97	343.61	4.59	2292.00

property `hasPart` which is defined in the ontology will automatically be added by its inverse, in this case the `isPart` relation. The consistency checks have been done as preparation to use the data in the multimedia retrieval engine. It allows to identify errors when transforming the XML structure of the flash content to an OWL graph based on the axiomatization of the ontology.

In Figure 4, the distribution of the media assets can be seen, which is given as a power-law distribution. This means that there are a lot of files with a few assets and only a few files with a huge amount of assets.



(a) Text assets

(b) Image assets

(c) Audio assets

(d) Video assets

Fig. 4. Distribution of assets of the four different media types, x-axis: Amount of asset, y-axis: Amount of files

6.2 Approaches for Representing Multimedia Documents

We investigate different strategies for representing multimedia documents using our M2DO ontology for the purpose of using this representation in a multimedia retrieval engine. The initial investigation focuses on the representation of time and space. The number of comparisons to be made between the media occurrences increases quadratic with the total number of assets in a document. If all relationships of every occurrence to all others is computed, a large data storage is required. For reasons of comparison calculating every relation is considered in variant (a). A lot of triples do not only lead to storage intensive use, but also to time-consuming queries. Therefore other approaches for representing multimedia documents without restricting or changing our ontology are investigated. We are exploring different strategies: We have decided to investigate a variant (b) that exploits the transitivity of properties such as `before` and `contains`. This idea is also considered in variant (c), where the transitivity of more complex relationships are exploited. Our last approach (d) is derived from the idea that the spatial relations in direction of the z-axis will not be needed. Following down the (negative) z-axis, one can only recognize if some object is behind some other. The user will not be able to distinguish if an object is covered by another objects that hold a before or a meets relation and that are smaller than the original object.

(a) Computing all possible relations In the first approach, we compute all relations between all media occurrences included in a multimedia document. While this approach is easy to implement, it is also the most data-intensive approach. It requires $\frac{n*(n-1)}{2}$ comparisons to be considered when building up the M2DO for n media occurrences. Only inverse relations are not computed like the relation A after B when already the B before A relation is stored. In Table 4, the amount of resulting relations is shown.

(b) Exploiting Transitivity on Temporal and Spatial Relations Using the axioms defined in the M2DO enables us to compute fewer relations when exploiting transitivity of temporal relations and spatial relations. For example, if A before B, B before C, and A before C, we can store the triples without the relation A before C, because it can be derived. This approach saves us one of three triples, if transitivity is given. The reduction of triples can be seen in Table 4.

(c) Exploiting Extended Transitivity on Temporal and Spatial Relations The transitivity property can be extended so that it is not only used for elements which are connected over the same relation, but for all of the 13 Allen relations which have logical similarities, in a sense that transitivity is given. For example, if A during B, B meets C, and A before C, we can again leave out the relation of A and C, because it can be derived by B meets C, which means that C is directly after B, and A is before C, as A is contained in B, which means that A is finished earlier than B. All possible combinations can be seen in [1].

(d) Reducing the Spatial Depth The x- and y-axis expand from 0,0 to the end of the screen. Thus, a user is able to see if two objects are next to each other or if the objects have some distance between them. That is why the spatial relations are important. In contrast, the z-axis has its extension in the direction to the user. This geometry leads to the fact that a user can only distinguish between three situations regarding the z-axis: A before B, A after B, and A equal B. Computing all spatial relations in direction of the z-axis would not give any benefit to the user, as he can only distinguish the three cases. Reducing the z-axis relation to the three cases reduces further the number of triples. In addition, before and after is a pair, which can be derived from each other. Thus, we only need to compute before and equal relations for the z-axis. Thus, only a few relations are needed to represent the whole z-axis relations.

In Table 4, the results for the different approaches described in the paragraphs *(a)* to *(d)* are shown, grouped by temporal and spatial relations. Comparing rows *(a)* and *(b)* shows a high improvement regarding storage capacity using the transitive property (mean of `(a)`: 10,016.74 and of `(b)`: 1,773.2). Comparing rows *(b)* to *(c)* shows a low improvement. As approach *(c)* leads to higher calculation costs while extracting and retrieving the data, the improvement of storing and the costs for conducting the calculations need to be compared. *(d)* has got benefits towards the other approaches such that the representation model can be computed efficiently (row Modeling Time in the table).

Table 4. Analysis results for different approaches `(a)-(d)`

Representation Approaches	μ	σ	min	max
`(a)` All relations				
Temporal Relations	10,016.74	95,454.66	0	1,209,364
Spatial Relations	645.33	7,238.10	0	126,318
Modeling Time	4.6	96.8	0	704.71
`(b)` Exploiting transitivity				
Temporal Relations	1,773.2	1,623.67	0	58,247
Spatial Relations	55.38	390.47	0	6,468
Modeling Time	7.71	558.3	0	2746.92
`(c)` Exploiting extended transitivity				
Temporal Relations	1,626.67	1,607.49	0	50,264
Spatial Relations	50.52	386.70	0	6,390
Modeling Time	12.5	648.62	0	3836.05
`(b)` Reducing Depth				
Temporal Relations	10,016.74	95,454.66	0	1,209,364
Spatial z-Relation	8.13	23.12	0	214
Modeling Time	3.24	41.79	0	296.44

7 Conclusion

We have shown with our M2DO a first step towards a whole multimedia retrieval engine. Implementing different approaches for representing the structural information of multimedia documents in an efficient way helps to improve the storage and query of documents in later steps. Taking various flash files into account, we can demonstrate that our approach is flexible w.r.t. the complexity of multimedia documents. In our future work, we will also consider interaction relations. In addition, the combination of reducing the z-axis information with the use of transitivity feature will be evaluated. We will develop a prototype using approaches `(c)` and `(d)` and conduct extensive evaluations.

Acknowledgments. This research is supported by the SeMuDocs project cofunded by the Ministry for Science, Research, and Arts, Baden-Wuerttemberg, Germany. We would like to thank D. Schmeiß, D. Mies, N. Kurz, and T. Thesing who implemented an early version of the multimedia extraction and representation engine.

References

1. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM 26(11), 832–843 (1983)
2. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)
3. Balbiani, P., Condotta, J., del Cerro, L.F.: A model for reasoning about bidimensional temporal relations. In: Principles of Knowledge Representation and Reasoning, pp. 124–130. Morgan Kaufmann (1998)
4. Boll, S., Klas, W.: ZYX - A Multimedia Document Model for Reuse and Adaptation. IEEE Trans. on Knowledge and Data Engineering 13(3), 361–382 (2001)
5. Boll, S., Klas, W., Westermann, U.: Multimedia document models - sealed fate or setting out for new shores? In: Int. Conf. on Multimedia Computing and Systems, p. 9604. IEEE, Washington, DC (1999)
6. Candan, K.S., Sapino, M.L.: Data Management for Multimedia Retrieval. Cambridge University Press, New York (2010)
7. Cao, Z., Wu, Z., Wang, Y.: UMQL: A unified multimedia query language. In: Signal-Image Technologies and Internet-Based System, pp. 109–115. IEEE (2007)
8. Celino, I., Valle, E.D., Cerizza, D., Turati, A.: Squiggle: a semantic search engine for indexing and retrieval of multimedia content. In: Semantic Enhanced Multimedia Presentation Systems. CEUR-WS.org (2006)
9. Clementini, E., Sharma, J., Egenhofer, M.J.: Modelling topological spatial relations: Strategies for query processing. Computers & Graphics 18, 815–822 (1994)
10. Diemert, B., Abel, M.-H., Moulin, C.: Semantic audiovisual asset model. Multimedia Tools and Applications 63, 663–690 (2013)
11. Döller, M., Kosch, H., Wolf, I., Gruhne, M.: Towards an mpeg-7 query language. In: Damiani, E., Yetongnon, K., Chbeir, R., Dipanda, A. (eds.) SITIS 2006. LNCS, vol. 4879, pp. 10–21. Springer, Heidelberg (2009)

12. Freksa, C.: Temporal reasoning based on semi-intervals. Artif. Intell. 54(1-2), 199–227 (1992)
13. Hardman, L., Bulterman, D.C.A., van Rossum, G.: The Amsterdam hypermedia model: adding time and context to the Dexter model. Communications of the ACM 37(2) (February 1994)
14. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Schneider, L.: The wonderweb library of foundational ontologies and the dolce ontology. Technical report, ISTC-CNR (2007)
15. Merzougui, G., Djoudi, M., Behaz, A.: Conception and use of ontologies for indexing and searching by semantic contents of video courses. International Journal of Computer Science Issues 8(3) (2012)
16. Saathoff, C., Scherp, A.: M3o: The multimedia metadata ontology. In: Proceedings of the Workshop on Semantic Multimedia Database Technologies, 10th International Workshop of the Multimedia Metadata Community (SeMuDaTe 2009) (2009)
17. Scherp, A.: Authoring of Multimedia Content: A Survey of 20 Years of Research. In: Semantic Multimedia Analysis and Processing. CRC Press (2013)
18. Scherp, A., Boll, S.: MM4U - A framework for creating personalized multimedia content. In: Managing Multimedia Semantics. Idea Publishing (2005)
19. Scherp, A., Boll, S.: Paving the Last Mile for Multi-Channel Multimedia Presentation Generation. In: Chen, Y.-P.P. (ed.) Proc. of the 11th Int. Conf. on Multimedia Modeling, Melbourne, Australia, pp. 190–197. IEEE (January 2005)
20. Scherp, A., Saathoff, C., Franz, T., Staab, S.: Designing core ontologies. Applied Ontology 6, 177–221 (2011)
21. Suarez-Figueroa, M., Gomez-Perez, A., Motta, E., Gangemi, A.: Ontology Engineering in a Networked World. Springer (2012)
22. Waitelonis, J., Sack, H.: Towards exploratory video search using linked data. In: Multimedia Tools and Applications, pp. 1–28 (2011)
23. Yang, J., Li, Q., Wenyin, L., Zhuang, Y.: Searching for flash movies on the web: A content and context based framework. World Wide Web Journal (September 2005)
24. Zillner, S., Westermann, U., Winiwarter, W.: EMMA – A query algebra for enhanced multimedia meta objects. In: Meersman, R. (ed.) OTM 2004. LNCS, vol. 3291, pp. 1030–1049. Springer, Heidelberg (2004)