Semantics, Content, and Structure of Many for the Creation of Personal Photo Albums

Susanne Boll
University of Oldenburg
boll@informatik.uni-oldenburg.de

Philipp Sandhaus, Ansgar Scherp
OFFIS - Institute for Information Technology
sandhaus|scherp@offis.de

Utz Westermann
Oldenburg, Germany
westermann@acm.org

ABSTRACT

Photos are often a means to remember personal events, and the creation of photo albums is the attempt to preserve our memories in a nice book. For a long time people have been creating such photo albums on the basis of prints from analog photos arranged in an album book with scissors and glue and annotated with comments and captions—a tedious task which in these days is getting support by authoring tools and digitally mastered photo books. Relying on the content of others such as printed travel guides, news papers, leaflets, but also friends and family the personal content often has been enriched, enhanced, and completed. This is the starting point of our work: with digital photography and the increasing amount of content-based and contextual metadata of personal photos we can now use this metadata to actually support the targeted and semi-automatic inclusion of interesting, related information from content of others, e. g., from Web 2.0 communities, and offer and add it at the right spot in the personal album. In this paper, we show how photo album creation can benefit from leveraging information learned from many users in regard of the album's content, structure, and semantics.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems

General Terms

design, experimentation, human factors

Keywords

personal photo collections, photo annotation, photo album authoring, multimedia semantics, Web 2.0

1. INTRODUCTION

Photos provide a visual handle to our memories. By arranging photos in albums we attempt to preserve for ourselves and convey to others the way we experienced essential events of our lives. In order to capture the various facets of events vividly, we not just include our photos in our albums but also add photos taken by friends or even provide text notes, snippets of news reports, or tickets if these are relevant for the moment. For a long time people have been creating such photo albums using prints from analog photos and arranging them carefully in a nice book with scissors and glue—a tedious task that these days is getting support by authoring tools for digitally mastered photo books [4]. Such authoring software allows to select and arrange digital images on the pages of a book, add textual annotations, and design the book with the preferred colors and style, and then order a print of it from a photo finishing company.

The digital photo collections on our hard drives, however, provide just an extremely undersampled snapshot of our experience. Other artifacts such as tickets for museums we visited, post cards of an art exhibition, or a map of a hiking trail we followed that offer great handles to our memories and that we would like to include in our albums are usually not part of our digital media collections. The user herself must actively collect and carry these materials home and either tediously scan them in order to be able to use them in the photo book authoring software or, after a book has been printed, anachronistically glue them in.

However, some of such additional media that might be of interest for a personal photo book is available on the Web today. With the advent of Web 2.0, we find communities that create large content sites from which it is also possible to "collect" related material from many other users for the personal event and use this to complement or extend the personal photo collection. Thus, the personal photo collection moves from a personal silo to a node in a connected Web of contents. This is the starting point of our work. With digital photography and the increasing amount of content-based and contextual metadata attached to personal photos we can support the collection of interesting, related information directly from the Web and add it right on spot to a photo album. In this paper, we identify different ways to leverage content, structure, and semantics of many users to facilitate and improve the authoring of a single user's personal photo album.

One source of this is the exploitation of the content of others to enrich and augment the personal collection. Starting out from the personal collection one can identify content of others from different sites which would well complement and enhance the personal collection. For example, photos from a trip to Paris might be augmented with a Wikipedia [1] article from the Web about the Eiffel tower or photos depicting the tower, searching for similar photos on Flickr [13]. We will discuss how we can use the personal collection for a rule-based retrieval of relevant Web 2.0 content to be added to the personal collection. With the content of many from Web 2.0 we not only aim to enhance the personal media collections but also import semantics in the sense of tagged and annotated content.

Another way of learning from many for photo book authoring is to analyze personal photo books for common layouting structures and for learning for an semi-automatic authoring wizard. For this, the books can be analyzed content-based, on structure, and on textual annotations. Exploiting large numbers of authoring processes and the resulting photo books, we can strive to understand the structure of photo albums and types or categories of albums. Furthermore, we can learn what kind of photos actually find their way into the album, which content features they carry to later learn how to pre-select photos from large sets for albums. The learning can well distinguish between deriving generic rules of album creation for every user as well as personal characteristics and preferences. We will discuss how much an analysis of many photo albums reveals for the individual photo album authoring process and which rules for an album wizard can be learned from many for one.

From the analysis of the authoring of photo books of large user collections, we also aim to derive semantics for the single user's album. For it, we define semantics derivation rules based on the analysis of existing photo albums and the long-term experience of our project partner CeWe Color[1] in the field. Thus, the semantics derivation rules can be considered as the accumulated knowledge of describing the semantics of a photo album that is derived from the project partner's expertise and the photo albums analyzed. The semantics derivation rules are then used and applied for semantically enriching newly created photo books. By this, the individual person benefits from the knowledge of many to semantically enrich and author his or her photo book.

The presented exploitation of content, semantics, and structure of many for the creation of photo albums are implemented in our context-driven smart photo book authoring tool xSMART [7, 8]. This tool supports the users in creating the presentation by a step-by-step wizard that guides the users through the photo book authoring process. The exploitation of the semantic, content, and structural information is fully integrated with the authoring tool and wizard. In our prototypes and first experimental results, we show the synergy of content, semantics, and structure from many and personal photo albums.

The remainder of the paper is organized as follows: Section 2 describes the photo book model we use throughout the paper. In Sections 3 to 5, we describe the different ways we follow to leverage external content and structural information to enrich and complement one's personal photo books and how we derive semantics from the structure of photo books. We discuss the approach and present examples and our prototypical implementation. In Section 6, we conclude the paper and summarize the challenges for the future work in content understanding, exploitation and usage that arises from personal and collaborative media.

[1] CeWe Color is the European leading photo finisher company.

2. PHOTO BOOK DOMAIN MODEL

Throughout the paper, we will be presenting rules illustrating the derivation of photo semantics and enrichment of photo albums from photo metadata, from the use of photos in photo albums of possibly different users, as well as from Web resources and community data. These rules are based on an abstract conceptual model of the problem domain, which is presented by Figure 1 in form of a Unified Modeling Language (UML) class diagram. Founding the conceptual model on UML allows us to apply Object Constraint Language (OCL) as the rule language. The rest of this section explains the conceptual model in more detail.

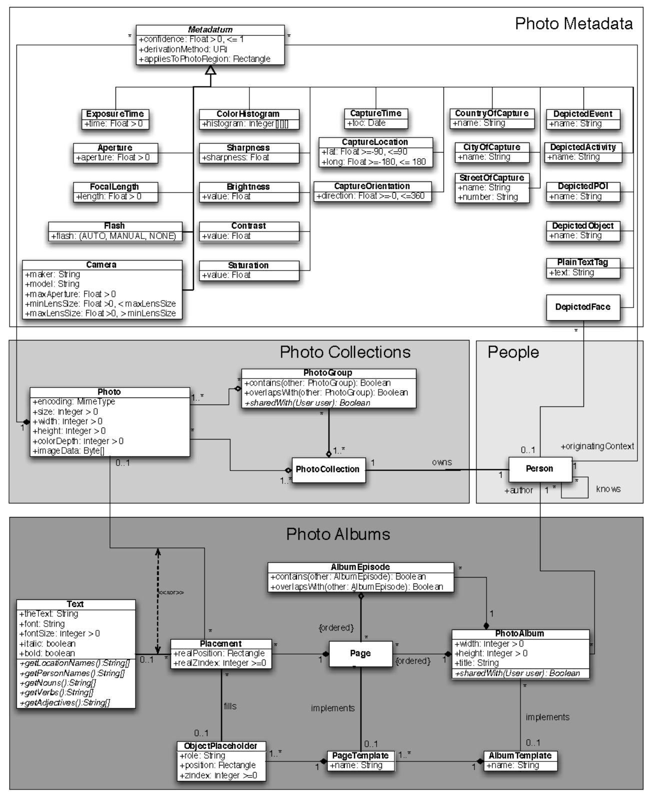


Figure 1: Photo Book Domain Model (UML class diagram)

The photo book domain model captures *people*, people's *photo collections* and *albums*, and *photo metadata and semantics*. The model considers photo collections of different people instead of implicitly focusing only on the photo collection and albums of one person. Reflecting communities, people in the model can be acquainted with each other via arbitrary direct or indirect relationships. Thus, the model takes account of one of the main hypotheses of this paper that the interaction of a user with his or her photo collection may impact the interaction of other users with their collections. Note that being an abstract model, the photo book domain model does not make any assumption on how acquaintance relationships between people are determined –

these could be obtained from a user's group membership on community platforms such as Yahoo Groups or taken from a user's e-mail address book.

The photo book domain model further assumes that each person in the problem domain owns a collection of an arbitrary number digital photographs. The model does not take account of whether such a collection is stored on a user's harddrive, in a photo management application, or on an online photo sharing platform. The model allows related photos to be arbitrarily clustered into possibly overlapping photo groups. Again, the model does neither make any assumption on how such a photo group is physically represented – it may be a simple folder on the user's hard drive or a stream on a photo sharing platform – nor does it make any assumption on how the group is derived – a user could manual group related photos together or a temporal clustering algorithm could dynamically create groups of likely related pictures.

Each photo can be arbitrarily augmented with metadata. A metadatum may apply to the whole picture or just a rectangular region of it. The model traces how each metadatum was created by storing a unique identifier of the applied creation or derivation method. In case that a metadatum was derived using an automatic, unreliable analysis method, the model allows one to capture a confidence value for the metadatum. The model further takes note of the originating context of a metadatum: i.e., the user in whose context a metadatum was derived. If a photo is traded between the collections of different users, it is thus possible to express that different users have derived different metadata about the photo.

The model is based on a very broad notion of metadata. Optical context data (such as camera model and exposure time), low-level content features (such as color histogram and photo sharpness), physical spatiotemporal context data (such as time and geographic position of the capture), logical context data (such as the country and city of the capture), photo semantics (such as depicted persons or points of interest), as well as simple plain text tags are all considered different types of metadata. Note that the metadata types depicted by Figure 1 are not considered complete. Further types can be easily introduced by adding further metadatum subclasses.

Finally, our photobook domain model captures the photo albums that have been authored by a person. Each person may own an arbitrary number of photo albums. Each of these albums consists of a sequence of pages, onto which photos from the author's collection or text fragments can be placed in a possibly overlapping manner. Albums may follow a template, which may again offer different page templates for different type of album pages. Each page template defines place holders for text fragments or photos. A place holder fulfills a certain role on a page, such as "heading" or "background picture". The model permits users to obey or not to obey a template. For example, it is possible for a user move a photo from the position originally reserved by a placeholder – a fact that may offer interesting insights and a further handle for the derivation of metadata. The model allows album pages to be grouped into (possibly overlapping) album episodes. Similar to photo groups, the model does not make any assumption on how these episodes are created: manually by the album author or by an automatic page clustering algorithms.

3. CONTENT OF MANY FOR ONE

With the increasing availability of online communities and the trend to collaboratively tag and organize shared multimedia content, a great potential lies in utilizing this ever growing pool of knowledge and media to augment personal photo books and add collaborative content and knowledge to the personal media collection. Imagine, e.g., having spent a nice holiday in Paris but having missed the chance to take a decent picture of the Eiffel tower. This missing photo might be searched on a photo community site like Flickr and added to the personal media set. For photo books, a day in Rome could be augmented by a description of the Colosseum from a Web 2.0 travel blog. So, the Internet can be used to search and find collaboratively created content to augment the personal media collection.

3.1 Web 2.0 for Photo Books

One of the key features of the Web 2.0 is the fact, that it contains a vast amount of user generated content and information about this content. Users are able to add, delete, comment, organize and alter content. A good example for this are photo communities: One can add or delete a photo, tag it in various ways add it to groups or comment it. Several of the provided Web 2.0 services are of interest for the creation of personal photo books.

One way to employ the Web 2.0 is to directly add content to a personal photo book. Semantics, like descriptions or locations, extracted from the photo book can be used to formulate adequate search requests to, e. g., photo community sites like Flickr or online encyclopedias like Wikipedia.

Besides photos, other visual media can be an attractive means to enrich personal photo books. With the availability of location information from the photo book, for example, geographical maps from services like Google Maps or Yahoo Maps can be retrieved. A map showing the route taken in a 4-week trip to Mexico can, e. g., be added to the beginning of the photo book.

Also textual content can augment one's personal photo book. For example, services like *geonames.org* provide a geographical name, like a city or country, for a given GPS-Position. This can be used to, e. g., be added as a label to a specific photo in the album, which is equipped with a GPS datum. But it can also act as a search request to other services. The result can be, e. g., a Wikipedia article describing the place the photos were shot. An excerpt, perhaps the first paragraph, can be used as a description, which can be added to the photo book. Also photo communities often provide rich textual descriptions of photos.

3.2 Rule-based Content Augmentation

To be able to perform adequate search requests on webservies a good semantic understanding of the photo book and its content is needed. One way to achieve this is to extract metadata from the photos itself. For this we employ our component-based metadata extraction architecture [3]. Semantics can also be derived from the structure of photo book as presented in Section 5. The more information is available the easier it is to formulate adequate search requests. To not blindly flood the photo book with content when certain information are available, it is necessary to limit the amount of added content by putting restrictions on the augmentation process. For this we follow a rule-based approach. Each rule stands for the definition of one specific kind of content which is subject to augment the photo book. A rule consists of a set of preconditions which have to be met in order to have the rule evaluated, a description of what content should be queried with what parameters and how this content should be integrated into the photo book. An example for such a precondition can be, e. g., that the amount of photos on a page should not exceed a certain number to prevent the photo book page from appearing too packed. For actually performing a request on a web service, definitions for such services have to be provided which have to be registered to the system.

Figure 2 shows an overview of how rules are used to integrate Web 2.0 content. An existing set of photos or a photo album together with relevant semantics is used to determine which rules are applicable for the photo book, that means for which rules the preconditions are met. These rules are evaluated, that means it is determined if a service is registered that provides the requested content. It is possible that more than one service has to be called for this. A good example for this is a query to Flickr to retrieve photos which are tagged with the name city other photos on a page are tagged with. When these tags are not available but the photos contain GPS data, the corresponding city name can be retrieved from a web service such as *geonames.org*.

The resulting content is integrated into the photo book. How this is done is defined by rules. We do not necessarily see this as a fully automated approach, but think that the photo book user can, and often has to be integrated in this process. He can, e. g., control which parameters are used for querying external sources, perhaps by manually adding annotations to a specific photo book page. It is also possible that rules do not define how content should be added to the photo book but that the users can choose from a list of automatically retrieved content.

3.3 Example Rules

For the definition of rules we distinguish between four granularity levels: A rule applies either to a photo, a photo book page, a photo book episode or a whole photo book. In the following sections we provide example rules for each of these levels. We briefly describe each rule and present a semi-formal definition based on OCL and the model defined in Section 2.

Evaluate Location.

A rule does not necessarily have to lead to instant visible changes in the photo book. The following rule enriches a photo which consists of a GPS information with the according location names.

```
context Photo::evaluateLocationEnrichmentRule()
pre: captureLocation -> notEmpty()
pre: countryOfCapture() -> empty()
pre: cityOfCapture() -> empty()
pre: streetOfCapture() -> empty()

body: let
        parameters = captureLocation()
      in
        countryOfCapture =
          ServiceConnector::
            getCountry(parameters).content)
        cityOfCapture =
          ServiceConnector::
            getCity(parameters).content)
        streetOfCapture =
          ServiceConnector::
            getStreet(parameters).content)
```

Evaluate Map.

This rule applies to a photo book page. It says that if the number of photos on a page is not larger than two and if at least one photo consists of a location information, a map should be retrieved with these locations as parameters and the map should be added to the page.

```
context Page::evaluateMapRule()
pre: photos -> size() < 3
pre: photos -> size() > 0
pre: photos -> exists(photo |
       photo.captureLocation -> notEmpty())
body: let
        parameters = photo.captureLocation()
      in
        photos = photos@pre.union(
          ServiceConnector::
           getMap(parameters).content)
```

Evaluate Location Label.

This rule employs information from the photo level rule defined above can use this information to add a descriptive label to the beginning of the album episode, if all pictures have been taken in the same city.

```
context AlbumEpisode::evaluateLocationLabelRule()
pre: pages.first().pageLabel -> empty()
pre: pages.photos ->
       forAll(p1, p2 |
         p1.cityOfCapture = p2.cityOfCapture)
body: let
         label = pages.photos.first().cityOfCapture
      in
        pages.first().pageLabel =
          'Visiting '.concat(label)
```

Evaluate Title Picture.

The following rule looks for an additional photo for the title page when a title is present in a photo album.

```
context Album::evaluateTitlePictureRule()
pre: episodes.first().
       pages.first().photos -> empty()
pre: title -> notEmpty()
body: episodes.first().
        pages.first().photos@pre.union(
          ServiceConnector::
            getPhoto(title))
```

Connectors to Web 2.0 sources.

The rules in the previous paragraph do not define, how content is actually retrieved. Instead a single service connector is called for this. The idea behind this is to keep the rules independent from specific services. Concrete definitions for services can be registered at this service connector. An example for such a service is:

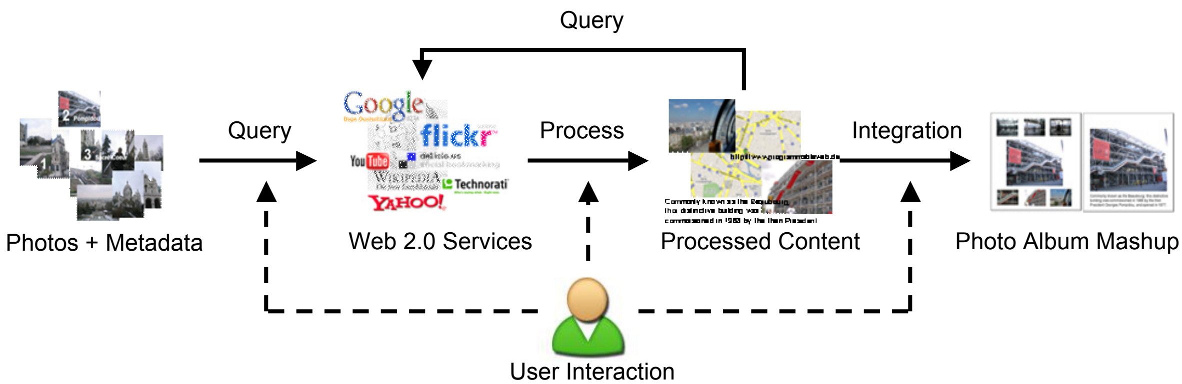


Figure 2: Steps of integrating Web 2.0 content into a photo book

```
context GoogleMapsService::getContent(parameter)
pre: parameter.captureLocation -> notEmpty()
post: result.type('Map')
```

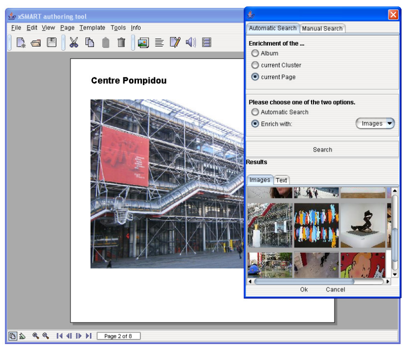
This definition specifies that the provided parameters should consist of at least one location metadatum and that content of the type *Map* is provided. Based on these preconditions and the provided output the right service for a rule can be chosen.

3.4 Photo Album Augmentation Prototype

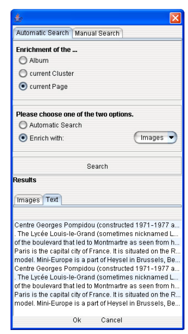
Based on our tool for semi-automatic authoring of multimedia presentations xSMART, we built a prototype application implementing our approach. We identified two possible spots where to integrate the augmentation process in the photo book authoring chain. Firstly, integrated in the automatic authoring process and secondly, interactively when the user is altering the generated photo book. For the first approach the user can chose within the wizard if the photo book should be enriched with additional content at all. The advantage of this approach is that no interaction with the user is necessary when doing the enrichment process, which means less effort for the user. The disadvantage is that there is far less information about the photo book semantics available than in later phases when users may have manually annotated or altered the photo album. For some kind of content it may also also not be possible or desirable to define how it should be integrated into the photo book.

Figure 3 shows how the rule evaluation with the second approach is implemented in xSMART. In this example a photo album from a vacation of two friends in Paris is being authored. The shown photo book page consists of a picture showing the Centre Pompidou and a matching page label. The user has chosen to have the current page enriched with additional content. By the evaluation of all rules for a page two rules were identified by the system, that are applicable to the current situation, one seeking for pictures containing the photo books page label in their description and another one querying Wikipedia for a describing text. The results of these queries are presented to the user and he is able to chose which content is of interest to him. She is then able to drag this content into his photo book. The result of such an augmented photo book page can be seen in Figure 4. The rules that were applied to the displayed page are:

- If the page has a text label (caption, title, ...) and consists of less than 3 pictures then use this text label to search for relevant photos at Flickr.
- If the page consists of at least one photo with a GPS position attached than look for a relevant Wikipedia article describing this location.



(a) Photos



(b) Texts

Figure 3: Offering relevant Web 2.0 content to augment the personal album

4. STRUCTURE OF MANY FOR ONE

Besides looking at *what* content is put into a photo book it is also interesting to know *how* this content is assembled and structured. Looking at this structural specifications of finished photo books can be way to better semantically understand it's content and to derive best practices for the assembly of new photo books. In this section, we describe how we are analyzing photo books which are ordered at our project partner and present some first results of our analysis. We also present ideas for future analysis and, based on this, how we want to improve automatic photo retrieval and automatic photo book layout.

4.1 Analysing of finished photo books

As a first step we are currently analyzing finished photo books that are ordered at our project partner CeWe Color. These photo books can be designed with the help of a custom photo book software developed by CeWe Color and ordered

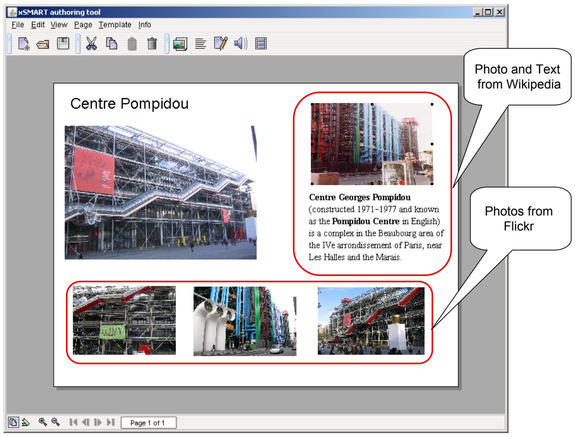


Figure 4: Photo book page augmented with additional content

via CD or internet. We are interested in the way how people actually style and structure their photo books. By analyzing these books we hope to be able to learn what a *good* photo book is. As a first step we have analyzed 512 photo book orders. At this time, unfortunately only about 10% of all randomly selected orders consist of photos that carry an EXIF (Exchangeable Image File Format [6]) header[2]. With each order of a finished photo book CeWe Color receives an order file that goes into the production of the printed photo book. From the photo book data structure we can derive statistical data about the size and distribution of photos in the book and also information we can derive from the mere photo content. For those photo books that comprised photos with EXIF header we are also looking into details about the photo and its context information. We are strictly not considering any personal information to keep the analyses completely anonymous.

Photo book structures.

The hoto book structure analysis aims to investigate how people distribute their photos over the book. The goal is in the long run to determine types of photo books and adapt the photo book authoring wizard to select and guide trough a suggested layout and structure. For this we assume that manually composed photo books will reveal sufficient information about this. Table 1 summarizes some of the analysis results. In general, photo book album sizes in the number of pages and number of photos, and text boxes per page are of interest. The first results from 512 photo books show that even though a medium size of around 17 pages has been found the variance σ is very high. The same is the case for the number of photos per page. Very interesting with regard to learning semantics from the photo book is that users almost always manually add exactly one text element per page. Therefore, we can use this to annotate the photos of this page with some confidence with the content of this text element. Also the users created albums that have an average of 30 percent photos with faces. For this we employed the face detection algorithm [10] implemented in [5].

[2]This is due to certain restrictions in the CeWe production chain and does not reflect the current share of digital photos with EXIF headers

	mean	σ
Perc. photos with faces	0.34	-
Nr. of album pages	16.96	13.54
Photos per page	6.69	12.60
Text boxes per page	1.45	0.74
Pages with text / all pages	0.87	0.91

Table 1: Some statistics about analyzed photo books

	mean	σ
Page time difference	33.95 h	56.34
Page time span	31.61 h	55.99
Album time span	578.85 h	709.03
Used cameras per album	2.27	1.52

Table 2: Statistics about analyzed photo books containing photo with EXIF headers

Results like this will help us to fine tune our "best of n" selection for photos.

Photo book time, persons, and cameras.

Looking at albums with photos that carry EXIF headers we have been looking only at small number of 55 photo albums but also found first interesting results. Table 2 summarizes the results from analyzes that employ information from the photos' Exif header. One point we were interested in is the time span that is covered by the photos in the album. Assuming that the time stamps in the EXIF header have some validity and, even if the time of the camera was not set correctly, the relative time stamps (of photos taken with the same camera) are valid. The mean album time span was almost 600 hours which is around 25 days, but one can also see, that the variance is quite high. We found that some photo books consist of photos with exactly the same time stamp, perhaps because the camera was not properly set up, but have also seen photo books with time spans of more than one decade. These statistics need to be treated with care. Also very interesting is the mean page time difference. This is the difference between the mean point in time of all photos on one page and the mean point in time of the following page. This is also more than a day with about 34 hours. However, it is not clear yet how clearly the temporal distance made the users cluster the photos on one page as also the page time span is pretty high with 31 hours.

Another very interesting finding was that most albums include photos from more than one camera. This shows that users are already augmenting their personal collection with additional photos which supports our idea of augmenting personal photo books described in Section 3. It also much supports the assumption of Section 5 that the inclusion of photos from others (other cameras) might allow to derive semantics for the other photos on the same page.

We have to note once more that these results have to be treated with care as some time stamps are obviously wrong and also only a small number of photo books have been analyzed. These circumstances notwithstanding, we see these first statistics as an interesting and encouraging result.

4.2 Future Analysis

Our goal is to learn from the many for the single photo book user and to better understand what a user regards a nice photo book. With the results from future analysis we are planning both to improve the (semi-)automatic photo selection, structuring and layout of photo books.

Improving automatic photo retrieval.

One very interesting aspect is to learn what kind of photos are chosen for the photo books. Imagine having taken about 700 photos on a holiday trip from which only 100 are chosen for the final photo book. When assembling a digital photo book, these 100 photo are manually chosen by the photo book software user. If we would also have information about the 600 remaining, not chosen photos, we would be provided with a very valuable ground-truth for an ideal selection of photos for a digital photo book. We think that it is not possible to formulate static rules for a *good* selection of photos. Some aspects which could affect this are:

- **Different user groups.** These could for example be distinguished by different kinds of used cameras. We assume that photo book preferences for users that take photos with a professional camera differ much from those of users with photos taken by an everyday consumer camera. Other groups could be distinguished by their layout preferences, photo book sizes, etc.
- **Different themes.** The kind of photos in a photo book could depend on the purpose they are assembled for. A photo documenting a birthday would probably consist of many photos taken at the same place in a rather short time (perhaps a few hours) and showing a lot of persons. A photo book documenting a holiday trip would be spanned over a longer time and would probably consist of a lot of photos showing landscape. Also the layout and kind of photos might be affected by the time of year,e.g., or special global events like christmas or easter.

Improving automatic photo book layout.

CeWe Color and other photo finishing companies follow a template-based approach in their photo album software, that means photos are arranged over the pages with the help of predefined page layouts. By analyzing photo books it is possible to determine, which layouts offered by the photo book software are really used, meaning which kind of layouts are accepted and which are not. But it is also interesting to know, for which photos the predefined layouts are altered on purpose. If perhaps the size of a photo might have been increased one could assume that this photo is important to the photo book user.

By analyzing photo books we want to investigate in more detail how photo books differ from each other depending on the kind of photos, cameras, users, and events.

5. SEMANTICS OF MANY FOR ONE

So far, we have presented how the authoring of an individual's photo album can leverage and benefit from the content and structure of many. In this section, we describe how the semantics of many photo albums can be applied to semantically enrich the photo album of the individual. For it, we first describe in Section 5.1 the definition of semantics derivation rules. They base on the semantics of many photo albums and are leveraged by xSMART to learn about the semantics of a newly created photo album. The semantics derivation rules are defined along two dimensions, the granularity and kind of derivation. In Section 5.2, we present representative examples of the semantics derivation rules along these two dimensions. As the derivation rules are of different reliability (some are more reliable and others less), we briefly discuss in Section 5.3 how we deal with uncertainty for semantics derivation in xSMART. In Section 5.4, we describe how the derived semantics influences the authoring of photo books, i. e., how it affects their content and structure.

5.1 Definition of Semantics Derivation Rules

As described in the previous sections, the photo albums our project partner CeWe Color receives every month are analyzed in regard of different characteristics like number of pictures per page, size of pictures, use of textual descriptions, and many more. Based on the analysis of existing photo albums and the long-term experience of our project partner in the field, we extracted common rules for semantically enriching the individual's photo album. Thus, the semantics derivation rules comprise the accumulated knowledge of our project partner and the analyzed data of a large set of actually printed photo books. Consequently, the rules can be considered as the semantics of many that is made available for semantically enriching the photo book of the individual. The characteristics analyzed and presented here is only a first step. They undergo a constant refinement and improvement and further characteristics are added.

As depicted in Figure 5, the semantics derivation rules are divided according to two dimensions. The first dimension is the scale of granularity in which the rule can be applied. The scale goes—like the content rules in Section 3.3—from a single medium asset, a page of a photo book, a set of photo book pages (sub-album), to an entire photo book. The second dimension defines three different kinds of semantics derivation. The rules can be either testing, valuating, or calculating kind.

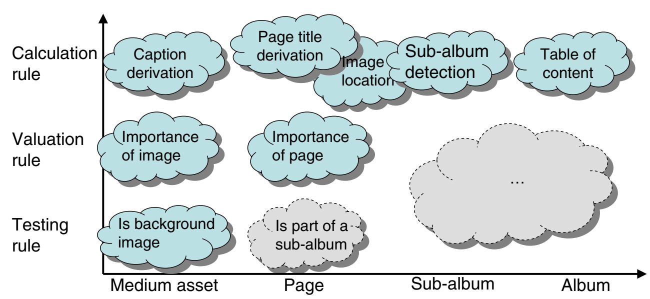


Figure 5: The two dimensions of semantics derivation rules and example rules

5.2 Examples of Semantics Derivation Rules

In the following, we present representative examples of semantics derivation rules along the three kinds: testing, valuating, and calculating. We briefly describe each rule and present a semi-formal definition in OCL based on the model defined in Section 2.

Testing Rules.

A testing rule is a derivation rule that is testing for the presence or absence of a specific value in the model of a photo album. For example, the rule “is background image” tests whether an image is used as background or not.

Is Background Image: A background image covers the entire page of an album. It is placed behind all other images and other visual media like texts and videos. In addition, the size of the image is the same as the photo album’s size. The following listing shows a semi-formal definition of this rule.

```
context Photo::isBackgroundImage() : Boolean
pre: self.size = photoAlbum.size
body:
  forAll(
    tempPhoto : page.current().getPhotos() |
      self.placement.realZIndex <
        tempPhoto.placement.realZIndex or
      self = tempPhoto )
```

Valuating Rules.

A valuating rule is a derivation rule that adds or removes *points of importance* to, e. g., images or pages of the photo album. For example, there is a rule that calculates the “importance of images”. Another rule determines the “importance of a photo book page”.

Importance of Image: From the domain knowledge of our project partner and the analysis of existing photo albums, one can derive that larger images are more important than smaller ones. Thus, the rule adds and removes points of importance to images used in the photo album based on their size. In order to determine whether images are small or large, we compare their spatial extension to the total size of the album. In addition, for electronic multimedia albums that can be created in xSMART in formats like Flash [2], SVG [11], and SMIL [12], we also consider whether the album is targeted at a Desktop PC with a large screen or at a mobile device like a cell phone with a small screen. For a Desktop PC size presentation and for a regular-size printed photo book, an image is considered large when it covers more then one third of the presentation page. An image that is larger than one third of a page is considered more important. If an image is smaller than one sixth of a page, it is considered small and less important. This valuation of importance is manifested by a point of importance that is added or removed to the image. For images of normal size between one sixth and one third of a page no point of importance is added or removed. The following listing shows this semantics derivation rule in OCL.

```
context Photo::imageImportanceRule1()
pre: photoAlbum.size = DESKTOP_PC or PHOTO_BOOK
pre: self.hasBeenApplied() -> false
body:
 if self.size > ( photoAlbum.size / 3 ) then
    self.importance = self.importance + 1
  else if self.size < ( photoAlbum.size / 6 ) then
    self.importance = self.importance - 1
  end if
post: self.basBeenApplied = true;
```

A multimedia album that is targeted at a cell phone changes the valuation rule of the importance of images significantly. On the small screen of a cell phone much less space is available compared to a Desktop PC screen or a regular-size physical photo book. Thus, fewer pictures will be selected for the album on the mobile screen. In addition, less images are placed per page (typically, not more that one or two images). It can be assumed that the selection of images for an album targeted at a mobile device is a subset of the images that would be selected for a Desktop PC screen size presentation or regular photo book. Consequently, all pictures selected for an album targeted at a mobile device are valuated with an additional point of importance. The following listing shows the semi-formal definition of this derivation rule.

```
context Photo::imageImportanceRule2()
pre: photoAlbum.size = MOBILE_DEVICE
pre: self.hasBeenApplied() -> false
body:
  self.importance = self.importance + 1
post: self.basBeenApplied = true;
```

Importance of Page: This valuation rule determines the importance of a photo album page. The rule counts the total number of images per page. Then, this number is compared with the average number of images in the album per page. If there are fewer images on a page than the average and these images are large, then this page is considered more important than if there are more images than average (which typically are small then).

Calculating Rules.

Finally, a calculating rule is a semantics derivation rule that is determining a specific, complex value based on given meta data. The calculated value adds semantic information to the photo album, like a “derived caption” of an image, the “title of an page”, “determining sub-albums”, estimating the “location of images”, and “determining the table of content” of the album.

Caption Derivation: A text that is placed directly under an image is considered caption of this media asset. However, the font size of the text is to be smaller than the font size of the (optional) page title. The following listing shows the caption derivation in OCL.

```
context Photo::captionDerivationRule() : String
pre:
  text.placement.realPosition.y -
  ( self.placement.realPosition.y +
  self.height ) < 20
derive:
  if page.title -> notEmpty() then
    if text.fontSize < page.title.fontSize then
      text.theText
    end if
  else
    text.theText
  end if
```

Page Title Derivation: If a text is placed at the top of a page (upper third of the page) and the text has a font size that is larger than 16 point, this semantics derivation rule considers the text as the title of the album page. In addition, no text element used in this page must have a larger font size than the title font size. The title must also not consist of more than ten words. The following listing shows the semi-formal definition of this rule.

```
context Page::pageTitleDerivationRule() : String
pre: text.placement.realPosition.y <
     ( photoAlbum.height / 3 )
pre: text.fontSize >= 16
pre: ServiceConnector::numerOfWords(
       text.theText ) < 10
pre:
  forAll(
    caption : ServiceConnector::getAllCaptions(
      self ) | text.fontSize > caption.fontSize )
derive: text.theText
```

Sub-album Determination: This semantics derivation rule for determining sub-albums bases on the "page title derivation" above. When a page title is detected, this can be the "title" of subsequent pages in the album. This is the case when subsequent pages do not have their own page titles. Then, the rule creates a new sub-album that starts with the page that has a page title and adds all subsequent pages without a page title until a page is found that again has a title. This rule is then recursively applied, with the page found that has a title. The reliability of this rule is reinforced, if the media assets on the album pages are arranged in chronological order and the time stamps of the pictures are close by. This is calculated by determining the temporal distance of the media assets on the page with the title. Then, the medium of all time stamps on the subsequent page must not be farer away then the temporal distance of time stamps on the page with the title. This algorithm is then applied to the next two pages (treating the first page as one that had a "title") and so forth.

Image Location: This rule allows to derive the location of an image from the location of the other images placed on the same page. If the other images used for the page have location information, e. g., in form of GPS signals, and the time stamps of all images (with and without location information) are similar (within a certain range r), there is a high probability that the location of the image without GPS signal is close to the location of the other images. This rule is applicable, if one uses a photo from a friend who accompanied the same vacation but this friend's camera does not have a GPS receiver. Ideally, the time stamps of all images are very close, i. e., within a few minutes.

Table of Contents: Based on the derived titles of the photo album pages and the determined sub-albums, this semantics derivation rule calculates a table of contents for the album. This table of contents is added to the beginning of the album. One could argue that his rule is not deriving semantics for the album but is creating new content. However, we consider the derived table of contents as an explicit representation of the album's semantics as it represents the narrative structure of the presentation.

Example Screenshot and Further Rules.

Figure 6 illustrates how the semantics derivation rules are applied. The screenshot shows a photo album page and the rules "image is background", "importance of images", "caption derivation", "page title derivation", and "image location" on a set of images taken by two friends on their vacation to Paris. Whereas the author of the photo book has a GPS camera, her friend has not. The screenshot shows the derived title, a more important picture in the center, and two less important pictures next to it on the left and right hand side. It also depicts the identified background image, two different captions, one for the two images outside the church and one for the inside. Finally, the xSMART tool could derive the location of the image that the author took from her friend.

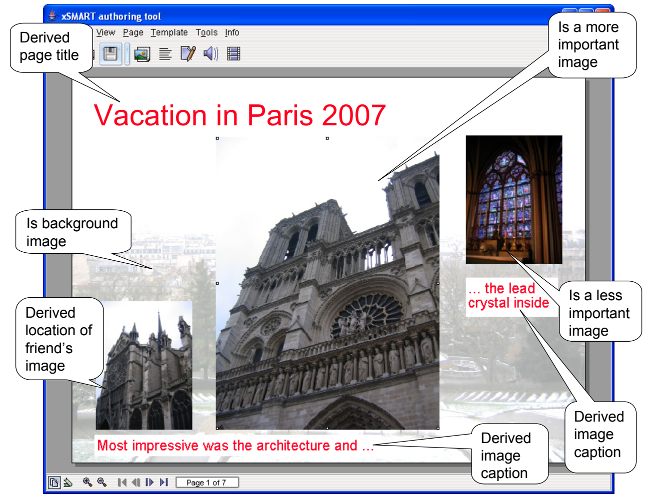


Figure 6: Application of semantics derivation rules

Besides the rules for semantics derivation presented so far, one can think of a multitude of further semantics derivation rules. These can be for language detection or extracting verbs, nouns, and names from titles and captions. For example, a rule could use a name vocabulary to extract names from the album and match the found names with a face detection algorithm applied on the photos. Another rule could automatically classify an album into categories like birthday, vacation, baby, and others. The latter is closely linked with the structure of a photo album presented in Section 4.

5.3 Reliability of Semantics Derivation Rules

None of the presented semantics derivation rules are hundred percent reliable. In order to reflect this fact, each derivation rule enriches the derived information with a reliability value. This reliability value lies in the interval of $]0, 1]$. A value close to 0 represents that the derived information is very unreliable. As a value of 0 would represent that the derived information is wrong, it is not represented in our model. A very reliable derived information is represented by a value close to 1.

Reliability depends on the rule itself (some are more some less reliable) and the input information provided to the rules (some derivations may base on the input of other rules). So far, we only roughly defined the reliability of our derivation rules. In future, we will conduct detailed analyses and evaluations of the actually printed photo albums and support the reliability values with statistical probability of real data.

5.4 Impact to Content and Structure

Besides "just" deriving new semantics about the created photo album, the semantics derivation rules can also have an impact to the exploitation of content and structure of the album. For example, an image that is marked as important may be ranked higher in future content selection processes. Thus, when it comes to a selection of content or enhancement of the album by additional content, these pictures will be advantageous and more often used than other pictures. An automatic layouting algorithm may choose a different layout to arrange the photos as some of them are marked to be more important than others. These two examples are considerably simple, however, they already illustrate how the semantics derivation may have an impact on the content and structure of the photo book creation for the individual.

6. CONCLUSION & CHALLENGES

In this paper, we showed how photo album creation can benefit from leveraging information learned from many users in regard of the album's content, structure, and semantics. The integration of external content is a good example of how the many can affect the individual to enhance one's personal photo albums. For the future we are thinking of much more sophisticated rules than the presented ones. We are hoping that much more Web 2.0 sites will provide interesting interfaces for retrieving content for the enhancement of personal photo books. As said, not only content is of interest for a photo book, but also web services that enrich the photo book with semantics that can lead to more focused search requests. For example, an event site such as SEraja [9] is an interesting means of enriching the photo book with structural information about a specific event, which could visually be represented in the photo book, but also by retrieving additional semantics, perhaps tags and place names, which enable for a more focused search at photo community sites.

We also see our analysis of the structure of existing photo books as a first step. By a more elaborated analysis, we hope to be able to derive layouting rules for automatically generating nicely assembled digital photo books.

Finally, also the semantics derivation rules defined in this paper are examples of our pursuit towards the development of more complicated and sophisticated rules. For example, we could analyze the use of photos over time in order to determine the change or stability of their semantics. When a photo is placed with other photos on an album page, it defines a semantic concept. This semantic concept can be annotated with the title found on the page. Whenever the same photo is used in another context, i. e., in other album pages with other photos, a new semantic concept is created. By analyzing the different concepts in which a photo has been used over time and in several albums, one can identify that the semantics of the photo changes. This is the case when the group of photos in which it is used is almost different in every page. However, if the photo is used in almost all album pages with all or an high overlap of the same photos, one may consider the semantics of the photo stable.

In future, we need to further elaborate the three identified areas of photo album authoring, namely the content, structure, and semantics. As we have presented, these areas are not independent of each other. Rather, there is a certain overlap between them and the areas influence each other. This needs to be more investigated.

However, conducting more analyses of photo albums printed by our project partner CeWe Color, deriving more characteristics, refining existing characteristics, and formalizing them into content, structure, or semantics derivation rules will be the central effort of our future work. With the leveraging of semantics, content, and structure of many for the creation of personal photo albums, we have achieved already very promising results. These results need now to be reinforced by a more detailed and elaborated analysis of our project partner's data. The potentials are high as CeWe Color receives and prints a large number of photo books every month. By analyzing this data, identifying and defining rules for photo album authoring, and leveraging this knowledge, we hope to be able to advance and ease the authoring of digital photo books and bring it into the next generation.

Acknowledgements

Herewith we thank our students Mareen Bantelmann and Holger Cremer who have implemented the described functionality in our *x*SMART tool. We also kindly thank our project partner CeWe Color for the always inspiring research cooperation, especially Sabine Thieme and her team. This research was partly supported by a Marie Curie International Fellowship within the 6th European Community Framework Programme.

7. REFERENCES

[1] Wikipedia, 2007. http://www.wikipedia.org.

[2] Adobe Systems, Inc., USA. Macromedia Flash (SWF) and Flash Video (FLV) File Format Specification (Version 8), 2007. http://www.adobe.com/licensing/developer/.

[3] S. Boll, P. Sandhaus, A. Scherp, and S. Thieme. Metaxa – context- and content-driven metadata enhancement for personal photo books. In *Proc. of the 13th international Multi-Media Modeling conference (MMM)*, Singapore, 2007. Springer.

[4] CeWe Color. CeWe Photobook, 2007. http://www.my-cewe-photobook.com.

[5] Intel Corp. Open Source Computer Vision Library (OpenCV). http://www.intel.com/technology/computing/opencv/index.htm.

[6] Japan Electronic Industry Development Association. Digital Still Camera Image File Format Standard (Exchangeable image file format for Digital Still Camera:Exif) Version 2.1, Dec. 1998. http://it.jeita.or.jp/document/publica/standard/exif/english/Exife.pdf.

[7] A. Scherp and S. Boll. Context-driven smart authoring of multimedia content with xSMART. In *Proc. of the 13th annual ACM Int. Conf. on Multimedia; Hilton, Singapore*, 2005.

[8] A. Scherp, S. Boll, and H. Cremer. Emergent semantics in personalized multimedia content. *J. of Digital Information Management*, 5(2), Apr. 2007.

[9] SEraja Technologies . SEraja EventWeb, 2007. http://www.seraja.com.

[10] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[11] W3C. Scalable Vector Graphics (SVG) Full 1.2 Specification, Apr. 2005. http://www.w3.org/TR/SVG12/.

[12] W3C. Synchronized Multimedia Integration Language (SMIL 2.1), Dec. 2005. http://www.w3.org/TR/SMIL2/.

[13] Yahoo Inc. Flickr, 2007. http://www.flickr.com.