

**12th International Workshop of the
Multimedia Metadata Community**

Workshop on
Semantic Multimedia
Database Technologies

co-located with SAMT2010

Saarbrücken, Germany

December 2nd, 2010

© 2010 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. Re-publication of material from this volume requires permission by the copyright owners.

For further information please contact us:

- **Department of Distributed Information Systems**

University of Passau
Prof. Dr. H. Kosch
Innstraße 43
94032 Passau
Germany
Tel: +49-851-509-3061
Fax: +49-851-509-3062
e-mail: smdt2010@easychair.org

- **Informatik 5**

RWTH Aachen
Dr. R. Klamma
Ahornstr. 55
52056 Aachen
Germany
Tel: +49-241-80-215-13
Fax: +49-241-80-223-21

- **Institute for Information Technology**

Klagenfurt University
Dr. Mathias Lux
Universitätsstraße 65-67
9020 Klagenfurt
Austria
Tel: +43-463-2700-3615
Fax: +43-463-2700-99-3615

- **Department D5: Databases and Information Systems**

Max-Planck-Institut für Informatik
Dr. M. Spaniol
Campus E1 4
66123 Saarbrücken
Germany
Tel: +49-681-9325-525
Fax: +49-681-9325-599

Technical Programme Committee

Anna Carreras (Universitat Politècnica de Catalunya, Spain)
Ansgar Scherp (University of Koblenz-Landau, Germany)
Antonio Penta (University of Naples Federico II, Italy)
Baltasar Fernández-Manjón (Complutense University of Madrid, Madrid, Spain)
Bill Grosky (University of Michigan, USA)
Christian Guetl (Graz University of Technology, Austria)
Christian Timmerer (Alpen-Adria-University Klagenfurt, Austria)
Chris Poppe (Ghent University - IBBT, Belgium)
Dominik Renzel (RWTH Aachen University, Aachen, Germany)
Francois Bry (LMU, University of Munich, Germany)
Giuseppe Amato (ISTI Pisa, Italy)
Jaime Delgado (Universitat Politècnica de Catalunya, Spain)
Laszlo Böszörményi (Klagenfurt University, Austria)
Lionel Brunie (INSA de Lyon, France)
Markus Strohmaier (Know Center, Graz, Austria)
Oge Marques (Florida Atlantic University, USA)
Prateek Jain (Wright State University, USA)
Richard Chbeir (Bourgogne University, France)
Savvas Chatzichristofis (Democritus University of Thrace, Greece)
Thierry Delot (University of Valenciennes, France)
Timothy Shih (NTUE, Taiwan)
Timo Ojala (University of Oulu, Finland)
Tobias Bürger (Salzburg Research, Austria)
Vincent Charvillat (ENSEEIH, France)
Vincent Oria (NJIT, USA)
Werner Bailer (Joanneum Research, Graz, Austria)
Yiwei Cao (RWTH Aachen University, Aachen, Germany)
Yu Cao (California State University at Fresno, USA)

Table of Contents

1	Introduction.....	5
2	“Video Retrieval from Few Examples Using Ontology and Rough Set Theory” by Kimiaki Shirahama and Kuniaki Uehara	6
3	“Collaborative Fake Media Detection in a Trust-Aware Real-Time Distribution Network” by Dominik Renzel, Khaled Rashed and Ralf Klamma	18
4	“Towards a User-aware Enrichment of Multimedia Metadata” by Ana-Maria Manzat, Romulus Grigoras and Florence Sedes.....	30
5	“A Model of Relevance for Reuse-Driven Media Retrieval” by Tobias Bürger	42
6	“Adapting Smart Graphics Behaviour to Users Characteristics” by Christophe Piombo, Vincent Charvillat and Romulus Grigoras	56
7	“Visualizing Mapping of Metadata Properties” by Martin Höffernig, Wolfgang Weiss and Werner Bailer	64

1 Introduction

We have the pleasure to organize the 12th Workshop of the Multimedia Metadata Community¹. This is the second workshop focusing on semantic multimedia database technologies and is held in conjunction with the 5th International Conference on Semantic and Digital Media Technologies (SAMT 2010), December 1-3, Saarbrücken, Germany.

Ontology-based systems have been developed to structure content and support knowledge retrieval and management. Semantic multimedia data processing and indexing in ontology-based systems is usually done in several steps. One starts by enriching multimedia metadata with additional semantic information (possibly obtained by methods for bridging the semantic gap). Then, in order to structure data, a localized and domain specific ontology becomes necessary since the data has to be interpreted domain-specifically. The annotations are stored in an ontology management system where they are kept for further processing. In this scope, Semantic Database Technologies are now applied to ensure reliable and secure access, efficient search, and effective storage and distribution for both multimedia metadata and data. Their services can be used to adapt multimedia to a given context based on multimedia metadata or even ontology information. Services automate cumbersome multimedia processing steps and enable ubiquitous intelligent adaptation. Both, database and automation support facilitate the ubiquitous use of multimedia in advanced applications.

We accepted five full papers and one demonstration paper. Our thanks go to the reviewers, who provided timely and thorough reviews. Their suggestions allowed authors to improve their contributions.

Our grateful thanks also go to the organizers of SAMT 2010. Their logistic support has been essential to the organization of our workshop. We wish a productive and enriching workshop and an excellent stay in Saarbrücken.

Your workshop co-chairs,
Harald Kosch and Florian Stegmaier, University of Passau, Germany
Ralf Klamma, RTWH Aachen, Germany
Matthias Lux, University of Klagenfurt, Austria
Marc Spaniol, Max-Planck-Institut für Informatik, Germany

¹ <http://www.multimedia-metadata.info>

Video Retrieval from Few Examples Using Ontology and Rough Set Theory

Kimiaki Shirahama¹ and Kuniaki Uehara²

1. Graduate School of Economics, Kobe University,
2-1, Rokkodai, Nada, Kobe, 657-8501, Japan
2. Graduate School of System Informatics, Kobe University,
1-1, Rokkodai, Nada, Kobe, 657-8501, Japan
shirahama@econ.kobe-u.ac.jp, uehara@kobe-u.ac.jp

Abstract. In query-by-example approach, a user can only provide a small number of example shots to represent a query. In contrast, depending on camera techniques and setting, relevant shots to the query are characterized by significantly different features. Thus, we develop a video retrieval method which can retrieve a large variety of relevant shots only from a small number of example shots. But, it is difficult to build an accurate retrieval model only from a small number of example shots. Consequently, the retrieval result includes many shots which are clearly irrelevant to the query. So, we construct an ontology as a knowledge base for incorporating object recognition results into our method. Our ontology is used to select concepts related to the query. By referring to recognition results of objects corresponding to selected concepts, we filter out clearly irrelevant shots. In addition, we estimate a parameter of a retrieval model based on the correlation between selected concepts and shots retrieved by the model. Furthermore, to retrieve a variety relevant shots characterized by different features, we use “rough set theory” to extract multiple classification rules for identifying example shots. That is, each rule is specialized to retrieve relevant shots characterized by certain features. Experimental results on TRECVID 2009 video data validate the effectiveness of our method.

1 Introduction

Recently, there is a great demand to develop a video retrieval method, which can efficiently retrieve interesting shots from a large amount of videos. In this paper, we develop a method based on “Query-By-Example (QBE)” approach, where a user represents a query by providing example shots. Then, QBE retrieves shots similar to example shots in terms of color, edge, motion, and so on. We consider QBE as very effective, because a query is represented by features in example shots without the ambiguity of semantic contents. In addition, QBE can retrieve any interesting shots as long as users can provide example shots.

However, QBE is challenging because in shots with similar features, semantic contents are not necessarily similar to each other. For example, when *Ex. 1* in

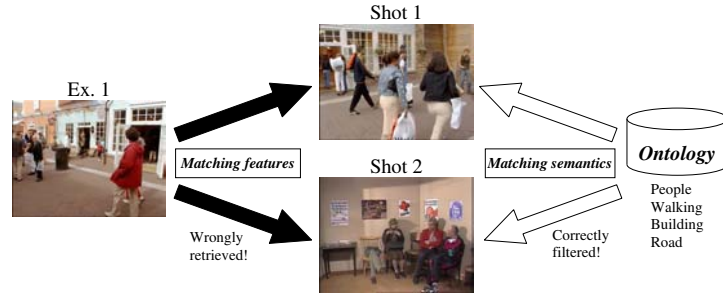


Fig. 1. Example of QBE using an ontology for the query “people walk in a street”.

Fig. 1 is provided as an example shot for the query “people walk in a street”, in addition to *Shot 1*, *Shot 2* is wrongly retrieved. The reason is that both of *Ex. 1* and *Shot 2* have red-colored and ocher-colored regions. Also, instead of rectangular buildings and windows in *Ex. 1*, rectangular posters are put up on the wall in *Shot 2*. Like this, one drawback of QBE is that it ignores semantic contents. Thus, we incorporate an ontology as a knowledge base into QBE.

We consider that an ontology is especially important for alleviating a lack of example shots in QBE. Usually, a user can only provide a small number of example shots (at most 10). In contrast, we represent each shot by using very high-dimensional features. For example, a popular SIFT feature leads to a shot representation with more than 1000 dimensions, where each dimension represents the frequency of a local edge shape (i.e. visual word). Generally, as the number of feature dimensions increases, the number of example shots needed to construct a well generalized retrieval model exponentially increases [1]. This means that the statistical information of features in a small number of example shots is not reliable. So, a retrieval model tends to be overfit to feature dimensions which are very specific to example shots but ineffective for characterizing relevant shots. For example, in Fig. 2, if *Ex. 1*, *Ex. 2* and *Ex. 3* are provided as example shots, the retrieval model is overfit to feature dimensions which characterize a small number of edges in the upper part (i.e. sky regions). As a result, it retrieves *Shot 1*, *Shot 2* and *Shot 3* which are clearly irrelevant to the query.

In order to filter out clearly irrelevant shots, we develop an ontology for utilizing object recognition results. Fig. 2 shows recognition results for three objects, *Building*, *Cityspace* and *Person*. Here, one shot is represented as a vector of recognition scores, each of which represents the presence of an object. In Fig. 2, we can see that *Building* and *Cityspace* are likely to appear in example shots while they are unlikely to appear in the other shots. Recently, researchers use object recognition results in video retrieval ¹. For example, research groups in City

¹ Objects are frequently called “concepts” in the field of video retrieval. But, some readers may confuse them with concepts which are hierarchically organized in an ontology. So, we use the term “concept” only when it constitutes an ontology.





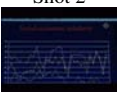
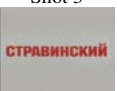
	Ex. 1	Ex. 2	Ex. 3	<i>Overfitting!</i> → <i>Filtered by ontology</i>		
				Shot 1	Shot 2	Shot 3
Building:	2.5	2.2	1.2	 -0.3	 -1.0	 -2.3
Cityspace:	1.1	2.6	1.5	-0.5	-1.5	-1.2
Person:	-1.0	-0.5	0.2	2.0	-1.3	-0.8

Fig. 2. An example of an overfit retrieval result for the event “tall buildings are shown”.

University of Hong Kong [3] and University of Amsterdam [2] build classifiers for recognizing 374 and 64 objects, respectively. In particular, such classifiers are built by using a large amount of training data (e.g. 61,901 shots in [3] and more than 10,000 shots in [2]). Thereby, objects can be robustly recognized independently of sizes, positions and directions on the screen. The effectiveness of using object recognition results is proved in TRECVID, which is a famous annual international workshop on video retrieval [4].

To utilize object recognition results, we port objects into concepts in an ontology. Specifically, we define a hierarchical structure of concepts and concept properties. Thereby, we can select concepts related to the query, and examine recognition scores of objects corresponding to selected concepts. For example, in Fig. 2, if *Building* and *Cityspace* are selected, *Shot 1*, *Shot 2* and *Shot 3* can be filtered out due to low recognition scores for *Building* and *Cityspace*. Also, filtering irrelevant shots reduces the computation time. Furthermore, we introduce a method for building an accurate retrieval model based on the correlation between concepts selected for a query and shots retrieved by the model. Note that we are not given the label of a shot (i.e. relevant or irrelevant), but given object recognition scores. Thus, we assume that an accurate retrieval model preferentially retrieves shots which have high recognition scores of objects, corresponding to concepts related to the query.

We address another important problem in QBE, where even for the same query, relevant shots are taken in many different shooting environments. As can be seen from example shots in Fig. 2, shapes of buildings and regions where they are shown are significantly different from each other. Additionally, in each of *Ex. 2* and *Ex. 3*, a road is shown while it is not shown in *Ex. 1*. So, shots relevant to the query are characterized by significantly different features. Regarding this, typical QBE methods only use one example and retrieve shots similar to it [5, 6]. As a result, many relevant shots are inevitably missed. Compared to this, we use multiple example shots and “Rough Set Theory (RST)” which is a set-theoretic classification method for extracting rough descriptions of a class from imprecise (or noisy) data [7]. By using RST, we can extract multiple classification rules which can correctly identify different subsets of example shots. Thereby, we can retrieve a variety of relevant shots where each classification rule is specialized to retrieve a portion of relevant shots characterized by certain features.

2 Related Works

2.1 Concept selection for Video Retrieval

The most popular ontology for video retrieval is “Large-Scale Concept Ontology for Multimedia (LSCOM) [8]”. It targets at broadcast news videos and defines a standardized set of 1,000 concepts. But, LSCOM just provides a list of concepts where no concept relation or structure is defined. So, many researchers explore how to appropriately select LSCOM concepts for a query.

Existing concept selection approaches can be roughly classified into three types, manual, text-based and visual-based selections. In manual concept selection, users manually select concepts related to a query [9]. But, different users select significantly different concepts for the same query. Specifically, [9] conducted an experiment where 12 subjects are asked to judge whether a concept is related to a query. As a result, only 13% of total 7,656 judgements are the same among all subjects. In text-based concept selection, WordNet is frequently used where words in the text description of a query are expanded based on synonyms, hypernyms and hyponyms [2, 3]. Then, concepts corresponding to expanded words are selected. But, WordNet only defines lexical relations among concepts, and does not define spatial and temporal relations among concepts. For example, from WordNet, we cannot know that *Building* and *Road* are frequently shown in the same shots. Finally, in visual-based concept selection, concepts are selected as objects which are recognized in example shots with high recognition scores [2, 3]. But, this approach relies on accuracies of object recognition. LSCOM includes concepts corresponding to objects, which are difficult to be recognized, such as *Dogs*, *Telephone*, *Supermarket*, and so on. So, visual-based concept selection may wrongly select concepts which are unrelated to the query.

To overcome the above problems, we manually organize LSCOM concepts into an ontology, which can capture both lexical relations among concepts and their spatial and temporal relations. To do so, we define several new concepts which are missed in LSCOM. For example, we define a new concept *Air_Vehicle* as a superconcept of *Airplane* and *Helicopter*, in order to explicitly represent that both of *Airplane* and *Helicopter* fly in the air or move in airports. Also, we introduce a method which can appropriately estimate parameters of a retrieval model based on concepts selected by our ontology. To our best knowledge, there are no existing parameter estimation methods based on ontologies.

2.2 Rough Set Theory

One of the biggest advantages of RST is that it can extract multiple classification rules without any assumption or parameter. Specifically, by combining features characterizing example shots based on the set theory, RST extracts classification rules as minimal sets of features, needed to correctly identify subsets of example shots. Compared to this, although a “Gaussian Mixture Model (GMM)” can extract multiple feature distributions of example shots, these shots are not necessarily distributed based on Gaussian distributions [18]. Also, the

“Genetic Algorithm” (GA) can be used to extract multiple classification rules, where sets of features useful for identifying example shots are searched based on the principles of biological evolution [21]. But, parameters in the GA, such as the number of chromosomes, the probability of crossover and the probability of mutation, cannot effectively be determined with no a priori knowledge. Furthermore, decision tree learning methods and sequential covering methods can be used to extract multiple rules, but several useful rules are not detected as these methods depend on the order of extracting rules [19].

In RST, it is very important to determine which features characterize example shots. In other words, we need to define the indiscernibility relation between two example shots with respect to features. A traditional RST can deal only with categorical features, where the indiscernibility relation can be easily determined by examining whether two example shots have the same value or not [7]. But, in our case, example shots are represented by non-categorical high-dimensional features. To apply RST to such features, we built a classifier on each feature, and define the indiscernibility relation by examining whether two examples are classified into the same class or not. Although this kind of classifier-based RST is proposed in [11], the high-dimensionality of features is not considered. Specifically, although [11] uses probabilistic classifiers such as naive bayes and maximum entropy, it is difficult to appropriately estimate probabilistic distributions only from a small number of example shots. Compared to this, we use Support Vector Machines (SVMs) which are known as effective for high-dimensional features [20]. [20] provides the theory that if the number of feature dimensions is large, SVM’s generalization error is independent of the number of feature dimensions. In addition, even when only a small number of examples are available, the margin maximization needs no probability distribution estimation. Therefore, we develop a classifier-based RST using SVMs.

3 Video Retrieval Method

First of all, we set the condition where our QBE method is developed. We use large-scale video data provided by TRECVID [4]. This data consists of 219 development and 619 test videos in various genres, like cultural, news magazine, documentary and education programming. Each video is already divided into shots by using an automatic shot boundary detection method, where development and test videos include 36, 106 and 97, 150 shots, respectively. Like this, TRECVID video data is sufficient for evaluating the effectiveness of our QBE method on large-scale video data.

In order to filter out clearly irrelevant shots to a query, we borrow recognition results of 374 objects, provided by the research group in City University of Hong Kong [3]. That is, recognition scores of 374 objects are associated with all shots in development and test videos. To utilize the above recognition results, we develop an ontology where LSCOM concepts corresponding to 374 objects are organized. Also, to extract features used in RST, we use the color descriptor software [13]. Specifically, we extract the following 6 different types of features

from the middle video frame in each shot: 1. *SIFT*, 2. *Opponent SIFT*, 3. *RGB SIFT*, 4. *Hue SIFT*, 5. *Color histogram* and 6. *Dense SIFT* (see [13] in more detail). For each type of feature, we extract 1,000 visual words by clustering 200,000 features sampled from development videos. That is, we represent a shot as a total 6,000-dimensional vector, where each type of feature is represented as a 1,000-dimensional vector. Finally, for a query, we manually collect example shots from development videos, and retrieve relevant shots in test videos.

3.1 Building Ontology for Concept Selection

Fig. 3 shows a part of our ontology. LSCOM concepts are represented by capital letters followed by lower-case letters, while concepts that we define are represented only by capital letters. Also, we represent properties by starting their names with lower-case letters. Our ontology is developed by considering the “disjoint partition” requirement. This is a well-known ontology design pattern for making our ontology easily interpretable by both human and machine [14]. The disjoint partition means that a concept C_1 should be decomposed into disjoint subconcepts C_2, C_3, \dots . That is, for $i, j \geq 2$ and $i \neq j$, $C_i \cap C_j = \phi$. So, an instance of C_1 cannot be an instance of more than one subconcept C_2, C_3, \dots . For example, we should not place *Vehicle* and *Car* in the same level of the concept hierarchy, because an instance of *Car* is an instance of *Vehicle*. Thus, we have to carefully examine whether a concept is a generalization (or specialization) of another concept.

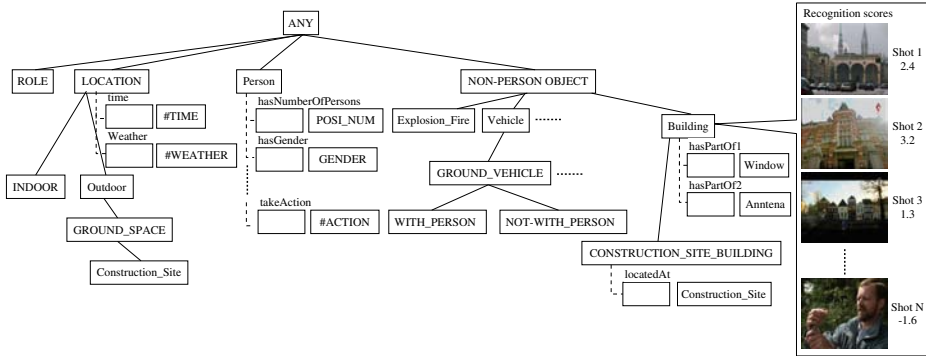


Fig. 3. A part of our ontology.

Furthremore, we consider visual characteristics to define our concept hierarchy. For example, as can be seen from Fig. 3, we define two subconcepts of *GROUND_VEHICLE*, *WITH_PERSON* and *NOT-WITH_PERSON*. We can induce that *Person* probably appears in shots containing subconcepts of *WITH_PERSON*, such as *Bicycle* and *Motorcycle*. On the other hand, it is uncertain that *Person* appears in shots containing subconcepts of *NOT-WITH_PERSON*.

Now, we explain how to select concepts related to a query. Basically, we firstly select concepts which match with words in the text description of the query. Then, for each selected concept, we select its subconcepts and concepts which are specified as properties. For example, for the query “buildings are shown”, *Buildings* and all of its subconcepts (e.g. *Office_Buildings*, *Hotel*, *Power_Plant* etc.) are firstly selected. Then, as shown in Fig. 3, *Windows* and *Antenna* are selected from *hasPartOf1* and *hasPartOf2* properties of *Building*. After that, from *locatedAt* property of *CONSTRUCTION_SITE_BUILDING* (a subconcept of *Building*), we select *Construction_Site* and all of its subconcepts (e.g. *City-space*, *Urban*, *Suburban* etc.). At this point, by tracing concept properties many times, we may select concepts which are unrelated to the query. For example, from the above *Construction_Site*, we can trace *ARTIFICIAL_ROAD*, *Sidewalk* and *Person*. But, these concepts are not related to the query. To avoid selecting unrelated concepts, we restrict the number of tracing concept properties to only one time. That is, for the above example, we finish concept selection after selecting *Construction_Site* and all of its subconcepts.

In Fig. 3, some concept properties are characterized by slots where # precedes concept names. We call such an operator “# operator” which represents a concept property, used only when it is specified in the textual description of a query. Let us consider the query “people are indoors”. For this query, we select *Person* and all of its subconcepts, and trace *Person*’s concept properties. But, for “takeAction” property, the current LSCOM only defines 12 concepts, such as *Singing*, *People_Crying*, *Talking* and so on. If these concepts are selected, shots containing them may be preferred. As a result, we may miss many shots where people take many other actions in indoor situations, such as eating and watching TV. Thus, only for queries like “people talking indoor”, we use the concept property “takeAction” to select concepts.

Since the textual description of a query is usually simple, we cannot select concepts which are definitely related to the query. For example, for the query “buildings are shown”, we select 55 concepts including *White_House*, *Military_Base*, *Ruins*, and so on. But, only a part of these concepts are truly related to the query. So, we validate selected concepts using example shots. Recall that all shots are associated with recognition scores of objects corresponding to LSCOM concepts, as shown in *Building* in Fig. 3. Based on such recognition scores in example shots, we validate concepts selected by our ontology. Specifically, for each object corresponding to a concept, we compute the average recognition score among example shots. Then, we rank concepts in the descending order. After that, we select concepts which are not only selected by our ontology, but also ranked in top T positions (we use $T = 20$). Like this, selected concepts are validated from both semantic and statistical perspectives.

Finally, we explain how to estimate classifier’s parameter based on object recognition scores. Note that this classifier is an SVM used to define indiscernibility relations among example shots in RST. Suppose that for a query, we have a set of selected concepts C , where each concept is represented as c_i ($1 \leq i \leq |C|$). Also, we have P parameter candidates for an SVM M , where the j -th parameter

is p_j and the SVM with p_j is M_{p_j} ($1 \leq j \leq P$). To estimate the best parameter, we temporarily retrieve S shots by using M_{p_j} (we use $S = 1,000$). Then, we compute the correlation between C and M_{p_j} as follows:

$$\text{Correlation}(C, M_{p_j}) = \sum_{i=1}^C \gamma(\text{rank}(M_{p_j}), \text{rank}(c_i)) \quad (1)$$

where $\text{rank}(M_{p_j})$ represents a ranking list of S shots according to their evaluation values by M_{p_j} . We obtain these evaluation values as SVM’s probabilistic outputs [17]. $\text{rank}(c_i)$ represents a ranking list of S shots according to recognition scores of the object corresponding to c_i . We compute $\gamma(\text{rank}(M_{p_j}), \text{rank}(c_i))$ as the Spearman’s rank correlation coefficient [15]. It represents the correlation between two ranking lists. If these are highly correlated, $\gamma(\text{rank}(M_{p_j}), \text{rank}(c_i))$ is close to 1, otherwise close to -1 . So, a larger $\gamma(M_{p_j}, c_i)$ indicates that M_{p_j} is more correlated with c_i . $\text{Correlation}(C, M_{p_j})$ represents the overall correlation over all concepts in C . Thus, we select the best parameter p_j^* where $\text{Correlation}(C, M_{p_j})$ is the largest among P parameters. In this way, we can estimate an SVM parameter which is semantically validated based on selected concepts.

3.2 Video Retrieval Using Rough Set Theory

We use rough set theory (RST) to extract classification rules, called “decision rules”, for discriminating relevant shots to a query from all irrelevant shots. To this end, we need two types of example shots. The first type of example shots are provided by a user and serve as representatives of relevant shots (“positive examples (p-examples)”). The second type of example shots serve as representatives of irrelevant shots (“negative examples (n-examples)”), but are not provided by the user. To overcome this, we have already developed a method which collects n-examples from shots other than p-examples [10]. Roughly speaking, our method iteratively enlarges n-examples by selecting shots which are more similar to already selected n-examples than p-examples. Thereby, our method can collect a variety of n-examples without wrongly selecting relevant shots as n-examples.

We discuss how to extract decision rules which can retrieve a large variety of relevant shots, only from a small number of p-examples. Note that decision rules are extracted by combining indiscernibility relations among examples, which are defined by SVMs. So, we need to build SVMs which can define various indiscernibility relations. To this end, we use “bagging” where SVMs are built on different sets of randomly sampled examples [16]. As described in [16], when only a small number of examples are available, SVMs’ classification results are significantly different depending on examples. Thus, we can define various indiscernibility relations by building SVMs based on bagging. However, due to the high-dimensionality of features, SVMs may be overfit and may not appropriately define indiscernibility relations. To alleviate this, we use the “random subspace method” where SVMs are built on different sets of randomly sampled feature

dimensions [12]. That is, the original high-dimensional feature is transformed into lower-dimensional features, so that we can alleviate to build overfit SVMs.

We regard SVM classification results as categorical features in RST, and extract decision rules for predicting the class of an unseen example. Let p_i and n_j be i -th p-example ($1 \leq i \leq M$) and j -th n-example ($1 \leq j \leq N$), respectively. a_k indicates the classification result by k -th SVM ($1 \leq k \leq K$). Here, $a_k(p_i)$ and $a_k(n_j)$ respectively represent class labels of p_i and n_j predicted by k -th SVM. That is, these are categorical features of p_i and n_j for a_k . In order to define the indiscernibility relation between each pair of p_i and n_j , RST extracts “discriminative features” which are useful for discriminating them. The set of discriminative features $f_{i,j}$ between p_i and n_j can be represented as follows:

$$f_{i,j} = \{a_k | a_k(p_i) \neq a_k(n_j)\} \quad (2)$$

That is, $f_{i,j}$ means that when at least one feature in $f_{i,j}$ is used, p_i can be discriminated from n_j .

Next, in order to discriminate p_i from all n-examples, we combine p_i ’s discriminative features. This is achieved by using at least one discriminative feature in $f_{i,j}$ for all n-examples. That is, we compute the following “discernibility function df_i ” which takes a conjunction of $\vee f_{i,j}$:

$$df_i = \wedge \{\vee f_{i,j} | 1 \leq j \leq N\} \quad (3)$$

Let us consider the discernibility function df_1 for one p-example p_1 and two n-examples n_1 and n_2 . Suppose that the set of discriminative features between p_1 and n_1 and the one between p_1 and n_2 are $f_{1,1} = \{a_1, a_3, a_5\}$ and $f_{1,2} = \{a_1, a_2\}$, respectively. Under this condition, df_1 is computed as $(a_1 \vee a_3 \vee a_5) \wedge (a_1 \vee a_2)$. This is simplified as $df_1^* = (a_1) \vee (a_2 \wedge a_3) \vee (a_2 \wedge a_5)$ ². That is, p_1 can be discriminated from n_1 and n_2 , by using a_1 , the set of a_2 and a_3 or the set of a_2 and a_5 . Like this, each conjunction term in df_i^* represents a “reduct” which is a minimal set of features needed to discriminate p_i from all n-examples

From each reduct, we can construct a decision rule in the form of *IF-THEN* rule. Since each feature in our RST is defined by an SVM, a decision rule represents a combination of SVMs. For example, the decision rule constructed from the reduct $(a_2 \wedge a_3)$ is “*IF* a shot s is classified as positive by both 2-nd and 3-rd SVMs, *THEN* its class label is positive”. That is, to match a decision rule with s , we examine whether s is classified as positive by all SVMs in the decision rule. In this way, we count how many decision rules match with s . Finally, we rank all shots in the descending order, and retrieve shots within top T positions (we use $T = 1,000$).

4 Experimental Results

We evaluate our method on the following 4 queries, *Query 1*: A view of one or more tall buildings and the top story visible, *Query 2*: Something burning with

² This simplification is achieved by using the distributive law $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$ and the absorption law $A \vee (A \wedge B) = A$.

flames visible, *Query 3*: One or more people, each at a table or desk with a computer visible, *Query 4*: One or more people, each sitting in a chair, talking. For each query, we run our method 9 times by using different sets of 10 p-examples. We evaluate the retrieval performance as the average number of relevant shots within 1,000 retrieved shots.

In Fig. 4 (a), we compare the following three types of retrieval, in order to evaluate the effectiveness of our ontology for filtering out irrelevant shots and estimating an SVM parameter. The first one is *Baseline* without using our ontology. The second type of retrieval is *Ontology1* which uses our ontology only for filtering out irrelevant shots. The final type of retrieval is *Ontology2* which uses our ontology for both irrelevant shot filtering and SVM parameter estimation. For each topic, performances of *Baseline*, *Ontology1* and *Ontology2* are represented by the leftmost red bar, the middle green bar and the rightmost blue bar, respectively.

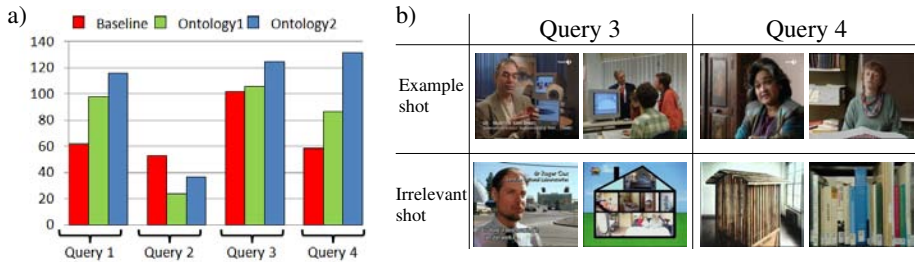


Fig. 4. (a) Performance comparison among *Baseline*, *Ontology1* and *Ontology2*, (b) Examples of shots filtered by our ontology.

As can be seen from Fig. 4 (a), except for *Query 2*, it is very effective to filter out irrelevant shots based on concepts selected by our ontology. The retrieval performance is further improved by estimating SVM parameters based on selected concepts. The reason for the low performance for *Query 2* is that, in each of 9 times retrieval, we can only select one or two concepts, that is, *Explosion_Fire* and *Earthquake*. What is worse, these concepts are not so effective for characterizing relevant shots. For example, 1,000 shots with highest recognition scores of *Explosion_Fire* and those of *Earthquake*, only characterize 37 and 12 relevant shots, respectively. As a result, we cannot appropriately filter out irrelevant shots and cannot appropriately estimate SVM parameters. To improve the performance for *Query 2*, other than *Explosion_Fire*, we need to recognize objects such as candle flame, bonfire, fire blasted from rockets, and so on.

In Fig. 4 (b), for each of *Query 3* and *Query 4*, we show two example shots and two clearly irrelevant shots which are filtered out by our ontology (for *Query 1*, see Fig. 2). For *Query 3*, *Baseline* without using our ontology wrongly retrieves shots where people just appear and shots which contains straight lines corresponding to computer shapes, and shapes of pillars and blinds in a background. For *Query 4*, *Baseline* wrongly retrieves shots which contain straight lines cor-

responding to shapes of background objects. By filtering out the above kind of shots, *Ontology1* and *Ontology2* can significantly outperform *Baseline*.

Finally, we examine whether filtering out irrelevant shots based ontology can reduce computation times. In Fig. 5, we show the average computation time of *Baseline* (left red bar) and *Ontology1* (right green bar) among of 9 times retrieval. As can be seen from this figure, filtering of irrelevant shots is useful for reducing computation times. Nonetheless, our method currently takes about 500 seconds, because it requires building of multiple SVMs and matching of many decision rules. So, from the perspective of computation cost, our method need to be improved. To this end, we are currently parallelizing processes of SVM building and decision rule matching by using multiple processors.



Fig. 5. Comparison between the computation time of *Baseline* and that of *Ontology1*.

5 Conclusion and Future Works

In this paper, we proposed a video retrieval method based on QBE approach. We construct an ontology to incorporate object recognition results into QBE. Specifically, our ontology is used to select concepts related to a query. By referring to recognition results of objects corresponding to selected concepts, we filter out clearly irrelevant shots. In addition, we estimate an SVM parameter based on the correlation between selected concepts and shots retrieved by the SVM. Also, to retrieve a large variety of relevant shots, we use RST for extracting multiple classification rules which characterize different subsets of example shots. Experimental results on TRECVID 2009 video data show that the retrieval performance can be significantly improved by using our ontology. Besides, our ontology is useful for reducing the computation time. Finally, by using RST, we can successfully cover a large variety of relevant shots.

Acknowledgments. This research is supported in part by Strategic Information and Communications R&D Promotion Programme (SCOPE) by the Ministry of Internal Affairs and Communications, Japan.

References

1. A. Jain, R. Duin and J. Mao: “Statistical Pattern Recognition: A Review”, *TPAMI*, Vol. 22, No. 1, pp. 4 – 37 (2000)
2. C. Snoek et al.: “The MediaMill TRECVID 2009 Semantic Video Search Engine”, In Proc. of TRECVID 2009, pp. 226 – 238 (2009)
3. C. Ngo et al.: “VIREO/DVM at TRECVID 2009: High-Level Feature Extraction, Automatic Video Search, and Content-Based Copy Detection”, In Proc. of TRECVID 2009, pp. 415 – 432 (2009)
4. A. Smeaton, P. Over and W. Kraaij: “Evaluation campaigns and TRECVID”, In Proc. of MIR 2006, pp. 321 – 330 (2006)
5. Y. Peng and C. Ngo: “EMD-Based Video Clip Retrieval by Many-to-Many Matching”, In Proc. of CIVR 2005, pp. 71 – 81 (2005)
6. K. Kashino, T. Kurozumi and H. Murase: “A Quick Search Method for Audio and Video Signals based on Histogram Pruning”, *TMM*, Vol. 5, No. 3, pp. 348 – 357 (2003)
7. J. Komorowski, A. Øhrn and A. Skowron: “The ROSETTA Rough Set Software System”, *Handbook of Data Mining and Knowledge Discovery*, W. Klösgen and J. Zytkow (eds.), chap. D.2.3, Oxford University Press (2002)
8. M. Naphade et al.: “Large-Scale Concept Ontology for Multimedia”, *IEEE Multimedia*, Vol. 13, No. 3, pp. 86 – 91 (2006)
9. M. Christel and A. Hauptmann: “The Use and Utility of High-Level Semantic Features in Video Retrieval”, In Proc. of CIVR 2005, pp. 134 – 144 (2005)
10. K. Shirahama, C. Sugihara, Y. Matsuoka, K. Matsumura and K. Uehara: “Kobe University at TRECVID 2009 Search Task”, In Proc. of TRECVID 2009 Workshop, pp. 76 – 84 (2009)
11. S. Saha, C. Murthy and S. Pal: “Rough Set Based Ensemble Classifier for Web Page Classification”, *Fundamenta Informaticae*, Vol. 76, No. 1-2, pp. 171 – 187 (2007)
12. T. Ho: “The Random Subspace Method for Constructing Decision Forests”, *TPAMI*, Vol. 20, No. 8, pp. 832 – 844 (1998)
13. K. Sande, T. Gevers and C. Snoek: “Evaluating Color Descriptors for Object and Scene Recognition”, *TPAMI*, Vol. 32, No. 9, pp. 1582 – 1596 (2010)
14. M. Horridge et al.: “A Practical Guide to Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0”, <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf> (2004)
15. R. Hogg and A. Crig: “Introduction to Mathematical Statistics (5th edition)”, Prentice Hall (1994)
16. D. Tao, X. Tang, X. Li and X. Wu: “Asymmetric Bagging and Random Subspace for Support Vector Machines-based Relevance Feedback in Image Retrieval”, *TPAMI*, Vol. 28, No. 7, pp. 1088 – 1099 (2006)
17. H. Lin, C. Jin and R. Weng: “A Note on Platt’s Probabilistic Outputs for Support Vector Machines”, *Machine Learning*, Vol. 68, No. 3, pp. 267 – 276 (2007)
18. J. Verbeek, N. Vlassis and B. Kröse: “Efficient Greedy Learning of Gaussian Mixture Models”, *Neural Computation*, Vol. 15, No. 2, pp. 469 – 485 (2003)
19. E. Alpaydin: “Introduction to Machine Learning”, MIT Press (2004)
20. V. Vapnik: “Statistical Learning Theory”, Wiley-Interscience (1998)
21. J. Han and M. Kamber: “Data Mining: Concepts and Techniques (Second Edition)”, Morgan Kaufmann Publishers (2006)

Collaborative Fake Media Detection in a Trust-Aware Real-Time Distribution Network

Dominik Renzel, Khaled A. N. Rashed, Ralf Klamma

Informatik 5 (Information Systems & Databases), RWTH Aachen University
Ahornstr. 55, D-52056, Aachen, Germany
{renzel,rashed,klamma}@dbis.rwth-aachen.de

Abstract. Due to the increased incorporation of external sources media agencies face the challenge of providing high-trust media to their customers. Automatic image processing approaches still do not bridge the semantic gap to identify fakes. Complementary community-based approaches lack real-time media distribution for improved awareness and base trust on subjective opinions instead of objective actions. In this paper we propose a collaborative fake media detection approach addressing these challenges in form of a federated, trust-aware media distribution network. Starting from a realistic use case scenario we elicit requirements and present an XMPP-based and Web service-enhanced multimedia distribution network as solution. Finally, we sketch a Web-based fake media detection application powered by our network and its services.

1 Introduction

Traditionally, people consider images as a means for true reproduction of real events and accepted as a proof of occurrence of such events. Recently, this consideration is not longer valid since fake images have a high occurrence especially now that images can be faked and distributed arbitrarily without much effort. Nowadays, news creation processes have taken significant distance from being conducted in isolation. Following the basic principles of the Open Innovation approach [3], in today's media distribution networks different communities are involved as both information providers and consumers. With the growing availability of low-cost high-quality multimedia processing and context sensor equipment in mobile devices, it has already become widespread practice to even have amateur reporters on site of interesting events serve as information sources. With such an inherently distributed approach, the authenticity of distributed multimedia is even more endangered than in previous more isolated approaches. Today's media thus face the challenge of deciding if media are real or faked, ideally before they are further broadcasted to their customers, who pay for high-trust media.

Consider the following infamous cases where faked media were finally published to information end-consumers. A recent example of a faked image manipulated by the newspaper Al-Ahram and published in international media is showing the Egyptian president Mubarak at the front of a group of world leaders, where in



Fig. 1. Image Fakery Examples

the original image he was lagging behind (cf. Figure 1). The fake thus tried to transport a subtle propagandistic message of a distorted reality. In turn, news papers and TV stations had to issue errata to recover their reputation.

Such events are eroding the public trust in media. Therefore, media agencies are required to make their distribution channels capable of identifying media fakery at the earliest stage possible not only to avoid reports of a distorted reality with possible negative consequences, but also to avoid additional costs due to the following correction means. The most desirable solution is automatic fake detection, but current methods still cannot identify semantic inconsistencies in media (cf. [29]). Thus, complementary Web 2.0 community-based approaches were developed to involve people in such processes. Systems such as NewsTrust (<http://newstrust.com>) pursue such an approach. However, information still has to be pulled by participants, although the current trend hints to real-time requirements and synchronous server side pushes [20] creating a new level of community awareness. Furthermore, the quality of authenticity judgements depends on the trustability of its judges. Current systems establish the trust level of a user by ratings of others which are often subjective and not based on objectively valuable contributions. Furthermore, the willingness to spend time on rating others is mostly not given. Instead of basing trust on subjective opinions, a method is required that objectively adapts trust levels depending on actions.

In this paper we overcome the above problems with an open standard-based collaborative image fake detection system distributed across various communities. The system operates in near real-time and complements traditional automatic approaches. Our approach is powered by a set of Web services based on the MPEG-7 standard as well as by services and infrastructure provided by the open standard Extensible Messaging and Presence Protocol (XMPP) [25, 26] and its extension protocols, in particular XMPP PubSub [17]. A media fake detection application connecting to our infrastructure is realized as a Web 2.0 application consisting of a set of OpenSocial Gadgets [19] for direct communication and the distribution of MPEG-7 [14] multimedia metadata across an XMPP network of media agents.

In Section 2 we first analyze the state-of-the-art of image fakery detection systems and technologies related to our approach. Then, in Section 3 we describe a

use case scenario where three media agencies have to detect a faked image, thereby identifying requirements for our system. In Section 4 we present the backend of our system as a multimedia distribution network including its individual parts in detail. In Section 5 we present a media fake detection application powered by our network. In Section 6 we conclude and provide an outlook to further work.

2 Related Work

Faked Image Detection: Faked image detection has been investigated for years and addressed by a number of approaches. Watermarking approaches [24] are based on imperceptibly embedding information within the image content. The requirements of embedding such information in digital images are specially equipped digital cameras. In addition, watermarking degrades the quality of the image content. In contrast to watermarking approaches, researchers in the field of digital image forensics have developed passive techniques which operate in the absence of any watermark or signature for image authentication (e.g. [6, 21, 11, 31]). They work on the assumption that although digital forgeries may leave no visual clues of having been tampered with, they may alter the underlying statistics of an image that can be detected using statistical models. The major drawback of such tools is that their use in public domains is computationally impractical.

Content based approaches (e.g.[18, 12]) aim at detecting all faked images produced from the original through active manipulation. They are based on similarity search and embed no additional information within the image content, thus considering the image itself as the watermark. The efficiency of such techniques is largely affected by the size of the reference image dataset [18]. Furthermore, current approaches lack discriminative power for fake detection due to the inability of capturing semantic aspects. Our collaborative fake detection system utilizes community aspects in addition to automatic content-based image similarity search techniques [4].

Collaborative Fake Detection: Sharing knowledge and control is the key idea of collaborative fake detection [22]. A Community of Practice [32] is the context where such collaborative activities can be achieved. Knowledge about media is exchanged within the communities of practice for example by the distribution of MPEG-7 metadata [14, 28]. Collaborative judgments and evidence against the suspected fake support the evaluation of semantic inconsistencies that cannot yet be detected with automatic approaches. The important problem faced in collaborative fake media detection is the assessment of trustable authenticity judgments that we address in the scope of this paper.

Trust Management: Trust management is a key issue in distributed networks, especially in sharing environments. Trust provides us with information about the people we should share content with and accept content from. There are some efforts to formalize trust. Massa et al. propose a trust-aware model in which the web of trust is explicitly expressed [16]. Golbeck analyzed and modeled the core characteristics of trust in collaborative social networks and developed several algorithms for computing trust on the example of the TrustMail application

[9]. In this work, we take into account the trust of information sources and the quality of their contributions using a simplified trust mechanism and present a modular trust-aware multimedia distribution network.

MPEG-7: MPEG-7 is a standard for the description of multimedia content. It provides descriptors for various data types - text, graphics, audio, video. In order to achieve interoperability and keep advantages of server side computation we have presented the Lightweight Application Server (LAS) [30] for MPEG-7 Web services. It provides communities with a set of core services and MPEG-7 semantic multimedia metadata and content processing services to connect to heterogeneous data sources [23]. In particular, the LIRE [13] library is used for automatic extraction and indexing of low-level features as well as content based image retrieval (CBIR).

Real-time federation: Due to frequent complaints about the intransparency and lack of control of private data storage with social networking platforms, there are already new alternative platforms emerging (e.g. Diaspora [10]), where the same functionality is offered in a way that anybody can run his instance in federation with others. At the same time, the demand for real-time application behavior [20] speeds up the information flow tremendously. Concepts such as security, privacy and trust have to be weaved in as unobtrusive, transparent, and least blocking as possible. In our approach we aimed to realize these requirements with a network of federation-enabled XMPP servers including respective services and data.

Publish/Subscribe: Nowadays, the *Publish/Subscribe (PubSub)*[2, 5] pattern is omnipresent (e.g. newspapers, blogs, even email lists). There is a *channel of communication* (resp. a *node*), *subscribers* receiving data sent on that channel, and *publishers* who send data *payloads* across the channel. The pattern was also described by Gamma et al. as the behavioural *Observer* pattern [7]. Until today, the pattern is applied successfully, sometimes working locally on one machine or remotely across whole networks. The *XMPP PubSub Extension Protocol*[17] supports the construction of remote PubSub systems transporting XML-based payloads. For this work we demonstrate the distribution of MPEG-7 multimedia content descriptions along with authenticity ratings.

3 Use Case Scenario & Requirements Analysis

In this section we first describe a scenario to understand a media fake detection process in a media distribution network such as in Figure 2. Afterwards, we derive a set of requirements for our system improving the process. Consider the following scenario. A government press agency sends a doctored picture of a successful long-range missile launch to Thompson Reuters as a demonstration of the country's military power, although the real outcome of the event was a crash of the missile. Despite the good cooperation with the government press agency in the past, the responsible media agent recognizes the image content as highly sensitive and thus decides to request expertise on its authenticity before further distribution. Although some trusted experts reviewed the image, the forgery is not discovered, and the picture distributed to customers. TV stations

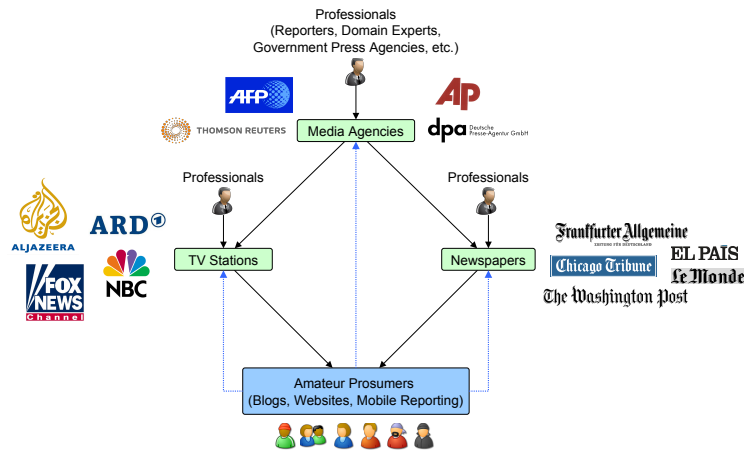


Fig. 2. Exemplary excerpt of a media distribution network

and newspapers around the world broadcast the sensitive information to their audiences. After the worldwide publication of the faked picture, a group of local dissidents who eye-witnessed the failed missile launch feel the urge to reveal the truth. In a message sent to Reuters they describe the real situation, send their own picture of the missile crash as proof for their statement, and state their willingness to help prevent such incidents in future. Further expert analysis on both pictures then reveals the fake. As a result, Reuters and all its customers issue a corrective statement to recover their public credibility. However, to prevent further occurrences of such situations, media agents decide to be more cautious towards their information sources or even decide for alternative sources. On the other hand, Reuters acknowledges the group of dissidents' help in discovering the fake and decides to involve their expertise for further authenticity judgement. From the above scenario we now derive a set of requirements to an information system supporting the process described above, before we explain our approach in the next sections.

- *media & metadata repository*: The first step is to make media and their metadata available for other parties. We base this work on our LAS MPEG-7 services and its repository [30].
- *federated multimedia distribution network*: The most important use case in the scenario is the transport of media (metadata) between entities in real-time. Here, PubSub is the main communication pattern. For a distributed approach, PubSub support is required in a remote and federated manner. The network should support arbitrary payload formats in order to stay generic. Here, we base our approach on the XMPP Protocol and its PubSub extension [17] fulfilling all these requirements.
- *authenticity rating service*: a service is required that allows the collaborative assignment of authenticity ratings to media as well as the computation

- and rendering of reasonable aggregates to create awareness for fakes and to support the decision of a media agency to publish a medium or not.
- *trust management service*: a service is required that manages trust relationships between entities again in a federated way and supports the dynamic evolution of trust. Since the service itself must be trusted by its users, privacy and security are non-functional requirements to be guaranteed.

4 A Trust-aware Multimedia Distribution Network

In this section we present a modular trust-aware multimedia distribution network based on the above requirements. In Section 4.1 we describe a basic network building block and its workflow. Each building block implies a simple trust protocol which is formalized in Section 4.2. Finally, we demonstrate the composition of complete information distribution networks of building blocks in Section 4.3.

4.1 The Basic Building Block

Conceptually, the basic building block of our architecture is a variation of the PubSub pattern (cf. Fig. 3). The central parts of this building block are an *untrusted in node* and a *trusted out node* with configuration under control of a *mediator*. For the in node, all of the mediator's *sources* are publishers and subscribers at the same time to support media distribution for collaboration. For the out node, only the mediator is allowed to publish. The list of subscribers reflects the mediator's consumers relying on the authenticity of the information published. First, a source introduces a new medium along with an authenticity

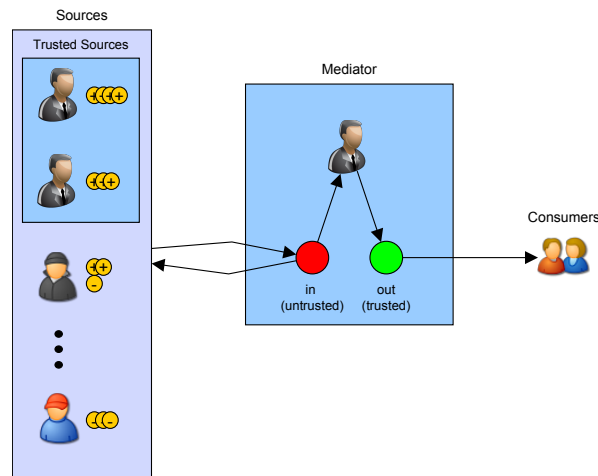


Fig. 3. Building block for media distribution network

rating by publishing it to the untrusted in node that immediately pushes it to all other sources, which in turn publish their authenticity ratings to the same node. Based on ratings from various sources accumulated over time, the mediator eventually decides the information to be trustworthy of being published to the out node or not. The decision depends on the individual levels of trust towards his sources. In Section 4.2 we provide a formalized description of our trust mechanism. Technically, each of the building blocks described above can be realized with a set of components depicted in Figure 4. Any XMPP Server hosting a PubSub service as specified in [17] realizes all necessary functionality regarding the management and configuration of nodes, in particular controlling node subscriber and publisher lists, as well as pushing arbitrary XML-based payloads to subscribers.

4.2 Authenticity Rating & Trust

In this section we formalize the relationship between authenticity ratings and trust used in our approach. Let $J = \{j_1, \dots, j_n\}$ be a set of unique identifiers for the entities involved in the fake detection process. For our approach we use JIDs (cf. [25]). Our basic notion of trust involves two entities, i.e. a *trustor* $tr \in J$, a *trustee* $te \in J$ and a *level of trust* $t(tr, te)$ between them. Although there exists work on sophisticated models such as [15], trust-aware social networks usually let users assign a single numerical rating for usability reasons [8]. In our model, the mediator m of a building block from Section 4.1 takes the role of the trustor of a *set of sources* $S_m \subset J$ as its trustees, that publish information payloads i of a certain domain I (in our case the domain of MPEG-7 descriptors).

For authenticity ratings, we define a function r that for a given i and a source s assigns a rating $\in R = \{true, fake\}$. In the following we describe the relationship between authenticity ratings and the dynamic adaptation of trust between involved entities.

Not only is $t(tr, te)$ depending on previous authenticity statements, but also should be adapted dynamically, either reinforcing desirable actions - in our case publishing a faked medium as fake resp. a real one as real - or punishing undesirable actions - in our case publishing a faked medium as real resp. a real one as fake. Thus, reinforcement consists in tr raising his trust level towards te , punishment in lowering it. Thus, each m must be enabled to update trust levels $\forall s \in S_m$. Listing 1.1 sketches an algorithm for updating trust values.

```

trust_update( $m \in J, i \in I, x$  action){
  for each  $s \in S_m$  {
    if  $r(i, s) = fake \wedge x = p_{fake}(m, i)$  then  $t(m, s)++$ ;
    else if  $r(i, s) = true \wedge x = p_{fake}(m, i)$  then  $t(m, s)--$ ;
    else if  $r(i, s) = fake \wedge x = p_{real}(m, i)$  then  $t(m, s)--$ ;
    else if  $r(i, s) = true \wedge x = p_{real}(m, i)$  then  $t(m, s)++$ ;
  }
}

```

Listing 1.1. Updating trust values after publication to trusted out node

Any trust update takes place whenever m feels confident to *publish* i as either *fake* ($p_{fake}(m, i)$) or *real* ($p_{real}(m, i)$). Furthermore, there is the option of *rejecting* any publication on the trusted out node ($rej(m, i)$). In this case, no trust update takes place. Since m in his role as trustor is interested in high-quality media (metadata) and reliable authenticity ratings, he can expose trust levels as an incentive to perform desirable actions only. To decide publication of an i , m relies on ratings from different $s \in S_m$, while using $t(m, s)$ as weighting factor. For a given $i \in I$, a function a returns an aggregate supporting m in his decision which action to take. For simplicity we chose $a(m, i)$ as weighted mean over all ratings on i by $s \in S_m$, where the weights are given by $t(m, s)$ (cf. Equation 1). The intuition behind choosing the weighted mean is that the higher a source’s trust value is the more influence his rating has on the resulting aggregate used by m to decide on publication.

$$a(m, i) = \frac{\sum_{j=1}^{|S_m|} t(m, s_j) * r(i, s_j)}{\sum_{j=1}^{|S_m|} t(m, s_j)} \quad (1)$$

Technically, the dynamic management of trust is realized as a service that maintains individual levels of trust between trustors and their trustees. Ratings of different sources for given information items are covered by another service.

4.3 Construction of a Network

A complete distribution network can now be modeled by reasonably connecting multiple building blocks. The intuition is that each mediator can act as a source for another mediator. Thus, information distribution networks can dynamically

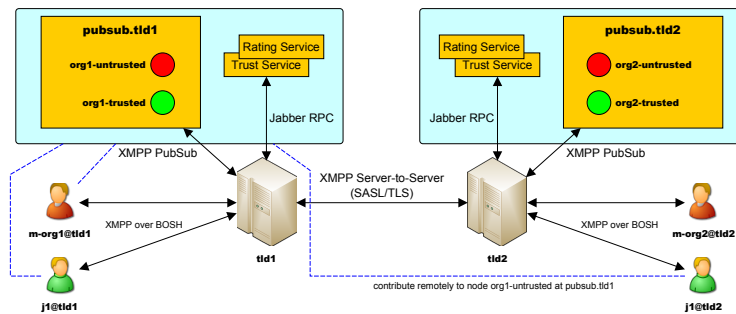


Fig. 4. A Trust-aware Federated Media Distribution Network

evolve over time by simple interactions with XMPP PubSub nodes. It should be noted that it is not necessary that each entity in the network maintains its own XMPP server, which would be acceptable e.g. for a high-profile media agency, but unacceptable e.g. for a freelancing information agent. For these purposes

it is possible to offer a building block from Section 4.1 as a service, which is hosted on one XMPP server or a whole cluster. On the technical level we realize a network of different interconnected building blocks by a network of XMPP servers in combination with the provision of services for the management of users, communities, MPEG-7 multimedia metadata, trust and authenticity rating as indicated in Figure 4. Given the inherent XMPP server-to-server communication [25, 26], all components are federated and accessible across the network via the protocol and its extensions [1, 17, 27]. In particular, [1] can be used to invoke services of our LAS.

5 A Fake Multimedia Detection Application

In this section, we briefly describe how to apply our trust-aware media distribution network from Section 4 for realizing a fake media detection application. Figure 5 shows a first mockup of such an application consisting of a set of three widgets. In the following we will briefly explain the interface for collaborative fake media

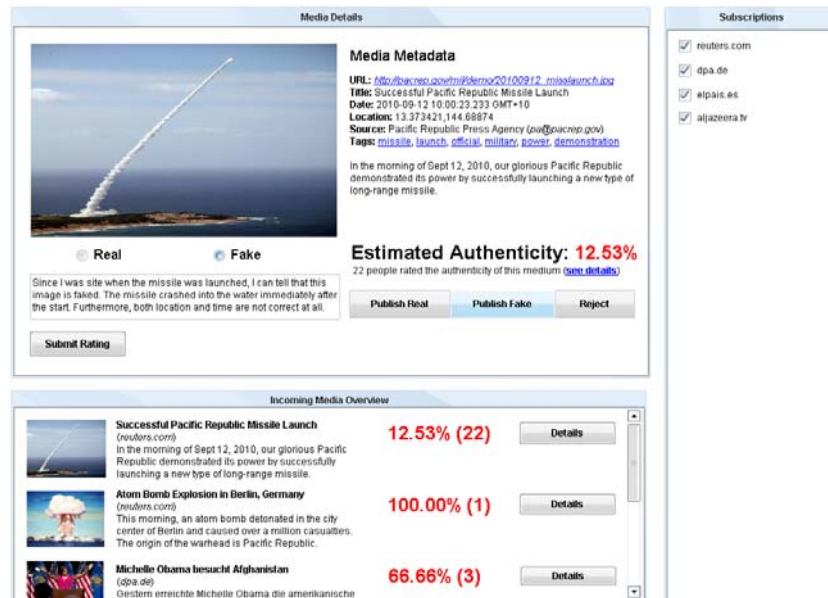


Fig. 5. Widget-based UI of a Multimedia Fake Detection Application

detection for both the mediator and his sources, which reflects the workflow from Section 4. In the *Incoming Media Overview*, the user gets an overview of media currently discussed on all untrusted in nodes he is subscribed to. Each element of the list provides a short summary of the medium and its metadata (cf. $i \in I$, Section 4.2) and the weighted authenticity ratings aggregate (cf. function

a, Section 4.2). From this list, the user can select any element, which is then rendered in more detail in the *Media Details* widget. Apart from the medium and its metadata, the user finds different buttons, depending on his role. As an information source, the user finds a rating interface, which allows him to choose between real or fake, add a comment and submit his rating. On submission, a triple consisting of a source identifier, a media identifier and a rating is encoded as an XML payload and published to the in node again. After automatic forwarding to all subscribers, their interfaces are updated with the new information. As a mediator, the user can decide on the three different actions *preal*, *pfake*, and *rej* (cf. Section 4.2) by pressing the respective buttons. A trust update (cf. Listing 1.1) is executed after any publication to the trusted out node by invoking the respective LAS service. Due to space restrictions, we will not elaborate here on further UI elements, such as media annotation (cf. [23]), advanced trust visualisation, etc.

Technically, the interface is realizable as a set of OpenSocial [19] gadgets using XMPP/LAS AJAX client libraries to connect to the XMPP server network and its services. For the access to PubSub nodes, we implemented an extension of the dojo XMPP library realizing the most important use cases of [17]. For the access to LAS Services, we implemented an AJAX connector client library. However, a further extension of the dojo XMPP library realizing the Jabber RPC extension protocol is a preferable alternative for the future.

6 Conclusions

In this paper we have demonstrated an approach for collaborative fake media detection based on a federated, trust-aware media distribution network with near real-time properties. We have presented an overview of related work in the domain of fake media detection, which is dominated by image processing approaches, that still do not bridge the semantic gap [29] and by community approaches lacking real-time communication and trust adaptations based on objective actions. Thus, we proposed our approach to overcome these challenges. Starting from a realistic use case scenario we elicited requirements and presented a realization as an XMPP-based and Web service-enhanced multimedia distribution network supporting arbitrary XML-based payload format. Finally, we sketched the design of a Web-based fake media detection application taking benefit from our network and its services.

At the time of writing this document many components of our multimedia distribution network as well as connector clients have been realized and evaluated. We already gained experience with XMPP-enabled OpenSocial Gadgets and therefore extended the well-known dojo JS library with support for PubSub, multi-user chats, etc. [33]. With these extensions, a real-time microblogging application was easily realizable. Although the XMPP standard provides detailed documentation about the protocol itself, there is not too much information which PubSub node topologies are suitable when scaling up to larger and highly distributed networks. Thus, we are currently evaluating architecture scalability

and performance in the context of the ROLE project (<http://role-project.eu>), where XMPP also serves as an open standard infrastructure for Widget-based PLE (Personal Learning Environments). Currently, we realize the fake multimedia detection application based on the design presented in the context of this work.

Acknowledgments. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 231396 (ROLE project) as well as DAAD.

References

1. D. Adams. XEP-0009: Jabber-RPC. Technical report, XMPP Standards Foundation, February 2009.
2. K. P. Birman and T. A. Joseph. Exploiting Virtual Synchrony in Distributed Systems. In *SOSP '87: Proceedings of the eleventh ACM Symposium on Operating systems principles*, pages 123–138, New York, NY, USA, 1987. ACM.
3. H. Chesbrough. *Open Innovation: The new imperative for creating and profiting from technology*. Harvard Business School Press, 2003.
4. R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40:5:1–5:60, May 2008.
5. P. T. Eugster, P. A. Felber, R. Guerraroui, and A.-M. Kermarrec. The Many Faces of Publish/Subscribe. *ACM Computing Surveys*, 35(2):114–131, June 2003.
6. J. Fridrich, D. Soukal, and J. Lukáš. Detection of Copy-Move Forgery in Digital Images. In *Proc. of DFRWS 2003*, pages 90–105, 2003.
7. E. Gamma, R. Helm, R. E. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Reading, MA, 1995.
8. J. Golbeck. Trust and Nuanced Profile Similarity in Online Social Networks. *ACM Transactions on the Web*. In press.
9. J. Golbeck. *Computing and Applying Trust in Web-based Social Networks*. PhD thesis, University of Maryland, 2005.
10. D. Grippi, M. Salzberg, R. Sofaer, and I. Zhitomirskiy. diaspora*. The privacy aware, personally controlled, do-it-all, open source social network, 2010.
11. M. Johnson and H. Farid. Exposing Digital Forgeries by Detecting Inconsistencies in Lighting. In *ACM Multimedia and Security Workshop*, pages 1–10, New York, NY, USA, 2005. ACM.
12. Y. Ke, R. Sukthankar, L. Huston, Y. Ke, and R. Sukthankar. Efficient Near-duplicate Detection and Sub-image Retrieval. In *MM '04: Proceedings of the ACM international conference on Multimedia*, pages 869–876, NY, USA, 2004. ACM.
13. M. Lux and S. A. Chatzichristofis. Lire: lucene image retrieval: an extensible Java CBIR library. In *MM '08: Proceedings of the 16th ACM international conference on Multimedia*, pages 1085–1088, New York, NY, USA, 2008. ACM.
14. B. S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7, Multimedia Content Description Interface*. John Wiley and Sons, Ltd., June 2002.
15. S. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, Department of Mathematics and Computer Science, 1994.
16. P. Massa and B. Bhattacharjee. Using Trust in Recommender Systems: An Experimental Analysis. In *Proceedings of iTrust2004 International Conference*, pages 221–235, 2004.

17. P. Millard, P. Saint-Andre, and R. Meijer. XEP-0060: Publish-Subscribe. Technical report, XMPP Standards Foundation, July 2010. Draft Standard.
18. S. Nikolopoulos, S. Zafeiriou, and N. Nikolaidis. Image replica detection system utilizing R-trees and linear discriminant analysis. *Pattern Recognition*, 43(3):636–649, March 2010.
19. OpenSocial Specification 1.0. Technical report, OpenSocial and Gadgets Specification Group, March 2010.
20. T. O’Reilly and J. Batelle. Web Squared: Web 2.0 Five Years On, 2009. Special Report Web 2.0 Summit.
21. A. Popescu and H. Farid. Statistical Tools for Digital Forensics. In *Proceedings of the 6th International Workshop on Information Hiding*, pages 128–147, Toronto, Canada, 2004. Springer-Verlag, Berlin-Heidelberg.
22. K. A. N. Rashed and R. Klamma. Towards Detecting Faked Images. In A. Carreras, J. Delgado, X. M. as, and V. Rodriguez, editors, *Proceedings of the 11th International Workshop on Interoperable Social Multimedia Applications(WISMA10), Barcelona, Spain, May 19-20*, May 2010. CEUR Workshop Proceedings, Vol. 583.
23. D. Renzel, R. Klamma, Y. Cao, and D. Kovachev. Virtual Campfire - Collaborative Multimedia Semantization with Mobile Social Software. In R. Klamma, H. Kosch, M. Lux, and F. Stegmaier, editors, *Proceedings of the 10th International Workshop on Semantic Multimedia Database Technologies (SeMuDaTe’09), Graz, Austria, 12 2009*. CEUR Workshop Proceedings, Vol. 539.
24. C. Rey and J.-L. Dugelay. A survey of Watermarking Algorithms for Image Authentication. *EURASIP J. Appl. Signal Process.*, 2002(1):613–621, 2002.
25. P. Saint-Andre. RFC 3920 – Extensible Messaging and Presence Protocol (XMPP): Core. Technical report, Jabber Software Foundation, October 2004.
26. P. Saint-Andre. RFC 3921 – Extensible Messaging and Presence Protocol (XMPP): Instant Messaging and Presence. Technical report, Jabber Software Foundation, October 2004.
27. P. Saint-Andre. XEP-0045: Multi-User Chat. Technical report, XMPP Standards Foundation, July 2008. Draft Standard.
28. G. Shih-Fu Chang, A. B. Chan, and P. J. Moreno. Overview of the MPEG-7 standard. *IEEE Trans. Circuits and Systems for Video Technology.*, 11(0):688 – 695, 2001.
29. A. Smeulders, M. Worring, S. Santini, A. G. A, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans Pattern Anal Mach Intell*, 22(12):1349 – 1380, 2000.
30. M. Spaniol, R. Klamma, H. Janßen, and D. Renzel. LAS: A Lightweight Application Server for MPEG-7 Services in Community Engines. In K. Tochtermann and H. Maurer, editors, *Proceedings of I-KNOW ’06, 6th International Conference on Knowledge Management, Graz, Austria, September 6–8, 2006*, J.UCS (Journal of Universal Computer Science)Proceedings, pages 592–599. Springer-Verlag, 2006.
31. W. Wang and H. Farid. Exposing Digital Forgeries in Video by Detecting Double Quantization. In *ACM Multimedia and Security Workshop*, pages 39–48, Princeton, NJ, September 2009. ACM.
32. E. Wenger. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge, UK, 1998.
33. M. Wolpers, M. Friedrich, R. Shen, C. Ullrich, R. Klamma, and D. Renzel. Early Experiences with Responsive Open Learning Environments. In H. Maurer, N. Kulathuramaiyer, and K. Tochtermann, editors, *Proceedings of I-KNOW 2010, 1-3 September 2010, Graz, Austria*, pages 391–402, 2010.

Towards a User-aware Enrichment of Multimedia Metadata

Ana-Maria Manzat, Romulus Grigoras, Florence Sèdes

Université de Toulouse – IRIT UMR 5505, 118 Route de Narbonne 31062 Toulouse, France
{Ana-Maria.Manzat, Florence.Sedes}@irit.fr, Romulus.Grigoras@enseeiht.fr

Abstract. A recent trend in multimedia information retrieval systems is the integration of users, by their preferences and interests, in the retrieval process. Generally, such systems consider the user only after the query's execution, while the results' presentation. We propose to consider the user as a source of metadata, by exploiting his behaviour and to enrich the document's metadata with a usage metadata. We introduce the concept of temperature, associated to each metadata descriptor, which denotes the popularity of the multimedia document's metadata. An algorithm for the computation, the increase and the decrease of this temperature is described in details. We present also how this algorithm can be used for the enrichment of each metadata descriptor according to the user's interactions with the multimedia content and the metadata.

Keywords: user's behaviour, multimedia metadata enrichment, metadata popularity, multimedia systems

1 Introduction

Nowadays, we are constantly surrounded by multimedia contents and devices. Thus, we are continuously creating and consuming multimedia data. Usually, before creating a multimedia document, the user has an idea of which kind of information he wants to include in his document and then he searches the multimedia contents that correspond to his needs [1]. Hence, the management of multimedia documents, which includes their storage, indexation and retrieval processes, is very important.

A recent trend in the information retrieval domain is the user's integration in the retrieval process. Thus, the user's preferences, interests and behaviour are analysed and modelled in order to improve the performance of the system. This improvement is realised by providing better results to a user query and by recommending him other interesting documents accessed by other users which have similar profiles [2].

In this context, we focus on the user's integration in the metadata management process. We want to provide a solution for the metadata enrichment through their usage and through the user's interaction with the multimedia document to which they are associated. This enrichment is accomplished through the concept of *temperature* which is associated to each metadata descriptor related to the multimedia document and to the multimedia document itself. This temperature can be considered as a popularity metadata that is updated each time the document or a part of it is

consumed. Thus, more a document is consumed, the hotter it and its metadata get. In this paper we focus on the presentation of: (1) an algorithm that exploits this concept by specifying the manner in which the temperature can be increased or decreased, and (2) the algorithm's application in several scenarios.

This kind of metadata can have several utilizations in: the recommendation systems of a certain document or only a part of it; the execution of the user's query, by taking into account the document's temperature in the computation of its score; the creation of the document's resume to be displayed in the results list; the selection of video's key-frames according to the user's profile.

The remainder of the paper is structured as follows. We begin with an overview of multimedia metadata and the user's interaction in the multimedia information systems, in Section 2. Then, in Section 3, we present a metadata framework that includes the concept of temperature. The proposed solution for the metadata enrichment according to their usage is described in Section 4. Finally, some preliminary results and conclusions are given.

2 State of the art

From our daily experience, we can deduce that the best way to find certain desired information from a huge collection of documents is to look not at the information itself but rather at a much smaller and more focused set of data. In the context of multimedia retrieval systems, this concise information is the *metadata*.

The metadata can be classified in: (1) content metadata (low-level, high-level, structure, life-cycle, identification and localization and management metadata) and (2) user metadata (user interaction and user context) [3]. During the last years, the number and the heterogeneity of metadata formats increased steeply. The majority of these standards are content centred, e.g., Dublin Core, XMP, MPEG-7, TV-Anytime. In general, an information system in charge with managing and retrieving multimedia contents is composed of [4, 5]: (1) a *multimedia collection* which contains several multimedia contents; (2) a *metadata collection* which contains information about the media characteristics (e.g., size, name) and their contents; (3) an *indexation engine* which includes several *indexing algorithms* to be applied on the multimedia collection in order to enrich the metadata collection. The indexing algorithms automatically applied on the multimedia contents produce metadata encoded into different standards and formats. These metadata are further employed in the retrieval process. This makes the management of the metadata and the query execution a very important task to be realised by a multimedia information retrieval system.

In [6], the metadata is presented in the centre of the multimedia document lifecycle, which makes the metadata creation and management a very important issue in the handling of multimedia documents. In addition, the metadata is consumed and produced at every stage of the document lifecycle [7]. This leads to a constant user interaction with the metadata, in a direct or indirect manner. Thus, the user can be considered as an auxiliary source of metadata, which could improve the metadata obtained from the indexation process. He can produce metadata in an explicit or implicit manner. By attaching annotations and tags [8] to multimedia documents the user is creating explicit metadata. The inconvenient of using this approach for

enriching the metadata is that, usually, the users are busy and annotating documents demands a lot of time and effort, and, consequently, the created metadata is very poor.

In order to obtain more information from users, some other strategies have been developed. One of them is to analyze the user's behaviour and to infer his interests [9] and his intentions [10]. These interests are used, for example, to adapt the presentation of the multimedia documents [11] and of the query results list [12] or to enrich the user query [13].

Apart from the implicit and explicit metadata we can consider also the *attention* [14] and *usage metadata*. This information is associated with the document and not with the user, as for the interests. In [15], the authors propose an algorithm for determining such metadata. The authors determine the popularity of multimedia documents in accordance with the number of users that access the documents. The authors attach this popularity information to entire documents, and not to parts of documents. Also, this information is computed in function of the number of users that access the document, and users' interests and preferences are not taken into account.

The behaviour of the user is also used in other domains, such as the adaptive hypermedia domain [16], where the presentation of the documents is modified according to the user, and the user-centric multimedia databases [17], where the user behaviour is captured through the analysis of the query logs.

As could be noticed, the research fields where the user is taken into account are very different and vast, from the presentation's adaptation to the multimedia information retrieval. The user's behaviour is studied in order to adapt the documents or the query's results, but the metadata associated to the multimedia contents are not enriched. Before presenting our approach for the metadata enrichment, we will describe in the next section the metadata framework developed in order to incorporate the notion of temperature.

3 Metadata Framework

In the domain of metadata interoperability many studies were carried out [18, 19, 20] in order to provide the possibility to use in the application the different and heterogeneous metadata standards and formats, and also to allow the exchange of metadata between systems and applications. All these approaches are focused on the interoperability problem, and they do not offer any possibility to enrich the metadata in function of their usage.

In this paper, we do not focus on the metadata model, but rather on the temperature concept. In order to illustrate this concept we present a preliminary metadata framework that allows the integration of existing metadata models and provides the possibility of enriching them through the usage. Our approach takes into account the users and their behaviour regarding the consumption of the retrieved documents and their associated metadata. The notion of temperature can be applied to any hierarchical metadata model.

In our model, Fig 1, we couple each multimedia content with a unique metadata file, that contains the whole set of metadata related to that document. The link between the two documents is done through the *documentSrc* attribute from the

Meta_Document metadata. As a multimedia document can be composed by different media types, its metadata can be formed by many *Meta_Documents*, each one corresponding to one media from the multimedia content. Each *Meta_Document* is divided in two parts: (1) *General_Metadata*, which corresponds to the general metadata, such as the life-cycle and the identification metadata (e.g., the creator, the description); (2) *Media_Metadata*, which corresponds to the media specific metadata.

In order to be as generic as possible and to allow the integration of different existing metadata standards, we decomposed the two parts presented above in *Units*. Each *Unit* represents a metadata element, e.g., the author. It has as attributes the name of the metadata, its type and, eventually, a definition or a reference to its definition that is provided into a thesaurus. Depending on the application's needs, a *Unit* can be decomposed in one or more *Units*. The actual value of each *Unit* is specified in a different element, *Value*, which has as attribute the *source* of the value, e.g., the metadata standard that provided the metadata element.

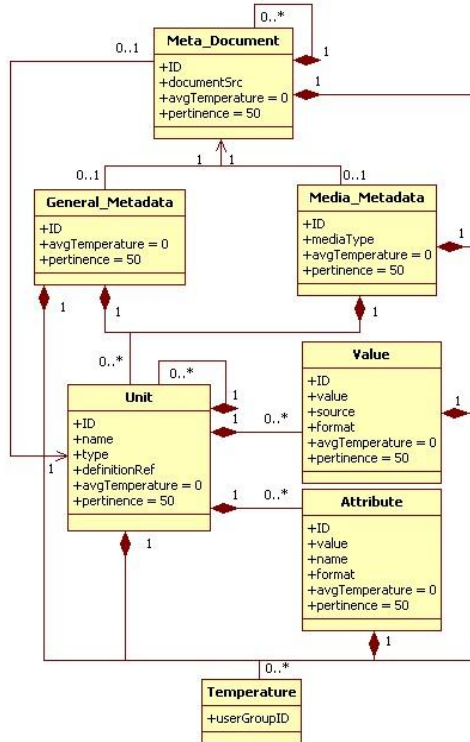


Fig. 1. Metadata framework

The usage metadata, *the temperature*, is associated to each element of the metadata format presented above. More precisely, every metadata element from the proposed framework has associated two kinds of temperature: (1) one computed for each group of users that interacts with the multimedia content, and (2) an average one for each metadata element, that is computed in function of each groups' temperatures.

In this paper, we do not focus on the determination of the users' groups that we use in our approach. We consider that these groups are already established and that they can evolve over time. In our work, the different groups can be disjoint or not, a user can belong to at least one group and over time he can migrate from one group to another. An approach for the creation of such groups, based on the users' interests, is defined in [21]. The advantage of using users groups is that in this way the temperature can be used for personalisation purposes. The algorithms presented in the reminder of the paper work regardless the number of user groups defined; it works as well for single users.

4 Metadata enrichment

We consider the user as an important source of implicit metadata, because he can produce metadata by interacting with the multimedia documents he obtains as results to his query. In our proposal we focus on exploiting the user's behaviour.

In order to be able to respond to as many users' queries as possible, in an information retrieval system many different indexing algorithms are applied. Thus, the multimedia metadata obtained are heterogeneous, from simple low-level features to more complex semantic high-level features. Usually, not all the generated metadata are used in the retrieval process. There are some metadata that are used more often than others. For this reason, we propose to enrich the metadata obtained after the indexation process with the concept of *temperature*. Thus, the more the documents or their associated metadata are used, the hotter they are.

We have attached the temperature to (1) the multimedia document (at the *Meta_Document* level in the metadata framework presented in the previous section) and also to (2) their associated metadata (the temperature attached to each metadata element in the proposed framework). This popularity metadata can be used, for example, in the query process. In the execution of a query, the popularity metadata is taken into consideration in the computation of the results' score. This way the popular documents and segments of documents are better ranked.

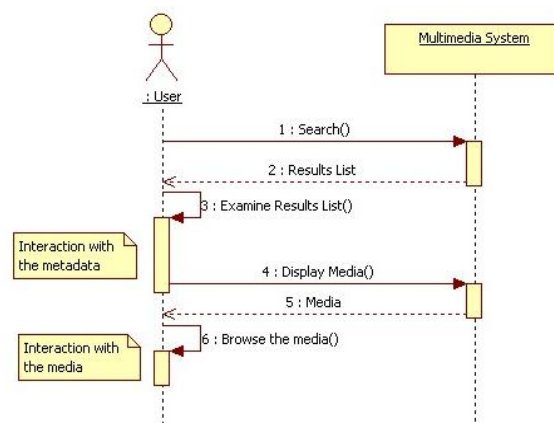


Fig. 2. User's actions in an information retrieval system

The above picture resumes the actions that a user makes when interacting with an information retrieval system. Based on these considerations, we propose to realise the metadata enrichment by taking into consideration the user's interaction with the metadata associated to query results (step 3 in Fig. 2) and with the multimedia document (step 6 in Fig. 2). First, we describe in Section 4.1 the metadata enrichment algorithm and then, in the next sections we present its concrete application based on the user's interaction with the results list, Section 4.2, and with the multimedia document, Section 4.3.

4.1 Metadata enrichment algorithm

Independently of the manner the decision of the increasing of the temperature is taken, the temperature is computed for each period of time Δt and it depends on the number of users that have consumed the metadata in that period. The temperature is defined as t in \mathcal{R} with $0 \leq t \leq 100$. The initial value of the temperature of all the documents and of their associated metadata is 0. The algorithm used for the increase of the temperature is presented in Table 1.

The parameters of the proposed algorithm are: the metadata whose temperature has to be increased, the number of users that consumed the metadata and the identifier of the group these users belong to. The first step of the algorithm is the computation of the metadata's temperature corresponding to the user group received as parameter. Afterwards, the average temperature of the metadata element is computed as an *arithmetic mean* of the temperatures associated to this metadata, corresponding to each user group in the system. For the computation of this average temperature can be use also *weighted mean*.

Each time the temperature of a metadata is modified using the *increaseTemperature* method, this modification is propagated to all its children metadata. The propagation method is presented in Table 2. It follows the same steps as the first algorithm. The temperature of each child metadata is changed with a value that is directly proportional with the variation of the temperature at the first level and with the level in the metadata hierarchy where the current element is. This propagation can be limited to a certain level in the hierarchy, specified by the *MAXLevel* constant

We apply the same reasoning for the propagation of the temperature to all the ancestors of the metadata element that initiated the process of temperature increasing. The propagation method is presented in Table 3. In the computation of the new temperature we follow the same rules as for the propagation to the child elements.

Table 1. The algorithm for the increase of the metadata's temperature

Algorithm 1: increaseMetadataTemperature

Input: The metadata, MD , whose temperature has to be increased, the number of users, n , who used the metadata, the identifier of the group, gID , to which the users belong.

Output: the metadata with the temperature increased, for all children and ancestors.

$\Delta temp \leftarrow computeGroupTemperature(MD, gID, n);$

$md \leftarrow setGroupTemperature(MD, gID, \Delta temp);$

$setHistory(MD, gID, \Delta temp);$

$avgTemp \leftarrow computeAvgTemperature(MD);$

```

md ← setAvgTemperature(md, avgTemp);
if MD has children then
  └ md ← propagateTemperatureDown(md, gID, Δtemp, l);
if MD has parent then
  └ md ← propagateTemperatureUp(MD.parent, gID, n, Δtemp, l);
return md;

```

Table 2. The algorithm for the propagation of the metadata’s temperature to all its children

Algorithm 2: propagateTemperatureDown

Input: The metadata, MD , for which we want to increase the temperature of the children; the identifier of the group, gID , for which the temperature has to be increased; the temperature $\Delta temp$, that is used for the computation of the new temperature; the $level$ of the recursive call

Output: the metadata with the temperature of all its children increased

```

inc ← computeTemperature(Δtemp);
foreach child of MD do
  └ md ← md U setGroupTemperature(child, gID, inc);
  └ setHistory(child, gID, inc);
  └ avgTemp ← computeAvgTemperature(child);
  └ md ← md U setAvgTemperature(child, avgTemp);
  └ if level < MAXLevel then
    └ └ md ← md U propagateTemperatureDown(child, gID, inc, level+1);
return md;

```

Table 3. The algorithm for the propagation of the metadata’s temperature to all its ancestors

Algorithm 3: propagateTemperatureUp

Input: The metadata, MD , for which we want to increase the temperature of the ancestors; the identifier of the group, gID , for which the temperature has to be increased; the temperature $\Delta temp$, that is used for the computation of the new temperature; the $level$ of the recursive call

Output: the metadata with the temperature of all its ancestors increased

```

if MD ≠ null then
  └ inc ← computeTemperature(Δtemp);
  └ md ← md U setGroupTemperature(MD, gID, inc);
  └ setHistory(MD, gID, inc);
  └ avgTemp ← computeAvgTemperature(MD);
  └ md ← md U setAvgTemperature(MD, avgTemp);
  └ if level < MAXLevel then
    └ └ md ← md U propagateTemperatureUp(MD.parent, gID, inc, level+1);
return md;

```

Table 4. The algorithm for the decrease of the metadata’s temperature

Algorithm 4: decreaseTemperature

Input: The metadata, MD , for which we want to decrease the temperature; the $history$ of the temperature increasing over time associated to the MD and its children; the number N of time intervals Δt that we want to undo from the temperature computation

Output: the metadata with the temperature of all its ancestors increased

```

foreach child of MD do
  └ for  $i=1$  to  $N$  do

```

```

    dec = getLastHistoryValue(child, history, gID);
    history ← removeLastHistoryValue(child, history, gID);
    md ← md U setGroupTemperature(child, gID, dec);
  | avgTemp ← computeAvgTemperature(child);
  | md ← md U setAvgTemperature(child, avgTemp);
  | md ← md U decreaseTemperature(child, history, N);
  return md;

```

If the temperature is augmented all the time, then after a certain period, the temperature of all metadata will attend the maximal value. In order to avoid this, the temperature of the unused metadata is reduced with a $\Delta temp$ proportional with the number of users who has consumed them in the period before. This is done for all the metadata elements and for all the users' groups. In order to realise this operation at each recalculation of the temperature of a metadata element, the variation is stored in a history file. The algorithm used for decreasing the temperature is illustrated in Table 4. This decreasing process can be applied after each time period Δt , or after a certain number of intervals Δt .

4.2 The interaction with the results list

In a classical information retrieval system the user sends his/her query to the system and retrieves some results. These results are ranked in function of the score they have obtained after the query execution over the metadata collection. In a typical results list, each result is composed of a link to the multimedia content and the metadata associated to this content. The system's human interface displays the results as a list. The metadata associated to each result is presented as a collapsed tree.

In order to find the most relevant document for him, the user examines first the metadata associated to the documents in the result list. This action consists, en fact, in the expansion of the displayed metadata tree, until a certain level. Another way of collecting this kind of information, in a less intrusive manner, could be the gaze tracking [22]. It can be considered as a metadata consumption and it has an influence on the metadata's temperature.

In order to illustrate the temperature's computation we take into consideration fist scenario of utilisation. Suppose that only a part of the metadata related to the results are displayed (e.g., the General_Metadata in the proposed framework) and that the user has the possibility to access the rest of the result's metadata. If for the same document, in a certain time interval Δt , several users, belonging to the same group, have accessed the same additional metadata by expanding it, then the temperature of the expanded metadata is augmented with a value $\Delta temp$ proportional with the number of users (n) which have consumed them. Only the temperature corresponding to the group to which the users belong will be recalculated. More precisely, the *Temperature* element with the *userGroupID* equal to the users' group ID will be modified for all the metadata elements displayed. According to the algorithm previously described, this change in temperature will be propagated to all the children of the expanded elements and to their ancestors as well.

In order to illustrate in more details this enrichment process, we consider the following situation: a multimedia information retrieval system where we have identified several groups of users. For readability reasons, in the examples we present in this paper we will consider only two groups. In this system, a user belonging to the

first group obtains the image DSC_2249.jpg as a result to a certain query. The system displays the metadata description in the form of a collapsed tree, as the one presented in Fig.3 a). This user expands the General_Metadata element until a certain level, as displayed in Fig.3. b). In the same time other 9 users from the same group access the same metadata. Thus the metadata associated to this image will be increased. The function *increaseTemperature* is applied for the following metadata elements: <Value source="DC">; <Value source="EXIF"> and <Unit name="creationDate">. For the last two elements, the temperature will be increased with a smaller value than the first one because they were not expanded until the last leaf.

```

- <Meta_Document documentSrc="DSC_2249.jpg"
  - <General_Metadata>
    - <Unit name="author">
      + <Value source="EXIF">
        - <Value source="DC">
          <value>Ana-Maria Manzat</value>
        </Value>
      </Unit>
    + <Unit name="creationDate">
  </General_Metadata>
+ <Media_Metadata mediaType="image">
</Meta_Document>
a)
- <Meta_Document documentSrc="DSC_2249.jpg"
  - <General_Metadata>
    - <Unit name="author">
      + <Value source="EXIF">
        - <Value source="DC">
          <value>Ana-Maria Manzat</value>
        </Value>
      </Unit>
    + <Unit name="creationDate">
  </General_Metadata>
+ <Media_Metadata mediaType="image">
</Meta_Document>
b)

```

Fig. 3. a) Metadata displayed with a query result; b) The same metadata after the user interaction with it

4.3 The interaction with the multimedia document

After the study of the results list, the user chooses a multimedia document and begins to interact with it: he explores the document, he studies in more details a part of the multimedia content, he spends an important period of time examining the document, etc.. This behaviour illustrates his interest in the document and in its compounds.

We augment the temperature of the metadata which correspond to the multimedia document's compound the user is interested in. When the temperature of a component is changed the temperature of the document and of the other metadata elements that describe the component are modified as well. The information is also propagated to the higher levels in the metadata hierarchy.

In order to illustrate this metadata enrichment, we can consider the SMIL presentation from Fig. 4. This presentation is composed of a video and an audio content and the presentation's slides as images. The organization in time of the presentation and the eventual audio and video segments are presented in Fig. 4.

For this example, we also consider an information retrieval system where two users' groups were identified. Several users belonging to the same group have used the system in the same time and they obtained the same SMIL presentation as a result to their different queries. They all have selected the presentation and have watched it from the 1'20'' until de 4'50''. In this case, the temperature of the metadata associated to all the multimedia contents displayed in this period of time will be modified. From the timeline presented in Fig. 4 we can deduce that the video segments *seg_Video₁* and *seg_Video₂*, the audio segments *seg_Audio₁*, *seg_Audio₂* and *seg_Audio₃* and the images *img₂.jpg* and *img₃.jpg* are candidates for having the

temperature modified. At this point several strategies can be established for choosing the segments to use for the metadata enrichment. For example, if the compound was watched for at least half of its length, then its temperature will be modified. In this case, the temperature of the audio segments seg_Audio_1 and seg_Audio_3 will not be modified, because they were listened for less than their length. Another possibility would be to increase the temperature for all the segments that were displayed, with a value proportional with the time that they were watched.

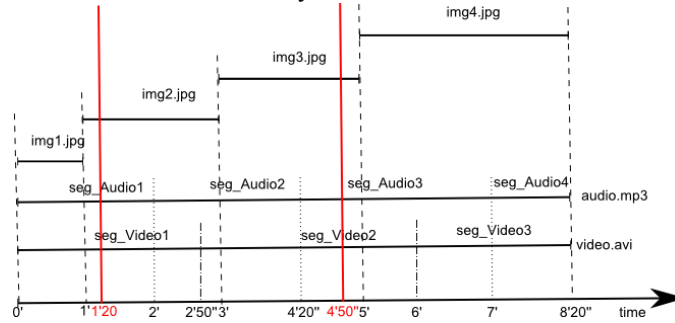


Fig. 4. The timeline structure for the SMIL presentation

Through the proposed algorithm for the temperature increasing, the temperature of the entire presentation will be increased, as a consequence of the consumption of a part of it. We can note that the more a document is watched, hotter it gets.

In the next section, we present some possible utilizations of the temperature concept in the context of a broadcast use case.

5 Implementation and discussions

In order to validate our proposal, we have applied the concept of temperature to a web site. In this case we used the algorithm in function of the users' interaction with the multimedia content. We consider a page of the site as a document and we associate to it metadata. The users' interactions with the web page (e.g., clicks) are collected into a database. For the tests effectuated we considered only a group of users. We have instantiated the metadata framework by using the XML and XSD technologies and the algorithm was implemented in Java.

The Fig. 5 shows the obtained results for a web page temperature computation. The results show that the time granularity is very important in the application of our algorithm. For the same users interactions with the page the temperature obtained for the whole page is different in function of the strategies employed: compute the temperature each 24 hours, each 12 hours, each hour or less than an hour. The decrease strategy is also important when the time granularity chosen for the computation of the temperature is small. These choices are use case dependent. The curves in the Fig. 5. show that these considerations have an influence on the evolution of the temperature. Thus, making the good choice is important in the progress of the temperature.

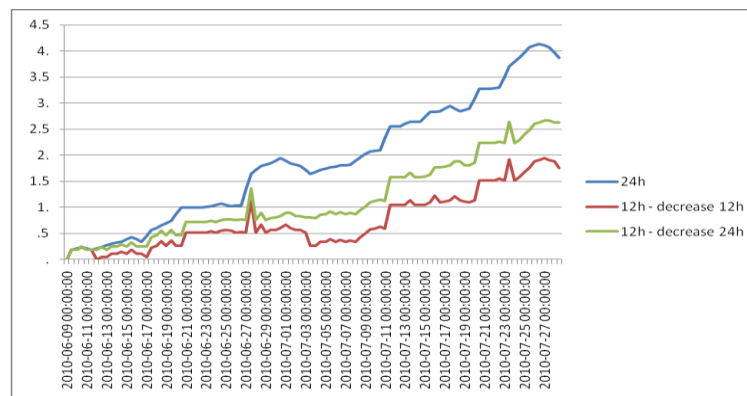


Fig. 5. Experiments results

6 Conclusion

In this paper, we have presented a modality of multimedia metadata enrichment based on the users' interaction with the multimedia content and with their associated metadata. This enrichment is done in two steps: (1) in function of the users' interaction with the metadata and the results list and (2) in function of the users' behaviour with the multimedia document.

We intend to implement and test our proposal in the context of the LINDO project (*Large scale distributed INDEXation of multimedia Objects*) (<http://lindo-itea.eu/>) in order to determine the best parameters of the algorithm (e.g., time granularity, decrease strategy, the level of propagation). These parameters cannot be set without the intervention of the user, thus we will realise some qualitative interviews with a set of volunteers. In a first time we will implement the second scenario for the computation of the temperature (presented in Section 4.3). After the specification of the parameters we will take the experiments a little further, by using the temperature for the metadata management in a distributed system [23].

Acknowledgments: This work has been supported by the EUREKA Project LINDO (ITEA2 – 06011).

References

1. Hardman L., Obrenovic Z., Nack F., Kerhervé B., Piersol K., Canonical Processes of Semantically Annotated Media Production. In: *Multimedia Systems Journal*, 14(6): 327-340, (2008).
2. Candillier L., Jack K., Fessant F., Meyer F., State-of-the-Art Recommender Systems. In *Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modeling*, Chevalier M., Julien C., Soule-Dupuy C. (Eds.), IGI Global, 1-22, 2008
3. Pereira F., Vetro A., Sikora T., Multimedia Retrieval and Delivery: Essential Metadata Challenges and Standards, *Proceedings of the IEEE* 96(4), 721 – 744, (2008)
4. Buckland M. K., Plaunt Ch. On the construction of selection systems. *Library Hi Tech*, 12, pp. 15-28 (1994).

5. Lancaster F. W.. Information Retrieval Systems. Wiley, New York (1979)
6. Smith J. R., Schirling P., Metadata Standards Roundup, In IEEE MultiMedia, 13(2): 84–88, (2006).
7. Kosch H., Boszormenyi L., Doller M., Libsie M., Schojer P., Kofler A., The Life Cycle of Multimedia Metadata, IEEE MultiMedia 12(1): 80-86, (2005)
8. Kahan, J., Koivunen, M.-R., Prud'Hommeaux, E., Swick, R. R., Annotea: an open RDF infrastructure for shared Web annotations. Computer Networks, 32(5):589–608, (2002).
9. Kelly D., Teevan J., Implicit feedback for inferring user preference: A bibliography, in *SIGIR Forum*, volume 37, pp. 18–28, (2003).
10. Lux M., Kofler Ch., Marques O. A classification scheme for user intentions in image search. In the 28th of the international conference extended abstracts on Human factors in computing systems (CHI EA '10), pp. 3913-3918, ACM, (2010)
11. Plesca C., Charvillat V., Grigoras R., User-aware adaptation by subjective metadata and inferred implicit descriptors, in *Multimedia Semantics—The Role of Metadata*, vol. 101 of *Studies in Computational Intelligence*, pp. 127–147, Springer, (2008)
12. Kofler Ch., Lux M., Dynamic presentation adaptation based on user intent classification. In the 7th Int. conference on Multimedia (MM '09). ACM, pp. 1117-1118, (2009)
13. Zayani C., Péninou A., Canut M.-F., Sèdes F. An adaptation approach: query enrichment by user profile. In *Signal-Image Technology & Internet-Based Systems (SITIS 2006)*, Hammamet - Tunisie, pp. 24-35, IEEE, (2006).
14. Memmel M., Dengel A., Sharing Contextualized Attention Metadata to Support Personalized Information Retrieval, In the Int. Workshop on Contextualized Attention Metadata: Personalized Access to Digital Resources, ACM/IEEE, pp. 19-26, (2007)
15. Brunie L., Pierson J.M., Coquil D., Semantic collaborative web caching, In the 3rd Int. Conference on Web Information Systems Engineering, (2002)
16. Brusilovsky P.. Adaptive hypermedia. in *User Modeling and User-Adapted Interaction*, 11(1-2):87–110, (2001).
17. Limam L. , Coquil D, Brunie L., Kosch H., Query Log Analysis for User-Centric Multimedia Databases, in *The 2008 International Conference on New Media Technology, I-Media'08*, pp.441-444, (2008).
18. Arndt R., Troncy R., Staab S., Hardman L., Vacura M., COMM: designing a well-founded multimedia ontology for the web. In *ISWC+ASWC*, pp.30-43,(2007).
19. Brut M., Laborie S., Manzat A.-M., Sèdes F., A Generic Metadata Framework for the Indexation and the Management of Distributed Multimedia Contents. In *New Technologies, Mobility and Security*, pp. 1-5, IEEE Computer Society, (2009)
20. Saathoff C., Scherp A, Unlocking the semantics of multimedia presentations in the web with the multimedia metadata ontology. In *WWW 2010*, pp. 831-840
21. Tchuenta D., Canut M.F., Baptiste Jessel N., Péninou A., El Haddadi A., Visualizing the evolution of users' profiles from online social networks », *The 2010 IEEE International Conference on Advances in Social Networks Analysis and Mining, ASONAM'10*.
22. Funahashi T., Fujiwara T., Koshimizu H., Face and eye tracking for gaze analysis. In *Int. Conference on Control, Automation and Systems (ICCAS '07)*, pp. 1337-1341, (2007).
23. Laborie S., Manzat A.-M., Sèdes F., Managing and querying efficiently distributed semantic multimedia metadata collections, In *IEEE MultiMedia special issue on multimedia-metadata and semantic management* 16, 4, 12-20, (2009)

A Model of Relevance for Reuse-Driven Media Retrieval

Tobias Bürger

Salzburg Research, Salzburg, Austria
tobias.buerger@salzburgresearch.at

Abstract. An often criticized fact in multimedia retrieval is, that user needs are not appropriately taken into account. Both knowledge about how end users search and how they assess the relevance of retrieved multimedia objects can provide invaluable hints for the design of multimedia retrieval systems. This paper reports on an end user study on multimedia retrieval behavior of media professionals who intend to reuse media objects in media productions. We present a conceptual model which contains empirically validated information on how users in the media production domain search for content to be reused and how relevance is assessed by them. Finally we sketch how this information can be used to improve ranking of media objects in multi-faceted retrieval scenarios.

1 Introduction

The amount of multimedia content available on the Web and the amount of professionally produced content stored in local or commercial databases grows every day: While there is a steady growth of professionally produced content available on the Web, a continuous blurred shift happens between consumers and producers of content, which share huge amounts of user generated content. This ever growing amount of content offers a great potential for reuse.

Reuse of multimedia content, i.e., every kind of use of content which has been used in a certain context before, is an ongoing challenge and is mostly not very well supported by existing tools and approaches. Supporting reuse can however provide significant improvements in the way how content is created, including increased quality and consistency, long-term reduced time and costs for development, maintenance or adoption for changing needs [25]. As our recent observations in the domain of media production reveal, only approximately 30 percent of the produced content is based on already existing content. We furthermore revealed barriers leading to this low figure which include reasons such as “*relevant content cannot be found*”, that “*it is sometimes faster to build content from scratch*”, that “*content is not adaptable to new situations*”, or that “*the legal situation is either unclear or does not allow reuse*”.

One of the identified barriers of reuse includes the problem of findability of content which maps the problem of reusability to the solution space of multimedia retrieval. One ongoing problem there is the gap between the research done

and the practical end user needs in different contexts as many approaches take a system-centric approach focusing on technical aspects of multimedia indexing and retrieval [11] and lack a theoretical background of the characteristics of users and their needs for the design of these systems [24]: In order to support efficient retrieval, matches to a request have to be presented in an appropriate order, minimizing the distance between actual features of the content and expected features by the user.

To bridge the aforementioned gap, user-oriented studies were conducted which analyzed the indexing practices and retrieval needs of typical end users. Some of these studies resulted in analytic models which formalized the characteristics of user requests and search patterns of these users (cf. Section 2). While the main aim of these studies was to conceptualize and bridge the Semantic Gap [4, 5], the judgement of relevance for the selection of media objects has so far not been researched to a great extent. In order to overcome this situation, we present a conceptual model which contains empirically validated information on how users in the professional media production domain search for content to be reused and how relevance is assessed by them.

In this paper we examined a typical retrieval task in the studied environment: A media professional is engaged in a design task and intends to search for images to reuse in his current production. He starts with formulating his needs in an image request and receives a result set of images. After that, he checks the topicality of the images in the retrieved result set and starts browsing. If either the topicality of the returned images does not match his needs or if he is unsatisfied with the results investigated during browsing, he reformulates his query. Otherwise he applies his relevance criteria and finally selects an image which he uses in his design task.

Understanding how and why users search for and select multimedia content to be reused can provide invaluable hints for the design of multimedia retrieval systems. Therefore the research leading to this paper aimed to address the following research questions:

1. *“Which factors do users use to search for reusable media objects?”* and
2. *“Which relevance criteria do users apply when searching for media objects to be reused?”*

In order to answer these questions, we built a basic model containing factors used in search and relevance assessment. To do so we analyzed prior literature and conducted interviews with design professionals. Subsequently we empirically validated the model through an end user survey and assessed the validity and importance of the factors in both tasks.

The remainder of this paper is structured as follows: Section 2 presents the background and motivation for the model. Subsequently the model is discussed in detail in Section 3: We present insights from prior literature and the basic mode. Section 4 details the validation of the model and presents the results of the conducted survey. Finally, Section 5 concludes the paper.

2 Research Background and Motivation

Since the 1960s research has been reported which analyzed the indexing practices and retrieval needs of typical end users. The work in this area can be divided into conceptual frameworks of image indexing (cf. [5, 9, 14, 22, 27]), which are mainly situated in cognitive psychology and models of user’s multimedia retrieval needs (cf. [1, 2, 10, 15, 16, 19, 29]). The intention of these conceptual frameworks was to provide groundings for the manual and automatic image indexing and the description of the semantics of multimedia content in general and images in particular.

The earliest model was provided by Panofsky [22] who recognized three types of subject matter, for instance, primary subject matter which requires no interpretative skills, secondary subject matter which necessitates an interpretation, and tertiary subject matter (“iconology”) demanding high-level semantic inferencing done by the user. Subsequent work by Shatford [27] simplified the three levels of Panofsky into generic, specific and abstract. Additionally Shatford introduced the distinction between “of-ness” and “aboutness” of a picture. A simpler model was provided by Greisdorf [9] who recognized three levels which correspond to visual primitives (e.g., color or shape), logical features (e.g., objects or events) and inductive interpretation (e.g., abstract features). Joergensen et al. have further refined the model by Shatford which resulted in the so-called visual indexing pyramid [14]. A newer model by Enser et al. builds on Joergensen’s notion of semantic facets of images and furthermore takes the combination of semantic content of an image and its context into account [5].

Besides the development of these models, studies were conducted which investigated user retrieval needs. Their intention was to inform other research strands which type of semantics can be extracted from multimedia content. Most of these studies revealed, that a user is typically interested in high-level semantics which are hard to derive based on automated approaches and which are often highly subjective. An analysis of early studies in this area by Jörgensen revealed a wide variation in subject foci and terminological speciality and also that the majority of requests were for specific events or objects, especially for specific, named features [16]. This observation was also made in [1, 19]. Other studies reported an emphasis on generic or affective visual features (cf. [2, 10, 15]). Validations and comparisons of these studies can be found in [1] and [29].

Especially in multimedia retrieval, uses and needs vary considerably, as media objects are used in a variety of domains (e.g., media production, art, journalism, or medicine) for different purposes. Furthermore, needs of professionals and needs of end users are in many cases different: End users are motivated by leisure, while professionals search for images for inspiration, reuse, or other reasons. As relevance differs considerable based on the situation of the user and his needs, domain specific investigations have been made: User needs in domain specific collections and for specific user groups have been conducted, e.g. for web images (cf. [6, 7, 15, 23]), for historical images (cf. [1, 2]), for medical images (cf. [17]), or for image retrieval in a journalistic context (cf. [11, 12, 18, 19, 29]). User needs of media professionals such as graphic-, or game- designers, and especially the

influence of the intention to reuse, were, however, rather unexplored up till now. Our aim was therefore to develop a conceptual model that dimensions relevance in reuse-driven multimedia retrieval in which people search for content to reuse.

3 A Model of Relevance in Media Reuse

This section presents a conceptual model which reflects factors which influence the relevance of multimedia content for end users in the particular situation in which they look for content to reuse. The model is based on insights from existing literature, on motivation and barriers for reuse, and on user studies which investigated multimedia retrieval in professional domains. Prior insights were validated and supplemented with expert interviews conducted with media professionals.

3.1 Insights from Existing Literature

Relevance is a central concept in information retrieval and there used as a measure for retrieval and to judge the effectiveness of an information system. Ingwersen and Järvelin suggest that relevance is a multidimensional cognitive concept whose meaning is largely dependent on searcher's perceptions of information and their own information need (cf. [26] as cited in [13]). While content features are in most situations the most appropriate indicators for relevance, non-content features of documents can give valuable hints, too. This is especially true for multimedia retrieval in professional environments in which relevance is not only based on topicality but also on visual, qualitative, situational and other contextual factors as reported in previous studies (cf. [1, 2, 15, 18, 19, 21, 28, 29]): Markkula investigated retrieval of images in a journalistic context [19]. His observations clearly indicated the diversity of relevance criteria which were applied and the situational nature of their relevance judgements. The primary criteria which is applied by journalists to assess relevance is topicality. Secondary criteria are technical properties, technical quality and biographical criteria. Images which are technically good, are current or were not recently published are being considered as relevant in this domain. The cost of images has also been identified as an important criterion. Even though expressive and also aesthetic criteria, such as color and composition, were used for search by journalists, they played the most important role in the final selection phase. The critical criteria to reject or to accept an image depended on earlier selections: An image already chosen for a page and nearby pages or used recently in a different or the same newspaper restricted the possibility to use other similar images. According to the journalists, the goal was to make the illustration of the page attractive, balanced, and dynamic. This was achieved by using images of different types (e.g., horizontal and vertical photos, portraits, group photos, action, or themes) and with different visual features.

Another study was conducted by Choi and Rassmussen who investigated relevance criteria in image retrieval of historic art images [2]. They identified

nine relevance criteria together with 8 non-visual (descriptive) attributes for highly relevant images:

- **Time frame:** The time period of the image.
- **Accuracy:** The image accurately presents what the user is looking for.
- **Topicality:** The image is related to the user’s task.
- **Completeness:** The image contains the necessary details.
- **Accessibility:** The availability of the image, as in the ease of obtaining the image and the means by which image information can be accessed.
- **Appeal of information:** The image is interesting and appealing to the user.
- **Novelty:** The image is new to the user.
- **Suggestiveness:** The image generates new ideas and insights for the user.
- **Technical attributes:** These attributes include mood, emotion, point of view, or color.

Furthermore the following descriptive attributes were identified as being important in the assessment of relevance in image retrieval in a historical art context: creation date, notes, subject descriptors, the title of the image, the source repository, the source collection, the medium, and the name of the creator.

Eakins’ study done in 2004 investigated features used in image search but not how these features affect relevance [3]. His results showed that despite of topicality, technical quality is the most important criterion in search. Besides that, he identified the following features as the most relevant:

- **Low level features** such as colour, texture or shape.
- **Technical quality features** such as sharpness.
- **Semantic content** containing general and specific semantic terms.
- **Abstracted features** such as *contextual abstraction* which refers to non-visual information derived from the knowledge of the viewer, *cultural abstraction* which refers to aspects which can only be inferred based on a cultural background, *Emotional abstraction* which refers to emotional responses triggered by the image, and *technical abstraction* which refers to aspects requiring specific technical expertise to interpret.
- **Metadata** such as the image type (e.g., photographic, painting, or scan).

Othman was the first to investigate image retrieval in a creative media-related context [21]. The relevance criteria she discovered for the domain of media production were similar to the ones by Choi and Rasmussen [2] but with a different mean importance of each criteria: Technical attributes were considered the most important, followed by completeness and topicality. Furthermore relevant images had to be qualified for processing and should not require any authentication. Most images in her study involved analysis and image manipulation, and thus the majority of users rated technical attributes as the most important relevance criteria. Technical criteria included resolution, size, color, and dimension. Topicality and completeness ranked second and third which indicated that images must be right on the topic and have all the objects specified. Time frame was

an important criterion for specific tasks. A further novel insight from her study was that the images which were judged as relevant and met their intended use ranged from one object in the image retrieved to the whole image itself.

The most recent study which reports on end user needs in image retrieval in a journalistic context was published by Westman and Oittinen [29]. It featured 47 criteria for image selection which were partially based on insights from Markkula [19]. The criteria were grouped along the following dimensions:

- **Information and content** (e.g., *information content* or *story of the photograph*)
- **Visual and compositional features** (e.g., *visual features, composition, or lighting*)
- **Technical features** (e.g., *technical error, sharpness, or physical size*)
- **Abstract and affective factors** (e.g., *movement and dynamicity, mood, expression of the person*)
- **Metadata and associated information** (e.g., *recentness, source, or importance*)
- **Publication context** (e.g., *compatibility with the headline, publication section, or importance of the article*)
- **Workflow and other actors** (e.g., *timetable or possible print quality*)
- **Practices and feedback** (e.g., *image selection practices or feedback from readers*)

According to Westman’s and Oittinen’s studies, several types of criteria were used in the relevance assessments made. Contextual factors (such as publishing section or layout of the page) formed a selection frame for suitable images. Topicality was identified as a necessary but insufficient criterion for relevance, used mostly as a starting point. Compositional and informational criteria followed in later stages of the process. The final selection criteria were dynamic, activated by comparisons of retrieved images and based on the characteristics and differences between them such as dynamic elements or sharpness. Final selection criteria also were preferential or reactive in some situations; selections were based on personal impressions of images being, for instance, more interesting than others. Furthermore several implicit criteria were employed in the image selection process. Unless otherwise asked, the image retrieved was as recent as possible and, if search is carried out across multiple archives, retrieval from the own archive was preferred. Constraints such as price, previous publication, recentness and presence of other images in a product also influenced the selection. Their results revealed that the most important relevance criteria were related to the informational content of the image. Several abstract and affective criteria also influenced the selection strongly. Least important were feedback and reactions from others. Various factors related to the eventual publication context of the image were considered important which means that often not the best matching image according to a query was used but the one matching the context most. A large number of individual criteria affected image selection strongly. Technical factors were identified as not being as crucial as previously thought by Markkula and Sormunen [19].

3.2 Basic Model

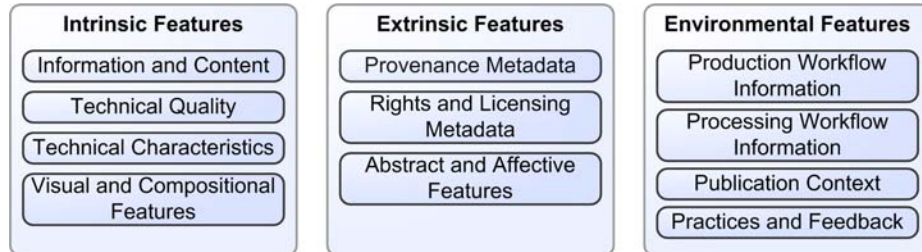


Fig. 1. A Conceptual Model for Reuse-Motivated Relevance Assessment

Our conceptual model, which is depicted in Figure 1, contains factors which are typically used by media professionals for search and/or to assess relevance of a media object. To create it, we refined and extended prior literature, validated and supplemented it with context specific interviews:

1. First interviews with designers, art directors, and researchers with experience in multimedia content creation and reuse were conducted.
2. Secondly, the qualitative data gathered from the expert interviews and prior observations were analyzed by transcribing and coding into meaningful expressions which were then classified.
3. Thirdly the results were systematically analyzed according to statistical theory.
4. Finally the analysis resulted in a cluster of themes, each containing a grouped set of factors that influence reuse. All of these factors were constantly emphasized by media professionals.

The derived factors are arranged into clusters which are partly based on the categorization schema from Westman [29] which we further extended and tested in our survey. The clusters are further arranged into *intrinsic and extrinsic features* and *environmental factors*. The following clusters are part of the model:

– **Intrinsic Features**

1. **Information and content (I)** which includes features such as *topicality, completeness, or conveyed message*.
2. **Technical Quality (TQ)** which contains features such as *sharpness, or technical error*.
3. **Technical Properties (TF)** which contains intrinsic technical features such as *resolution or size*.
4. **Visual and compositional features (V)** contain features such as *composition, color, angle, or other visual features*.

– **Extrinsic Features**

5. **Abstract and affective features (A)** include *expression, dynamicity, eye catching ability, or mood*

6. **Provenance and bibliographic metadata (PM)** includes features capturing *previous uses* of the media object such as *popularity, recentness, previous publishing*, etc.
 7. **Rights and licensing (RM)** includes information regarding the *rights holder(s), permitted use*, etc.
- **Environmental Factors**
8. **Production workflow (PDW)** captures features related to the overall production and its requirements.
 9. **Processing workflow (PCW)** bundles features regarding the actual processing of the media objects, such as *if it is adaptable* or *how it can be processed*.
 10. **Publication context (PC)** refers to the concrete location into which the media object should be integrated to and by that provides additional constraints such as *available space, consistency with the layout*, or *publishing history* of the item to be reused.
 11. **Practices and feedback (PF)** refers to work related factors such as *typical habits* of the designer itself, *typical guidelines from the company*, or *social recommendations* from colleagues, customers or experts.

Table 1 presents the dimensions investigated in the different models in different domains as reported in the literature. In order to compare previous studies, we mapped it in the clusters used by our model: *M1* refers to the model from Markkula [19], *M2* to the model from Choi [2], *M3* to the model from Othman [21], *M4* to the model from Westman [29], and *MR* to our model. An “x” means that the category has been confirmed to be relevant in the domain investigated in the respective model.

Category	M1 [19]	M2 [2]	M3 [21]	M4 [29]	MR
Information and content (I)	x	x	x	x	x
Visual and compositional Features (V)		x	x	x	x
Technical features (TF)				x	x
Technical quality (TQ)	x	x		x	x
Abstract and affective factors (A)	x	x	x	x	x
Provenance and bibliographic Metadata (PM)	x	x	x	x	x
Rights and licensing Metadata (RM)	x				x
Publication context (PC)	x			x	x
Publication workflow (PUW)				x	x
Processing workflow (PRW)		x	x	x	x
Practices and Feedback (PF)				x	x

Table 1. Dimensions investigated in different relevance models

4 Validation

To test the proposed research model and to gain insights on the degree of influence of the different factors, we adopted the survey method for data collection,

validated it using statistical methods, and compared our insights to the results derived from previous related work in this area.

4.1 Methodology and Data Collection

We conducted an end user survey to collect data and designed a questionnaire reflecting the factors in order to assess the conceptual model. The resulting questionnaire was first tested in a small group of members of a media design online forum. After analyzing the results from the test phase one item from *Abstract and affective factors* has been dropped because it has not been used by any of the participants. A further refined version was then sent as a self-administrated questionnaire to 150 media design professionals in Europe. The data was collected via an online survey in two languages (German and English)¹. The back translation approach was applied in order to ensure consistency between both language versions of the questionnaire [20]. The questionnaire consists of five parts: The first part contains general questions regarding reuse such as how much content is reused on a personal, company- and production-oriented level, and asked for reasons and barriers for reuse. The second part contains questions regarding factors used in search for media objects to reuse and the third part contained questions regarding selection criteria of content for reuse. In parts two and three the participants were asked to indicate the frequency of the use of the factors in search and for the assessment of the relevance of a media object on a five-point scale. The fourth part includes questions about the actual use of reused content (e.g., if it is adapted, or used as is). The fifth part contains concluding questions regarding the demographics of the participants.

31 responses which makes a response rate of 21 percent were returned, from which two responses with incomplete data were eliminated from further analysis. The gathered data reflects habits of media professionals spanning the domain of print and Web design over game design to the design of learning material. The majority of the participants had an experience from 3 – 5 years in their domain (32.14%), followed by 25% which had more than 10 years of experience and 21.34% which had 5 – 10 years of experience. The remaining respondents had up to 3 years of experience.

In order to check the internal consistency of the model, it was assessed using factor analysis [8]. Further structured relationships between the variables were examined.

4.2 Survey Results

The gathered data revealed several interesting insights on barriers and motivations for reuse of media, how people search for and assess the relevance of media objects in a particular situation (cf. Section 4.2 and 4.2), and how they finally use the selected media objects (cf. Section 4.2).

¹ The online questionnaire used in the survey is available at <http://www.tobiasbuerger.com/reusesurvey/>

Factors Affecting Search In this section we provide answers to the question "Which factors do users use to search for reusable media objects?" based on insights from the conducted survey.

In the first part of our questionnaire, participants were asked to indicate the importance for each feature on a five-point scale: 1 means that the factor is *never used*, 2 that it is used *very infrequently*, 3 that it is used *infrequently*, 4 that it is used *frequently*, and 5 that the factor is used *very frequently*. Based on that, Figure 2 shows the mean importance of the factors used in search by media professionals grouped into the relevant clusters:

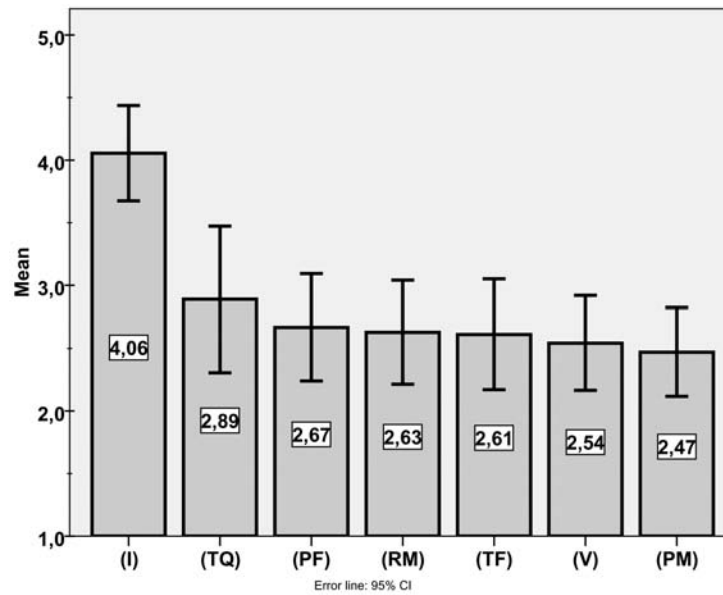


Fig. 2. Mean Importance of Factors in Search

Not surprisingly, the cluster with the highest frequency is *Information and Content (I)* meaning that users use keywords or classification information to search for content very frequently. Following this cluster are six clusters having almost equal frequency. The first one is *Technical Quality (T)* which includes factors such as *clarity of structure, sharpness, brightness, or technical error*. This is followed by *Practices and Feedback (PF)* including *feedback from experts and colleagues* which has the highest impact. After that, *Rights and Licensing (RM)*, *Technical Features (TF)*, *Visual and Compositional Features (V)* and finally *Provenance and Bibliographic Data (PM)* follow which are used rather infrequently. Users are typically not using affective factors, context or workflow information for search. A further observation from related work which has been confirmed by the expert interviews is, that media professionals typically start with a keyword query and then extensively use browsing facilities. This confirms

earlier insights from [21]. Furthermore it seems to be appropriate to present images with rather differing visual properties in some situations.

The results from this part of the survey are in line with results from Eakins [3] who identified topicality and technical quality as the most important criteria used in search.

Factors Affecting Selection and Assessment of Relevance Our survey revealed, that users use different factors from all clusters of the model to assess the relevance of media objects. Figure 3 shows the mean importance of the factors used for the assessment of relevance grouped into the clusters of the model:

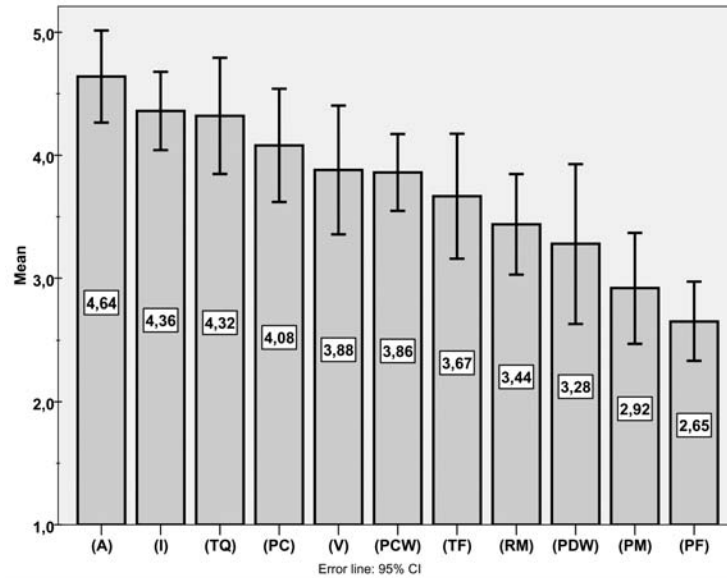


Fig. 3. Mean Importance of Factors in Relevance Assessment

Our results show only small differences to the study done by Westman et al. [29] in which the authors reported similar mean values for different factors in relevance assessment. Their study revealed that the cluster *Information and Topicality (I)* is the most important one, followed by *Abstract and Affective Factors (A)* and by *Visual and Compositional Features (V)*. Our results however suggest that *Abstract and Affective Features (A)* are most important even before *Information and Topicality (I)*. This can be explained by differences in the domain that we investigated, as a media object in media production only seems to be relevant if the aesthetics and other abstract features are compatible with the intended usage. This comes even before *Information and content (I)*. This observation can partly be explained by the fact how media professionals reuse media objects; media professionals retrieve media objects for inspiration very frequently (in 30% of all cases) meaning that they may create media objects

which reflect their thought topicality based on an aesthetically pleasing artwork. The other clusters had a similar mean importance as reported in previous work: *Technical Features (TF)* and *Technical Quality (TQ)* is followed by *Publication Context (C)* and other metadata. The smallest mean value is assigned to *Workflow Related Issues (W)* and *Practices and Feedback (PF)*.

It should be noted that some of the factors from clusters which are ranked lowest such as *Rights and Bibliographic Metadata (RM)* are in the top-10 of most important factors such as price or usage rights (cf. Table 2.)

Factor (Cluster)	Mean
Aesthetic compatibility (A)	4.67
Topical compatibility with the usage context (I)	4.46
Mental associations (I)	4.39
Technical quality (TQ)	4.32
Price (RM)	4.15
Technical adaptation possibilities (PCW)	4.11
Consistent layout (PC)	4.11
Usage rights (RM)	4.08
Technical format compatibility (TF)	4.04
Adaptation effort (PCW)	4.00

Table 2. Mean Importance of Relevance Criteria (Top-10 Factors)

This explains the difference to the clusters from Westman [29] in that category. The importance of rights can be explained by the fact that the Internet is the most frequent source for reusable media objects followed by the local hard-disk as our study indicated; on the Internet stock image sites are used most frequently followed by specialized image search engines such as Google image search². Company wide content management systems are ranked even after social media sharing sites such as Flickr³. The importance of adaptability can be explained by the differences in the domains investigated. In the journalism domain, which was investigated by Westman, images or photos are typically used as is and only marginally adapted, whereas in media production aesthetics and other abstract features have to be compatible with the intended usage. Furthermore novelty of created media objects is a very important criterion especially in games, animation or film production, which makes the need for bigger adaptations evident (cf. Section 4.2).

Usage of Selected Media Objects The fourth part of our study revealed interesting insights into how people reuse media objects that they select. The types of reuse can be grouped according to the definition provided in Section 1:

² <http://images.google.com>

³ <http://www.flickr.com>

(i) content is either reused as is, (ii), only parts of it are reused, (iii) it is reused after being adapted, or (iv) it is only reused for “inspiration”.

In most cases, media objects are only retrieved for inspirational purposes which can be explained by the fact that work has to be original in the investigated domain. A media object as is is only reused infrequently, parts of media objects are in contrast to that reused frequently. Results of our study furthermore reveal that content is being adapted very frequently before use. The adaptations range from basic features like resolution, contrast, brightness to the extraction of the background or parts of the content such as objects or areas.

5 Conclusions and Future Work

In this paper we reported on a conceptual model which dimensions relevance assessment in multimedia retrieval scenarios in which people search for content to reuse. The model is grounded on prior literature, completed with insights gained from expert interviews and validated based on empirically gathered results from an end user survey. The model captures factors used for search and relevance assessment, proposes a clustering of these factors, and assigns a mean importance value to each factor based on the results of the reported survey.

Our next steps contain the realization of a hybrid image search engine which integrates content based search with semantic search and which takes the results reported in this paper into account in order to re-rank the fused result lists from both search engines. We believe that the values assigned to the factors can be used to rank results which were retrieved in multi-faceted search including keywords related to the topic of images but also metadata such as rights, pricing information, or visual features. We plan to perform a second validation and calibration of the model based on end user experiments using the search engine.

References

1. H. Chen. An analysis of image queries in the field of art history. *Journal of the American Society for Information Science and Technology*, 52(3):260–273, 2001.
2. Y. Choi and E. M. Rasmussen. Users’ relevance criteria in image retrieval in american history. *Information Processing & Management*, 38(5):695–726, 2002.
3. J. P. Eakins, P. Briggs, and B. Burford. Image retrieval interfaces: A user perspective. In *Image and Video Retrieval*, LNCS, pages 628–637. Springer, 2004.
4. P. Enser. The evolution of visual information retrieval. *Journal of Information Science*, 34(4):531–546, 2008.
5. P. Enser, C. Sandom, J. Hare, and P. Lewis. Facing the reality of semantic image retrieval. *Journal of Documentation*, 63(4):465–481, 2007.
6. A. Goodrum and A. Spink. Image searching on the excite web search engine. *Information Processing & Management*, 37(2):295–311, March 2001.
7. A. Goodrum, M. Bejune, and A. C. Siochi. A state transition analysis of image search patterns on the web. In *Image and Video Retrieval*, Springer, 2003.
8. R. L. Gorsuch. *Factor Analysis*. Lawrence Erlbaum Publ., 1983.

9. H. Greisdorf and B. O'Conner. Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation*, 58(1):6–29, 2002.
10. L. Hollink, A. T. Schreiber, B. J. Wielinga, and M. Worrying. Classification of user image descriptions. *Int. J. Hum.-Comput. Stud.*, 61(5):601–626, 2004.
11. T.-Y. Hung. Search moves and tactics for image retrieval in the field of journalism. *J. of Educational Media & Library Sciences*, 42(3):329–346, 2005.
12. T.-Y. Hung, C. Zoeller, and S. Lyon. *Digital Libraries: Implementing Strategies and Sharing Experiences*, chapter Relevance Judgments for Image Retrieval in the Field of Journalism: A Pilot Study, pages 72–80. LNCS. Springer, 2005.
13. P. Ingwersen and K. Jaervelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005.
14. C. Joergensen, A. James, A. B. Benitez, and S.-F. Chang. A conceptual framework and empirical research for classifying visual descriptors. *J. Am. Soc. Inf. Sci. Technol.*, 52(11):938–947, 2001.
15. C. Joergensen and P. Joergensen. Image querying by image professionals. *J. of the American Society for Information Science and Technology*, 2005.
16. C. Joergensen. *Image Retrieval: Theory and Research*. The Scarecrow Press, 2003.
17. L. Keistler. User types and queries: Impact on image access systems. In R. F. et al., editor, *Challenges in Indexing Electronic Text and Images*, pages 7–22. Medford, NJ: Learned Information Inc., 1994.
18. M. Laine-Hernandez and S. Westman. Image semantics in the description and categorization of journalistic photographs. In *Proc. 69th Annual Meeting of the American Society for Information Science and Technology, Austin (US)*, 2006.
19. M. Markkula and E. Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1(4):259–285, 2000.
20. M. R. Mullen. Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 26(3):573–596, 1995.
21. R. Othman. A model for an image retrieval tasks for creative multimedia. *Performance Measurement and Metrics*, 6(2):115 – 131, 2005.
22. E. Panofsky. *Studies in iconology*. Harper & Row, New York., 1962.
23. H.-T. Pu. An analysis of failed queries for web image retrieval. *J. Inf. Sci.*, 34(3):275–289, 2008.
24. E. M. Rasmussen. Indexing images. *Annual Review of Information Science and Technology*, 32:169–196, 1997.
25. A. Rockley. *Managing Enterprise Content*. New Riders, 2002.
26. T. Saracevic. Relevance reconsidered. In *Information Science: Integration in perspectives. Proceedings of the Second Conference on Cerpceptions of Library and Information Science*, pages 201–218, 1996.
27. S. Shatford. Analyzing the subject of a picture: a theoretical approach. *Cataloging & Classification Quarterly*, 6(3):39–62, 1986.
28. S. J. Westerman et al. Creative industrial design and computer-based image retrieval: The role of aesthetics and affect. In *Affective Computing and Intelligent Interaction*, chapter Creative Industrial Design and Computer-Based Image Retrieval: The Role of Aesthetics and Affect, pages 618–629. Springer, 2007.
29. S. Westman and P. Oittinen. Image retrieval by end-users and intermediaries in a journalistic work context. In *IiX: Proceedings of the 1st international conference on Information interaction in context*, pages 102–110, New York, NY, USA, 2006.

Adapting Smart Graphics' Behaviour to Users' Characteristics

Christophe Piombo, Romulus Grigoras, Vincent Charvillat

IRIT - University of Toulouse, 2 rue Charles Camichel, 31071 Toulouse Cedex 7, France
{christophe.piombo, romulus.grigoras, vincent.charvillat}@enseeiht.fr

Abstract. Many existing-web based systems aim at making interfaces more user-friendly. Web content designers commonly use graphical components to illustrate concepts or to present numerical data. Adapting dynamically these components to the context in which they are used, has lead to the development of *smart graphics*. Some common context features are encountered such as platforms and network capabilities. Few systems consider users characteristics in order to provide more interactivity and flexibility. The objective of our work is to investigate this latter issue. We are currently developing a user model based on several characteristics that include preferences and motivation factors. To structure the user model data and support knowledge retrieval, we propose an ontology-based smart graphics framework. The methodology includes validation of this model through experimental study and developing an adaptive hypermedia e-commerce system that automatically learns users' characteristics and adapts graphical content accordingly. This paper presents an overview of the objectives and the methodology of this work.

Keywords: adaptation, user model, ontology, framework, smart graphics

1 Introduction

Web designers have used rich graphical components for such purposes as illustrating concepts in a web site, visually depicting numerical data, or making interfaces more user-friendly. However, the graphics themselves were static, which has limited their usefulness. A convergence of computer graphics and artificial intelligence technologies is leading to the development of *smart graphics* [1], which recognize some basic user environment characteristics such as platforms and network capabilities to adapt themselves accordingly.

Today, the smart graphics community enriched of researchers and practitioners from the fields of cognitive sciences, graphic design and user interface, have raised a new challenge: framing their investigations in human-centred way, presenting content that engages the user, effectively supports human cognition [2], and is aesthetically satisfying [3]. The ultimate objective is to prove the utility of adapting graphical object behaviours and visual display to individual users. For example, in [19], authors discussed about the usefulness of considering sequence and timing for improving the effectiveness of ad banners on a commercial web site. Results show that varying the format of banner and its display in a session has an impact to the level of users' interest and session duration.

The advent of the Internet has improved delivery and management issues. Considering the evolution of the web technology, powerful CPUs and graphics accelerators, as well as abundant memory, it becomes possible to envisage adaptive hypermedia systems that allow web content designers to develop graphical components that can be personalised to users' profiles. User adaptive systems have been largely studied by the user modelling community in the field of adaptive hypermedia [9] and traditional [10] web site. Some researches have considered the problem of adapting Web 3D content and presentation [11] in virtual environment context [13] to different web application areas [14], such as education and training [15], e-commerce [16], architecture and tourism, virtual communities and virtual museum [12]. Today, smart graphics based web systems inherit of user model representation techniques used in 2D web site and 3D worlds [1] improving organization and presentation of the content to the end-user. Therefore, implementing smart graphics facilitate users' understanding and assimilation.

Such smart components have inherited architectures of agent and smart object which are composed of many parts like action model [4] or domain model [5]. A standardisation effort has been started to develop marketable and interoperable smart graphics systems [6] [7].

This paper is composed of two parts. The first one presents an overview that shows the different use cases of smart graphics and a second one in which we will describe the objectives and the methodology of our approach.

2 Using Smart Graphics

Smart graphics are used in different domains but have the same objective: offer to the end-user the best way to accomplish a task with a tool (**Fig. 1**). In data intensive decision-making processes, end-users have to make effort to craft a meaningful visualization. The users are usually domain experts with marginal knowledge of visualization techniques. When exploring data, they typically know what questions they want to ask, but often do not know how to express these questions in a form that is suitable for a given analysis tool, such as specifying a desired graph type for a given dataset, or assigning proper data fields to certain visual parameters. In [18], authors proposed a semi-automated visual analytic model: Articulate. This smart graphics-based system is guided by a conversational user interface to allow users to verbally describe and then manipulate what they want to see. Natural language processing and machine learning methods are used to translate the imprecise sentences into explicit expression. Heuristic graph generation algorithm is then used to create a suitable visualization.

In other applications like tutoring or e-commerce, smart graphics aim to increase user satisfaction and to build customer loyalty, addressing the interests and preferences of each individual user. We find in the literature systems with different levels of adaptation. *Customisable* systems offer basic forms of personalization. Users were limited to setting user interface parameters and some other preferences such as platforms and network capabilities. This type of *adaptation* requires explicit choices from the user which are considered as a user profile or model. They are stored within the system and used to adapt its environment. This technique assumes that all adaptable aspects are understandable to the user who can clearly identify his/her preferences, and that all preferences can be derived from a questionnaire [8]. Obviously, this approach cannot cope with complex user models and systems in which behaviours must be embedded within each component distributed by the web.

Consequently, a new generation of *adaptive* systems, based on the use of smart components, is being developed. These systems have the *ability* to adapt the behaviours of each component to every individual user needs by analysing logs or by monitoring user interactions [5][26]. 3D content is increasingly employed in these systems that authors in [14] divided into two broad categories:

- sites that display interactive 3D models of objects embedded into web pages, such as e-commerce sites allowing customers to examine 3D models of products,
- sites that are mainly based on a 3D Virtual environment which is displayed inside the web browser, such as tourism sites allowing users to navigate inside a 3D virtual city.

They use essentially two adaptation techniques: adaptive navigation support and adaptive presentation [9]. Systems that support adaptive navigation structure their contents to allow the user to navigate through 3D objects that are most suitable. The system therefore grabs users' attention by visually highlighting those 3D objects. Two techniques inherited from adaptive hypermedia systems are used to implement adaptive navigation: adaptive annotation and curriculum sequencing. The first technique changes the order or availability of objects inside a 3D scene. Whereas, the second makes decision about which object (or details of an object) to display next depending on prerequisites and achievement. For example, in the Educational Virtual Environment proposed by [17], the student is assessed against learning objectives which evaluate the level of knowledge of an X3D language feature. Failing to pass the test, the user is not allowed to browse 3D objects with more complicated features. The results of such assessment are also used to update the student's profile. Most of these approaches focus exclusively on the level of knowledge of the student. They do not consider other factors, especially cognitive, that differentiate learners. Systems that support adaptive presentation offer often choices between different media when presenting materials (such as text and audio), but related to 3D objects technology, adaptive presentation consists to remove or add visual details and behaviours to an object.

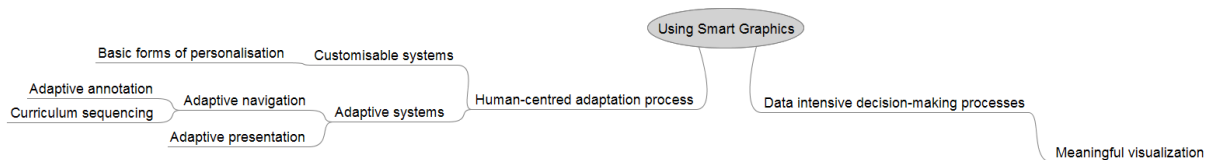


Fig. 1. Using Smart Graphics

Most of these techniques are limited when applied to advanced smart-graphics-enabled systems. The human-centred adaptation process is complex and requires taking into consideration various individual parameters that go beyond the assessment of user's achievements and simple user preferences.

3 The Proposed Approach

We address the problem of adapting smart graphics behaviours and visual display to the users' profile. Estimating user characteristics is essential for systems that require adaptation. For example, in adaptive tutoring systems, the learning style influences the learning behaviour [20] and in e-commerce the style of buying influences the buying behaviour [16]. Therefore, we define users' profile as being the way an individual tackles a contextual task with a specific tool. This profile depends on various factors including cognitive, preferences, motivations, interests, skill and social aspects. Three main aspects will be considered in this work: modelling the users' profile using ontology representation (see 3.1), developing a smart graphics framework that automatically assesses and uses such profile (see 3.2), contribute to the standardisation effort started within the smart graphics community by proposing smart graphics ontology to increase interoperability aspect (see 3.3).

3.1 Users' Profile Ontology

Semantic web made it possible to have the necessary tools to handle computer-understandable semantics. These tools, generally evolving from XML are used to enrich the description of web-pages, giving a deeper understanding of the relations between the concepts. OWL (Ontology Web Language) and RDF (Resource Description Framework) are some of the most widely used representations. Various definitions and models have been proposed for users' profile.

The Digital Item Adaptation part of the MPEG-21 Multimedia Framework provides a rich set of standardized tools such as the Usage Environment Description Tools to depict user characteristics. But usually, the users' profile describes mainly preferences about the various properties of the usage environment, which originate from users, to accommodate transmission, storage and consumption. For example, in [25], authors consider that user characteristics parameters represent the user's quality preferences on graphics components of geometry, material and animation as well as 3D to 2D conversion preference.

Recently, some researchers have started using ontology formalism to investigate how user preferences, interests, disinterests and personal information could be stored into a semantic user profile [23]. They argue that techniques like RDF and OWL together with ontology are the key elements in the development of the next generation user profiles. In this approach, the user profile is divided into particular domain sub-models and conditional sub-models, each containing particular information about the users' behaviour or context where a set of preferences should be applied. These kinds of models are named User-Profile Ontology with Situation – Dependent Preferences Support (UPOS).

Our objective is to develop a users' profile ontology based on UPOS which integrates various individual characteristics such as perception, thinking style, social aspects, and motivation factors associated to a context (e.g. platforms, activity...). Using a context-aware semantic reasoning, we will be able to adapt some features of the smart graphics. For example, when a user look at a camera inside a training activity on his laptop or inside a trading activity on his smart phone, the smart graphic used does not offer the same features and functionalities. In the first case, a user would like to learn to manipulate the device. In the second one, the user would like to know the price and camera zoom compatible.

The objective of this phase is to propose general user ontology for web site using smart graphics that can dynamically author materials depending on the user characteristics (e.g. thinking style, preferences...) and some context features such as web site domain area and activities (e.g. training, simulation, trading...) or material capabilities (e.g. platforms, network...). This will lead to the creation of a semantic description of a user environment model.

3.2 Smart Graphics Framework

We will design a component architecture based on the concept of smart component that can adapt its behaviour to individual users. Smart components are often represented as being able to interact with its environment through sensors and actuators (**Fig. 2**). Sensors cause perceptions that update smart component's beliefs compliant with its environment model. The smart component can reason about its beliefs and plan its optimal actions sequence to achieve a given goal. Based on its actions model, the smart component adapts the actions sequence to play.

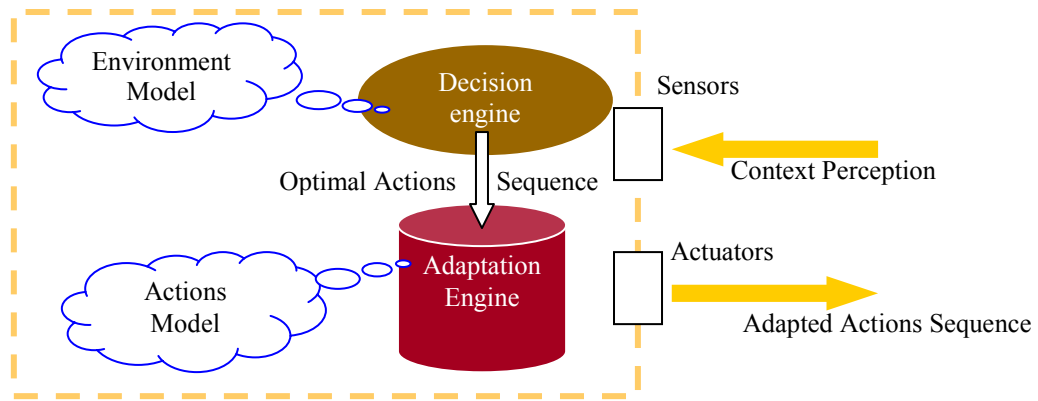


Fig. 2. Smart Component Schema

The main advantage of this approach is that all the information needed to interact with the component is located at the component level and not at the application level [4]. We argue that this solution could be used to design the architecture of web site using smart graphics facilitating the reuse of the component to deal with marketable aspects. In addition, we believe that defining a framework is needed to facilitate software development by allowing designers and programmers to devote their time to meeting software requirements rather than dealing with the more standard low-level details of providing a working system, thereby reducing overall development time.

In [5], authors propose an enhancement of MVC architecture for smart graphics. This approach enables interactive systems to use different views of the same model at the same time and to keep them synchronously updated. The visual display evolves from a simple presentation to an intelligent visualization that values data and presents only the result relevant to the user. Today, 3D objects are often used as visual display of a smart component. 3D computer graphic description languages (e.g. X3D) are used to describe their characteristics (e.g. shape, position, orientation, appearance...). Encoding X3D content using a XML-based syntax offers the possibility to transform them into smart graphics more suitable for visualization using XSL transformation [15].

A smart visualization framework, called IMPROVISE has been proposed to tailor system visual responses to a user interaction context [21]. The system catches a user request and dynamically decides the proper response content. Using an example-based visualization sketch design, the proper visual metaphor for the given content is decided. An adaptation layer transforms the display using constraints associated with a context model (user, environment...).

These approaches lack a high-level semantic description needed to enable smart graphics to interact with their environment. Thus preventing the necessary interoperability used in smart web based system to share or to reuse smart components. Some authors [22] propose to use semantic web technology to create a formal specification of smart components leading to increase the perception, understanding and interaction with their environment.

The **Fig. 3** presents our ontology based smart graphics framework. The main idea of the framework is to use semantic web technology to semantically enrich the pure geometric data with information about how to interact with the smart graphic based on the knowledge of the user environment model. We propose to consider smart graphics component as an agent related to its virtual representation: an avatar. So, two parts will be designed. A smart graphics core which encompasses the core functionality provided by an agent and a smart graphics avatar which is its virtual representation defining a visual display and behaviours. The interface of the smart graphics to the environment is realized by sensors and actuators. Sensors provide context perception from its current environment. Actuators are behaviours offered by the component.

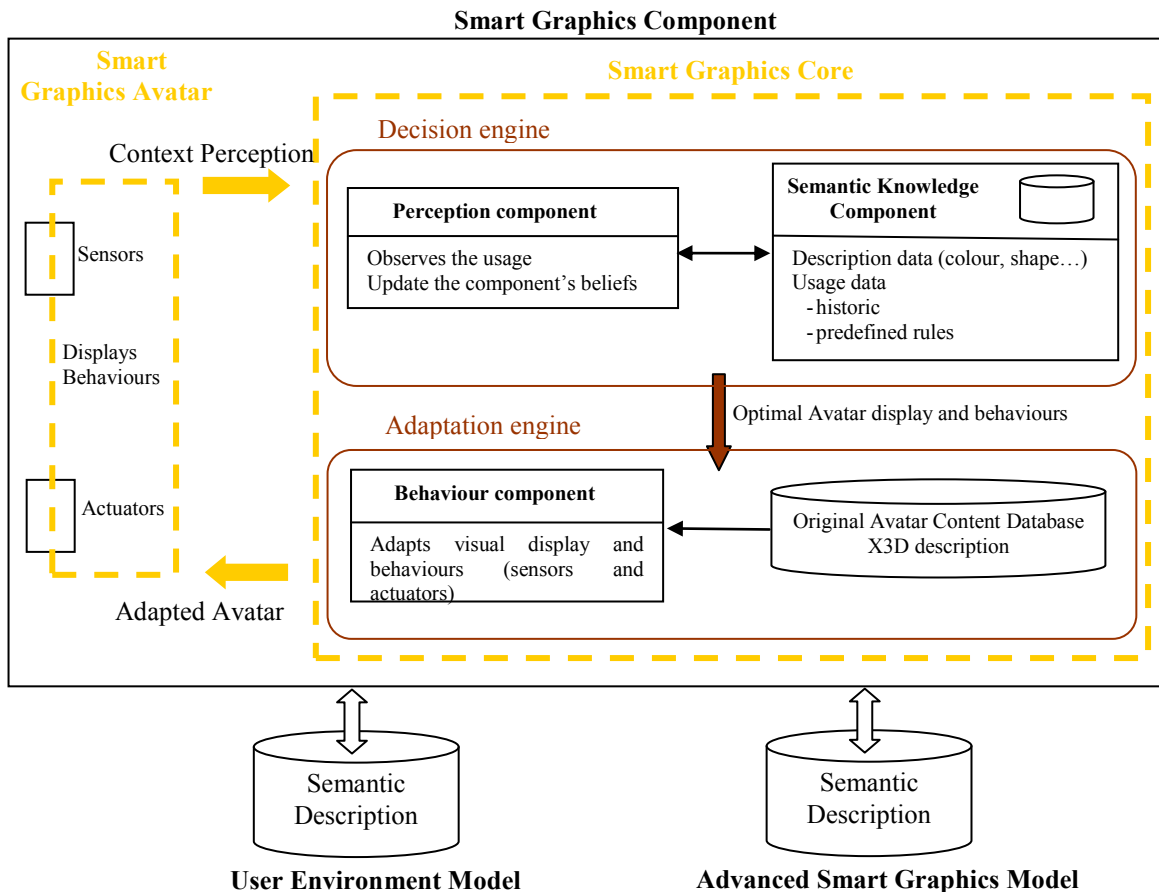


Fig. 3. Ontology based Smart Graphics Framework

Considering a web-site with smart graphics components embedded in web pages. When a user connects for the first time to the website, the decision engine retrieves semantic knowledge of the user environment model (e.g. platform and network capabilities, user preferences...) and uses predefined rules maintaining by the semantic knowledge component to define the optimal avatar display and behaviours. The adaptation engine makes an adapted avatar of the original avatar stored in content database using adaptation rules.

While manipulating the smart graphics, the user is monitored by the perception component of the decision engine, that observes the usage and update the component's beliefs. The component will be able to dynamically learn the user preferences. The automatic learning process will be continuous and by reinforcement. During the user activities, the semantic knowledge component maintains an historic of user usage and the perception component updates the user environment model information such as user preferences.

The decision engine will used an adaptation algorithm to match the user preferences to the web site objectives (e-commerce, training, simulation) and environment. Among other aspects, basics interactions (e.g. zoom, editing, querying, tutoring), the level of the object details, the control of camera path (e.g. freely, constraint, predefined), lighting a region of interest, overall navigation to related object and the mode of presentation (e.g. 2D image, 3D object, 3D meshes, sound, video) will be decided as the optimal avatar. An adaptation engine will generate dynamically the adapted avatar content compliant with original avatar content.

3.3 Smart Graphics Ontology

Semantic representations are usually distinguished by the use of ontology, which aims at specifying concepts. Some research has been conducted in the autonomous agents or avatars community to describe these smart objects using regular vocabulary and simplified representation [24]. **Fig. 4** shows a restrictive view about a smart object.

The objective in this work is to find out how features of Virtual Humans considered as a kind of smart object, can be "labeled" in computational systems in order to facilitate their interchange, scalability, and adaptability according to specific needs. In addition, the authors demonstrated that it is possible to construct the graphical representation of a Virtual Human from its semantic descriptors.

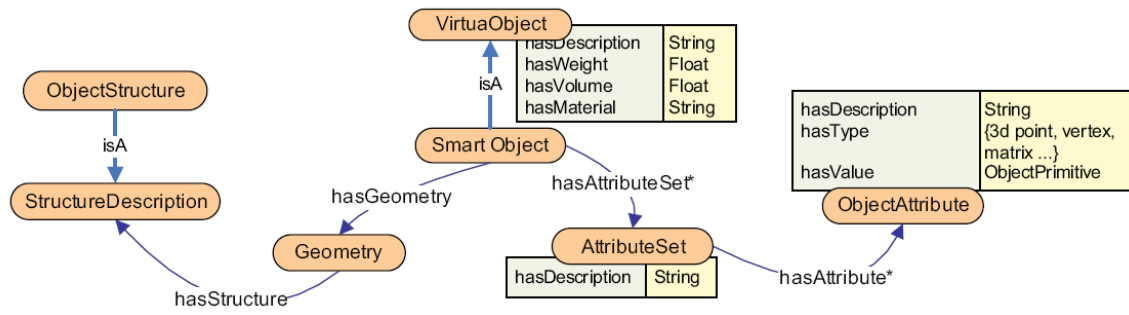


Fig. 4. Semantic for Smart Object

Semantic description of multimedia items has been mainly developed for audio, video and still images. These descriptions are defined in order to categorize, retrieve and reuse multimedia elements. The MPEG-7 standard, formally named Multimedia Content Description Interface, provides a rich set of standardized tools to describe multimedia content but only small attention has been given to interactive 3D items.

In [6] [7], authors propose a set of metadata to describe smart graphics in a standard way. The Smart Graphics data model based on these metadata describe the configurations of a set of Smart Graphics, whether they are in a single file or in multiple files. It includes some basics tags values such as ID, name, Description and highlights.

This description is not rich enough to manage a smart adaptation of the graphics like a control on camera path, light sources or behaviours.

Our aim is to pursue and extend these works and then contribute to the upcoming standardisation effort that aims to develop marketable and interoperable smart graphics systems. We propose to define ontology of smart graphics (Fig. 5). The semantic description will consider several field of knowledge such as geometry, behaviour, display and sensor among others. This semantic description of smart graphics will be compliant with our smart graphics framework Fig. 3. It will contribute to a common understanding among different research fields that aims at creating an advanced smart graphics model.

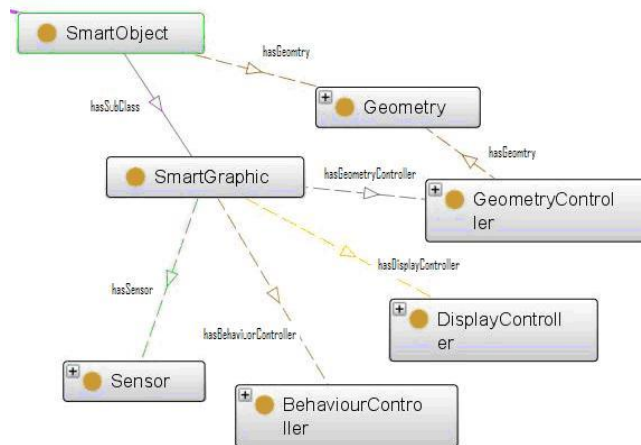


Fig. 5. Smart Graphics Ontology

The Fig. 6 shows a partial view of an OWL version of our smart graphics ontology. We can see that a smart graphic is a subclass of a smart object defining by [24]. The smart graphic class has several properties such as behaviour controller which will be used to manage both object animations and interactive functionalities that are offered to the user. The sensor will interact with the user environment model through an event model to adapt the display of the 3D item. For example, the display controller will be associated with a camera path manager that produces relevant camera paths around the target object (camera pose and zoom sequence). A good path may chain good viewing positions learnt by crowdsourcing. Different user profiles might lead to learn and then select different relevant camera paths. This principle will also be used to manage light sources and the object geometry in order to highlight regions of interest strategically.

```

<Ontology xmlns="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.semanticweb.org/ontologies/2010/9/SmartGraphics.owl"

  <Declaration>
    <Class IRI="#BehaviourController"/>
  </Declaration>
  <Declaration>
    <Class IRI="#SmartGraphic"/>
  </Declaration>
  <Declaration>
    <Class IRI="#SmartObject"/>
  </Declaration>
  <Declaration>
    <ObjectProperty IRI="#hasBehaviourController"/>
  </Declaration>

  <SubClassOf>
    <Class IRI="#SmartGraphic"/>
    <Class IRI="#SmartObject"/>
  </SubClassOf>

  <ObjectPropertyDomain>
    <ObjectProperty IRI="#hasBehaviourController"/>
    <Class IRI="#SmartGraphic"/>
  </ObjectPropertyDomain>

  <ObjectPropertyRange>
    <ObjectProperty IRI="#hasBehaviourController"/>
    <Class IRI="#BehaviourController"/>
  </ObjectPropertyRange>

</Ontology>

```

Fig. 6. Partial view of Smart Graphics ontology with OWL format

On today's e-commerce sites, the integration of interactive 3D objects into web pages, rather full 3D store environment is a common approach. Therefore, we will conduct an experimental study on e-commerce web sites to evaluate the sale performance of our ontology based smart graphics framework.

Our study will be conducted on a significant number of participants to help us:

- Develop and validate the user environment model based on the use of a questionnaire filled by each participant. This questionnaire will measure user's characteristics as perception, thinking style, social aspects, motivation factors and purchasing behaviour,
- Assess the pertinence of our framework to detect users' characteristics and to adapt the 3D objects' visual display and behaviours during a shopping session. To support this experiment, we will use our platform presented in [19] that enables to conduct a multivariate tests on web site.

The target population will be chosen to be as diverse as the audience of an e-commerce: wide age range, males/females, socio-professional categories etc.

To make our platform as interoperable as possible, we will base our work on standards whenever possible. For example, we will use OWL to describe the semantics aspects of smart graphics and users' profile using ontology formalism and X3D to manage visual display and behaviours of a 3D objects. Web technologies will be used to develop engine and ontology management system appearing in the framework architecture.

4 Conclusion

This paper has first presented a survey about different use cases of smart graphics. We also introduced a new framework to both describe and use smart graphics in many applications including e-commerce. This work ultimately aims at adapting graphics to individual user profile by using web usage mining techniques. Three complementary aspects are addressed. First we model the users using user profile ontology with situation-dependent preferences support. Second we defend a smart graphics framework that automatically *learns* the user profile and adapt visual display and behaviours of the smart graphics. Last, but not least, this proposal could contribute to an upcoming standardisation effort and bring an advanced smart graphics ontology that meets the interoperability challenges.

References

1. Edwards, J., Dailey Paulson, L.: Smart graphics: a new approach to meeting user needs, *Computer*, vol. 35, no. 5, 18--21 (2002)
2. Hammond, T., Prasad, M., Dixon, D.: Art 101: Learning to Draw through Sketch Recognition, *Smart Graphics*, vol. 6133, 277--280 (2010)
3. Kairi, M., Kenichi, Y., Shigeo, T., Masato, O.: Automatic Blending of Multiple Perspective Views for Aesthetic Composition, *Smart Graphics*, vol. 6133, 220--231 (2010)
4. Jorissen, P., Lamotte, W.: A Framework Supporting General Object Interactions for Dynamic Virtual Worlds, *Smart Graphics*, vol. 3031, 154--158 (2004)
5. Mahler, T., Fiedler, S., Weber, M.: A Method for Smart Graphics in the Web, *Smart Graphics*, vol. 3031, 146--153 (2004)
6. Jack, H. : Content & Smart Graphic Communication, AICC Management and Processes Subcommittee (2004)
7. Fraysse, S.: Designing Smart Graphics "simple scenarios" with IMS Simple Sequencing, AICC Management and Processes Subcommittee (2006)
8. Piombo, C., Batatia, H., Ayache, A.: Réseau bayésien pour la modélisation de la dépendance entre complexité de la tâche, style d'apprentissage et approche pédagogique, SETIT2005, Tunisie (2005).
9. Brusilovsky, P. : Adaptive hypermedia. *User Modeling and User Adapted Interaction*, vol. 11, 87--110 (2001)
10. Perkowski, M., Etzioni, O.: Adaptive Web Sites, *Communication of the ACM*, vol. 43, 152--158 (2000)
11. Chittaro L., Ranon R., Dynamic Generation of Personalized VRML Content: a General Approach and its Application to 3D E-Commerce, *Proceedings of Web3D 2002: 7th International Conference on 3D Web Technology*, pp. 145-154, ACM Press, New York (2002)
12. Chittaro L., Ieronutti L., Ranon R., Navigating 3D Virtual Environments by Following Embodied Agents: a Proposal and its Informal Evaluation on a Virtual Museum Application, *PsychNology Journal (Special issue on Human-Computer Interaction)*, Vol. 2, No 1., 24--42 (2004).
13. Chittaro L., Ieronutti L., Ranon R. Adaptable visual presentation of 2D and 3D learning materials in web-based cyberworlds. *The Visual Computer*, Vol. 22, No. 12, pp. 1002--1014 (2006)
14. Chittaro L., Ranon R. Adaptive 3D Web Sites. In Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): *The Adaptive Web: Methods and Strategies of Web Personalization*, Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, (2007)
15. Chittaro L., Ranon R. Web3D Technologies in Learning, Education and Training: Motivations, Issues, Opportunities, *Computers & Education Journal*, Vol. 49, No 2, 3--18 (2007)
16. Chittaro L., Ranon R., New Directions for the Design of Virtual Reality Interfaces to E-Commerce Sites, *Proceedings of AVI 2002: 5th International Conference on Advanced Visual Interfaces*, ACM Press, 308--315 (2002)
17. Chittaro L., Ranon R., Adaptive Hypermedia Techniques for 3D Educational Virtual Environments, *IEEE Intelligent Systems*, vol. 22, no. 4, 31--37 (2007)
18. Sun, Y., Leigh, J., Johnson, A., Lee, S.: Articulate: A Semi-Automated Model for Translating Natural Language Queries into Meaningful Visualizations, *Smart Graphics*, vol. 6133, 184--195 (2010)
19. Baccot, B., Choudary, O., Grigoras, R., Charvillat, V.: On the impact of sequence and time in rich media advertising, *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, 849--852 (2009)
20. Moebs S., Piombo C., Batatia H., Weibelzahl S.: A Tool Set Combining Learning Styles Prediction, a Blended Learning Methodology and Facilitator Guidebooks – Towards a best mix in blended learning, *ICL (2007)*
21. Wen, Z., X Zhou, M.: IBM Research Center, http://domino.research.ibm.com/comm/research_projects.nsf/pages/ria.Focused%20Areas.html
22. Nesbigall, S., Warwas, S., Kapahnke, P., Schubotz, R., Klusch, M., Fischer, K., Slusallek, P.: Intelligent Agents for Semantic Simulated Realities - The ISReal Platform, *ICAART*, vol. 2, 72--79 (2010)
23. Stan, J., Egyed-Zsigmond, E., Joly, A., Maret, P.: A User Profile Ontology For Situation-Aware Social Networking, *3rd Workshop on Artificial Intelligence Techniques for Ambient Intelligence (AITAmI)*, (2008)
24. Garcia-Rojas Martinez, A.: Semantics for virtual humans, thèse n° 4301, école polytechnique fédérale de Lausanne (2009)
25. Kim, H.K., Lee, N.Y., Kim, J.W. : 3D Graphics Adaptation System on the Basis of MPEG-21 DIA, *Smart Graphics*, vol. 2733, 283--313 (2003)
26. Vincent Charvillat, Romulus Grigoras: Reinforcement learning for dynamic multimedia adaptation. *J. Network and Computer Applications* 30(3): 1034-1058 (2007)

Visualizing Mapping of Metadata Properties

Martin Höffernig, Wolfgang Weiss, Werner Bailer

JOANNEUM RESEARCH, DIGITAL –
Institute of Information and Communication Technologies, Graz, Austria
{firstName.lastName}@joanneum.at

1 Introduction

Millions of hours of audiovisual content are held by collections of dedicated broadcast, film and sound archives, institutional or corporate archives, libraries and museums. There is a large heterogeneity between the different audiovisual archives resulting from their history and tradition but also from cultural differences of the countries where those archives reside. Consequently metadata models covering the workflows and necessities in the archives differ as well. This fact and the need for metadata for various different use cases in the archives lead to a number of metadata models and standards. Thus mapping between different metadata models is inevitable in practical applications.

We are currently developing a system for automating metadata mapping by formalizing semantics of properties in the different formats and their relations [4], based on an intermediate ontology, namely the *meon* ontology [3]. In order to enable users to validate the automatically determined mappings visualization functionalities are required in the system. This paper describes the integration of the ontology visualization developed in [6] into our mapping system prototype.

Creating comprehensive, clear and intuitive visualizations of ontologies and RDF graphs is an ongoing challenge. Different approaches can be found in applications for Semantic Web engineers. An example is Protege¹, which is an open, platform independent environment for creating and editing ontologies and knowledge bases. The application is extensible by its plug-in architecture and thus provides several visualizations. IsaViz² is a visual tool for browsing and authoring of RDF models. Resource nodes are represented by ellipses, literals as rectangles and properties are displayed as lines with arrows. OntoSphere 3D [1] uses a collection of three-dimensional visualization techniques displaying ontologies. gFacet [2] combines the graph visualization with facet search in the graph.

These applications use different kinds of visualization techniques to present the user a possibly easy to understand and complete overview of the whole RDF graph. Using graph visualizations of RDF data especially for end users has a number of drawbacks. For example, these visualizations are flat and every node is treated as a primary node. Also, displaying a graph with hundreds of nodes and edges results in a cluttered visualization (cf. [5]). Nonetheless graph visualizations have their place, especially for Semantic Web engineers [6].

¹ <http://protege.stanford.edu/>

² <http://www.w3.org/2001/11/IsaViz/>

2 Implementation

The prototype³ helps users finding, validating, and understanding metadata mappings by automatic metadata matching and appropriately visualizing mapping relations. Figure 1 shows the textual part of the user interface where the user creates a query. The first step of the user is to select an input- and output metadata format. Via the *Load button*, the application lists all available concepts from the selected formats. The next step is to select one or more concepts for which the mapping relations should be found. After confirming the selected concepts by clicking the *Ask button*, dependencies between the input and output concepts according to the defined rules are calculated and displayed. For each selected output format the information whether a mapping is feasible or not is displayed: *True*, if the output concept can be mapped from one or more of the selected input concepts, *False* otherwise. In case that there are output concepts without corresponding selected input concepts, the *Find requirements* option can be used. After selecting this option additional necessary input concepts are computed in order to establish mapping relations to the selected output concepts.

In addition to the boolean information about the feasibility of the mapping explained above, possible mapping relations are visualized in a graph visualization. The RDF-like graphs include the selected input concepts as yellow nodes, the selected output concepts as red nodes, the related *meon* concepts as green nodes and potentially missing input concepts as white nodes. It focuses on the current task of the user by displaying only necessary nodes and edges. This kind of visualization supports the user in understanding and validating the found mapping relations between input and output concepts of the metadata formats. An example of the graph visualization of the selected concepts shown in Figure 1 is depicted in Figure 2. The graph representation reveals that the input concept `mpeg7:Height` together with `mpeg7:Width` can be mapped to the selected output concept `ma:FrameSize` via `meon:Resolution`, which is part of the intermediate ontology (*meon* ontology). Beside this positive mapping relation, no appropriate mapping relation can be established to the remaining output concept `ma:Creator` from any of the selected input concepts. However, `mpeg7:UnqualifiedCreator` is a possible input concept to map to `ma:Creator`.

The mapping prototype is a Web application using standard Web technologies such as HTML and JavaScript for the user interface as well as Scalable Vector Graphics (SVG) for the graph visualization. To generate the graph we use the Java Universal Network / Graph Framework (JUNG)⁴, which provides a number of layout algorithms and mechanisms to manipulate graphs. An internal evaluation has shown that the “self-organizing map layout for graphs” produces the best results for our requirements. However, this layout algorithm generates a different layout at every single run. Therefore, it is necessary to animate the graph visualization for the user. The animation helps the user to follow how

³ <http://prestoprime.joanneum.at>

⁴ <http://jung.sourceforge.net/>

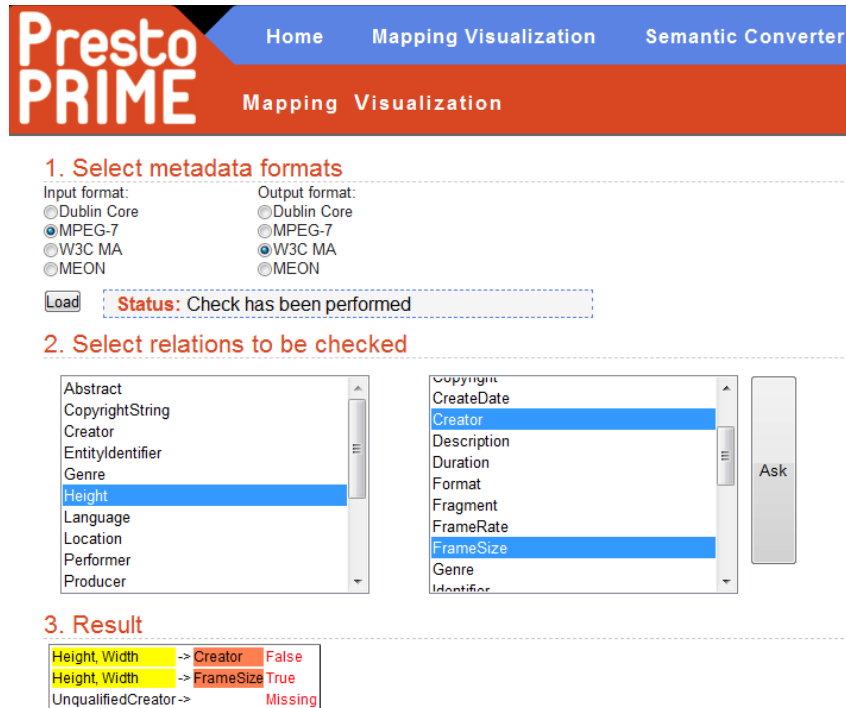


Fig. 1. Visualization interface.

the layout changed since the last run. For processing the RDF data the Jena Semantic Web framework⁵ has been used.

3 Conclusion and Future Work

In this paper we have presented the visualization functionality of our metadata mapping prototype⁶ which helps users finding, understanding and validating metadata mappings by automatic metadata matching and appropriately visualizing mapping relations. The visualization shows mapping relations between input and output metadata formats which are determined by the system via an intermediate ontology. It uses coloured nodes and focuses on the current user task by displaying only nodes which are necessary for the current task. The system is able to find direct metadata mappings as well as to suggest further input concepts to satisfy the desired metadata mappings. In the future the system shall support the definition of mapping rules by the user in order to improve the results in cases where incomplete or ambiguous mappings between pairs of formats exist.

⁵ <http://jena.sourceforge.net/>

⁶ <http://prestoprime.joanneum.at>

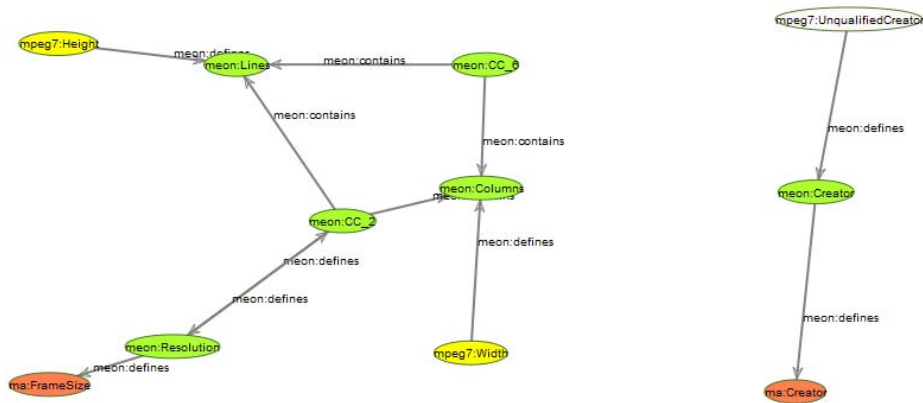


Fig. 2. Example of mapping visualization.

Acknowledgments

The research leading to these results has received funding from the European Union’s Seventh Framework Programme under grant agreement nr. FP7 231161, “PrestoPRIME” (<http://www.prestoprime.eu>).

References

1. Alessio Bosca, Dario Bonino, and Paolo Pellegrino. Ontosphere: more than a 3d ontology visualization tool. In *Proceedings of the 2nd Italian Semantic Web Workshop, 14-15-16 December 2005*, number 166, University of Trento, Trento, Italy, December 2005.
2. Philipp Heim, Thomas Ertl, and Jürgen Ziegler. Facet graphs: Complex semantic querying made easy. In *7th Extended Semantic Web Conference (ESWC2010)*, June 2010.
3. Martin Höffernig and Werner Bailer. Formal metadata semantics for interoperability in the audiovisual media production process. In *Workshop on Semantic Multimedia Database Technologies (SeMuDaTe)*, Graz, Dec. 2009.
4. Martin Höffernig, Werner Bailer, Günter Nagler, and Helmut Mülner. Mapping audiovisual metadata formats using formal semantics. In *5th International Conference on Semantic and Digital Media Technologies*, Saarbrücken, DE, Dec. 2010. To appear.
5. David Karger and M.C. Schraefel. The pathetic fallacy of rdf. http://swui.semanticweb.org/swui06/papers/Karger/Pathetic_Fallacy.html, 2006. Retrieved 06 February 2007.
6. Wolfgang Weiss, Michael Hausenblas, and Gerhard Sprung. Visual exploration, query, and debugging of rdf graphs. In *Proceedings of the Fifth International Workshop on Semantic Web User Interaction*, April 2008.